WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

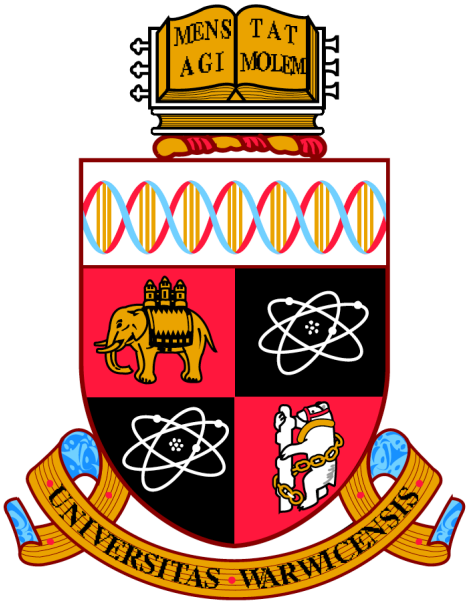
Permanent WRAP URL:

http://wrap.warwick.ac.uk/164025

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications



Structural and functional features of the SLC10 membrane protein transporters

by

Aurélien Anthony Grob

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

School of Life Sciences - Department of Structural Biology

December 2021

Abstract

The solute carrier transporter family 10 (SLC10), comprising more than 50 members across plants, animals and bacteria, encodes for a multitude of fascinating influx transporters, playing distinct roles in the transport of bile acids, steroidal hormones, drugs, and many other unknown solutes.

SLC10A7, classified as an "orphan transporter", is a unique and significantly abnormal member of this family, exclusively seen to play a distinct role in the negative regulation of intracellular calcium homeostasis, among humans, mice, zebrafish and yeast. This PhD study attempted to primarily characterise the structural features of bacterial homologues of this unique ten transmembrane transporter, whilst functional traits were also investigated. Membrane protein overexpression, detergent solubilisation, thermostability investigations as well as optimised protein purifications, from a multitude of bacterial SLC10A7 homologues, have all provided significant novel findings, whilst protein crystallisation and 3D structure determination remain challenging.

The apical sodium dependent bile acid transporter (ASBT) is involved in the uptake and transport of bile acids in the enterohepatic circulation of the intestines. Novel structural studies of a nine transmembrane *Leptospira biflexa* ASBT bacterial homologue were characterised, with successful crystallisation optimisations yielding a 2.2 Å crystal dataset. Along with functional thermostability and optimised purification studies, this novel transporter exhibits both structural and functional differences to previously crystallised bacterial ASBT homologues ($ASBT_{NM}$ and $ASBT_{Yf}$).

Together, both these projects further the knowledge of the functional and structural classifications of the SLC10 protein family.

Declaration

I hereby declare that I personally have carried out the work submitted in this thesis under the supervision of Prof. Alex Cameron at the School of Life Sciences, University of Warwick. Where work has been contributed to by other individuals, it is specifically stated in the text.

No part of this work has previously been submitted to be considered for a degree or qualification. All sources of information are specifically acknowledged in the form of references.

Acknowledgements

My utmost thanks, appreciation and gratitude is attributed to my supervisor, Professor Alex Cameron, for his resilient help and guidance among the past 4-5 years of this PhD study. His persistent support and excellent suggestions have allowed me to successfully grow as a more matured and developed research scientist, acquiring a plethora of transferable skills within the fields of Biochemistry, Molecular Biology and Structural Biology, which I will apply further in my future.

Special recognition is also addressed to Dr. Deborah Brotherton for her immensely generous help and knowledgable advice in the laboratory, which have strongly encouraged the successful completion of many challenging experiments.

My sincerest thanks and admiration are also allocated to my thesis viva examiners, Professor Roslyn Bill and Dr. Allister Crow, who have both significantly enhanced the quality and story-telling abilities of this thesis, via their detailed feedback and comments. Special gratefulness in particularly is directed to Dr. Allister Crow for his patience and fruitful guidance and support of the edited and finalised thesis.

My thanks go to colleagues of the C10 department.

Particular acknowledgements are assigned to my funding partners and PhD course coordinators: notably the BBSRC and MIBTP.

I would also like to thank my family: my parents for their constant support during my PhD studies and older brother for his undivided help and guidance with the writing of this thesis in LaTex!

Finally, exceptional appreciation is also directed to my partner, Charlotte, and her family, for their undivided support; endless thesis suggestions/feedback; immense help in the generation of certain figures; and constant belief in my studies in the completion of this PhD thesis. Without their constant support and encouragement, finalisation of this thesis would not have been possible!

Contents

Abbreviations

β-OG n-Octyl-β-D-glucopyranoside

AI Amelogenesis imperfecta

AS Alternative splicing

ASBT Apical sodium dependent bile acid transporter

BA Bile acid

BSEP Bile salt export pump

CA Cholic acid / Cholate

CDCA Chenodeoxycholic acid / Chenodeoxycholate

CMC Critical micellar concentration

CPM N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide

CRACR2A CRAC channel regulator 2A

CS Chondroitin sulfate

CYP7A1 7α-hydroxylase

CYP8B1 7α-hydroxycholest-4-en-3-one 12-α-hydroxylase

DCA Deoxycholic acid / Deoxycholate

DDM n-Dodecyl-β-D-maltopyranoside

DHEAS Dehydroepiandrosterone sulfate

DM n-Decyl-β-D-maltopyranoside

E3S Estrone-3-sulfate

EHC Enterohepatic circulation

ER Endoplasmic reticulum

FSEC Fluorescence size exclusion chromatography

FSEC-TS Fluorescence size exclusion chromatography - thermostability

FXR Farnesoid X receptor

GAG Glycosaminoglycan

GCA Glycocholic acid / Glycocholate

GCDCA Glycochenodeoxycholic acid / Glycochenodeoxycholate

GFP Green fluorescent protein

GFP-TS Green fluorescent protein - thermostability

HACS High-affinity Ca^{2+} uptake system

HS Heparan sulfate

ILBP Ileal lipid binding protein

IMAC Immobilised metal affinity chromatography

IPTG Isopropyl β-D-1-thiogalactopyranoside

kDa Kilo dalton

LACS Low-affinity Ca^{2+} uptake system

LB Lysogeny / Luria broth

LCA Lithocholic acid / Lithocholate

LCP Lipidic cubic phase

LDAO N,N-Dimethyldodecylamine-N-oxide

LMNG Lauryl maltose neopentyl glycol

Ni-NTA Nickel nitrilotriacetic acid

NM n-Nonyl-β-D-maltopyranoside

NTCP Sodium taurocholate co-transporting polypeptide

OATP Organic anion transporting polypeptide

OSTα/β Organic solute transporter α/β

PDC Protein-detergent complex

PS Pregnenolone sulfate

RCH1 Regulator of Ca^{2+} homeostasis 1

RFU Relative fluorescence unit

RMSD Root mean square deviation

SARAF SOCE-associated regulator factor

SBF Sodium bile acid co-transporter family

SD Skeletal dysplasia

SDS Sodium dodecyl sulphate

SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC Size exclusion chromatography

SERCA Sarcoplasmic/endoplasmic reticulum ATPase

SHP Short heterodimer partner

SLC Solute carrier transporter

SLC10 Solute carrier transporter family 10

SOCE Store-operated calcium entry

STIM1 Stromal interaction molecules

T_m Melting temperature

T7RNAP T7 RNA Polymerase

TCA Taurocholic acid / Taurocholate

TCDCA Taurochenodeoxycholic acid / Taurochenodeoxycholate

TMD Transmembrane domain

UDCA Ursodeoxycholic acid / Ursodeoxycholate

UDM n-Undecyl-β-D-Maltopyranoside

UV Ultra violet

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Biological Cell Membranes

Every living cell is enclosed by a biological membrane, critical for the compartmentalisation of the inside and outside of the cell/organelle. These mobile and dynamic enclosures allow for the free movement of molecules across the membrane plane, but are specialised to prevent the leakage of molecules generated from within and the unwanted diffusion from outside. This leads to the fundamental functions of transport systems, in order to exchange molecules across the membrane successfully. Indispensable for life, these biological barriers establish cell boundaries and serve to organise and regulate enzyme activity, maintain electrical potential, and store energy in order to inform signal transduction, thereby monitoring crucial differences within the cytosol and extracellular space of living cells (Alberts et al., 2008; Berg et al., 2012).

The plasma membrane is an intricately formed biological barrier, comprising a specialised arrangement of two very thin continuous fatty lipid and protein encompassed sheets. Assembled on an active plane of non-covalent attractive forces, the biological nature of the lipid molecules create a highly impermeable hydrophobic interior core of the cell membrane, prohibiting the passage of almost all water-soluble molecules. Specialised membrane embedded proteins are therefore utilised for the transmission of such molecules which are vital to all living cellular processes, and indeed life itself (Korn, 1968; Berg et al., 2012; Goñi, 2014).

These membrane proteins vary widely in structure and function, ranging from channels, pumps (energy transducers) and solute transporters, to receptors and enzymes. It has been estimated that almost 30 % of all proteins encoded within animal cell genomes are membrane proteins, highlighting their fundamental roles in biological systems and abundance in regulating cell molecular homeostasis across all different cell types (Wallin and von Heijne, 1998).

This introductory chapter will explore in greater detail the biological, functional and structural aspects of cell membranes and their individual counterparts, including membrane lipids and proteins. Membrane protein passive and active transport is also explored, shedding necessary light on their vast, but fascinating abilities within the lipid bilayer, both structurally, and functionally. Finally, in-depth analysis of the SLC10 membrane protein family, with specific focus on the human SLC10A2 and SLC10A7 membrane proteins, is also explored.

1.1.1 Membrane Lipids

Lipids are defined as water-insoluble biomolecules, that are exclusively soluble in organic solvents. Largely, lipid molecules consist of long fatty acid hydrocarbon chains, which vary in both length and degree of unsaturation. They contain between 14 and 24 carbon atoms, 16 to 18 being the most common (Alberts et al., 2008).

There are three major types of membrane lipids: phospholipids, glycolipids (sphingolipids) and cholesterol (sterols); although hundreds of minor classes also exist. These lipid molecules individually vary in levels of abundance in the cell membrane, but together form 50 % of the total mass of most animal cell membranes. The remaining mass is occupied by a multitude of varying types of membrane proteins (Berg et al., 2012).

Phospholipids are the most abundant form of lipid found in the cell membrane and consist of one or more fatty acids, a platform for adhesion of fatty acids (either glycerol, a three carbon alcohol or sphingosine), a phosphate and an alcohol. These lipid molecules are amphipathic/amphiphilic, consisting of both hydrophilic ("water-loving", polar) and hydrophobic ("water-fearing", non-polar) moieties. The fatty acid "tails" are the culprits for the highly hydrophobic interior environment of the lipid bilayer, whereas the remaining components ("head" group) constitute the hydrophilic nature of lipids, which interact with the aqueous surroundings. The charged hydrophilic domains tend to orientate themselves to favour contact with water, forming favourable and stable electrostatic interactions and hydrogen bonds, whereas the hydrophobic hydrocarbon tails of lipid molecules, prefer instead to aggregate and clump together, forming non-polar hydrophobic interactions among themselves. Typically, of the two fatty acid chains, one will contain one or more cis-double bonds (i.e. be unsaturated), whilst the other will be saturated, significantly affecting the fluidity of the cell membrane. Phosphoglycerides are the most common form of phospholipids within animal cells, consisting of a three-carboned glycerol backbone. The two long-chained fatty acids are attached to two glycerol carbon atoms via ester bonds and the third carbon

is linked to a phosphate group, which is further attached to a specialised head group, typically consisting of either ethanolamine, serine, or choline. An explicit illustration of phosphatidylcholine is demonstrated in figure 1.1 below (Harayama and Riezman, 2018).

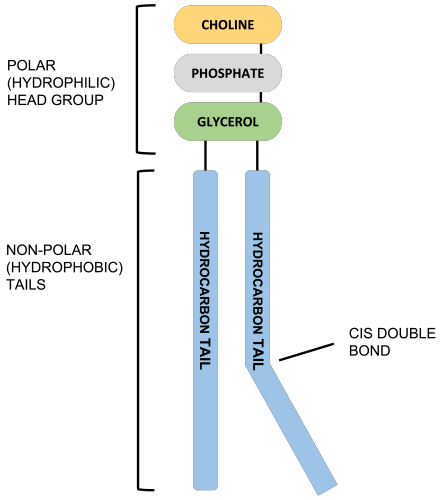


Figure 1.1: Schematic representation of a phosphatidylcholine molecule. Depiction explicitly highlights both the hydrophobic and hydrophilic moieties of the lipid molecule, with a "kink" in one of the fatty acid tails, caused by the inclusion of one or more cis-double bonds.
This figure was generated in Powerpoint

Glycolipids possess the same backbone as phospholipids, but contain an added sugar component derived from sphingosine (in animal cells). Accounting for roughly 5 % of all lipid molecules in the membrane, and found in all animal plasma membranes, glycolipids are typically located on the non-cytosolic monolayer (i.e. extracellular side) of the cell membrane, exhibiting the greatest asymmetry of all cell membrane lipid types. Having the tendency to self-associate via hydrogen bonds and van der Waals forces between their sugar and hydrocarbon chain components, glycolipids instead sometimes prefer to partition into lipid rafts (van Meer et al., 2008).

Cholesterol is the third and final major type of membrane lipid, found in extensive amounts in eukaryotic cell membranes (typically 1:1, cholesterol:phospholipid molecule ratio respectively). Cholesterol is classified as a sterol molecule, containing a rigid ring structure of four linked hydrocarbon rings. A non-polar hydrocarbon chain is attached to a steroid from one side, and a polar hydroxl group is added to the hydrocarbon ring on the other. The hydroxyl end of cholesterol interacts and orientates itself with neighbouring phospholipid head groups in the lipid bilayer. Interestingly, cholesterol molecules are entirely exempt in prokaryotic cell membranes, but found in varying degrees in almost all animal biological cell membranes, highlighting significant differences in the composition of cell membranes across all biological species (Yeagle, 2016).

1.1.2 The Plasma Membrane

Biological membranes are sheet-like structures ranging from 60-100 Å in thickness. They are composed primarily of lipids and proteins, stabilised in a fluid-like environment, to form non-covalent attractive interactions between lipid-lipid, lipid-protein and lipid-water assemblies. Hydrophobic interactions are the major driving forces in the creation of bimolecular sheets. Upon formation, the hydrocarbon tails release water molecules due to the sequestered environment of the non-polar interior of the lipid bilayer. Close packing of the membrane lipid tails are established via van der Waals forces, whereas electrostatic and hydrogen bonding forces allow for the attraction of lipid polar head groups and water molecules in the surrounding aqueous environments. As previously mentioned, the amphipathic nature of lipids creates a highly selective permeability barrier within the plasma membrane, forming a highly hydrophobic interior core. Due to this phenomenon, the transport of water-soluble molecules across the membrane is scarcely possible without the aid of membrane embedded proteins. As such, these lipids are directly responsible to monitor the stability, oligomerisation and functional activation of embedded membrane proteins, as will be discussed in greater detail below (Korn, 1968; Goñi, 2014; Bernardino de la Serna et al., 2016; Nji et al., 2018).

In aqueous solutions, phospholipids prefer to arrange themselves in a double layered bimolecular sheet rather than a spherical lipid micelle. This is due to both the amphiphilic nature and overall shape of the fatty acid chain of phospholipids. Comprising of two fatty acid chains, these are too bulky to arrange into a micelle and instead favour the formation of a lipid bilayer. In contrast, many detergent molecules as well as salts of fatty acids readily form micelles due to their single fatty acid chain. A micelle has a limited structure of usually less than 20 nm diameter, whereas a lipid bilayer can extend to macroscopic dimensions of up to a millimetre. Spherical micelles are formed with the hydrophobic compartments of lipids facing inwards, whereas lipid bilayers prefer hydrophobic tails to be sandwiched between hydrophilic head groups (Alberts et al., 2008; Goñi, 2014). A very basic illustrative comparison between a micellar and lipid bilayer arrangement is displayed in figure 1.2 below.

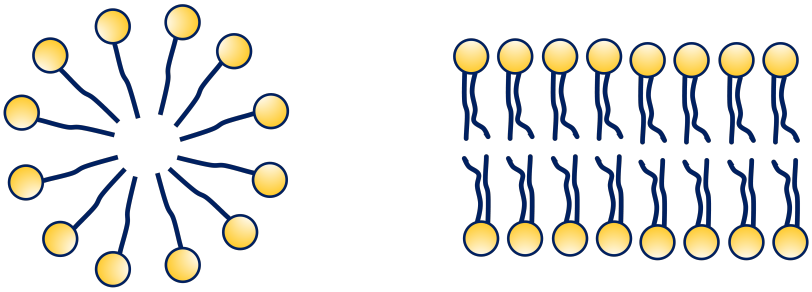


Figure 1.2: Comparison of a micellar (left) and lipid bilayer (right) arrangement of phospholipid molecules. Single fatty acid chains prefer to be arranged into micelles, whereas double fatty acid chains are strictly confined to a lipid bilayer conformation.
This figure was generated in Powerpoint.

Membranes and their constituents (comprising of a myriad of different lipids and proteins), are dynamic and fluid structures. This suggests that individual lipid and protein molecules possess the ability to constantly diffuse freely within the membrane, rotating rapidly in a perpendicular (rotational diffusion), and lateral manners to the planar axis of the membrane. This movement, coupled with the natural disordered environment of the cell membrane, presents an irregular surface for hydrophilic interactions and increases the overall cellular membrane fluidity (Los and Murata, 2004; Nji et al., 2018).

Fatty acid and cholesterol molecules specifically conduct the fluidity of the cell membrane. As mentioned above, fatty acid chains differ widely in both length and degrees of saturation, influencing their ability to pack together successfully. They can be arranged in either an ordered and rigid state; or a disordered and fluid state. Presence of a cis double bond in the fatty acid, bends the hydrocarbon chains, thus favouring a disordered assembly, and rendering a more fluid-like environment. Cholesterol also modulates cell membrane fluidity, by encompassing a bulky steroid nucleus with a flexible hydrocarbon tail and hydroxyl group. The hydroxyl group associates with the carbonyl oxygen atom of the phospholipid head group to form a hydrogen bond, and the hydrocarbon tail embeds itself in the non-polar core of the lipid bilayer. Cholesterol's insertion amongst phospholipids enhances the permeability-barrier characteristics of the cell membrane. Due to its bulky, irregular structure (compared with phospholipids), its insertion in the cell membrane perturbs the regular interactions of the fatty acid chains. Cholesterol therefore tightens the unordered packing of lipid molecules in the plasma membrane and renders the membranes more fluid, allowing for greatly enhanced transduction rates across the membrane bilayer. Since the cell membranes of bacteria cells contain no cholesterol molecules, these are instead stabilised by their overlying cell walls. In animal cells however, membrane fluidity is predominantly modulated by cholesterol (Los and Murata, 2004; Berg et al., 2012; Goñi, 2014; Yeagle, 2016; Bernardino de la Serna et al., 2016).

Every biological membrane is structurally and functionally asymmetric. This means that both the inner and outer lipid compartments of the membrane are strikingly different, resulting in a plethora of crucially different functional characteristics of both sides of the membrane, which is essential for the transfer of information from the extracellular space, to the cytosol of the cell. Membrane proteins also exhibit asymmetry, as they are orientated into the membrane in a unique manner upon synthesis. This enhances specific structural and functional characteristics that are fundamentally different on either side of the membrane, and due to the inability to rotate from side to side, this asymmetry is conserved in the cell membrane (Rothman and Lenard, 1977; van Meer et al., 2008; Lorent et al., 2020).

Finally, membranes typically form compartments of entire cells, but also of specialised organelles, such as the nucleus or the mitochondria of eukaryotes. The membrane proteins in such compartments are crucially involved in mediating molecular and ionic concentrations. In both the cell nucleus and the mitochondria, for instance, they are surrounded and enclosed by a double membrane, highlighting the functional importance of the gatekeeper of genetic material and the site of ATP production respectively (Alberts et al., 2008; Berg et al., 2012; Krapf, 2018).

Figure 1.3 below, presents a basic overview of the biological layout of the eukaryotic cell membrane, including its many membrane lipid and protein components.

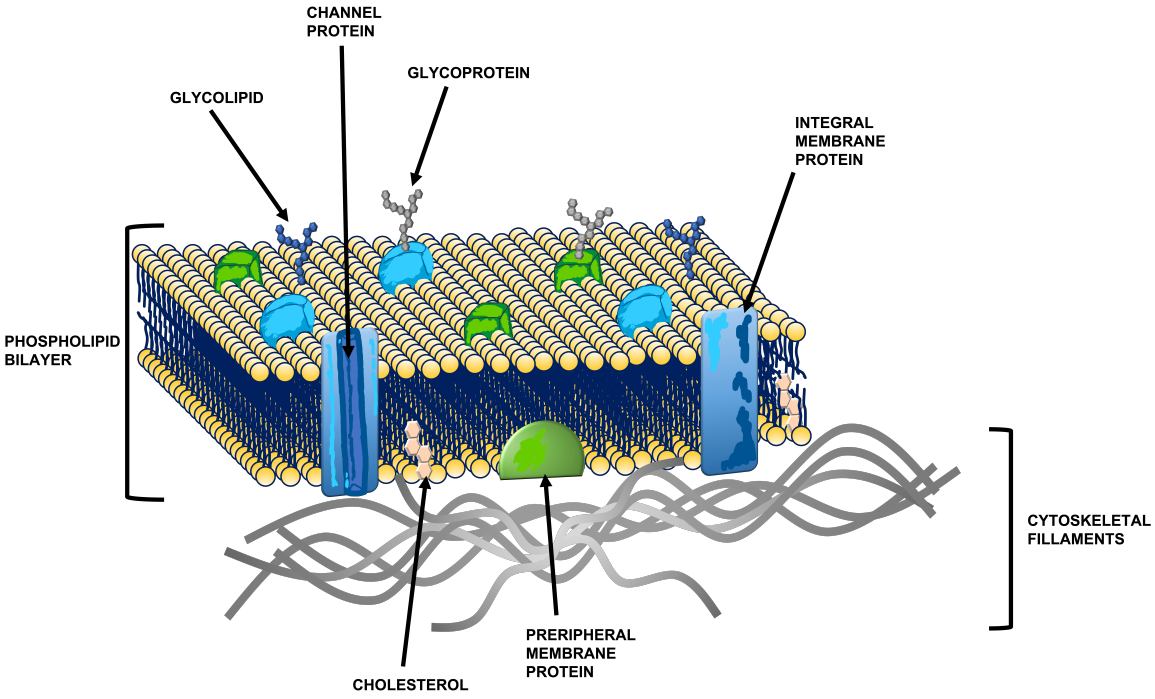


Figure 1.3: The fluid mosaic model of the cell membrane, entailing a multitude of macromolecular molecules, including phospholipids, glycolipids, cholesterol and many specific membrane proteins. This figure was adapted from OpenStax College, 2013, and generated in Powerpoint.

1.2 Membrane Proteins

In general, the lipid bilayer of all living cells provides the overall biological cell structure and barrier to the external environment, where the fundamentally functional characteristics are performed by the membrane proteins embedded within. There are a variety of classes of membrane proteins, each responsible for a plethora of biological functions. Roughly, the ratio of lipid to membrane protein molecules is 50:1 respectively, in a typical cell comprising of 50 % total protein mass (Laganowsky et al., 2014; Yeagle, 2016).

Membrane proteins, much like membrane lipids, are also amphipathic molecules, comprising of both hydrophobic and hydrophilic moieties. Their hydrophobic regions transition through the membrane, associating within the hydrophobic core of the cell membrane, sequestered from water molecules. The hydrophilic regions, in turn, prefer to interact with the water molecules on both extracellular and cytosolic sides of the biological membrane (van Meer et al., 2008; Sadaf et al., 2015; Harayama and Riezman, 2018).

As mentioned previously, membrane proteins are embedded in the bimolecular sheet in an asymmetric fashion, undertaking a unique structural orientation. This in turn means different protein functions are executed on either side of the cell membrane, maintaining substrate homeostasis of the cell (Lorent et al., 2020).

1.2.1 The Different Types of Membrane Associate Proteins

There are several ways in which a protein can interact with the plasma membrane. These are outlined in more detail in figure 1.4 below, but largely comprise of (Heijne, 2007):

1. Single-pass integral membrane proteins. A single-pass alpha helix passing the bilayer just once (such as α helix recognition receptors).

2. Multi-pass integral membrane proteins. A multi-pass alpha helical arrangement spanning the plane of the cell membrane several times (such as α helical bundle enzymes, transporters and receptors). This is typically the most common type of membrane protein involved predominantly in transport systems.

3. β-barrelled integral membrane proteins. Multiple beta sheets spanning the hydrophobic core of the bimolecular sheet, rolled to form an overall β-barrelled assembly (such as β barrelled transporters (most typically channels)).

4. Monotopic peripheral membrane proteins. A single alpha helix protein anchored to just the cytosolic compartment of the membrane, by weak hydrophobic attractive forces (such as enzymes and anchorage proteins).

5. Extrinsic peripheral membrane proteins. A special scenario where only the N- or C-terminal non-helical (loop) domain of the protein embeds itself into the extracellular side of the plasma membrane (such as enzymes and anchorage proteins).

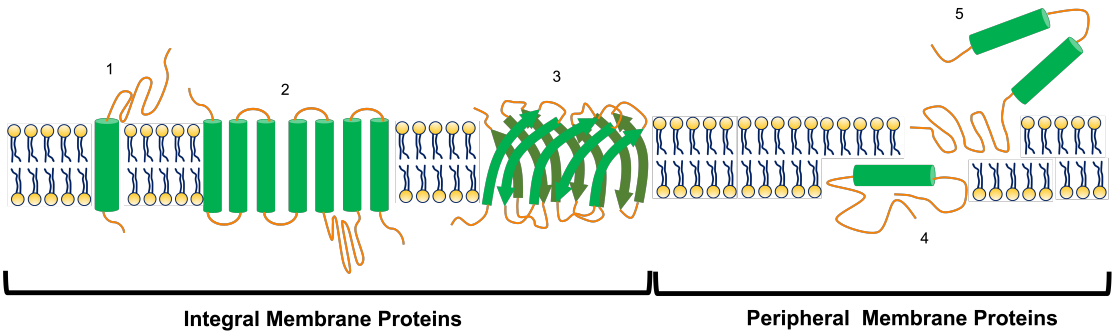


Figure 1.4: The 5 types of membrane proteins.
1. Single-pass α helix, 2. Multi-pass α helix, 3. β-barrelled, 4. Monotopic, 5. Extrinsic.
This figure was adapted from Alberts et al., 2008, and generated in Powerpoint.

1.2.2 Peripheral and Integral Membrane Proteins

Membrane proteins can be typically categorised into two main classes: peripheral and integral membrane proteins, based primarily by their dissociability properties (Alberts et al., 2008; Berg et al., 2012).

Peripheral membrane proteins are typically bound by electrostatic and hydrogen forces to the polar head groups of membrane lipids. These proteins can also additionally be attached via specific interactions to the surface of integral membrane proteins on either side of the cytoplasmic and extracellular environments. These interactions can be easily perturbed by the addition of high or low ionic salts, or extreme changes of pH, allowing for easy extraction from the cell membrane. This therefore illustrates the overall weak bindings of the protein-protein and protein-lipid interactions of this subclass (Whited and Johs, 2015; Boes et al., 2021).

Integral membrane proteins, in comparison, are transmembrane proteins which span the entirety of the lipid bilayer and whose segments associate with both the highly hydrophobic interior core, and intra- and extracellular hydrophilic compartments of the lipid bilayer. These membrane proteins produce multiple strong,

attractive, non-covalent forces, rendering protein extraction much harder. Competition and disruption of same non-polar interactions holding these bonds together must occur, which is usually possible with specialised biological detergents (as is explored later) (Odahara, 2004; Whitelegge, 2013; Cymer et al., 2015).

These outlined differences in membrane protein subclasses, can be seen visually in greater detail in figures 1.4 and 1.3, presented above.

Within the class of integral membrane proteins, exist a further 2 structural motif categories: α-helical and β-stranded assemblies, which are discussed further below.

1.2.3 Alpha Helical Integral Membrane Proteins

Membrane proteins which span the entire thickness of the cell membrane tend to be arranged by α-helices. This is typically the most common structural motif, due to the increased efficiency in forming hydrogen bonds between hydrogen and oxygen atoms in amino and carboxyl groups, of amino acids spaced four residues apart. As such, all internal backbone hydrogen bonds become fulfilled, guaranteeing a lack of exposed amino acid polar groups in the hydrophobic core of the cell membrane. In fact, almost all multi-pass transmembrane proteins found in animal cells are predominantly formed of α-helical assemblies.These membrane proteins entail domains located within both the cytosolic and extracellular membrane compartments, and are separated by long hydrophobic, largely non-polar and non-charged polypeptide chains, located within the highly hydrophobic interior core domain of the lipid bilayer (Alberts et al., 2008; Berg et al., 2012).

Within this helical motif, amino acid residues are connected in series by polar peptide bonds, which in the absence of water, form hydrogen bonds with other neighbouring peptide bonds. This repeated hydrogen bonding, as well as established hydrophobic interactions with lipid molecules, leads to the formation of α-helical domains, which span both lipid layers of the membrane. Typically, the hydrocarbon core of a membrane lipid bilayer is 30 Å wide, suggesting it only takes α-helices approximately 18-25 amino acid residues to reach either side of the membrane (i.e. the extracellular or cytosolic compartments). These helices are mobile in nature and slide along each other in specialised and regulated manners, allowing for conformational changes to take place, where substrate and/or ion binding pockets are either exposed or hidden, resulting in the successful transfer across the cell membrane (Doig et al., 2001; Eisenberg, 2003).

Only multi-pass α-helical transmembrane proteins undergo substrate/ion translocation across the cell membrane, since these form specific non-covalent interactions with neighbouring helices, allowing for specific structural conformations to occur.

Changes in these conformations typically occur upon binding of substrates and/or ions to exposed binding sites, resulting in the transport across the cell (as is explored in greater detail below). Without these exquisitely specific and crucial non-covalent interactions, the structural and functional integrity of transmembrane proteins would be severally impacted, often resulting in denatured, aggregated and highly deactivated protein assemblies (Doig et al., 2001; Eisenberg, 2003; Alberts et al., 2008; Berg et al., 2012).

An example of the structural moeity of an α-helix integral membrane protein is seen below in figure 1.5.

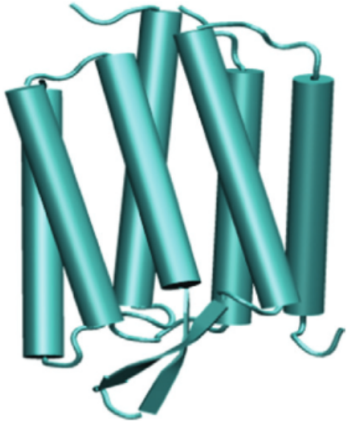


Figure 1.5: The α-helix structural protein motif, as seen in the integral membrane protein Bacteriorhodopsin, from *Halobacterium salinarum* (PDB (Berman et al., 2000): 2BRD). This figure was taken from (Yeagle, 2016)

1.2.4 Beta Barrelled Integral Membrane Proteins

Certain membrane proteins, such as pores and channels, are typically formed from the specialised assembly of β-stranded sheets. In comparison to α-helices, these structures are usually very rigid, and can be formed by as few as 8 to as many as 22 β-sheets. β-strands are formed when the protein polypeptide chain is passed through the cell membrane multiple times and is folded back on itself in an antiparallel manner. This creates multiple transmembrane hydrogen bonds between neighbouring strands, via certain polar peptides but also polar amino acid side chain residues, within the highly hydrophobic interior core of the membrane. This in turn produces multiple β-sheets, which are repositioned and rolled together, via an energetically favourable manner, to create a closed and hollow cylindrical assembly known as a β-barrelled membrane protein pore or channel (Eisenberg, 2003; Fairman et al., 2011; Chaturvedi and Mahalakshmi, 2017).

Interestingly, β-barrelled integral membrane proteins are predominantly found in the outer cell membranes of chloroplasts, mitochondria and almost all bacteria.

Some can form water-filled channels (i.e. pores), which allow for the passage of small hydrophilic molecules across the lipid bilayer. The interior aqueous lining of these channels are exclusively made of polar amino acid side chains, whereas the outer lining of the channel is comprised mostly of non-polar side chains, interacting with the hydrophobic core of the membrane. Due to the rigidly bound hydrogen bonds formed between neighbouring β-sheets, it is typically unlikely for β-barrels to undergo conformational changes. In light of this, it is common for β-barrelled integral membrane proteins to behave as solute and ion transporters operating in a passive manner, as is explored in further detail below (Eisenberg, 2003; Fairman et al., 2011; Alberts et al., 2008; Berg et al., 2012; Chaturvedi and Mahalakshmi, 2017).

Some β-barrelled membrane proteins also serve other functions such as receptors or enzymes. These types are preferentially held within the membrane by the rigidly anchored β-barrel shape, and allow for specialised intra- and extracellular molecular binding via the cytosolic loop segments (Fairman et al., 2011; Chaturvedi and Mahalakshmi, 2017).

An example of the structural assembly of an integral β-barrelled membrane protein is seen below in figure 1.6.

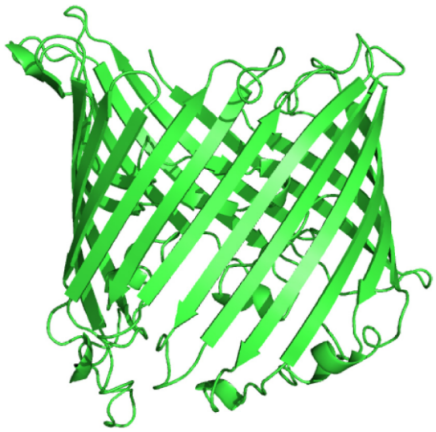


Figure 1.6: The β-barrelled structural protein motif as seen in the integral membrane protein ScrY porin, from *Salmonella typhimurium* (PDB (Berman et al., 2000): 1A0S).
This figure was taken from (Yeagle, 2016).

In both α-helical and β-barrelled membrane protein arrangements, the protein structures are positioned within the cell membrane in a highly regular arrangement. This suggests that all backbone hydrogen bond donors and acceptors must participate in the formation of hydrogen bonds, and since there is no water present in the hydrophobic core to compete for polar groups, their breakage is highly unfavourable (Alberts et al., 2008; Berg et al., 2012).

1.3 Membrane Transport

Membrane protein transporters serve as the gatekeepers of living cells, allowing for the exchange (i.e. uptake and efflux) of vital components, including amino acids, drugs, vitamins, various cationic and anionic ions, hormones, dyes, inorganic molecules, nucleotides, lipids, sterols, sugars, signalling molecules, bile acids, neurotransmitters, carbon compounds and metabolic end products. Functionally, they are responsible for a plethora of crucial biochemical regulations including the maintenance of ion homeostasis (both intra- and extracellularly), the acquisition of organic nutrients, the environmental sensing and cell communication, as well as the excretion of toxic metabolic waste products in eukaryotes, bacteria and archaea (Hediger et al., 2004; Balakrishnan and Polli, 2006; Hassan et al., 2014).

As previously mentioned, only multi-pass transmembrane proteins, which are fully embedded across the whole lipid bilayer, are able to transport small hydrophilic molecules, from one side of the membrane to the other. Certain non-polar, non-charged lipophilic (and sometimes water molecules) can pass through membranes by dissolving within the lipid bilayer (via simple diffusion). However, almost all small hydrophilic, polar and charged molecules (such as ions and polar substrates) require the aid of specialised transport proteins to be successfully transferred (via either passive or active diffusion), across the cell membrane. This is in large part due to the highly impermeable intrinsic selective barrier, originating from the hydrophobic characteristics of lipid molecules. These macromolecular transport assemblies must therefore provide a specialised mechanical transmission pathway, for molecular transfer to be initiated (Bassham and Raikhel, 1996; Quick and Javitch, 2007; Alberts et al., 2008; Berg et al., 2012).

Membrane transport is crucial for all living organisms, highlighted by the fact that 30 % of their genomes encode for membrane proteins in all cells. Furthermore, certain specific mammalian cells often use up to two thirds of their metabolic energy consumption to initiate membrane transport of crucial substrates (Wallin and von Heijne, 1998).

The metabolic activities of a certain cell type is largely determined by the expression of its transporters in the lipid bilayer. These transporters typically define the working ins and outs of the cell, specifically regulating its ionic and molecular substrate compositions, by executing the relevant downstream biochemical reactions for cell survival (Hassan et al., 2014).

Finally, the transmission of molecules across a cell membrane would not be possible without the production of its ion gradients. These gradients allow for the flow of essential molecules across the membrane, leading to fascinating downstream biochemical processes, such as the creation of electrical signals in nerve cells, or

the synthesis of adenosine triphosphate (ATP) to drive substrate transport in primary active transporters (see below for more detail) (Paterson, 1968; Stillwell, 2016; Alberts et al., 2008; Berg et al., 2012).

1.3.1 The Different Classes of Membrane Transport Proteins

Transport systems vary considerably, due to their specific substrate binding selectivity, their putative membrane topology, as well as their mechanism for energy coupling. Largely, transporters are classed into one of four categories, based on the latter criteria (Hassan et al., 2014):

1. Membrane channel proteins. These are pore-forming membrane proteins, which allow for the rapid transmission of ions (ion channels), water (aquaporins), and other small non-polar, non-charged lipophilic solutes, across a cell membrane, via a thermodynamically downhill action. These act as facilitated/passive diffusion systems.

2. Membrane pump proteins. These typically use a free energy storage source (such as the absorption of light or the hydrolysis of ATP), in order to initiate the thermodynamic uphill transport of ions or solutes across the cell membrane. Pumps are prime examples of primary active transporters.

3. Membrane solute carriers. These modulate the transfer of small hydrophilic molecules across a lipid bilayer, via a thermodynamically uphill manner, and are initiated by the favourable thermodynamic downhill transfer of specific ions across the membrane. Solute carriers behave exclusively as secondary active transporters.

4. Membrane group translocators. These are unique macromolecules, which specifically modify the molecular structure of transported substrates upon transmission across the cell membrane.

An illustrative view of all 4 classes is presented in figure 1.7 below

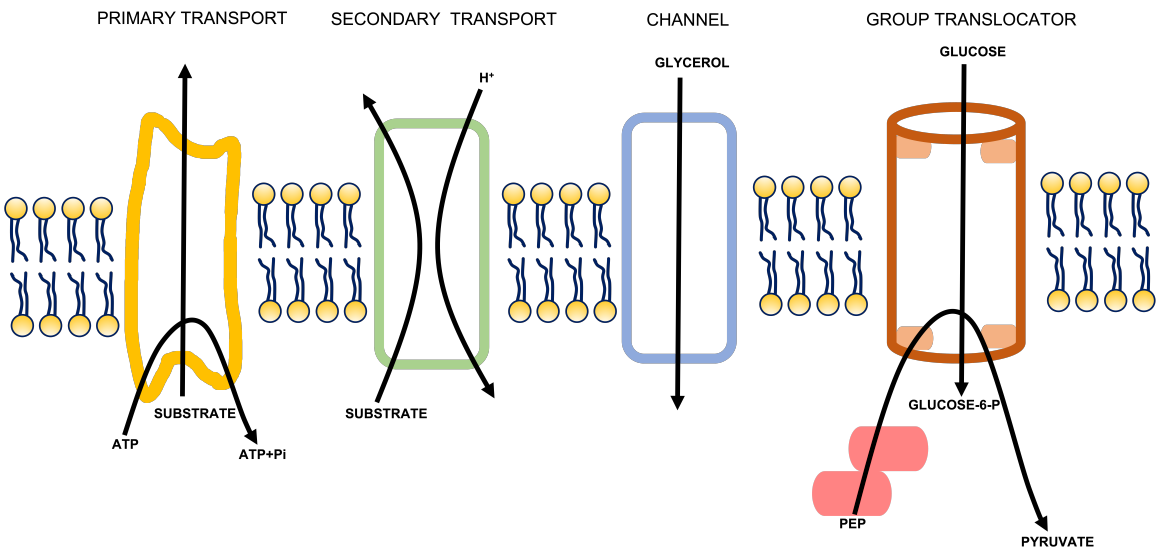


Figure 1.7: The 4 distinguished classes of membrane transport proteins: Primary active transporters (such as pumps), Secondary active transporters (such as carriers), Passive transporters (such as channels) and Group translocaters.
This figure was adapted from Hassan et al., 2014, and generated in Powerpoint.

These four classes of macromolecules, transfer small hydrophilic substrates and/or ions across the plasma membrane, in either a passive or active fashion, typically determined by the permeability features of the substrate within the membrane, as well as the abundance of an available energy source (Alberts et al., 2008; Berg et al., 2012; Hassan et al., 2014).

1.3.2 Passive/Facilitated Transport

The transfer of certain ions (such as sodium (Na^+), calcium (Ca^{2+}), chloride (Cl^-) and potassium (K^+) ions) across a cell membrane, usually occurs in a passive manner. This is achieved through membrane channels which form highly specialised pores across the lipid bilayer, allowing for the facilitated, rapid transfer of small hydrophilic molecules, ions or water, acting down their concentration and electrical gradients (Feher, 2012; Zhernenkov et al., 2016).

Ions, for example, are facilitated to diffuse through specific channel proteins, from a high to low concentration, regulating the cell's ionic homeostasis on both sides of the biological cell membrane. These channel proteins consist of highly selective, narrow, hydrophilic interior pores, which are typically gated (i.e. not permanently open), allowing for rapid changes to the biochemistry of the cell via changes in cell electrical potentials (voltage-gated channels), or changes in response to specific stimuli (ligand-gated channels). Since transport is modulated by the substrates' own energy gradient, and the flow is successfully conducted without the use of any other metabolic energy source, the transport is termed

facilitated/passive diffusion (Zeuthen, 1995; Feher, 2012; Zhernenkov et al., 2016).

Ion flow through channels is very rapid, and can reach rates of more than 1000 times faster than active transporters (see below for pumps and carriers). This means that up to 100 million ions can be transported across the lipid bilayer every second, via an opened channel. These rates closely resemble rates of freely diffusing ions in aqueous solution, emphasising the rapid transmission characteristics of channels. These channels though are not simple "tubes" spanning the membrane, allowing for rapid ion flow, but rather highly intricate molecular assemblies, responding to both chemical and physical changes, which undergo specifically timed structural modifications to establish flow across the lipid bilayer. The narrow hydrophilic pores of the channel restrict the entry of specific substrates, based on their size and charge. In contrast to active transporters however, the interaction between the interior channel pore and its specific substrate is much weaker (Zeuthen, 1995; Alberts et al., 2008; Berg et al., 2012; Feher, 2012; Zhernenkov et al., 2016).

1.3.3 Active Transport

Active transport allows for the transfer of ions and small hydrophilic molecules, across a cell membrane, via the use of energy from a specific source. Typically, these molecules are transferred against their electrochemical gradients (i.e. from a low to high concentration), resulting in a decrease in entropy. Because of this, a free energy source is required, usually actively generated by the transporter in a variety of ways (Alberts et al., 2008; Berg et al., 2012).

Typically, active transporters (such as pumps and carriers), are able to transfer ions across the cell membrane at rates of several thousand ions every second. Compared to passive diffusion (i.e. channels), this is more than 1000 times slower and largely due to the abundance of ions, as well as energy sources available for aided molecular transmission. Furthermore, many structural moving parts are involved to transfer molecules from one side to the other of a cell membrane, limiting the available binding sites available to both ions and molecules. Active transporters however, exhibit a higher substrate stereospecificity in comparison to passive transporters (such as channels) (Ghosh, 1963; Pardee, 1968).

Almost every lipid bilayer, in all species, possess an electrical potential difference (voltage gradient), whereby the interior of the cell exhibits a net negative charge with respect to the extracellular environment. This voltage difference usually favours the influx of cations (i.e. sodium, calcium or potassium ions), and opposes anions (i.e. chloride ions) (Skou, 1965; Berg et al., 2012).

When a substrate carries a net charge, both its concentration differences across

the cell membrane, as well as its electrical potential gradient, govern its transport. Assembled together, these gradients establish the net driving force of substrate transmission, known as the electrochemical gradient. These substrate electrochemical gradients are crucial for the initiation of various transport systems, as is explored in greater detail below (Ghosh, 1963; Skou, 1965; Alberts et al., 2008).

Active transport proteins can be classified further as primary or secondary, both undergoing specific structural conformational changes, in order to allow for the transport of small hydrophilic molecules across the cell membrane. As such, active transport is driven either by a primary chemical energy source (primary active transporters) or by a pre-existing ionic/solute source (secondary active transporters) (Ghosh, 1963; Skou, 1965; Pardee, 1968; Alberts et al., 2008; Berg et al., 2012).

Primary Active Transport

Primary active transporters are typically represented by specific membrane pumps. These can either be by light activated (found primarily in bacteria and archaea, initiated from an input of light energy), or by ATP driven pumps. The latter couples the "uphill" transport of specific molecules, via the energy generated by the hydrolysis of ATP. As part of the process, ATP is broken down to ADP and a phosphate molecule, generating free energy to drive the transport of ions or substrates against their electrochemical gradients (i.e. "uphill"), across the biological cell membrane. Often termed energy transducers, primary active transporters convert one form of free energy into another, in order to drive solute transport (Berg et al., 2012; Stein and Litman, 2015).

There are three major types of primary active pumps that utilise an ATP energy source for assisted transport of molecules: P-type pumps, F-type pumps and ATP-binding cassette (ABC) transporters. These exhibit specific structural conformational changes upon the binding and hydrolysis of ATP, resulting in the transport of a bound ions or molecules across the plasma membrane (Stein and Litman, 2015; Chen and Lui, 2021).

P-type ATPases pump ions against their electrochemical gradient, via the hydrolysis of ATP molecules, resulting in the transient phosphorylation at specific aspartic residues. This category of ion pumps is largely responsible for the maintenance of K^+, H^+, Ca^{2+} and Na^+ gradients, across all biological cell membranes. Regulating the voltage gradient of all cells is crucial, and is primarily mediated by this subclass of primary active transporters. The Na^+-K^+ ATPase ionic pump, for example, is an exchanger, actively extruding three Na^+ to the exterior cell environment, against its electrochemical gradient, for every two K^+ imported into

the cell cytosol. This leads to the cell becoming electrogenic, where a net current across the cell membrane is enforced, causing the cell's interior to be negatively charged with regards to the cell exterior. Transfer of these ions is only achieved via the use of ATP as an energy source, and this also provides a conserved mechanism of transport for other cationic molecules including Ca^{2+} (Ghosh, 1963; Alberts et al., 2008; Stein and Litman, 2015; Chen and Lui, 2021).

F-type pumps consist of turbine-like proteins, entailing several different subunits. They are structurally very different from both P-type pumps and ABC transporters, and are predominantly found in the cell membranes of mitochondria, chloroplast thylakoids, and bacteria. In comparison to P-type ATPases, F-type pumps instead function in reverse, as ATP synthases, using the H^+ gradient to drive the production of ATP molecules, from both ADP and phosphate. This H^+ gradient is exhibited typically during photosynthesis (in chloroplasts), by light (in *Halobacterium*), or by the electron-transport steps of the oxidative phosphorylation mechanism (in mitochondria or aerobic bacteria). In light of this, F-type pumps can be seen as key regulators of the cell's pH levels (Alberts et al., 2008; Berg et al., 2012; Stein and Litman, 2015; Chen and Lui, 2021).

ATP-binding cassettes (ABC transporters) are by far the largest membrane transport protein family (in both pro- and eukaryotes), consisting of four major structural domains: two membrane-spanning hydrophobic domains (each entailing six segments, providing substrate specificity) and two conserved hydrophilic ATPase domains (protruding into the cytosol and responsible for ATP binding and hydrolysis). Compared to both P-type and F-type ATPases, ABC transporters preferentially transfer small polar molecules across the lipid bilayer, instead of inorganic ions. Furthermore, these pumps themselves do not become phosphorylated, but rather utilise the energy generated from the hydrolysis of ATP to drive structural conformational changes, resulting in the transfer of specific substrates across the cell membrane. ABC transporters can either perform as export systems (found in all domains of life), or import systems (specific to just prokaryotes). In eukaryotes and prokaryotes, substrates exported out of the cell typically include toxins, signal molecules, various drugs, metabolites, and lipids. In eukaryotes, the ABC efflux transport mechanism occurs as follows: the substrate-binding site is exposed to the cytosol, when no ATP molecule is bound (inward-facing conformation). Upon ATP binding, dimerisation of the ATP-binding domains occur, exposing the substrate-binding pocket to the extracellular side of the membrane (outward-facing conformation). Following ATP hydrolysis, the ATP-binding domains dissociate, resulting in the formation of the original (inward-facing) transporter conformation. A large majority of multi-drug and antimicrobial resistant proteins operate as ABC membrane proteins, highlighting their fundamental im-

portance in biological processes and specific use as drug targets (Ghosh, 1963; Alberts et al., 2008; Berg et al., 2012; Stein and Litman, 2015; Chen and Lui, 2021).

The differences and similarities among these transporters as explored above, are illustrated below in figure 1.8.

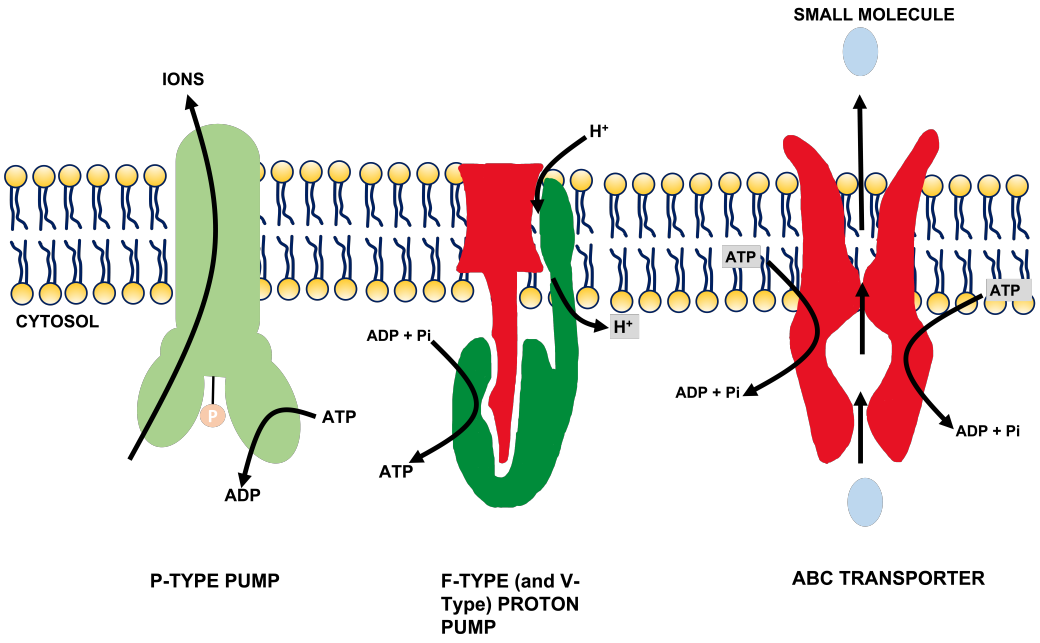


Figure 1.8: The three types of ATP-driven primary active transporters. P-type pumps (left) use the hydrolysis of ATP (itself becoming phosphorylated in the process), to export ions across the cell membrane. F-type pumps (middle) work in reverse, synthesising ATP via the favourable influx of H^+ ions. ABC transporters (right) specifically target small hydrophilic molecules instead of ions, using the hydrolysis of ATP to generate energy, to export the molecules across the plasma membrane.
This figure was adapted from Alberts et a.,2008, and generated in Powerpoint.

Secondary Active Transport

In comparison to primary active transporters, secondary active transporters translocate ions and molecules across the lipid bilayer, without the requirement of ATP hydrolysis/synthesis. Instead, carriers specifically couple the thermodynamically favourable transmission of one molecular species (usually ions) down its electrochemical gradient, to generate an energetic "boost" required for the thermodynamically unfavourable "uphill" transport of molecular substrates against their electrochemical gradients (Tanford, 1983; Krishnamurthy et al., 2009). This form of molecular transport is very common among most species, functioning in either direction across the biological cell membrane. This has extreme beneficial implications for the cell, since the combined workings of an ionic and solute translocation results in an increased concentrated transfer of solute across a membrane by 10^6

fold and occurring 10^5 times faster, than translocation by simple diffusion (i.e. diffusion across the cell membrane unassisted by integral or peripheral membrane proteins) (Krishnamurthy et al., 2009).

There are typically three types of classifications of secondary active transporters: antiporters, symporters and uniporters; categorised by their vectorial mechanism of transport (i.e. the directional mode of translocation of solutes and ions). Antiporters (also known as exchangers) couple the "uphill" transfer of one species in one direction, to the downhill flow of another, in the opposite direction, across the plasma membrane. Symporters (also known as co-transporters) work in the same way, but transfer both sets of species across the lipid bilayer in the same direction. Uniporters are specific molecular transporters, able to transfer specialised species in either direction, across the plasma membrane, executed solely by differences in its own substrate concentrations, determined by Vmax and Km rates (Forrest et al., 2011; Perland and Fredriksson, 2017; Beckstein and Naughton, 2021). Figure 1.9 below illustrates these 3 classes in further detail.

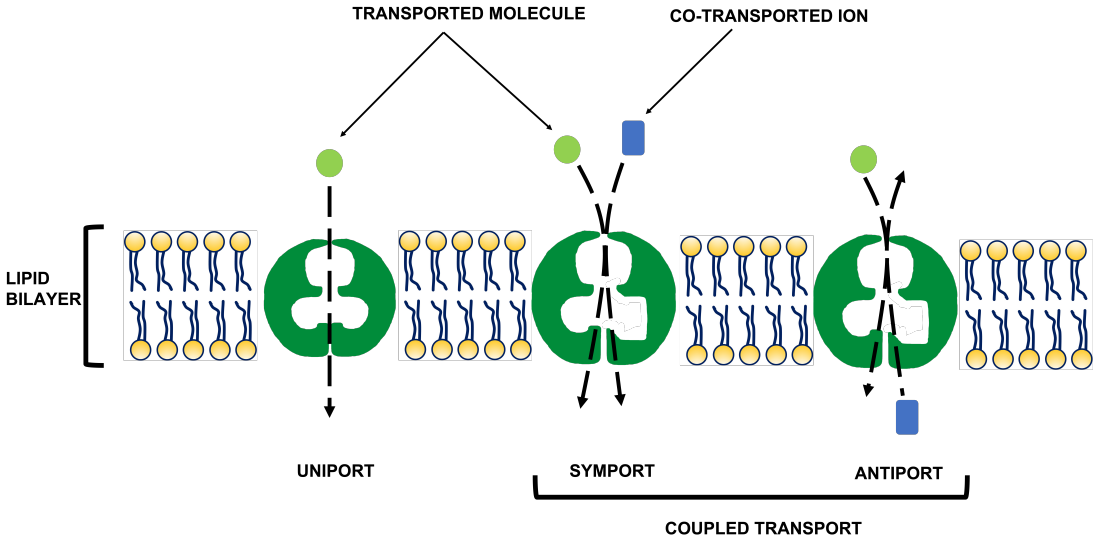


Figure 1.9: The three types of secondary active transport. Uniporters (left) transport substrates in both directions, across the cell membrane, in an active manner executed solely by the substrates own electrochemical gradient. Both symporters (middle) and antiporters (right) require the energy generated by the favourable coupled transport of ions, to translocate substrates across membranes in the same or opposite directions, respectively.
This figure was adapted from Alberts et al., 2008, and generated in Powerpoint.

This mechanism of transport utilises many large structural conformational changes, which limits the maximal rate of transmission to just a few molecules per transport cycle. Interestingly, in animal cells, Na^+ is usually the preferred co-transported ion, its electrochemical gradient serving as the major net driving force for the active transfer of coupled substrates. Furthermore, the Na^+ which subsequently enters the cell, via secondary active transport, is rapidly pumped

back out again via a P-type ATPase Na^+ pump, in order to regulate the voltage gradient across the plasma membrane. Indirectly, by maintaining the Na^+ gradient across the cell membrane, this further drives the secondary active transport and the cycle repeats itself. Unsurprisingly, the greater the Na^+ electrochemical gradient, the greater the rate of substrate transfer across the cell membrane (Boudker and Verdon, 2010; Shi, 2013; Bosshart and Fotiadis, 2019).

A summary of the different passive and active solute translocation processes observed above, can be summarised again below in figure 1.10.

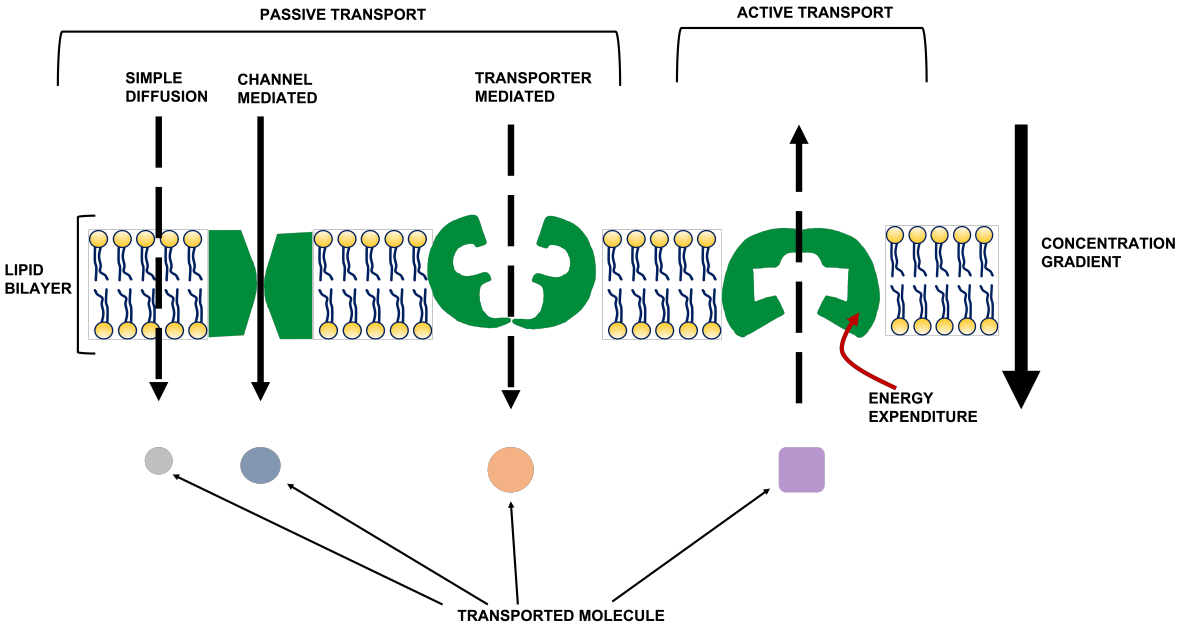


Figure 1.10: A basic illustration highlighting the differences in solute and ion translocation across the cell membrane, in both passive and active transport systems.
This figure was generated in Powerpoint.

1.4 The Alternating Access Mechanism

Since almost 30 % of all genome sequences encode membrane proteins, and with an ever increasing number of diseases associated, investigation of the protein structures, and better understanding of specific translocation mechanisms, via the alternating access mechanism, are of the utmost importance (Wallin and von Heijne, 1998).

Generally, the transport of solutes across a biological membrane, can be explained and the conditions satisfied, by the alternating access mechanism, first proposed by Oleg Jardetzky in 1966. In this model, four conditions must be met for a protein to be termed a "transporter", as outlined below by (Jardetzky, 1966; Tanford, 1983):

1. The transporter protein must have at least two different conformational arrangements, such that one side is open (to allow for solute binding) and the opposite side closed, to ensure that no two sites are accessible from both sides of the membrane simultaneously.

2. The transporter protein must contain a binding site for the solute to bind and be transported across the biological plasma membrane. The thermodynamic binding affinity of this solute-bound site at the uptake location of the membrane must be greater than that of the discharge site.

3. In order for transport of the solute to occur, a conformational change in the transporter structure must follow, allowing for the movement of the binding site itself (with the substrate still bound).

4. A mechanism must be available for the transporter to return to its original native state, with no substrate bound. Therefore, for one successful cycle of this mechanism to occur, a conformation transition of the transporter must occur twice: once with the substrate bound, and once without.

Comprehending how specialised ligands bind to intricately formed membrane transporters, and thereby assessing in detail how this can lead to membrane protein structural conformational changes, allowing for the discharge of the transported ligand to the opposite side of the membrane, is pivotal in order to understand the alternating access mechanism of membrane transporters and their contribution to devastating diseases (Jardetzky, 1966; Tanford, 1983).

1.5 The Solute Carrier Superfamily

Within humans, the solute carrier (SLC) superfamily entails membrane transport proteins from over 450 members, categorised and allocated into 65 families (SLC1 - SLC65). This is a rapidly expanding superfamily, comprising of members with fascinating diversity, responsible for the transport of a multitude of substrates, both charged and uncharged as well as organic and inorganic. Organisation of these members is achieved via amino acid sequence identity, whereby only members sharing 20-25 % are grouped together within the same family. Although this grouping relies on low percentage sequence identity between members, typically these will express similar biological functions, with conserved amino acid residues, suggesting this grouping to provide significant accuracy (Hediger et al., 2004).

The entire SLC superfamily consists of membrane transporters which undergo either passive/facilitative or secondary active transport, via ion-coupled symport and antiport processes. Primary active transporters as well as ion or aquaporin channels are not associated with any grouping of this SLC superfamily (Geyer et al., 2006). An up-to-date list of all 65 SLC families (as of September 2021) is detailed in table 1.1 below.

Table 1.1: The 65 human SLC families. Adapted from http://slc.bioparadigms.org/

Family	**Family Name**	**Total Members**
SLC1	High-affinity glutamate and neutral amino acid transporter family	7
SLC2	Facilitative GLUT transporter family	18
SLC3	Heavy subunits of the heteromeric amino acid transporters	2
SLC4	Bicarbonate transporter family	10
SLC5	Sodium glucose cotransporter family	12
SLC6	Sodium-and-chloride-dependent neurotransmitter transporter family	22
SLC7	Cationic amino acid transporter/glycoprotein-associated family	15
SLC8	Na^+/Ca^{2+} exchanger family	4
SLC9	Na^+/H^+ exchanger family	18
SLC10	Sodium bile salt cotransport family	7
SLC11	Proton-coupled metal ion transporter family	2
SLC12	Electroneutral cation-coupled Cl^- cotransporter family	9
SLC13	Human Na^+-sulfate/carboxylate cotransporter family	5
SLC14	Urea transporter family	8
SLC15	Proton oligopeptide cotransporter family	4
SLC16	Monocarboxylate transporter family	14
SLC17	Vesicular glutamate transporter family	9
SLC18	Vesicular amine transporter family	4
SLC19	Folate/thiamine transporter family	3
SLC20	Type III Na^+-phosphate cotransporter family	2
SLC21	Organic anion transporter family	11

SLC22	Organic cation/anion/zwitterion transporter family	29
SLC23	Na^+-dependent ascorbic acid transporter family	4
SLC24	$Na^+/(Ca^{2+}$-$K^+)$ exchanger family	5
SLC25	Mitochondrial carrier family	60
SLC26	Multifunctional anion exchanger family	11
SLC27	Fatty acid transporter family	6
SLC28	Na^+-coupled nucleoside transport family	3
SLC29	Facilitative nucleoside transporter family	4
SLC30	Zinc efflux family	10
SLC31	Copper transporter family	3
SLC32	Vesicular inhibitory amino acid transporter family	1
SLC33	Acetyl-CoA transporter family	2
SLC34	Type II Na^+-phosphate cotransporter family	3
SLC35	Nucleoside-sugar transporter family	31
SLC36	Proton-coupled amino acid transporter family	4
SLC37	Sugar-phosphate/phosphate exchanger family	4
SLC38	System A and System N sodium-coupled neutral amino acid transporter family	11
SLC39	Metal ion transporter family	14
SLC40	Basolateral iron transporter family	1
SLC41	MgtE-like magnesium transporter family	3
SLC42	Rh ammonium transporter family	3
SLC43	Na^+-independent, system-L-like amino acid transporter family	3
SLC44	Choline-like transporter family	5
SLC45	H^+/sugar cotransporter family	4
SLC46	Folate transporter family	3
SLC47	Multidrug and Toxin Extrusion (MATE) family	2
SLC48	Heme transporter family	1
SLC49	FLVCR-related transporter family	4
SLC50	Sugar efflux transporters	1
SLC51	Transporters of steroid-derived molecules	2
SLC52	Riboflavin transporter family	3
SLC53	Phosphate carriers	1
SLC54	Mitochondrial pyruvate carriers	3
SLC55	Mitochondrial cation/proton exchangers	3
SLC56	Sideroflexins	5
SLC57	NiPA-like magnesium transporter family	6
SLC58	MagT-like magnesium transporter family	2
SLC59	Sodium-dependent lysophosphatidylcholine symporter family	2
SLC60	Glucose transporters	2
SLC61	Molybdate transporter family	1
SLC62	Pyrophosphate transporters	1
SLC63	Sphingosine-phosphate transporters	3
SLC64	Golgi Ca^{2+}/H^+ exchangers	1
SLC65	NPC-type cholesterol transporters	2
Total		**458**

1.5.1 The SLC10 Protein Family

A large sub-family of the SLC superfamily, is the sodium bile acid cotransport family, SLC10. This family name is misleading, as certain members translocate solutes other than bile acids, including the transport of steroidal hormones, drugs, and several unknown solutes (Claro da Silva et al., 2013; Geyer et al., 2006). Exclusively operating as influx transporters, more than 50 members have been discovered, among multiple organisms, including animals, plants and bacterial species (Hagenbuch and Dawson, 2004). In humans, seven members have been identified (SLC10A1 - SLC10A7), which differ substantially from one another, in both functional and structural features (Hagenbuch and Dawson, 2004). Their individual substrate specificities, cellular expressions and gene loci are highlighted below in table 1.2 below.

Table 1.2: The seven human members of the SLC10 protein family.
N/A: Not Available, C: Cotransporter, O: Orphan Transporter
Table adapted from (Hagenbuch and Dawson, 2004; Geyer et al., 2006; Döring et al., 2012; Claro da Silva et al., 2013)

Family Member	**Protein**	**Substrates**	**Transport and Coupling Ions**	**Tissue Expression**	**Gene Locus**	**References**
SLC10A1	NTCP LBAT	Bile acids Steroids Xenobiotics HMG-CoA inhibitors	C/Na^+	Liver Pancreas	14q24.1	Hagenbuch and Meier, 1994
SLC10A2	ASBT IBAT ISBT NTCP2	Bile acids	C/Na^+	Ileum Kidney Biliary tract	13q33	Craddock et al., 1998
SLC10A3	P3	N/A	O	Ubiquitous	Xq28	Geyer et al., 2006
SLC10A4	P4	N/A	O	Brain Placenta Liver	4p12	Geyer et al., 2008 Schmidt et al., 2015
SLC10A5	P5	N/A	O	Liver Kidney	8q21.13	Fernandes et al., 2007
SLC10A6	SOAT	Estrone-3-sulfate Pregnenolone-sulfate Dehydroepiandrosterone-sulfate	C/Na^+	Testis Placenta Pancreas	4q21.3	Geyer et al., 2007
SLC10A7	C4orf13 P7	N/A	O	Liver Testis	4q31.22	Zou et al., 2005 Godoy et al., 2007

Functional roles of certain SLC10 members have been extensively and conclusively investigated (including NTCP, ASBT, and SOAT), whilst other members (including SLC10A3, SLC10A4, SLC10A5 and SLC10A7) have merely had their functional roles hypothesised and suggested, with no additional evidence provided to support identification of novel substrates involved in translocation. The human biological functional attributes for all members of the SLC10 family are outlined in greater detail below:

SLC10A1 / NTCP

Exclusively expressed at the basolateral sinusoidal lipid bilayer of hepatocytes, the Na^+/taurocholate cotransporting polypeptide (NTCP) serves a pivotal role in the regulation of the enterohepatic circulation (EHC) of bile acids (BAs) in the liver (Kullak-Ublick et al., 2000). BAs are extracted from the portal blood, via NTCP, and re-secreted into the canalicular membrane for the cycle to recommence. Uptake is sodium activated with a stoichiometry of 2 Na^+ ions for every BA molecule (Weinman, 1997). Evidence has shown that NTCP provides most, if not all, transport of BAs across the hepatocyte, highlighting its essential participation in BA homeostasis (Hagenbuch and Meier, 1994; Kullak-Ublick et al., 2004). NTCP also exhibits transport of other substrates, including steroid sulfates, such as oestrone-3-sulfate, but also provides a mechanism of entry for hepatitis B and D viruses (Yan et al., 2012).

SLC10A2 / ASBT

The apical sodium-dependent bile acid transporter (ASBT), expressed at the apical membranes of ileal enterocytes, serves to recycle BAs from the intestinal lumen, allowing for their transport to the liver via the portal circulation. In similar fashion to NTCP, ASBT translocates BA molecules in a sodium-dependent manner (Craddock et al., 1998), in a 2:1 Na^+/BA stoichiometry (Weinman et al., 1998). Functioning as a BA salvation mechanism, ASBT plays a central role in the EHC in the intestine, as has been outlined by the consequences of primary bile acid malabsorption, interruption of the enterohepatic BA circulation, and lowered low density lipoprotein (LDL) cholesterol levels, upon loss-of-function mutations of the membrane protein (Wong et al., 1995; Oelkers et al., 1997).

SLC10A3 / P3

The function of SLC10A3 is currently unknown, even though it is expressed ubiquitously in humans and shares high conservation across many animal species (Hagenbuch and Dawson, 2004). It is believed however, to be involved in housekeeping functions, due to the presence of an unmethylated CpG island (Geyer et al., 2006). Even though SLC10A3 shares at least 25 % sequence identity with both ASBT and NTCP, no transport of bile acids is identified.

SLC10A4 / P4

A study in 2013 (Abe et al., 2013) showed SLC10A4 to be able to transport bile acids, but only upon cleavage of part of its N-terminal, thereby acting as a protease activated bile acid transporter. However, the majority of SLC10A4 associated research suggests its participation in the storage of vesicles and/or the exocytosis of neurotransmitters, highlighting its dominance in the nervous tissue (Geyer et al., 2008; Schmidt et al., 2015). In mice, SLC10A4 co-localises with synaptic vesicle transporters involved with dopamine and acetylcholine, highlighting SLC10A4 as a potential role in cholinergic systems (Zelano et al., 2013; Melief et al., 2016). In light of this, SLC10A4 could provide a crucial target for the interventions of both Alzheimer's and Parkinson's disease (Popova and Alafuzoff, 2013).

SLC10A5 / P5

SLC10A5 is an orphan transporter, with no known substrate. Although sharing similar tissue expression sites to ASBT and NTCP (in both the liver and kidneys), SLC10A5 fails to transport any of the known substrates associated to the SLC10 family, including bile acids and steroidal hormones (Fernandes et al., 2007).

SLC10A6 / SOAT

The sodium-dependent organic anion transporter (SOAT) is strictly not a bile acid transporter, but instead translocates specialised steroidsulfates including oestrone-3-sulfate and dehydroepiandrosterone sulfate (Geyer et al., 2004, 2007). Like both NTCP and ASBT, this transporter also operates via the coupled transport of Na^+ ions. Involved in the production of androgens (such as oestrogen) as well as the synthesis of progesterone, SOAT has been shown to play critical physiological functions in the cellular delivery of the above mentioned pro-hormonesd to the testis, adrenal gland, placenta and probably peripheral tissues (Geyer et al., 2006; Claro da Silva et al., 2013).

SLC10A7 / P7

SLC10A7, much like SLC10A3, SLC10A4 and SLC10A5, is also an orphan solute carrier transporter. There have been no substrates associated as of yet (September 2021), but recent research suggests an involvement in the negative regulation of intracellular calcium signalling amongst humans and yeast (Jiang et al., 2012; Alber et al., 2013; Zhao et al., 2016; Karakus et al., 2020). This also correlates with other research, where an SLC10A7 knockout mutant among humans, results in a drastic disturbance in calcium homeostasis, leading to skeletal dysplasias, amelogenesis imperfecta and a disruption in the mediation of GAG biosynthesis (Dubail et al., 2018; Ashikov et al., 2018; Laugel-Haushalter et al., 2019).

As has been explored above, substantial functional and physiological differences are present among the seven human SLC10 members (Claro da Silva et al., 2013). To examine these further, an amino acid sequence alignment plot, depicted in figure 1.11 below, displays detailed comparisons of these members, as well as the previously crystallised bacterial $ASBT_{NM}$ (Hu et al., 2011) and $ASBT_{Yf}$ (Zhou et al., 2014) homologues.

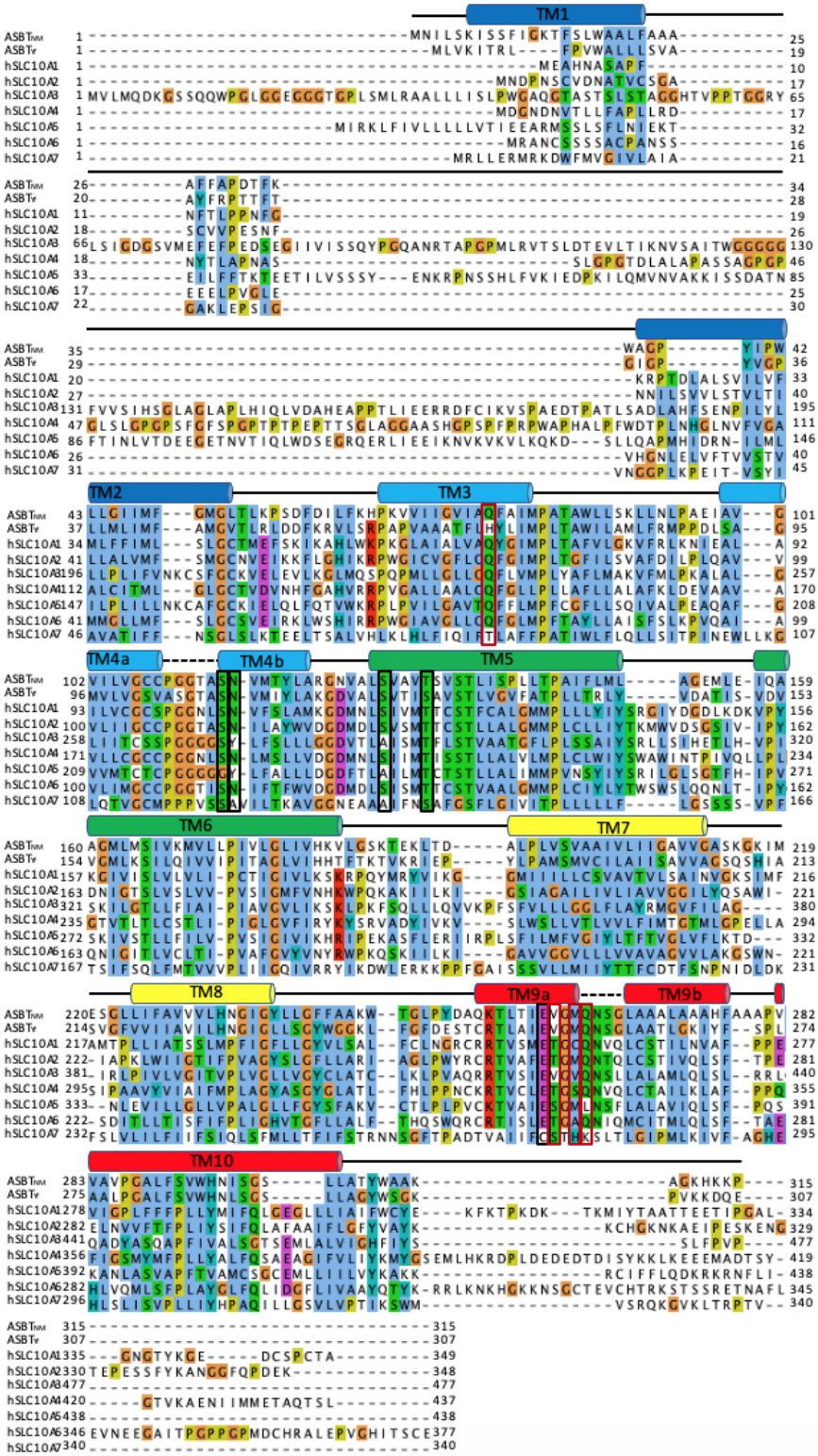


Figure 1.11: The seven SLC10 human members and $ASBT_{NM}$ and $ASBT_{Yf}$ bacterial homologue multiple sequence alignment. Respective amino acid sequences were analysed in Jalview (Waterhouse et al., 2009), and specifically aligned with Clustalx (Larkin et al., 2007) and MuscleWS (Edgar, 2004). The established $ASBT_{NM}$ transmembrane (TM) domains (as previously reported by Hu et al., 2011) are presented above the sequences in cylinders, depicting the structural TM1-TM10 domains from the N-terminal (blue) to C-terminal (red). Na^+ binding site amino acid residues, implicated for Na1 and Na2, are depicted in red and black boxes, respectively.

Clustalx (Larkin et al., 2007) notation of residues: Blue: Hydrophobic, Red: Positive charge, Magenta: Negative charge, Green: Polar, Pink: Cysteines, Orange: Glycines, Yellow: Prolines, Cyan: Aromatic, and White: Unconserved.

As can be seen in figure 1.11 above, overall, the analysed human and bacterial homologue SLC10 members share good amino acid residue conservation, although key important differences arise within the human SLC10A7 member which will be discussed in greater detail below.

This PhD study specifically focusses on the bacterial homologues of two members of the SLC10 family: ASBT (SLC10A2) and SLC10A7. A greater analysis of their physiological and functional roles in humans are explored below.

1.6 The Human Apical Sodium Dependent Bile Acid Transporter (ASBT) - SLC10A2

As described above, the sodium taurocholate co-transporting polypeptide (NTCP - SLC10A1 gene) and apical sodium dependent bile acid transporter (ASBT - SLC10A2 gene) are well-established members of the SLC10 protein family in humans. Together they transport and regulate bile acids (BAs) in the enterohepatic circulation (EHC) pathway, and are primarily involved in their translocation across the liver and small intestine tissues. In-depth analysis of this mechanistic pathway and intricate role in maintaining BA homeostasis is portrayed below.

1.6.1 Bile Acids

BAs are classified as negatively charged, stable, rigid and planar amphipathic (consisting of both hydrophilic and hydrophobic features) signalling molecules, which carry out essential regulatory roles in energy expenditure, inflammation, glucose and cholesterol/lipid homeostasis (Hofmann and Hagey, 2008; Dawson, 2012).

At low concentrations, these BAs exist in monomeric states, but form simple micelles when high intra-luminal concentrations exceed the critical micellar concentration (CMC) threshold (2-20 mM, depending on the BA species). These micelles undergo solubilisation of large cholesterol molecules (via the incorporation into their hydrophobic core), resulting in a total diameter of 5 Å (Hofmann, 2009; Slijepcevic and van de Graaf, 2017). When excreted into bile, BAs can further associate with lipid bilayers to form mixed micelles, containing phosphatidylcholine, partially ionised fatty acids and monoglycerides; resulting in an increase in their total micellar diameter to 30 Å (Slijepcevic and van de Graaf, 2017). In comparison to simple BA micelles, these significantly larger mixed micelles provide a three-fold greater solubilisation, digestion and transportation ability of dietary lipid molecules (including fatty acids, cholesterol, fat-soluble vitamins and monoglycerides), in the enterocytes of the small intestine, for their subsequent degradation and excretion in faeces (Hofmann, 2009; Dawson, 2012; Slijepcevic and van de Graaf, 2017; Baiocchi et al., 2019).

Once synthesised, BAs are secreted into bile fluid in order to circulate the body and solubilise fatty dietary molecules. In addition to conjugated and unconjugated BAs, bile fluid also consists of a multitude of compounds, including organic substrates, cholesterol, plasma electrolytes, glucose, water, phospholipids, calcium, bicarbonate, xenobiotics, conjugated bilirubin, amino acids, conjugates of metabolites, glutathione and other low molecular weight substrates. Its primary

role is to excrete and eliminate heavy metals, toxic/exogenous and endogenous compounds (e.g. bilirubin), and drug metabolites (Dawson, 2012; Slijepcevic and van de Graaf, 2017).

In total, BAs have shown to be involved in six fundamental biological functional processes, as outlined below (Hofmann, 2009; Dawson, 2012; Baiocchi et al., 2019):

1. BAs stimulate the release of biliary lipids (such as cholesterol and phospholipids) and maintain the canalicular bile flow;

2. BAs aid in the metabolisation, absorption and digestion of dietary fats, including fatty vitamins (such as vitamin A, D, E and K) and cholesterol, which result in the production of mixed micelles, to facilitate the aqueous solubility in the lumen of the gut and increase diffusion and absorption across the intestinal epithelium;

3. BAs regulate cholesterol homeostasis, whereby cholesterol from the diet is metabolised, and its products solubilised and absorbed through BAs in the intestines. Cholesterol molecules are then expelled through the faecal route of the body. Since BAs are the amphipathic by-products of the catabolism of cholesterol, increasing its elimination in the faeces, increases the biosynthesis process to replenish lost BAs, which thereby raises the rate of conversion and elimination of cholesterol;

4. BAs promote direct and indirect antimicrobial protection of the gut, via the bacteriostatic features of the BA fatty acid mixed micelles, as well as the stimulation of antimicrobial genes and transcription factors in the small intestine;

5. BAs inhibit the production of calcium gall- and kidney-stone enucleating precipitates;

6. BAs feature as hormonal signalling molecules, to maintain the functionality of hepatocytes; regulating energy, fat and glucose homeostasis; modulating the EHC; and sustaining gut motility, via the specific interactions with G-protein-coupled receptors (such as the formyl-peptide receptors, Gpbar1/TGR5, Gαi-coupled receptors, and muscarinic receptors), nuclear receptors (such as the pregane X receptor (PXR); vitamin D receptor (VDR); and farnesoid receptor X (FXR), and cell-signalling pathways (such as the serine/threonine protein kinase (AKT/PKB), protein kinase C (PKC), p38 mitogen-activated protein kinase (p38 MAPK), c-jun N-terminal kinase 1/2 (JNK 1/2), epidermal growth factor receptor (EGFR), and extracellular signal-regulated kinase (ERK)).

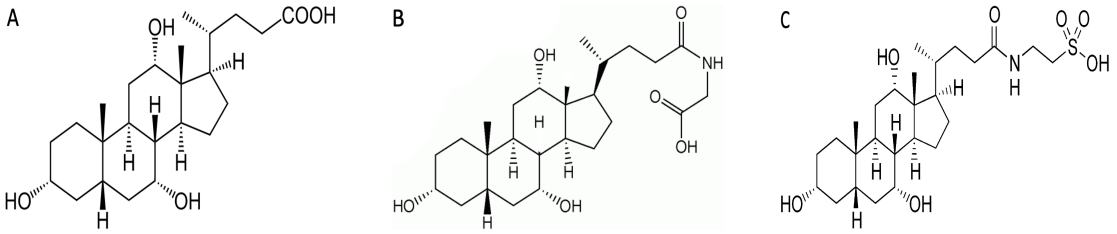
In addition to the above mentioned functional attributes of BAs, in the biliary epithelium, key roles of bile acids in differentiation, apoptosis, proliferation and secretion have also been previously reported (Baiocchi et al., 2019).

1.6.2 Bile Acid Biosynthesis and the Catabolism of Cholesterol

In the human liver pericentral hepatocytes, the catabolism of highly hydrophobic cholesterol molecules, via a multitude of 17 complex enzymatic reactions (consisting of multiple oxidation and conjugation reactions, including the oxidation of side chains and hydroxylation of the sterol ring), result in the production of amphipathic, water-soluble BAs. Specifically, this biosynthesis is established via two separate complex pathways, including (1) the classical/neutral (cholesterol 7α-hydroxylase) and (2) the alternative/acidic (oxysterol 7α-hydroxylase) pathways. Key differences arise in the initial cholesterol transformation reactions, whereby the classical pathway targets the transformation of the cholesterol sterol ring to 7α-hydroxycholesterol, via the rate-limiting hepatic cholesterol 7α-hydroxylase (CYP7A1) enzyme; whereas the alternative pathway targets the transformation of cholesterol side chains via the C-24, C-25 and C-27 sterol hydroxylase enzymes, and oxysterol 7α-hydroxylation in the liver. After complex biochemical transformations, the classical pathway, accounting for >90 % of total BA production, predominantly produces cholic acid (CA); whilst the alternative pathway favours the production of chenodeoxycholic acid (CDCA) (Dawson, 2012; Slijepcevic and van de Graaf, 2017; Xiao and Pan, 2017).

Once produced in the liver, these BAs become immediately conjugated to either glycine or taurine, via N-acyl amidation. This conjugation significantly increases the acidic strength of BAs, as well as their hydrophilic polar attributes, rendering them highly impermeable to the epithelium, and hence less passive. Furthermore, complete ionisation occurs in both the small intestine and biliary tracts, ultimately leading to increased resistance and tolerance to calcium induced aggregation and precipitation in the gall bladder, preventing the formation of gallstones. Finally, this conjugation also allows for the regulation of high intraluminal BA concentrations, thereby facilitating the solubilisation of cholesterol molecules and digestion of fatty lipids in the small intestine (Balakrishnan and Polli, 2006; Dawson, 2012; Slijepcevic and van de Graaf, 2017).

Since the CA and CDCA BA products were directly formed as a result of the catabolism of cholesterol, they are classified as primary BAs. Chemical structures of these primary BAs, including their conjugated glycine/taurine forms, are depicted below in figure 1.12.



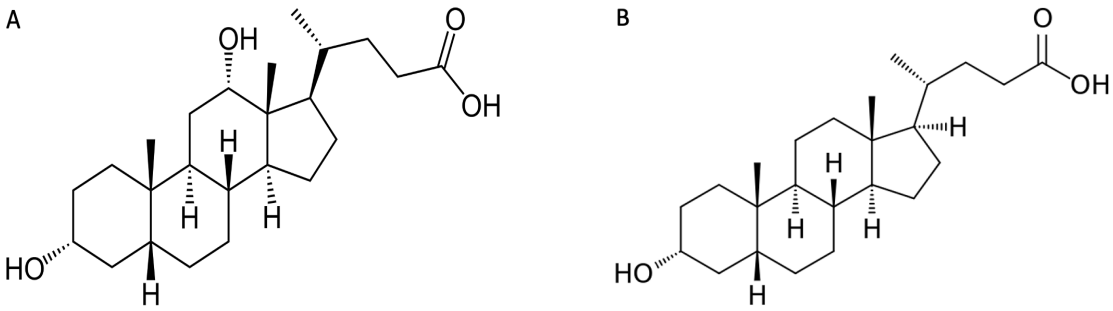
Cholic Acid (CA) | Glycocholic Acid (GCA) | Taurocholic Acid (TCA)

Chenodeoxycholic Acid (CDCA) | Glycochenodeoxycholic Acid (GCDCA) | Taurochenodeoxycholic Acid (TCDCA)

Figure 1.12: Chemical structures of the cholic acid (CA) and chenodeoxycholic acid (CDCA) primary bile acids, in their unconjugated (A and D) and conjugated glycine (B and E) and taurine (C and F) states.

Interestingly, primary BA species residing in the gut can subsequently become de-conjugated, as a result of specific interactions from endogenous bacteria. In light of these structural alterations, BA hydrophobicity levels are increased, significantly reducing their aqueous solubility and affecting their uptake and flux across the intestinal epithelium. As a result, increases in the excretion rate of BAs in faecal waste are observed. Furthermore, these de-conjugated BAs, from the colon and ileum, heavily impact on the composition of the total circulating BA pool, having considerable effects on cellular toxicity, detergent features and hormonal signalling abilities. In order to prevent such drastic alterations to the human physiological conditions, these newly modified un-conjugated BA members are actively re-absorbed in the small intestine and transferred to the liver for re-conjugation, before being expelled to the bile flow once again for circulation and solubilisation of fatty molecules (Dawson, 2012).

The products formed from these BA bacterial modifications are termed secondary BAs, and consist of the drastic structural modifications (including epimerization (α-hydroxyl group becomes a β-hydroxyl group or vice versa); dehydroxylation; and dehydrogenation (formation of an oxo group via the oxidation of a hydroxy group)) of circulating primary BAs. In humans, CA is transformed to deoxycholic acid (DCA), and CDCA is transformed to lithocholic acid (LCA), via 7α-dehydroxylation mechanisms (Balakrishnan and Polli, 2006; Dawson, 2012). Both the chemical structures of these secondary bile acids are portrayed in figure 1.13 below.



Deoxycholic Acid (DCA) Lithocholic Acid (LCA)

Figure 1.13: Chemical structures of the deoxycholic acid (DCA) and lithocholic acid (CDA) secondary bile acids.

1.6.3 The Enterohepatic Circulation of Bile Acids in Humans

BAs produced as a result of the catabolism of cholesterol, undergo an efficiently regulated recycling process know as the enterohepatic circulation (EHC). BAs are first secreted across the hepatic canalicular membrane into bile, via the bile salt export pump (BSEP - ABCB11 gene), before being transported to the terminal ileum of the small intestine. Here they are re-absorbed by the apical sodium-dependent bile acid transporter (ASBT - SLC10A2 gene), and bound to the ileal lipid binding protein (ILBP - FABP6 gene), for their transcellular transport across the ileal enterocyte. These BAs are shuttled and excreted to the portal circulation at the ileal basolateral membrane via the heterodimeric organic solute transporter α/β (OSTα/β - SLC51A/SLC51B genes). Travelling through the portal circulation, BAs are then reabsorbed at the hepatic basolateral/sinusoidal membrane, via the sodium-dependent taurocholate co-transporting polypeptide (NTCP - SLC10A1 gene) and the sodium-independent organic anion transporting polypeptides (OATP-A - SLC21A3 gene, OATP-C - SLC21A6 gene; and OATP8 - SLC21A8 gene) transporters, where the cycle then recommences (Lin et al., 1990; Kramer et al., 1993; Oelkers and Dawson, 1995; Kramer et al., 1997; Ballatori, 2005; Ballatori et al., 2005; Dawson et al., 2005; Balakrishnan and Polli, 2006; Dawson et al., 2009; Dawson, 2012; Slijepcevic and van de Graaf, 2017; Baiocchi et al., 2019).

When fasting, circulating BAs in the portal vein, small intestine, liver and serum are all significantly reduced, due to the seclusion of roughly half the total pool of BAs to the gallbladder, for storage. Upon digestion of a meal, these stored BAs are released into the blood circulation, thereby initiation the BA EHC mechanism, in order to solubilise and digest fat-soluble dietary molecules (including vitamins, cholesterol and lipids) in the small intestine (Hofmann and Hagey,

2008; Dawson, 2011).

Fascinatingly, >95 % of the total pool of BAs are rescued and recycled, via the ASBT transporter, from the intestine to the liver, through the venous portal circulatory system. This highly efficient EHC system, ensures the conservation of the BA pool to reside between 3-5 g in humans. Undergoing several rounds of cycling every day, it is estimated that in the intestines of humans alone, up to 30 g of bile acids are reabsorbed, with each BA molecule being estimated to be salvaged by ASBT 20 ×, before its expulsion in faeces (Hofmann et al., 1983; Hulzebos et al., 2001; Balakrishnan and Polli, 2006; Dawson, 2011; Chothe et al., 2018; Xiao and Pan, 2017).

Amazingly, through this tightly regulated and efficient BA recycling system, only 3-5 % (accounting for <0.5 g per day) fail to be reabsorbed by the ASBT transporter of the terminal ileum and are instead excreted in faeces. Furthermore, in hepatocytes, due to some leakiness of the system, 10-50 % of BAs (dependent on species) can evade extraction, and escape into the systemic circulation. However, the glomerular filtration system of kidneys are tasked to rescue these aberrant molecules, and are reabsorbed by the Na^+-dependent coupled transport of ASBT, expressed in renal tubules. As a result, only 1-2 µmol is eliminated in urine. Finally, any BA loss in either urine or faeces is efficiently accounted for, by the neo-synthetic conversion and elimination mechanisms of cholesterol in the liver, presenting the primary path of cholesterol extrusion from the body (Wilson et al., 1981; Dietschy et al., 1993; Dietschy and Turley, 2002; Balakrishnan and Polli, 2006; Dawson, 2011; Slijepcevic and van de Graaf, 2017; Baiocchi et al., 2019).

To summarise, this highly regulated BA recycling process, whilst minimising loss through both urinary and faecal routes, also restricts the cytotoxic accumulation of BAs in both hepatocytes and intestine enterocytes. An illustrative overview of this impressive recycling mechanism of BAs in humans is presented in figure 1.14 below.

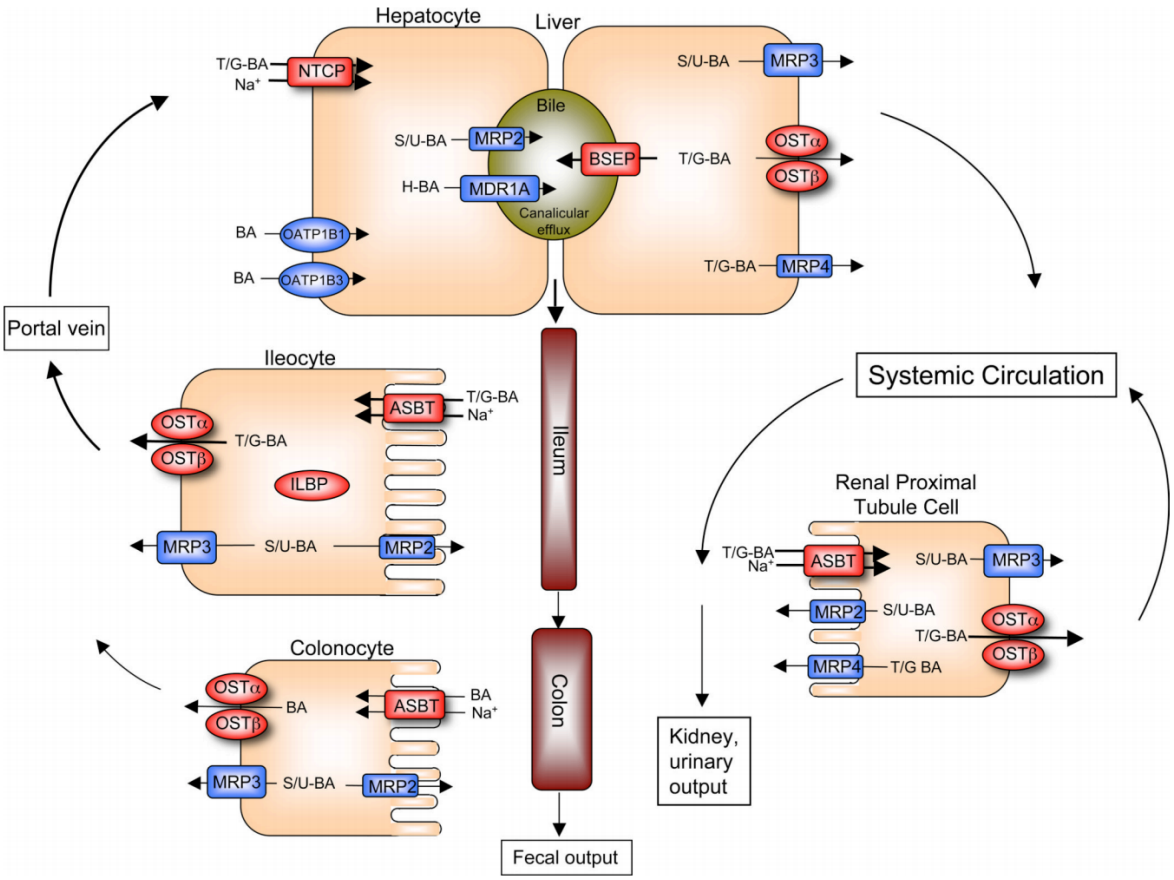


Figure 1.14: The enterohepatic circulation of bile acids in the hepatocytes, ileocytes and renal proximal tubule cells in humans. Synthesised glycine/taurine conjugated BAs are secreted to the bile fluid via BSEP at the canalicular membrane. Minor sulfate/glucuronide modified BAs and additionally hydroxylated BAs are secreted by the multidrug resistance-associated protein-2 (MRP2) and multidrug resistance protein 1 (MDR1), respectively. The bile flows to the small intestine, where BAs are then re-absorbed by ABST and bind to cytosolic ILBP. The BAs are then shuttled to the basolateral membrane and exported into the portal circulation via OSTα-OSTβ. The mutidrug resistance-associated protein 3 (MRP3) exports glucuronidated or sulfated BAs, as does MRP2. Transported BAs then travel back to the liver tissue, and are re-absorbed by NTCP. OATP1B1 and OATP1B3 facilitate the uptake of unconjugated BAs.
Under cholestatic conditions, all forms of BAs can be additionally exported via OSTα-OSTβ, MRP3 and MRP4, to the systemic circulation, reaching the renal tubular cells. These BAs are filtered in the renal glomerulus and salvaged by ASBT, to be reintroduced to the systemic circulation. Failure to reabsorb BAs in ileal and renal cells, results in their expulsion from the body in faeces and urine.
This image was taken from Dawson, 2011.

1.6.4 The Negative Feedback Regulation of the Biosynthesis of Bile Acids

In order to maintain normal BA homeostasis levels, regulation of the biosynthesis of BAs, and their circulation from the liver to small intestines and back again, must be accounted for. In the instances when excessive BAs are in circulation and/or production, an essential regulatory feedback system must be available, for its successful biosynthesis depletion and removal from the body (Xiao and Pan, 2017).

Fascinatingly, BAs possess the ability to regulate their own biosynthesis, via an efficient negative feedback signalling mechanism. Both unconjugated/conjugated BAs stimulate the intrahepatic bile sensing farnesoid receptor X (FXR) on ileal enterocytes, thereby activating the production of fibroblast growth factor 19 (FGF19). The latter is extruded from the enterocyte and transported to the hepatocyte, initiating a cascade of signalling events via the stimulation of the fibroblast growth factor receptor-4/β-Klotho (FGFR4/β-Klotho) complex. The outcome of multiple signalling events leads to the inhibition of the expression of CYP7A1, known to be a rate limiting enzyme in the biosynthesis of BAs. In addition, other negative feedback targets include the BA induced modulation of the CYP8B1 enzyme expression, resulting in significantly reduced production levels of the primary bile acid, CA (Dawson, 2012; Slijepcevic and van de Graaf, 2017; Baiocchi et al., 2019).

Elsewhere, expression levels of the human ASBT transporter in intestines, can also be targeted and downregulated, via a separate negative feedback mechanism, induced by specific transcription factors. The FXR receptor again activates the FGF19 (FGF15 in mice) signalling molecule to bind and activate the FGFR4/β-Klotho complex, resulting in the stabilisation and upregulation of the short heterodimer partner (SHP). This SHP stimulation results in the repression of both the retinoic acid receptor α (RARα) and liver receptor homolog-1 (LRH-1) genes, ultimately leading to de-activation of ASBT expression levels (Dawson, 2012; Slijepcevic and van de Graaf, 2017).

An illustrative overview of this negative feedback mechanism in humans is represented below in figure 1.15.

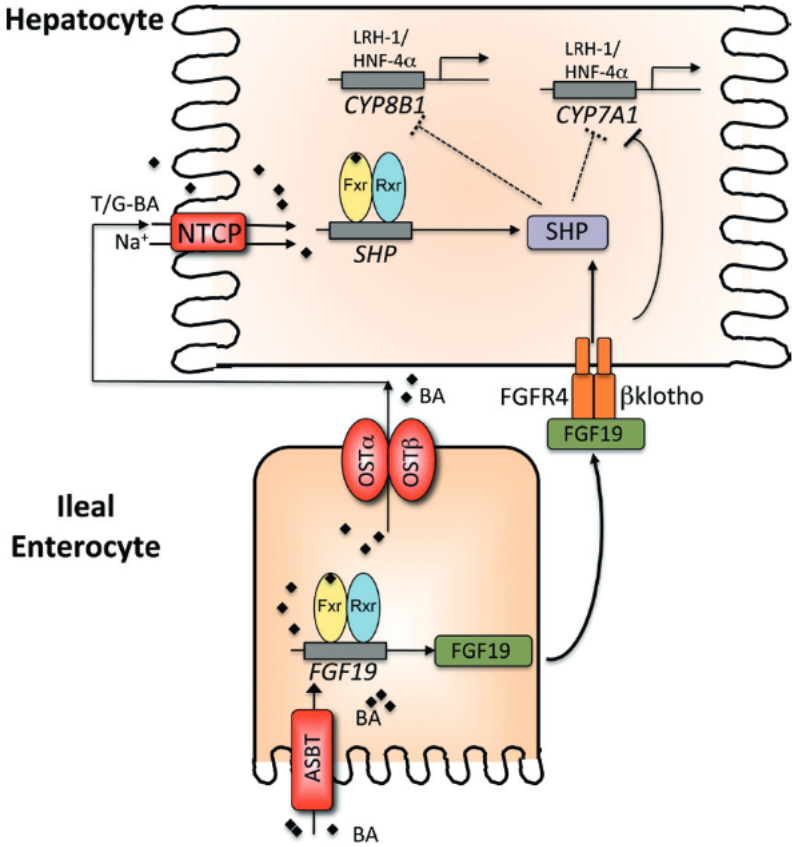


Figure 1.15: The negative feedback signalling pathway of bile acid biosynthesis. In hepatocytes, increased influx rates of bile acids from the portal circulation stimulates the FXR receptor which activates expression of SHP. The latter prevents the expression of both CYP8B1 and CYP7A1 rate limiting bile acid biosynthesis enzymes, via the inhibition of both HNF4α and LRH-1 transcription factors.
In the ileal enterocytes (small intestine), BAs are re-absorbed by ASBT and stimulate FXR, resulting in increased expression of FGF19. The latter migrates to the liver and initiates a signalling cascade at the FGFR4/βKlotho complex receptor. This leads to the initiation of the MAPK/ERK1/2 signalling pathway, resulting in complete inhibition of CYP7A1 transcription. This image was taken from Dawson, 2012

1.6.5 The Human Apical Sodium-dependent Bile Acid Transporter (ASBT/SLC10A2)

As seen in the EHC of BAs, the ASBT transporter (also known as ISBT, IBAT, ABAT and SLC10A2), primarily located at the distal ileum enterocyte brush border membrane (and accounting for 6 % of all expressed transporters at the small intestine), is essentially required to actively salvage circulating BAs, that are otherwise destined for excretion from the body in faeces. All major conjugated/unconjugated BA species are substrates for this transporter, although preferential transport of trihydroxy over dihydroxy, and conjugated over unconjugated BAs have been previously reported (Lack, 1979; Craddock et al., 1998; Kramer et al., 1999; Ho et al., 2004). In stark contrast to the NTCP transporter, which is known to additionally transport rosuvastatin (an HMG CoA reductase inhibitor) and the estrone-3-sulfate steroidal hormone, no non-bile acid solute has yet been identified for the translocation across biological membranes, for the ASBT transporter (Kramer et al., 2001; Ho et al., 2006; Dawson et al., 2009; Dawson, 2012; Slijepcevic and van de Graaf, 2017; Xiao and Pan, 2017; Baiocchi et al., 2019).

The human SLC10A2 gene, entailing 6 exons and localised to chromosome 13q33, translates a 348 amino acid ASBT transporter, with a predicted nine TMD assembly, consisting of an extracellular amino and cytosolic carboxyl termini. With a proposed molecular weight of 93kDa, and the requirement of multimeric complexes for functional BA transport, human ASBT has been strongly associated to function as a dimer, although monomeric states of the transporter are possible. It is hypothesised these monomers can self-associate to form homomultimers or can form heteromultimers with the ILBP protein (Dawson, 2011; Xiao and Pan, 2017).

Interestingly, in 2018, Chothe and colleagues generated a cysteine depleted (cysless) human ASBT variant, in order to assess its ability to form oligomeric states in the absence of structurally essential disulfide bonds. All 13 endogenous cysteine residues were mutated, which ultimately led to significantly reduced levels of protein expression with decreased transport functionality (significantly reduced dose-dependent uptake of [3H]-TCA). In addition, increase in protein instability was observed, resulting in the cysless variant's quick degradation by cellular proteasomes. However, the authors concluded the human ASBT transporter was still able to form active higher ordered oligomeric states, in the presence and absence of cysteine residues (i.e. cysteine-independent homodimerisation) (Chothe et al., 2018).

As is the case with many transporters associated to the EHC, ASBT is a secondary active transporter. This transporter requires an electrogenic sodium-

coupled co-transportation of two Na^+ ions to generate sufficient energy to transfer an unfavourable single BA molecule. The driving force for the supply of available Na^+ relies on the internal flow of the Na^+ gradient, which is governed by the negative intracellular potential and basolaterally expressed Na^+/K^+ATPase transporter (Dietschy, 1968; Aldini et al., 1996; Hu et al., 2011; Xiao and Pan, 2017). Due to the restrictions of sodium electrical gradients, the ASBT transporter is only seen to operate for the internal intestinal translocation of substrates (i.e. BAs can only be absorbed and not released from the ileum) (Lewis and Root, 1990; Marcus et al., 1991; Weinman et al., 1998; Dawson, 2012).

Other than its localisation to the ileal enterocyte, the ASBT transporter has also been shown to be moderately expressed in a multitude of other human tissues, including large cholangiocytes, biliary epithelium, the apical membrane of the renal proximal convoluted tubule cells, and gallbladder epithelial cells; all of which play fundamentally conserved roles in transporting and absorbing circulating bile acids, in order to maintain the EHC, and prevent loss of BAs to faecal and urinal waste (Wong et al., 1994; Christie et al., 1996; Alpini et al., 1997; Lazaridis et al., 1997; Craddock et al., 1998; Chignard et al., 2001; Hagenbuch and Dawson, 2004; Dawson, 2012).

Fascinatingly, upregulated and downregulated expression of the ASBT gene is seen to be under strong controlled regulation of a multitude of molecular members, including transcription factors (such as the hepatocyte nuclear factor-1 α (HNF-1α); peroxisome proliferator-activated receptor-α (PPAR-α); vitamin D receptor (VDR); glucocorticoid receptor (GR); GATA binding protein 4 (GATA4); sterol response element binding protein-2 (SREBP2); and retinoicacid receptor (RAR)), hormones, substrates, sterols, BAs and cytokines (Bosse et al., 2006; Battle et al., 2008; Dawson, 2012; Xiao and Pan, 2017).

1.6.6 Targeted Pharmaceutical Interventions of the ASBT Transporter - Novel Treatments of Severe Diseases

Significant disruptions and alterations to the homeostatic regulation of the EHC of BAs in humans, can lead to severe physiological disorders, including primary BA malabsorption, formation of calcium and cholesterol induced gallstones, irritable bowel syndrome, hypercholesterolaemia, chronic ileitis, cholestatic liver damage, Crohn's disease and hypertriglyceridemia, to name but a few (Dawson, 2012).

The aberrant expression and mutational inhibition of the human ASBT transporter has been seen to detrimentally affect its functional attributes resulting in severe disruptions of cholesterol homeostasis and BA EHC, causing a multitude of lipid metabolism and gastrointestinal diseases (Xiao and Pan, 2017). Primary BA malabsorption, classified as a rare idiopathic gene alteration, presenting with chronic diarrhea, fat-soluble vitamin malabsorption, black pigment (bilirubin) gallstones, intracranial hemorrhage, steatorrhea, and reduced plasma levels of cholesterol; was seen to be cause by disruptive L243P and T262M ASBT mutations (Heubi et al., 1982; Brink et al., 1996; Oelkers et al., 1997; Brink et al., 1999; Vítek and Carey, 2003). Additional intestinal disorders, which may also be associated to ASBT mutations and disruptions, include Idiopathic Chronic Diarrhea (Schiller et al., 1987), Cholesterol and Black Pigment Gallstone disease (Vítek and Carey, 2003; Holzer et al., 2008), Crohn's disease (Krag and Krag, 1976; Tougaard et al., 1986; Nyhlin et al., 1994), Colon Cancer (Wang et al., 2001), Familial Hypertriglyceridemia (Angelin et al., 1987; Duane et al., 2000), Postcholecystectomy Diarrhea, Chronic Ileitis (Meihoff and Kern, 1968), and Irritable Bowel Syndrome (Camilleri et al., 2009).

Due to the cytolytic detergent toxicity traits of BAs, their accumulation beyond acceptable levels must be strongly regulated and maintained, in order to prevent serious cellular damage. This can be achieved typically via the inhibition of ileal and hepatic transport and/or the administration of a sequestrant (Hofmann, 2009; Baiocchi et al., 2019).

Specific tailored inhibition of the ASBT transporter, has resulted in novel beneficial treatment options for cholestasis, hypercholesterolaemia and diabetes, due to the inflicted reduction in levels of both circulation BAs and plasma cholesterol (Xiao and Pan, 2017).

In light of this, the ASBT transporter can be targeted for a multitude of pharmaceutic benefits, including prodrug design (to raise oral bioavailability), inhibition (to provide novel therapeutic strategies for hepatocyte protection and hypercholesterolaemia); drug-drug interactions and drug absorption (Dawson, 2011; Hu et al., 2011).

Targeting the ASBT Transporter to design prodrugs for increased Oral Bioavailability

Pharmaceutical interest has been shown for the targeting of the ASBT transporter, to enhance the oral bioavailability of difficult to solubilise drugs, thereby enhancing a drug's intestinal permeability. The affinity (efficiency in substrate absorbance), specificity (broad availability of transportable substrates) and capacity (levels of expression and turnover rate of transport), are all key features governing the successful design of prodrugs. Concerning ASBT, its distinct specificity for BAs and unique localisation to the intestinal brush border membrane provides a unique situation, whereby BA substrates and derivatives can be presented as "Trojan Horses", allowing for the design and attachment of pro-drugs (Balakrishnan and Polli, 2006; Kramer and Glombik, 2006; Dawson, 2011).

Prodrug design can involve one of two strategies:

1. Modification of a transporter's naturally occurring solute to be strongly bound with an additional drug candidate of interest. Since the transporter's endogenous substrate is used, this method offers low levels of cytotoxicity.

2. Known as "substrate mimicry", applies the known three-dimensional structure of a natural substrate, to generate a novel drug of greatest resemblance by serendipity or design.

Specific drugs of interest can be manufactured to bind to BAs (at the steroid ring/nucleus C-3, C-7, or C-12 regions, as well as C-24 carboxylate and C-17 region), due to their unique chemical structure and sustained stability (Balakrishnan and Polli, 2006; Kuhajda et al., 2006; Davis, 2007). Chlorambucil (an anti-neoplastic alkylating nitrogen mustard), oligonucleotides, oxaprolylpeptide (inhibiting synthesis of hepatic collagen), and inhibitors of the HMG CoA reductase, are all but a few examples of drugs which have been successfully coupled to BAs, and imported into hepatic/ileal cells (Dawson, 2011). However, valacyclovir bound to CDCA, via a valine linker in the side chain, presented with the most successful pro-drug report to date. Being administered in rats led to a 2-3-fold increase in the oral bioavailability of the drug (Tolle-Sander et al., 2004).

Although seeming to be the perfect candidate for the generation of novel pro-drugs, the human ASBT transporter has been insufficiently targeted to generate better oral bioavailability of drugs, resulting in the generation of only a very limited set of FDA-approved ASBT inhibitors, including the HMG CoA reductase inhibitors and dihydropyridine calcium channel blockers (Balakrishnan and Polli, 2006; Dawson, 2011).

Inhibition of the ASBT transporter can lead to beneficial interventions for Hypercholesterolaemia

Hypercholesterolaemia, presenting with raised plasma cholesterol levels, is believed to result from imbalanced degrees of cholesterol uptake and secretion. This disorder can influence atherosclerotic deposits in arteries, thereby restricting the flow of blood, which can result in serious heart attacks and other severe cardiovascular diseases (Xiao and Pan, 2017).

It has previously been reported in mice that inhibition of the ASBT transporter leads to reductions in low-density lipoprotein (LDL) cholesterol and hepatic triglyceride levels, as a result of increased rates of the catabolism of cholesterol (and BA biosynthesis), to replenish bile acids lost by faecal expulsion. Elsewhere, ASBT inhibition resulted in elevated 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) levels of stimulation, which produced greater expression of LDL receptors (Xiao and Pan, 2017). Together, these interesting outcomes suggest a potential targeting and inhibition of the ASBT transporter, for the treatment of hypercholesteroleamia (Slijepcevic and van de Graaf, 2017; Xiao and Pan, 2017).

Specific targeting of the ASBT transporter with potent, non-absorbable non-bile acid and bile-acid derivative inhibitors, were trialled in animal models, for the potential discovery of novel treatments of hypercholesterolaemia (Kramer and Glombik, 2006). In rats, TA-7552 ([1-(3,4-dimethoxyphenyl)-2,3-bis(methoxycarbonyl)-4-hydroxy-6,7,8-trimethoxynaphthalene]), inhibited the ASBT intestinal absorption of BAs, which resulted in increased BA faecal elimination, and the stimulation of hepatocytes to biosynthesise BAs from the catabolism of cholesterol, thereby significantly reducing the latter's plasma concentration (Kramer and Glombik, 2006). Elsewhere, in monkeys and hamsters, R-146224 ([1-7-[(1-(3,5-Diethoxyphenyl)-3-[(3,5-difluorophenyl)(ethyl) amino]carbonyl-4-oxo-1,4-dihydroquinolin-7-yl)oxy]heptyl-1-methylpiperidinium bromide]) was also seen to inhibit ASBT, resulting in drasticly decreased levels of non-HDL cholesterol members (Kitayama et al., 2006). In rabbits, the ASBT inhibitor S-8921 ([methyl 1-(3,4-dimethoxyphenyl)-3-(3-ethylvaleryl)-4-hydroxy-6,7,8-trimethoxy-2-naphthoate]) displayed significantly decreased cholesterol serum levels, with a preventative effect on the prognosis of atherosclerosis (Hara et al., 1997; Higaki et al., 1998). Finally, the benzothiazepine derived ABST competitive inhibitors 2164U90 ([(-)-(3R,5R)-trans-3-butyl-3-ethyl-2,3,4,5-tetra-hydro-5-phenyl-1,4-benzothiazepine1,1-dioxide]) and 264W94 ([(-)-(3R,5R)-trans-3-butyl-3-ethyl-2,3,4,5-tetrahydro-7,8-dimethoxy-5-phenyl-1,4-benzothiazepine1,1-dioxide]), blocked intestinal BA re-absorption, stimulated hepatic biosynthesis of BAs, and reduced LDL and VLDL plasma cholesterol levels, in both mice and rats (Lewis et al., 1995; Root et al., 2002). Additionally, the LUM002 and Resveratrol

ASBT inhibitors, also explicitly depicted cholesterol-lowering abilities (Xiao and Pan, 2017).

Although moderate success in reducing cholesterol levels in animal models were well-established, the large majority of novel ASBT inhibitors entered in human clinical trials failed to progress past the clinical trial 1 phase, due to efficacy and drug dose-related side effect constraints, including gallstone disease, increased susceptibility to colon cancer, diarrhea, and hypertriglyceridemia (Zheng et al., 2009; Dawson, 2011).

Inhibition of the ASBT transporter can lead to beneficial interventions for Cholestasis

When the flow of bile is impeded, cytotoxic BAs are recalled and accumulated in hepatocytes, resulting in cholestatic liver damage, fibrosis and inflammation. Current treatments with administrations of ursodeoxycholic acid (UDCA) and/or CA, have shown to yield limited improvements and efficacy. Therefore, a potential gain of hepatoprotection, via the inhibition and downregulation of both intestinal and hepatic BA transporters, is of clinically pharmacological importance and interest. Indeed, inhibition of the ASBT transporter has been investigated by pharmaceutical companies as a novel treatment for cholestasis. The LUM001 ASBT inhibitor, directly prevents the reabsorption of BAs in the intestine, resulting in increased levels of their elimination in faeces. As a result, up to 80 % of the total bile acid pool was seen to be depleted, freeing up the accumulated bile acids localised to the hepatocytes, and improving intrahepatic cholestasis. In addition, extra-hepatic cholestasis was also seen to be improved by the inhibition of ASBT via the S0960 inhibitor (Xiao and Pan, 2017; Slijepcevic and van de Graaf, 2017).

Inhibition of the ASBT transporter can lead to beneficial interventions for Type II Diabetes

Diabetes mellitus (also known as type II diabetes) is categorised by an imbalance in the production of insulin, resulting from raised levels of BA reabsorption in the intestine (due to increased ASBT expression levels), leading to an increased pool of circulating bile acids and raised plasma cholesterol levels. Hypercholesterolaemia and dyslipidaemia can be regarded as downstream knock-on disorders originating from the onset of diabetes mellitus, highlighting the seriousness of this disorder. Inhibiting the transport function of ASBT in the ileum, prevents the reabsorption of circulating BAs, thus stimulating the export of the enterohepatic glucagon-like peptide-1 (GLP-1) hormone from L-enteroendocrine cells. This release promotes increased cellular sensitivity to insulin and raised secretion levels of insulin from

pancreatic β cells, subsequently regulating the glucose concentration levels of the bloodstream. The GSK2330672 ASBT inhibitor is a non-absorbable and highly potent molecule, which provides a clinically interesting therapeutic treatment for human individuals presenting with type II diabetes (Hofmann, 2009; Xiao and Pan, 2017; Slijepcevic and van de Graaf, 2017).

Stimulation of the ASBT transporter can lead to beneficial interventions for Crohn's Disease

Crohn's disease (also known as inflammatory bowel disease), is a chronic inflammatory disorder of the intestines, presenting with weight loss, malnutrition and diarrhoea. Specific inhibition of the ASBT transporter, via inflammatory cytokines, has shown to be the cause for BA malabsorption and watery diarrhoea (due to raised colonic BA levels inducing chloride channels), in patients affected by IBD. Therefore, the stimulation and increased expression of ASBT (via the administration of glucocorticoids, for example) could help with the treatment and maintenance of individuals suffering with Crohn's disease (Xiao and Pan, 2017).

1.6.7 Crystal Structures of the Bacterial Homologues of the Human ASBT Transporter

The *Neisseria meningitidis* ASBT Crystal Structure ($ASBT_{NM}$)

In 2011, Hu and colleagues successfully solved the crystal structure of the *Neisseria meningitidis* bacterial homologue of the human ASBT bile acid transporter. Solving the structure to 2.2 Å resolution, the transporter presented with an inward-open facing topology of ten transmembrane domains (TMDs), orientated in two inverted structural assemblies of five TMDs respectively, with cytoplasmic amino and carboxy termini. The two repeating units formed a flat N-terminal V-shaped "panel" domain consisting of TM1-TM2, and TM6-TM7; and a core domain comprising of two essential Na^+ binding sites, formed by TM3-TM5 and TM8-TM10. In addition, a taurocholate molecule was also seen bound in an intracellular cavity. Figure 1.16 below depicts this ten TMD topology, highlighting both the core and panel domains, sodium binding sites and bound taurocholate.

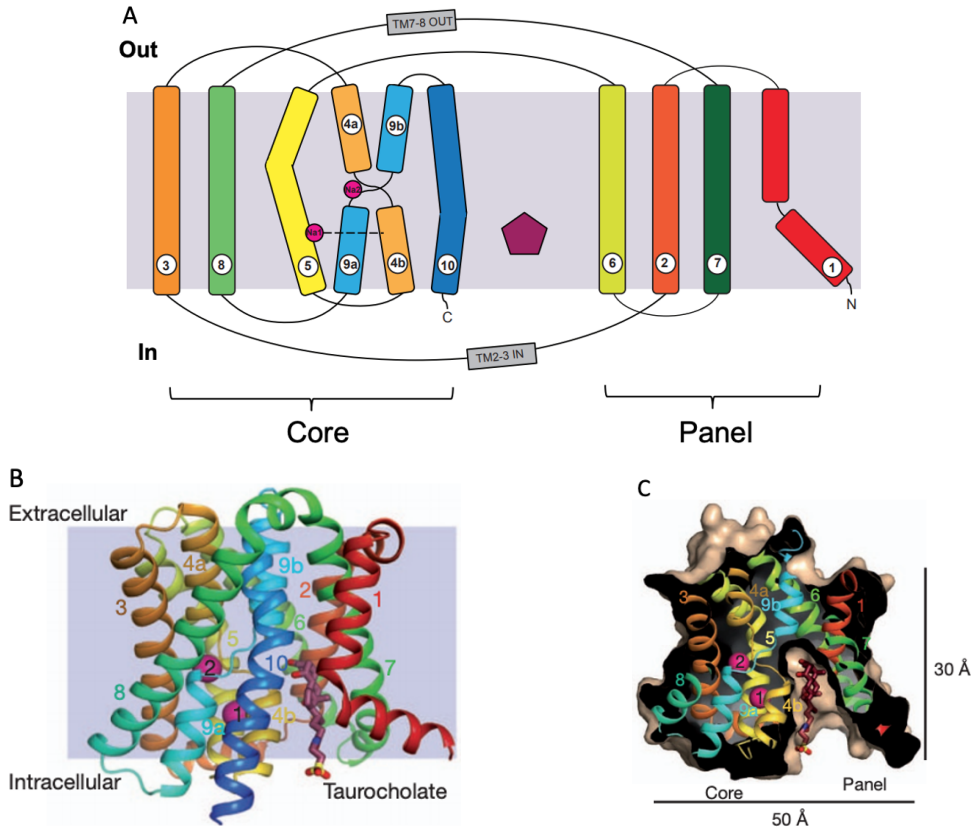


Figure 1.16: The $ASBT_{NM}$ ten TMD topology. (A) Illustrated representation of TM1-TM10 of $ASBT_{NM}$, coloured from red at the N-terminus to blue at the C-terminus, with Na1 and Na2 depicted as pink circles and bound taurocholate represented as a wine-red pentagon. (B) Ribbon depiction of the $ASBT_{NM}$ seen in the grey plane of the membrane. Transmembrane domains 1-10 have been coloured from red at the N-terminus to blue at the C-terminus, with sodium ions presented as pink spheres and taurocholate bound shown as a stick model. (C) Surface depiction outlining the intracellular taurocholate binding cavity of the $ASBT_{NM}$ transporter.
All images were taken from Hu et al., 2011.

Interestingly, both TM4 and TM9 domains form broken/discontinuous helices (TM4a, TM4b, TM9a, and TM9b), which cross over each other, to form kinked helical hairpins with TM5 and TM10, respectively. These discontinuous motifs seem to be a common occurrence of secondary active transporters, with their crossover also seen elsewhere in structurally unrelated transporters, including uraA (Lu et al., 2011), NapA (Lee et al., 2013), NhaA (Lee et al., 2014b), AE1 (Arakawa et al., 2015), UapA (Alguel et al., 2016), and SbtA (Fang et al., 2021).

As previously mentioned, both NTCP and ASBT transporters require two Na^+ ions for the successful translocation of one bile acid substrate. Within the $ASBT_{NM}$ core domain, both sodium-binding sites were successfully identified, with Na1 located between TM4b and TM5 and Na2 located at the point of the discontinuous helices crossover of TM4a-TM4b and TM9a-TM9b. Being separated by just 8 Å and occluded from the bulk solvent, both Na1 and Na2 associate with the carboxylate region of Glu 260 from TM9a. Na2 additionally coordinates with the carbonyl oxygen side chain groups of Gln 264 of TM9a and Gln 77 of TM3. The Glu 260 residue, seen to be interacting with both sodium ions, has been shown to be vital for transport and activity of the ASBT transporter, whilst all remaining interacting residues are also seen to be crucially conserved among both NTCP and ASBT transporters. Finally, the presence of Na2 seems to be essential for the mechanic drive of the transporter, as it stabilises the association with TM4a, by neutralising the negative dipole of TM9a (Hu et al., 2011).

The authors additionally reported on the successful uptake of taurocholate via the $ASBT_{NM}$ transporter, by radioactive uptake studies, and as such were able to model a taurocholate (TCA) molecule into the electron density of the inward-facing hydrophobic cavity of the protein structure, formed between both core and panel domains. The modelled TCA was orientated perpendicular to the cell membrane, with its cholesterol ring interacting with the crossover region of the discontinuous helices and taurine tail extending towards the intracellular opening of the cavity. Most interestingly, amino acid residues located near this bound bile acid have also previously been implicated in the specific binding of human ASBT inhibitors. However, the presence of this taurocholate molecule in the structure is only weakly bound, forming just one hydrogen bond with the $ASBT_{NM}$ transporter between TM10 Asn 295 residue and 7α-hydroxyl group taurocholate molecule. Mutation of this amino acid residue however, significantly impacted the translocation of taurocholate, highlighting its vital association in transport. The taurine moiety of the bile acid is seen to be located between TM1 and TM10, with the latter being categorically implicated in the translocation of BA substrates. In addition, the Asn 265 residue, located between TM4b and TM9b is also suspected to be involved in the translocation mechanism, whereby its mutation to alanine

in $ASBT_{NM}$ resulted in a decreased transport activity of 80 % (Hu et al., 2011).

Figure 1.17 below, highlights the crucial binding pockets explored by Hu and colleagues for both sodium ions and taurocholate molecule.

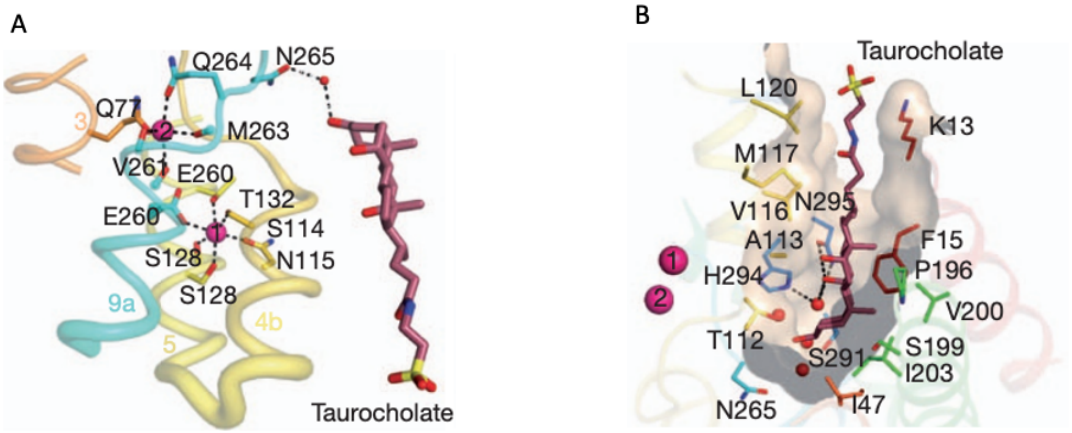


Figure 1.17: The $ASBT_{NM}$ sodium and taurocholate binding sites. (A) The Na1 and Na2 binding sites and their interactions transporters residues. Na1 is octahedrally positioned with Ser 114 and Asn 115 from TM4b; Thr 132 and Ser 128 from TM5; and Glu 260 from TM9a. Na2 forms a square pyramidal interaction with Glu 260, Val 261, Met 263 and Gln 264 from TM9; and Gln 77 from TM3. Both Na1 and Na2 ions have been displayed as pink spheres. (B) The inward-facing hydrophobic intracellular cavity of the taurocholate binding site. Amino acid residues near the BA molecule, lining the cavity are shown, representing weak interactions of the substrate to the $ASBT_{NM}$ structure. The taurocholate substrate has been represented as a stick model in the cavity.
All images were taken from Hu et al., 2011.

In order to transport bile acids and other substrates, conformational transformations from outward to inward-facing orientations must occur. The authors proposed a rearrangement of the sodium binding sites, specifically at the crossover regions of the discontinuous helices (TM4a-TM4b and TM9a-TM9b), to be the origin of conformational changes driving the mechanistic alterations from outward-facing to inward-facing assemblies. This hypothesis was founded on similar helical arrangements being the driving factor for solute translocation in the sodium/proton NhaA transporter (Lee et al., 2014b). However, as $ASBT_{NM}$ transports bulky bile acid solutes, additional conformational changes are certainly likely (Hu et al., 2011).

The generation of a predicted outward-facing model for the $ASBT_{NM}$ structure was able to distinctly highlight possible alterations in the location of the panel domain with regards to the core domain, presenting with the greatest noticeable conformational shifts. Figure 1.18 below, outlines this putative mechanism for BA transport.

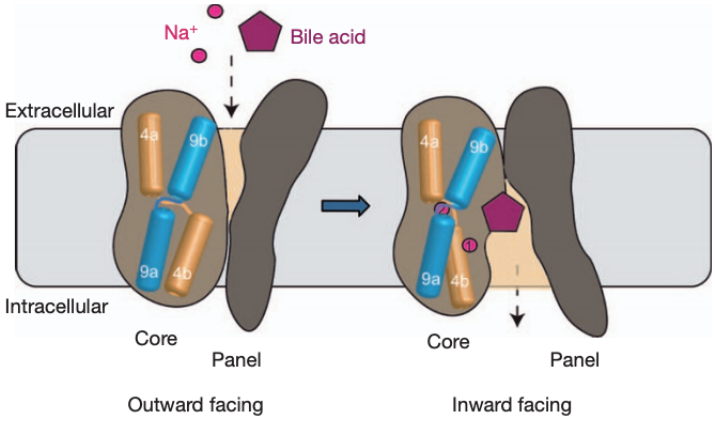


Figure 1.18: The proposed mechanism of ASBT$_{NM}$ bile acid transport. Illustration depicting alterations to the panel domain with respect to the core domain, upon binding and translocation of two sodium ions and one bile acid molecule from the extracellular to intracellular regions. This image was taken from Hu et al., 2011.

Fascinatingly, although no conserved sequence identity or homology has previously been reported, the overall conformation of the ASBT$_{NM}$ symporter similarly reiterated the structural architecture exhibited by the sodium/proton antiporter NhaA (Lee et al., 2014b), thereby acknowledging en extraordinary plasticity of transporters to exhibit common scaffolds, whilst transporting distinctly separate substrates (Hu et al., 2011).

To conclude, the authors proposed the architecture of the ASBT$_{NM}$ core domain to be strongly dictated by the sodium binding sites. This strongly influenced the activity of the panel domain, which alterd its conformation, with regards to the core domain, in order to drive the accessibility of substrate binding to the internal cavity. With the generation of this novel inward-open ASBT$_{NM}$ structure, significant advances can be made to generate potent human ASBT inhibitors for potential treatments for hypercholesterolaemia (Hu et al., 2011).

The *Yersinia frederiksenii* ASBT Crystal Structure ($ASBT_{Yf}$)

In 2014, Zhou and colleagues successfully solved two crystal structures of the *Yersinia freriksenii* bacterial homologue of the human ASBT bile acid transporter. Sharing 40 % sequence identity with the $ASBT_{NM}$ transporter, Zhou et al., (2014) elucidated further the possible mechanism of transport of bile acids and Na1/Na2 molecules across the cell membrane, with specific focus on structural alterations. These novel $ASBT_{Yf}$ structures uncovered significant rigid-body rotations of the substrate-binding domain, significantly affecting the conserved crossover region of the TM4a-TM4b and TM9a-TM9b discontinuous helices, thereby resulting in altered accessibility of the transporter from both extracellular and intracellular sides of the lipid bilayer. These altered structural conformations fundamentally targeted the orientation and localisation of bound bile acids and sodium ions, playing major roles in their translocation into the cytosol (Zhou et al., 2014).

Similarly to the $ASBT_{NM}$ structure, the 1.95 Å $ASBT_{Yf}$ crystal structure exhibited a ten transmembrane domained transporter, with a two-fold pseudosymmetry axis, generating the conserved panel (TM1-2, and TM6-7) and core (TM3-5 and TM8-10) domains, and intracellular N- and C-terminii. As had originally been reported in $ASBT_{NM}$, the $ASBT_{Yf}$ TM4 and TM9 domains, crucially located in the core domain, also formed discontinuous helices and unwound at their point of crossover. In addition, the authors also reported on four amphipathic helices (AH1-4) localised to the interface of the bulk solution and the cell membrane (Zhou et al., 2014). Figure 1.19 below depicts an illustration of this $ASBT_{Yf}$ ten transmembrane domain topology, highlighting similar architecture to $ASBT_{NM}$.

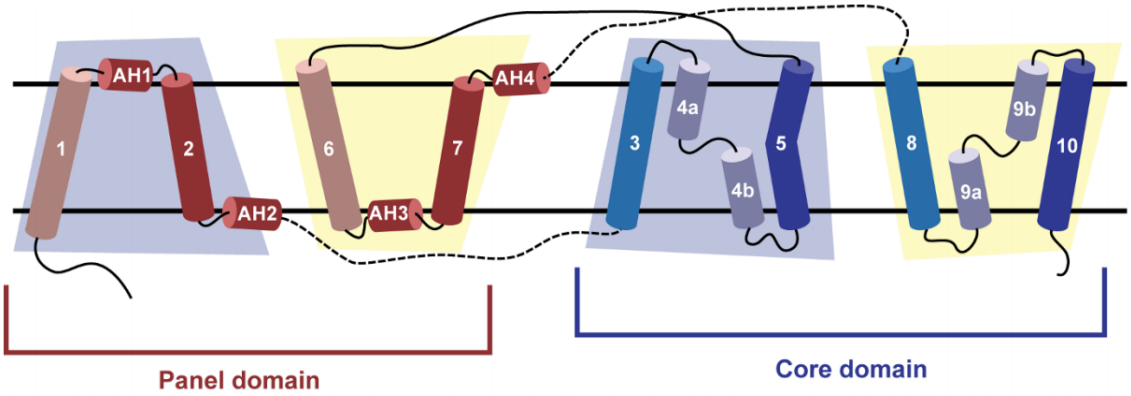


Figure 1.19: The $ASBT_{Yf}$ ten TMD topology. Illustrative diagram depicting the two-fold pseudosymmetry representation of TM1-TM10, with the panel and core domains highlighted in red and blue respectively. First and second inverted repeats are denoted by blue and yellow trapezoids, respectively.
This image was taken from Zhou et al., 2014.

The panel and core domains of the structure were seen to be orientated in an inward-open facing, un-liganded, conformation, resulting in a large solvent-accessible cavity from the cytosol. Indeed, access from the extracellular side was not possible, due to strong interactions of TM2, TM7 and TM9b, sealing the outward-facing solvent-cavity. Interestingly, although strong amino acid sequence conservation is seen between both $ASBT_{NM}$ and $ASBT_{Yf}$ structures for the Na1 (located at TM4b, TM5 and TM9a) and Na2 (located at TM4a and TM9a) binding sites, only very weak electron density for Na2 of the $ASBT_{Yf}$ was reported. The authors proposed a tilt of the TM4b by 11° from the crossover region to be the culprit for the lack of Na1 in the model, resulting in unwinding of TM1. In light of this, both Asn109 and Ser108 residues were moved out of range to form bonding with Na1, thereby also affecting the orientation and localisation of neighbouring amino acid residues involved in Na2 binding (Zhou et al., 2014).

In order to generate alternative structural conformations of the $ASBT_{Yf}$ transporter, Zhou and colleagues mutated the highly conserved Glu254Ala residue involved in Na1 binding. A 2.5 Å E254A $ASBT_{Yf}$ crystal structure resulted in an outward-open facing, un-liganded, conformation of the transporter, with significant alterations to both core and panel domains, indicating relative motion. The authors concluded the core domain of this mutant to exhibit rigid-body rotations of the TM4b and TM9b domains, resulting in their interface with the panel domain to translate by 6 9Å towards the periplasm. In addition, AH2, AH4 and the TM5-6 loop were all seen to behave as hinges to enhance this rigid motion. Fascinatingly, the TM2, TM7 and TM4b domains of this E254A mutant structure all produced novel contacts at the intracellular side, resulting in a large solvent-accessible cavity to form on the extracellular side, at the crossover region. However, similar to the wild-type $ASBT_{Yf}$ structure, this mutant structure failed to show any binding of Na1 and/or bile acid substrates (Zhou et al., 2014).

Comparing both solvent-accessible cavities of the wild-type inward-open and E254A outward-open facing conformations, revealed a narrow region of the core and panel domains accessible to solvent in both structures. Specifically, this region concerned the crossover region of the discontinuous helices as well as additional residues strongly conserved across many ASBT homologues. Figure 1.20 below, represents the structural differences between both these inward and outward 3D conformations for the $ASBT_{Yf}$ transporter.

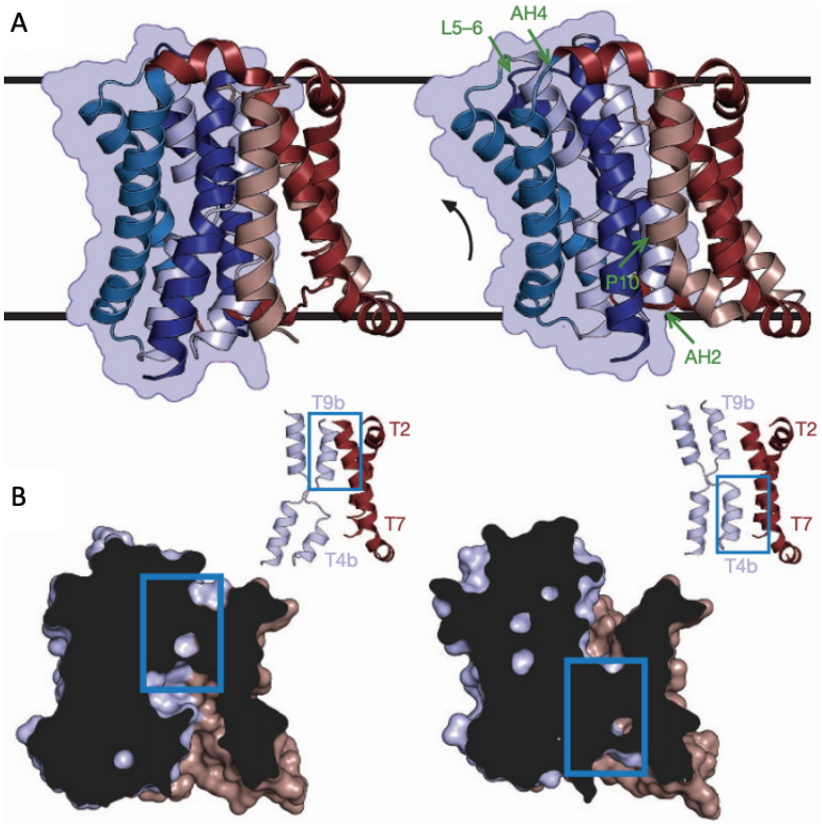


Figure 1.20: Structural comparisons of the ASBT$_{Yf}$ wild-type inward-open and E254A outward-open facing conformations. (A) Cartoon representations of the wild-type (left) and E254A mutant (right) 3D structures. Black lines roughly display the extracellular and intracellular regions of the lipid bilayer. The core domain is contoured in blue and regions acting as hinges in the E254A structure are marked in green arrows. (B) Surface cutaway representations for the wild-type (left) and E254A mutant (right) 3D structures, outlining the intracellular and extracellular solvent-accessible cavities. Cartoon representation of the TM4-TM9 discontinuous helices are displayed to highlight the interface between the core and panel domains. Their respective location in each structure is denoted by blue rectangular boxes.
This image was taken from Zhou et al., 2014.

The physiological relevance of this ASBT$_{Yf}$ E254A mutant outward-facing structure was strongly questioned and criticised, due to its elimination of Na1 and severe crippling of the transporters functional ability. In addition, no substrate was co-crystallised with the transporter, which is surprising considering a higher affinity for substrate binding is expected in outward-open conformations, although this could be explained by the abolishment of Na1 (Zhou et al., 2014). In order to investigate this further, Wang and colleagues introduced novel disulfide bridges to "trap" the ASBTYf transporter in an outward-facing conformation (Wang et al., 2021a). Neither Na^+ or solute binding sites were affected in the process, thereby generating both functional and structural transporter integrity, with binding and transmission rates of bile acids yielding similar affinities to the wild-type protein. These novel cysteine mutations affected the panel-core domain interfaces of TM4 and TM7, respectively, but allowed for natively folded core and panel domains

to perform natural physiological motion, until restricted by inter-domain disulfide cross-linking. Fascinatingly, this novel cysteine-induced outward-open crystal structure strongly supported the E254A mutant structure obtained previously, concluding the latter to be an accurate outward functional representation of the ASBT$_{Yf}$ transport cycle (Wang et al., 2021a). Figure 1.21 below, displays the "trapped" outward-facing crystal structure as a result of disulfide cross-linking.

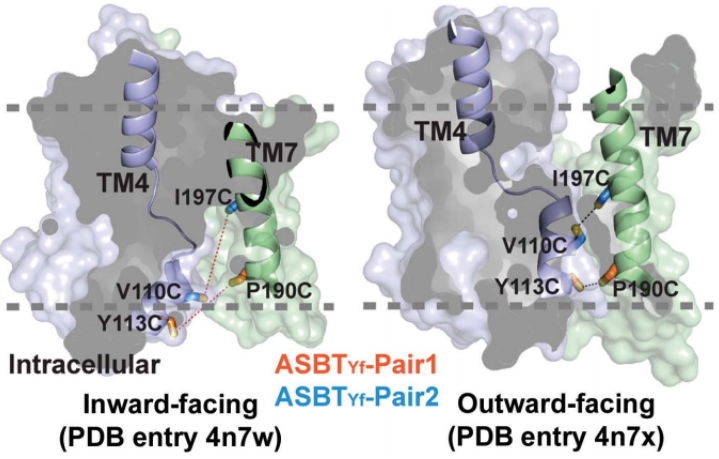


Figure 1.21: Novel engineered disulfide bridges "trap" the ASBT$_{Yf}$ transporter in an outward-facing conformation. (A) Surface representation of the wild-type ASBT$_{Yf}$ inward-facing structure (PDB entry 4n7w; Zhou et al., 2014) and (B) the ASBT$_{Yf}$ E254A mutant outward-facing structure (PDB entry 4n7x; Zhou et al., 2014). Both core and panel domains are highlighted in blue and green, respectively. Cartoon representations of TM4 and TM7 are shown, with the engineered Cys pairs displayed as sticks. The Y113C/P190C (Pair1) and V110C/I197C (Pair2) disulfide bridges are displayed in orange and blue, respectively. The grey dashed lines indicate relative positioning of the transporter structures in the lipid bilayer.
This image was taken from Wang et al., 2021a

The proposed taurocholate (TCA) binding site of the ASBT$_{NM}$ transporter was seen to be accessible only from the inward-open conformation (Hu et al., 2011). However, upon modelling to the outward-open state of the ASBT$_{Yf}$ E254A mutant structure, Zhou and colleagues noticed a clash of the TCA molecule in the protein matrix, resulting in a lack of access to the periplasm (Zhou et al., 2014). Due to distinct disagreement with the alternating-access mechanism, in which solute translocation across a cell membrane occurs at ligand-binding sites accessible from both intracellular and extracellular sites, via distinct structural conformational changes, the authors suggested an alternative TCA binding site in a lateral orientation (with the hydrophobic cholesterol ring orientated to the hydrophobic panel domain, and hydrophilic taurine orientated to the polar core domain), in oder to satisfy alternating access to both sides of the membrane. In this novel configuration, the translocation of bile acids could sufficiently be accounted for by the rigid body motion explained by the two ASBTYf crystal structures. This TCA molecule, seen to be bound vertically in the ASBT$_{NM}$

structure, therefore likely presents an exit position rather than the substrate-binding site (Zhou et al., 2014).

Via molecular docking, structural and functional studies, Wang and colleagues investigated this substrate binding site in further detail (Wang et al., 2021b). Upon solute translocation, the core domain moves relative to the panel domain, in an elevator-like alternating access fashion. ASBT substrates are thought to bind to a horizontal central cavity, localised between both the core (affecting helices TM4, TM5, TM9, and TM10) and panel (affecting helices TM2 and TM7) domains, accessible to both sides of the cell membrane. APBS electrostatic analysis elucidated a negatively charged region in the proximity of TM5, and a positively charged pocket near TM10. With TCA molecules seen to interact near the TM4/TM9 crossover, the authors concluded the steroid head to interact with TM5, and the taurine tail to associate with TM10. Fascinatingly, amino acid residues within 3.5 Å of the bound TCA molecule were localised to the core domain, associating with the TM4, TM5, TM9, and TM10 helices, highlighting the significance of this domain in solute translocation. More specifically, Ser104, Thr106 from TM4, Thr130 from TM5 and Asn259 from TM9 formed strong hydrogen bonds with the three hydroxyl regions of the steroid head, whilst Asn259, Ser260, and Gly261 from TM9, and His286 from TM10 produced hydrogen bonds with the sulfonic region of the TCA tail (Wang et al., 2021b).

Furthermore, co-crystallisation of ligand-like acid molecules, including glycine, citrate and sulfate, were all seen to bind in the same substrate acid tail region of the central cavity, forming strong interactions near the TM4/TM9 discontinuous crossover of the transporter, as well as TM10. This highlights the conservation of this substrate binding site to both positive/negative and hydrophilic/hydrophobic orientations, which is strongly compatible and representative in the alternating access mechanism (Wang et al., 2021b).

The TM4/TM9 crossover region is suspected to perform conformational alterations in order to accommodate differently sized substrates with varying properties, and ensure their translocation across the cell membrane. Therefore, this feature is highly important in substrate binding and translocation (Wang et al., 2021b). Figure 1.22 below summarises this conserved $ASBT_{Yf}$ substrate binding site, emphasising the locations and orientations of both the head and tail bile acid moieties.

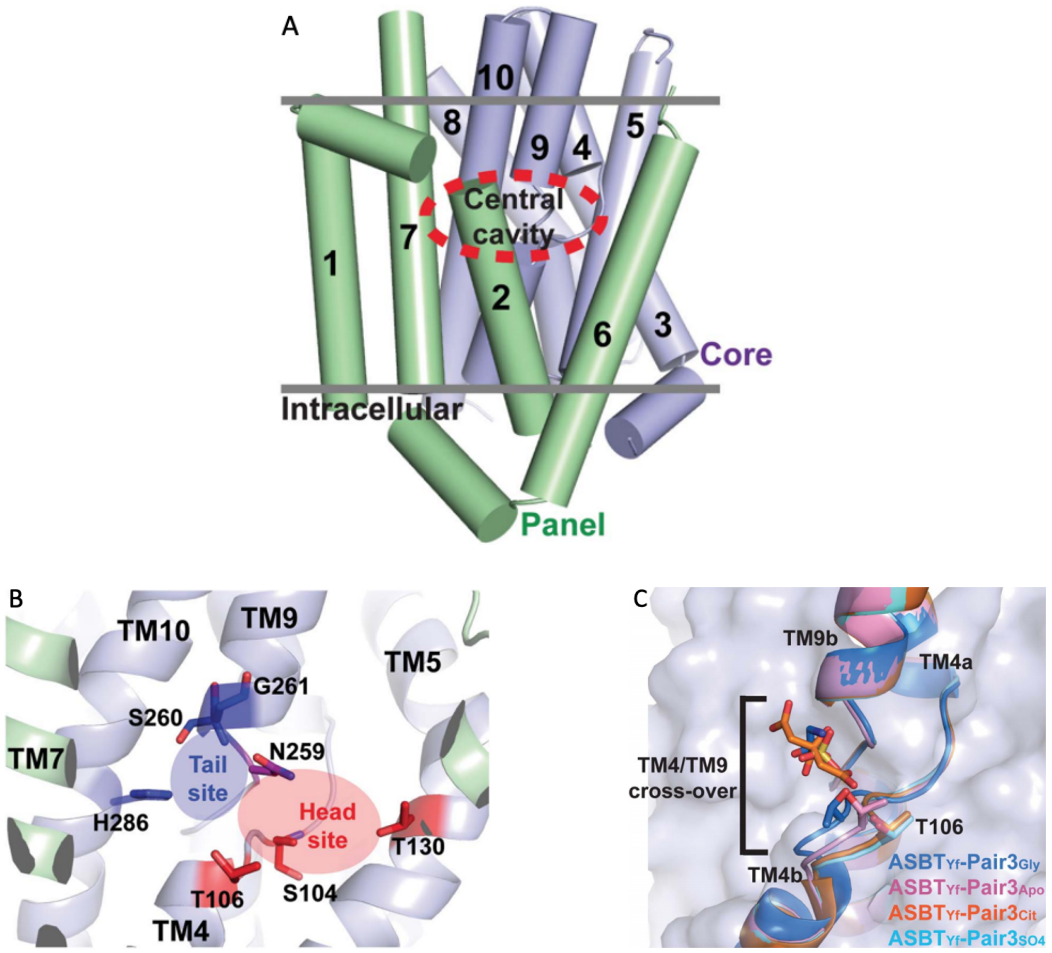


Figure 1.22: The substrate binding central cavity of the ASBT$_{Yf}$ transporter. (A) Cartoon depiction of the outward-facing ASBT$_{Yf}$ E254A mutant structure (PDB entry 4n7x), with panel and core domains coloured green and blue, respectively. The numbers represent corresponding helical TMs, with the central substrate binding cavity highlighted in a red dashed oval. (B) The substrate head and tail binding sites are displayed in the cavity and coloured by red and blue spheres respectively. Associating amino acid residues are shown in stick form and also coloured in red and blue, to depict corresponding head/tail interactions. Asn259 is coloured purple, as it interacts with both the head and tail regions of bound substrates. (C) Cartoon representation displaying the binding location of the ligand-like acids near the discontinuous TM4/TM9 crossover helices. Superposition of all four outward-facing ASBT$_{Yf}$ cysteine engineered structures (Apo, Glycine, Citrate and Sulfate), with bound ligands displayed as sticks present in the blue coloured core domain.
All images were taken from Wang et al., 2021b.

To summarise the structural findings obtained from both ASBT$_{NM}$ and ASBT$_{Yf}$, from both outward (PDB entry 4n7x) and inward-facing (PDB entries 4n7w and 3zuy) conformations, figure 1.23 below, depicts the proposed elevator-style alternating access mechanism for ASBT mediated bile acid translocation across the cell membrane (Hu et al., 2011; Zhou et al., 2014; Wang et al., 2021a,b).

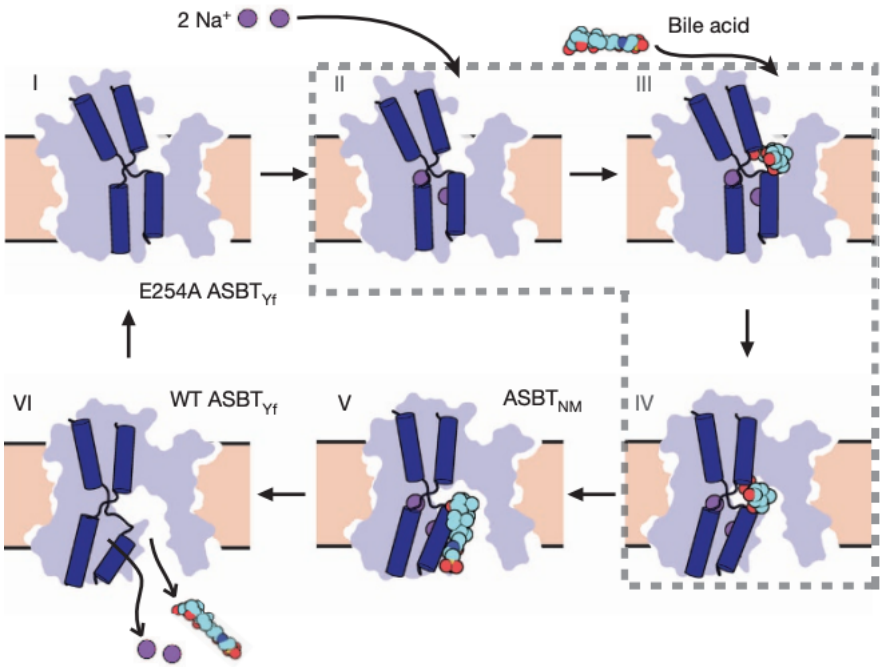


Figure 1.23: The proposed elevator-style $ASBT_{Yf}$/$ASBT_{NM}$ alternating access mechanism. (I) The outward-facing state of the transporter exposes the TM4/TM9 crossover region, allowing for the (II) binding of Na^+ to Na1 and Na2 binding sites. (III) Bile acid substrates subsequently bind to the central cavity between panel and core domains. (IV) Structural conformation alterations of the TM4/TM9 discontinuous helices result in the generation of an inward-open facing conformation, allowing for the (V) migration of the bile acid into an exit position (seen in $ASBT_{NM}$), and (VI) subsequent expulsion of both sodium ions and bile acid substrate into the cytoplasm, due to the low intracellular Na^+ concentrations. the grey dashed line depicts hypothetical structural states of the alternating access mechanism.
This image was taken from Zhou et al., 2014

Finally, like $ASBT_{NM}$, $ASBT_{Yf}$ also shares similar structural folds with structurally unrelated transporters, including GltPh (Crisman et al., 2009), LeuT (Krishnamurthy and Gouaux, 2012), XyIE (Sun et al., 2012), NapA (Lee et al., 2013) and NhaA (Lee et al., 2014b). The two latter Na1/H1 antiporters similarly present with inverted pseudosymmetry repeats, forming panel and core domains with a substrate-binding site. With the solved crystal structures of the NhaA (inward-open state) and NapA (outward-open state) antiporters, a remarkably similar rigid body movement of the crossover region of the core domain is required to translocate protons across the cell membrane NapA (Lee et al., 2013, 2014b). Although significant differences in the size of the substrate transported are observed, the suggested translocation mechanism for the $ASBT_{Yf}$ structures, closely resemble those for the NhaA/NapA antiporters (Zhou et al., 2014; Wang et al., 2021a,b).

1.7 The SLC10A7 Orphan Transporter in Humans, Mice, Zebrafish and Yeast

1.7.1 Molecular and Phylogenetic Characterisation

SLC10A7 is characterised as an orphan transporter, meaning its translocated solute to this day remains unknown. Over the past 15 years, however, exciting research has been conducted, yielding very interesting, novel outcomes. These reported findings are examined in greater detail below:

SLC10A7 was first classified as a membrane protein and part of the SLC10 family by Zou and colleagues in 2005 (Zou et al., 2005). A large-scale sequencing analysis of human fetal brain cDNA yielded a novel 2706bp long sequence, termed C4orf13. This sequence translated to a 340 amino acid polypeptide chain, entailing a sodium bile acid co-transporter family (SBF) domain (residues 44-225), with a topology of ten TMDs with intracellular cis orientating N- and C-terminal ends (Zou et al., 2005). Subsequently attributed as the seventh member of the SLC10 family, it is the only member in humans to present with a ten TMDs topology and has been localised to chromosome 4q31.2, containing 12 exons and expected to yield a predicted molecular mass of 37.4kDa with an isoelectric point of 9.67 (Zou et al., 2005; Godoy et al., 2007).

Reverse transcription polymerase chain reaction (RT-PCR) analysis revealed C4orf13 to be widely expressed in human tissues: with high expression observed in the liver and lung; moderate expression seen in the spleen, kidney, thymus and placenta; and low expression detected in the testis, heart and prostate. The highest expression of C4orf13 was found to be in the liver, suggesting a potential role in liver function, which would correspond well with the classification of this protein as a member of the sodium bile acid co-transporter family (Zou et al., 2005).

In 2007, Godoy and colleagues were able to examine further the molecular and phylogenetic properties of the SLC10A7 membrane protein (Godoy et al., 2007). Among mice and rats, SLC10A7 was also found to be expressed broadly across several tissues including the small intestine, liver, lung, brain, colon, adrenal gland and heart, but a lack of expression in skeletal muscle was also witnessed. In adult humans, the highest mRNA expression was identified in the liver and testis; and in frogs, expression of SLC10A7 was solely localised to the spleen, small intestine and skeletal muscle (Godoy et al., 2007).

To investigate potential substrates involved in transport, SLC10A7 membrane protein transport studies were conducted in both Xenopus laevis oocytes and

HEK293 cells. Using NTCP (SLC10A1) as a positive control, a lack of transport activity of multiple bile acids and steroid sulfates (including cholate/cholic acid (CA), taurocholate/taurocholic acid (TCA), glycocholate/glycocholic acid (GCA), chenodeoxycholate/chenodeoxycholic acid (CDCA), dehydroepiandrosterone (DHEAS), estrone-3-sulfate (E3S) and pregnenolone sulfate (PS)) were reported in SLC10A7 expressing cells; further highlighting the classification of this SLC10A7 member as an orphan transporter (Godoy et al., 2007).

Human and rat SLC10A7 proteins yielded a strong SDS-PAGE separation of molecular weight bands at 27kDa, and a further fainter band at 54kDa. This latter band most likely depicts a dimeric complex rather than a glycosylated state, since a lack of N-linked or O-linked glycosylation sites were identified. However, in comparison to amino acid sequence prediction programmes reported by Zou et al., 2005, this monomeric 27kDa molecular mass band was significantly smaller than the reported 37.4kDa. This change can be accounted for as a result of the hydrophobic features of membrane proteins, whereby >60 % amino acid residues are localised to the ten transmembrane α helical domains, causing a 60-75 % migration of the protein's predicted molecular weight on SDS-PAGE (Zou et al., 2005; Godoy et al., 2007).

Being expressed among humans, rats, mice and frogs, an overall SLC10A7 amino acid sequence identity of >85 % was identified. Typically, SLC10 members are seen to be taxonomically restricted to vertebrate species, however, it was identified that SLC10A7 also existed in vertebrates, bacteria, yeast and plant species, making this member the most widespread taxonomical membrane protein transporter of the SLC10 family. In addition, sequence identity of >20 % was exhibited between vertebrate and bacterial SLC10A7 proteins, which increased to 60 % when filtering for highly conserved residues and sequence domains in TMDs 5, 7, 9, and 10. This was significantly greater than any other sequence identity observed between SLC10A7 and any other SLC10 member (SLC10A1-SLC10A6) (Godoy et al., 2007).

Finally, a phylogenetic tree revealed SLC10A7 to cluster independently of the remaining SLC10 members (SLC10A1-SLC10A6), showing sequence identities of >20 % within each branch and <15 % between each branch. This further acknowledges the atypical attributes of this newly identified SLC10 family member, highlighting fundamental differences in tissue expression, transporter function, membrane topology and genomic organisation (Godoy et al., 2007).

1.7.2 SLC10A7 in Humans

From the years 2018-2020, further research into the possible functional roles of SLC10A7 have been suggested and successfully investigated among several research groups:

In 2018, Dubail and colleagues identified five unique novel homozygous mutations in the SLC10A7 gene, consequently leading to the phenotypic presentation of skeletal dysplasias (SD) with multiple dislocations, as well as amelogenesis imperfecta (AI). These deficiencies resulted in disruptions in intracellular calcium homeostasis and defective biosynthesis of the glycosaminoglycan (GAG) mechanism (Dubail et al., 2018).

Also in 2018, Ashikov and colleagues further classified this novel SLC10A7 protein, highlighting its unique involvement in the trafficking and sorting of glycoproteins and proteoglycans from the Golgi to the extracellular matrix, outlining fundamental roles in both bone mineralization and Golgi homeostasis. After the profiling of unique genomic and glycomic signatures of plasma glycoproteins in human individuals with hybrid variant N-glycans, two additional novel heterozygous mutations in SLC10A7 were uncovered, which resulted again in the severe phenotypes of AI and SD (Ashikov et al., 2018).

Later, in 2019, Laugel-Haushalter and colleagues, additionally analysed a novel SLC10A7 mutation which led to a milder and less grievous display of SD with AI (Laugel-Haushalter et al., 2019).

Most recently, in 2020, Karakus and colleagues outlined a possible novel mechanistic molecular role of SLC10A7 on the calcium homeostatic pathway in humans (Karakus et al., 2020).

The Reported Human SLC10A7 Mutations

The five newly identified homozygous SLC10A7 mutations uncovered by Dubail and colleagues, included two splice site deficiencies (at exon 9 (donor site) and 10 (acceptor site)); two missense substitution mutations in exon 3 (Leu74Pro) and 4 (Gly130Arg); and a premature stop codon in exon 7 (Gln172*). Three of the five gene disruptions led to null alleles, including both splice sites and premature stop codons. Both missense mutations led to severe SLC10A7 encoding disruptions including TM3 from a highly conserved amino acid missense mutation at Leu74Pro; and a non-specific protein domain from a highly conserved amino acid residue at Gly130Arg (Dubail et al., 2018).

Ashikov and colleagues additionally reported on two novel heterozygous SLC10A7 mutations, including a paternal genetic variant of a splice accepter site, resulting in the skipping of SLC10A7 exon 9, leading to a frameshift mutation of 29 amino

acid residues forming a premature stop codon (Ile241Arg*). A maternal missense mutation (Gly112Asp), causing a disruption in the hydrophobicity of TM4, and its deletion entirely (as predicted by TMHMM (Krogh et al., 2001)), was also discovered (Ashikov et al., 2018).

Moreover, Laugel-Haushalter and colleagues located a homozygous missense mutation (Pro303Leu) affecting exon 11, corresponding to TM10 of the SLC10A7 membrane transporter. This mutated proline residue was highly conserved among both vertebrate and non-vertebrate species, and its mutation to leucine was predicted to be both deleterious and disease causing by PPH2 (Adzhubei et al., 2010) and SIFT (Vaser et al., 2016) prediction programmes respectively. This was the first variant to be identified residing towards the end of the protein amino acid sequence, providing a possible explanation for a milder display of AI with SD in the affected individual (Laugel-Haushalter et al., 2019).

Finally, certain individuals presented with a complete lack of SLC10A7 cDNA expression, resulting in the complete loss of function of the SLC10A7 transporter (Ashikov et al., 2018).

A summary of these hSLC10A7 mutations are depicted in table 1.3 below.

Table 1.3: The eight reported novel human SLC10A7 mutations/variants.
N/A: Not Available, *: Premature Stop Codon
Table summarised and created from findings reported in Dubail et al., 2018; Ashikov et al., 2018; Laugel-Haushalter et al., 2019.

	Nucleotide Mutation	**Status**	**Protein Mutation**	**Location**	**Reference**
1	c.773+1G>A (Exon 9 skipping)	Homozygous	N/A	Intron 9	Dubail et al., 2018
2	c.774-1G>A (Exon 9+10 or 10 skipping)	Homozygous	N/A	Intron 9	Dubail et al., 2018
3	c.221T>C	Homozygous	Leu74Pro (L74P)	Exon 3	Dubail et al., 2018
4	c.335G>A	Heterozygous	Gly112Asp (G112D)	Exon 4	Ashikov et al., 2018
5	c.388G>A	Homozygous	Gly130Arg (G130R)	Exon 4	Dubail et al., 2018
6	c.514C>T	Homozygous	Gln172* (Q172*)	Exon 7	Dubail et al., 2018
7	c.722-16A>G (Exon 9 skipping)	Heterozygous	Ile241Arg* (I241R*)	Intron 8	Ashikov et al., 2018
8	c.908C>T	Homozygous	Pro303Leu (P303L)	Exon 11	Laugel-Haushalter et al., 2019

The Phenotypic Presentations of SLC10A7 Mutations in Humans

As a result of the specific SLC10A7 gene aberrations mentioned above, all affected patients developed severe phenotypes, including but not limited to: evolved carpal and tarsal bone ossification; altered vertebrae; pre- and postnatal short stature; spinal hyperlordosis; dislocated large joints; shortened long bones; facial dysmorphisms; below average intellectual retardation; abnormal head, spine and skeleton shape and sizes; luxation of knees; microretrognathia; SDs; monkey wrench display; moderate hearing disabilities (bilateral hypacusis); joint laxity; small epiphyses; altered cartilage and bone formation (reduced bone mineralisation (osteoporosis)) in the femoral neck and proximal femur; and severe disruptions in cardiac, respiratory and neurological systems. Furthermore, all patients suffered from a conserved hypomineralized/hypoplastic AI disorder, which presented with rough surfaced yellow-brown teeth and brittle/soft enamel (due to a disruption in enamel formation), with shortened and widely-spaced tooth crowns, exhibiting microdontia. Protruded incisors, a narrowed pharygeal tract and a lingual dysfunction were additional traits observed (Dubail et al., 2018; Ashikov et al., 2018; Laugel-Haushalter et al., 2019).

Additional debilitating disorders, specific to only certain individuals, included myopia gravis; chronic malabsorption; genua valga; long philtrum; submucous cleft palate; phimosis; mandibular hypoplasia; bilateral coxa valga; webbed neck; progressive scoliosis and kyphoscoliosis; round face; microetrognathia; dolichocephaly; brittle hair; short neck, nose and limbs; horizontal acetabulum; intrauterine and postnatal growth retardation; hypermetropia; strabismus convergens; brachymetacarpia; clubfeet; short tubular bones; macrocephaly; ptosis; sepsis; small thorax; inguinal hernia; flat face; large lilac wings; micrognathia; brachydactyly and rhizomelia hands and feet; cervical stenosis; hypoplastic ribs; femoral necks and clavicles; hypertelorism; joint hypermobility; prognathism; astigmatism; ballooned vertebral bodies; blue sclerae; short neck; pedes planovalgi; prominent eyes; pectus excavatum; and glaucoma (Ashikov et al., 2018; Laugel-Haushalter et al., 2019).

Figure 1.24 below outlines some of the above mentioned disastrous SLC10A7 deficient human phenotypic attributes.

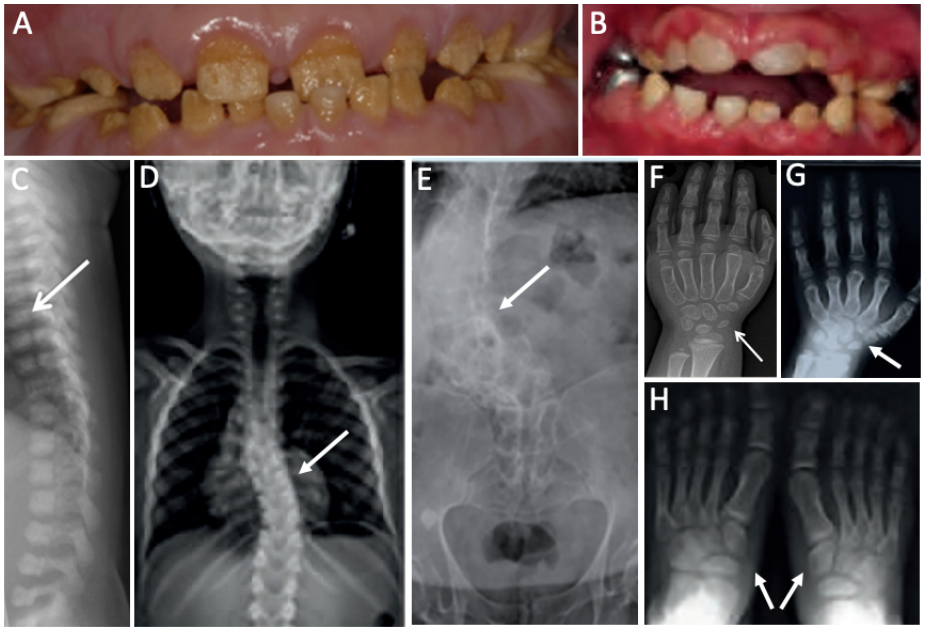


Figure 1.24: Mutations in human SLC10A7 result in severe phenotypes, including the development of skeletal dysplasias (SD) and amelogenesis imperfecta (AI). (A and B) Intra-oral photography highlighta detrimental disruptions in enamel formation (hypomineralised AI) and reduced tooth development, leading to the appearance of rough surfaced yellow-brown teeth in two separate 9 year old patients. (C) Spinal X-ray of a 1 month old patient presenting a curved spine with coronal clefts (arrow). (D and E) Radiographs of 9 (D) and 31 (E) year old individuals, displaying progressive scoliosis and severe right convexity thoraco-lumbar scoliosis, respectively (arrows). (F-H) Hand x-rays of 3 (F) and 4 (G) year old patients, and feet of a 9 year old patient (H), all displaying evolved and progressed carpal and tarsal ossification (arrows) This figure was generated from images included in Dubail et al., 2018 (A, C and F), Ashikov et al., 2018 (E), and Laugel-Haushalter et al., 2019 (B, D, G and H).

Disruptions in Protein Glycosylation and Glycosaminoglycan (GAG) Biosynthesis

In human cells, SLC10A7 has been observed to be localised to the cell membrane and Golgi secretory pathway, housing delicate molecular machinery for protein glycosylation processes. Specifically, SLC10A7 co-localised with cis-, medial- and trans Golgi markers, expanding on a fundamental association on post-Golgi vesicular transport of proteins involved in the evolution of bone and skeletal tissues, as well as a vital role in the mineralisation and regulation of the extracellular matrix (Ashikov et al., 2018).

Whereas perinuclear Golgi fluorescence was almost entirely localised to the biological cell membrane in control fibroblasts; fibroblasts exhibiting disturbed SLC10A7 expression barely flagged any fluorescence in the plasma membrane and instead were more prominently expressed in the intracellular membranes of the Golgi and vesicles. In SLC10A7 dysfunctional fibroblast cells, these findings suggest an aberration in the post-Golgi transport of glycoproteins in the secretory pathway (Ashikov et al., 2018).

In SLC10A7 mutated patient fibroblasts, a cytosolic mislocalisation of specific glycoproteins was identified, as well as an altered post-Golgi traffic of glycoproteins to the plasma membrane. These mutations were discovered to disrupt various stages of the N-glycosylation process of proteins, including a remodelling of glycans, leading to truncated, aberrant, N-glycoprotein structures. As a result this, there was an increase of truncated glycan levels lacking GlcNAc, as well as high-mannose glycans. Sialylation levels however, were decreased, even though no obvious deficiency in efficacy was discovered (Ashikov et al., 2018; Dubail et al., 2018).

Dubail and colleagues also identified novel disruptions to the SLC10A7 gene which led to alterations in the biosynthesis of GAGs. Although there were no real differences in the complete GAG levels observed in both control and SLC10A7 disrupted fibroblast samples, mutated SLC10A7 patients reported 2-fold lower heparan sulfate (HS) levels, which were also similarly reported in Slc10a7 -/- mice, presenting with 2.5-fold lower levels in the cartilage. The patterns of sulfation, as well as chain length, of these HS proteins, were seen to be unaffected in both Slc10a7 -/- and Slc10a7 +/+ mice, indicating that Slc10a7 -/- mice fundamentally result in alterations only to the production of HS proteins. Finally, chondroitin sulfate (CS) intensity in growth plates of Slc10a7 -/- and wild type controls, showed no differences. However, reduced homogenous levels in chondrocytes were noted in Slc10a7 -/- mice (Dubail et al., 2018).

1.7.3 SLC10A7 in Mice

The Phenotypic Presentations of SLC10A7 Deficient Mice

In mice, SLC10A7 mRNA was identified in all evolving stages of skeletal tissues, including the cartilage and growth plates in long bones. From early embryonic (E12.5) to postnatal (P0) stages, SLC10A7 was expressed in a multitude of locations including the heart trabeculae; cartilage; mandible; digits; spine; lung; epithelium of vertebrae; growth plate; incisors and molars; humerus and femur; chondrocytes; ossifying bones and teeth; oral mucous membrane; inner and outer dental epithelium cells; and ameloblasts and odontoblasts of developing teeth. Fascinatingly, analogous expression patterns were also observed in humans with specific SLC10A7 mRNA expression located in the heart, vertebrae, and cartilage of long bones; representing conserved localisation of SLC10A7 expression between vertebrate species (Dubail et al., 2018; Laugel-Haushalter et al., 2019).

Comparing mice born with an Slc10a7-/- double homozygous mutation to wild-type littermates highlighted the former to express considerably smaller morphologies, with lowered overall body mass and dysmorphic faces, including rounded heads and reduced naso-occipital attributes. Hindfeet were also smaller and presented with increased levels of evolved ossification. Although there were no presence of large joint dislocations, the lengths of frontal, nasal and occipital bones of the skull were significantly reduced, leading to detrimental reductions in the volume of all dental anatomical structures, including the incisors, mandible and molars, all seen to be missing the aprismatic external enamel layer. Furthermore, hypoplastic areas were also reported in the external prismatic enamel layer. However, although serious deficiencies have been reported, interestingly, no decrease in the thickness of enamel was observed, when again compared to wild-type littermates (Brommage et al., 2014; Dubail et al., 2018).

Figure 1.25 below highlights some of the above mentioned SLC10A7 deficient mouse phenotypic attributes.

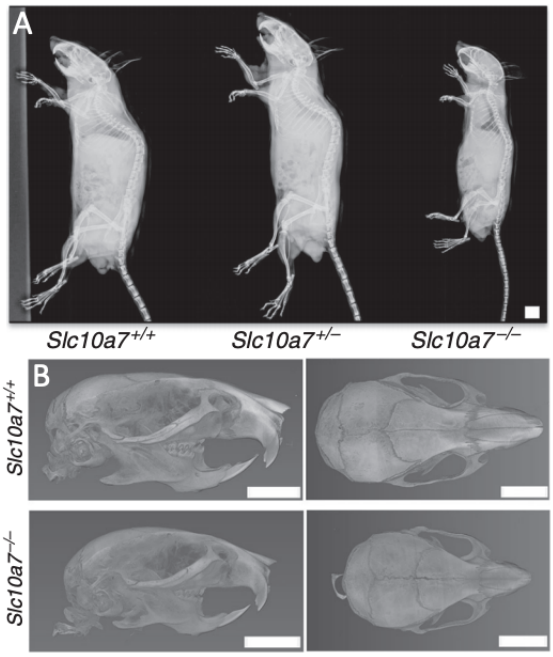


Figure 1.25: Slc10A7-/- double homozygous mice mutations result in skeletal dysplasias with skull aberrations. A) Naso-occipital lengths and body weight measurements of Slc10a7+/+, Slc10a7+/- and Slc10a7-/- mice at 8 weeks, illustrating a much smaller body and more rounded skull of the Slc10a7-/- phenotype. B) Skull width and length of 8 week old mice expressing Slc10a7+/+, and Slc10a7-/- phenotypes. Slc10a7-/- mice skulls and bodies are considerably smaller and present with reduced elongation compared to the Slc10a7+/+ phenotype.
This figure was generated from images displayed in Dubail et al., 2018.

Moderate SDs, ligamentous laxity, and reduced bone mass with shortened limbs were also noted. Long-bone microarchitechture was also deformed in mice lacking the SLC10A7 gene. These mice developed shorter and thicker long bones when compared to wild-type littermates. Enlarged distal condyles, shorter necks and prominent proximal trochanters were also exhibited by Slc10a7-/- femurs, as well as thinner growth plates. Since both the bone growth and growth plate of Slc10a7-/- mice were disrupted, the authors suggested distinct aberrations of the compositions of the extracellular matrix to be possible culprits; most probably resulting from reduced collagen and/or proteoglycan levels (Brommage et al., 2014; Dubail et al., 2018).

Figure 1.26 below displays some of the above mentioned SLC10A7 deficient mouse phenotypic attributes.

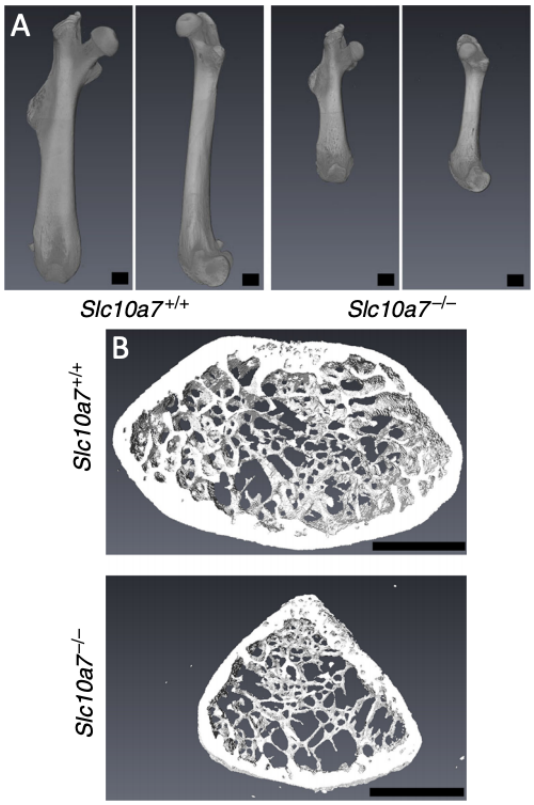


Figure 1.26: Mice with disrupted SLC10A7 genes portray long-bone micro- and macrostructure degenerations. A) A 3D reconstruction of femurs from 8-week old mice. Slc10a7-/- mice femurs strongly exhibit thicker and shortened lengths in comparison to their wild-type counterparts. B) A 3D section of distal femur metaphyses from 8-week old mice exhibiting both Slc10a7+/+ and Slc10a7-/- gene phenotypes. Again, significant differences in the width of the femurs are observed.
This figure was generated from images displayed in Dubail et al., 2018.

1.7.4 SLC10A7 in Zebrafish

The Phenotypic Presentations of SLC10A7 Deficient Zebrafish

Investigating SLC10A7 disruptions in zebrafish, reiterated the same phenotypes seen in human individuals, especially SDs. Here, Ashikov and colleagues identified a key association of SLC10A7 in developing bone mineralisation and cartilage, as well as the regulation of glycosylation, by motivating the transport of glycoproteins and proteoglycans to the extracellular matrix (Ashikov et al., 2018).

Zebrafish slc10a7 gene morpholino morphants, affecting the donor site of exon 2 of the solute carrier transporter, were created. Compared to control wild-typed zebrafish, morphant zebrafish injected with 8 and 12 ng/nl slc10a7-sp, exhibited a downwards bend of the teeth cartilage. In 12 ng/nl injected slc10a7-sp morphant zebrafish, a detrimental display of whole-bodied edema, curved eyes and tail, and a decreased head size were reported (Ashikov et al., 2018).

The subsequent addition of Alcian blue stain, highlighting GAG expression, depicted a complete loss of cerathobranchial four and decreased levels of Meckel's cartilage in morphant slc10a7-sp injected zebrafish. Alizarin red stain was also added, outlining tissue calcium deposits, and presented a loss of operculum and palate cartilage, decreased cerathobranchial 5, cleithrum, entopterygoid and notochord levels; and widened palatal skeleton, in 8 ng/nl injected morphants. In higher dosed 12 ng/nl slc10a7-sp injected zebrafish morphants, no noticeable bone mineralisation was identified (Ashikov et al., 2018).

Figure 1.27 below displays some of the above mentioned SLC10A7 deficient zebrafish phenotypic attributes.

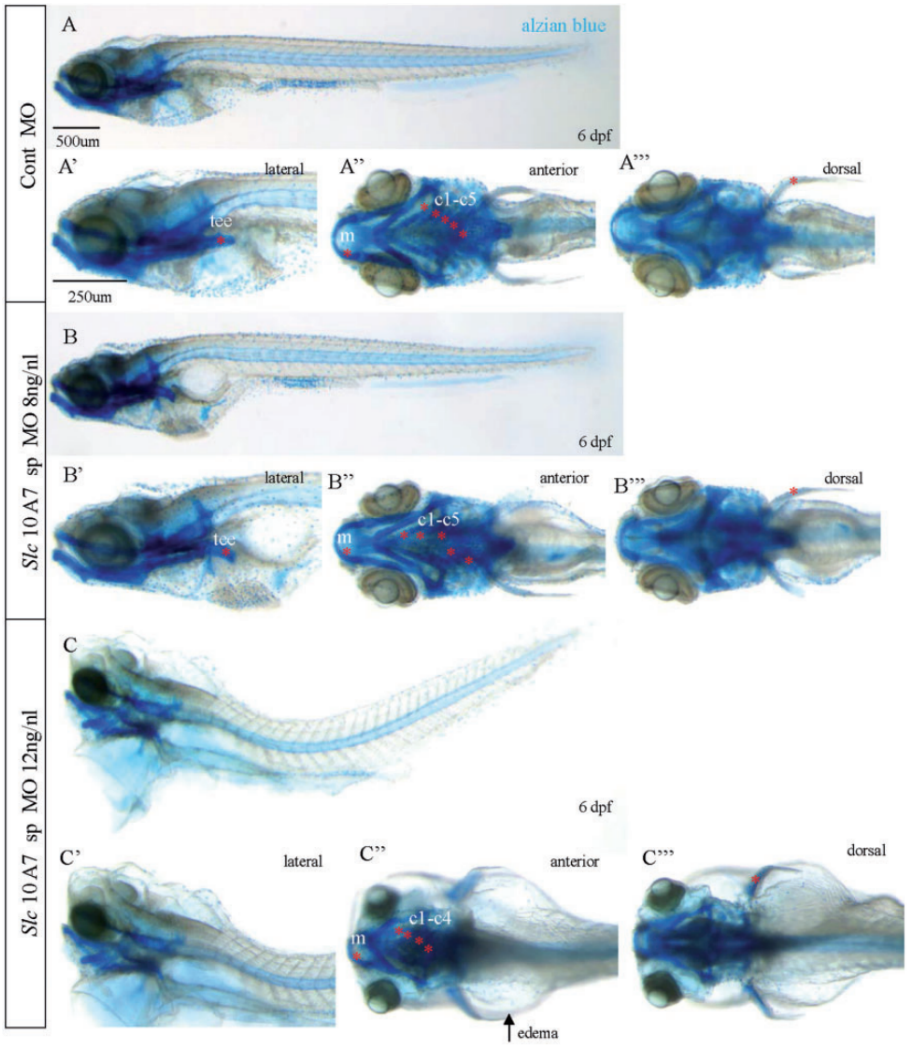


Figure 1.27: 6dpf embryo zebrafish, deficient in slc10a7, present with severe skeletal phenotypes. slc10a7-sp morpholino morphants and a control morpholino were injected with Alcian blue stain at 8 and 12 ng/nl. (A-C) Lateral view of complete embryo. (A'-C') Lateral view close-up of head. (A"-C") Anterior view close-up of head. (A"'-C"') Dorsal view close-up of head.
This figure was taken from Ashikov et al., 2018.

1.7.5 SLC10A7 in Yeast

A primary physiological and functional role of the SLC10A7 transporter in the regulation of intracellular calcium homeostasis was first proposed by Jiang and colleagues in 2012 (Jiang et al., 2012) regarding a *Candida albicans* SLC10A7 transporter homologue: Rch1p. In 2016, Zhao and colleagues reiterated the same physiological presentation in *Saccharomyces cerevisiae* (Zhao et al., 2016).

The RCH1 (regulator of Ca^{2+} homeostasis 1) gene, contains a 411/434 amino acid protein with 33.7 % sequence identity to human SLC10A7 and is characterised by ten TMDs with a cis-facing intracellular localisation of both N- and C-termini. This protein is homologous to SLC10A7 and similarly localised to the cell membrane (Jiang et al., 2012; Zhao et al., 2016).

An amino acid sequence alignment plot depicted below in figure 1.28 analyses in greater detail the significant similarities and differences of these novel *C. albicans* and *S. cerevisiae* SLC10A7 yeast homologues. Here, detailed comparisons to the human NTCP, ASBT and SLC10A7 members, as well as the crystallised bacterial $ASBT_{NM}$ and $ASBT_{Yf}$ homologues, show an overall strong conservation of amino acid residues, albeit significant differences in residues involved in the binding of sodium ions to the crystallised bacterial ASBT members are also present.

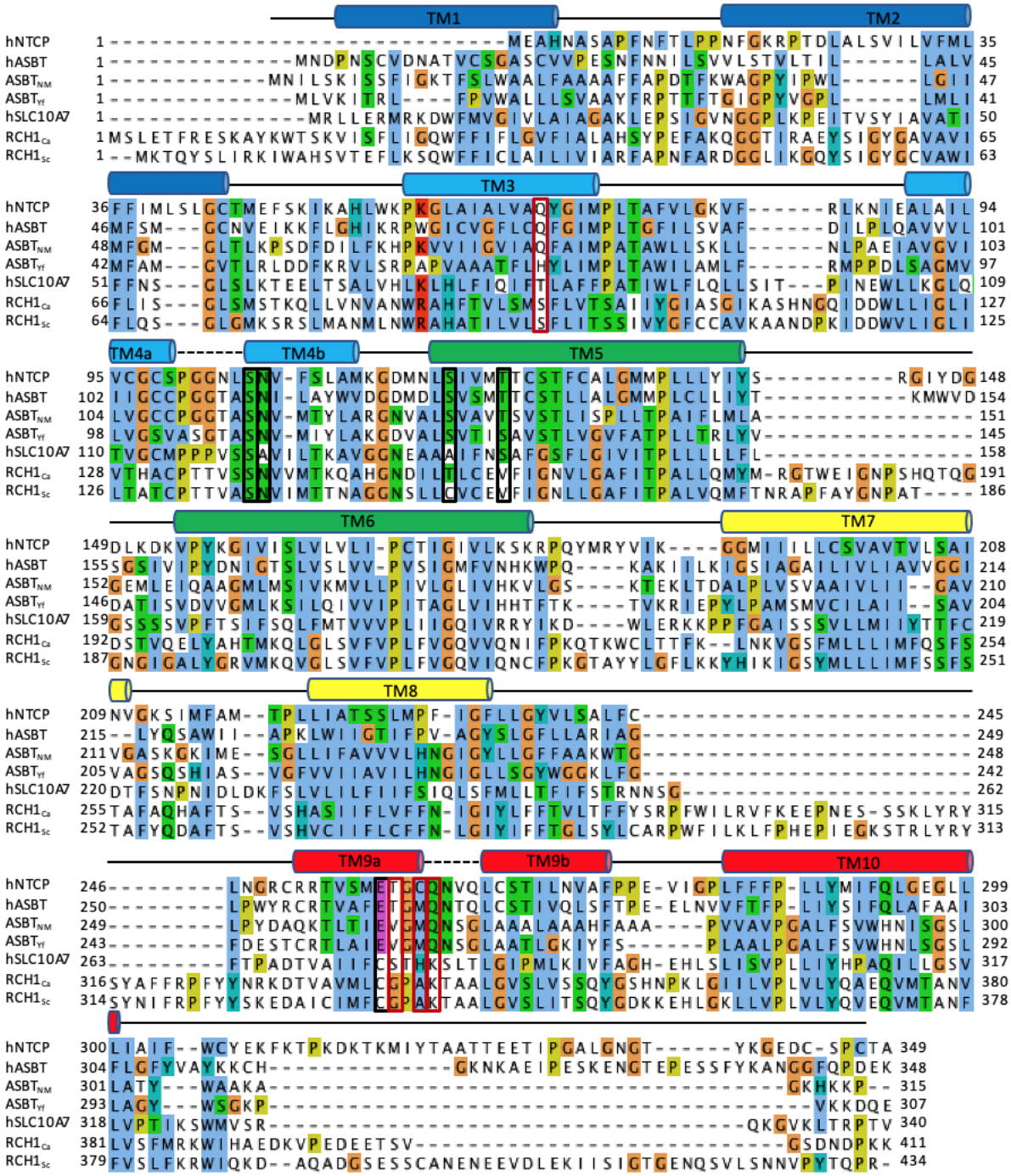


Figure 1.28: RCH1 *C. albicans* and *S. cerevisiae* bacterial homologues; human NTCP, ASBT and SLC10A7; and ASBT$_{NM}$ and ASBT$_{Yf}$ bacterial homologues multiple sequence alignment. Respective amino acid sequences were analysed in Jalview (Waterhouse et al., 2009), and specifically aligned with Clustalx (Larkin et al., 2007) and MuscleWS (Edgar, 2004). The established ASBT$_{NM}$ transmembrane (TM) domains (as previously reported by Hu et al., 2011) are presented above the sequences in cylinders, depicting the structural TM1-TM10 domains from the N-terminal (blue) to C-terminal (red). Na^+ binding site amino acid residues, implicated for Na1 and Na2, are depicted in red and black boxes, respectively.
Clustalx (Larkin et al., 2007) notation of residues: Blue: Hydrophobic, Red: Positive charge, Magenta: Negative charge, Green: Polar, Pink: Cysteines, Orange: Glycines, Yellow: Prolines, Cyan: Aromatic, and White: Unconserved.

Typically, in yeast cells, low Ca^{2+} cytosolic concentrations are regulated via Ca^{2+} sequestrators and transporters (such as the Vcx1p Ca^{2+}/H^{+} exchanger and Pmc1p pump), located in the cell and organelle membranes. However, when intracellular concentrations of Ca^{2+} are too elevated, cells respond to this stress by activating the Ca^{2+}/calcineurin signalling pathway. This regulation of internal Ca^{2+} concentration is imperative to a multitude of cellular systems (including cell survival, morphogenesis, cell growth, antifungal drug tolerance, and virulence), whereby highly unregulated levels are extremely toxic and detrimental to cell survival. Therefore, an effective and crucial auto-regulatory system is necessary to efficiently coordinate and maintain calcium homeostasis (Jiang et al., 2012).

Through mutational gene knockout studies, it was seen that a complete homozygous mutation in RCH1 resulted in increased hypersensitivity of *C. albicans* cells to high extracellular concentrations of Ca^{2+}, as well as an increased tolerance to Li^{+} and azoles/antifungals (including econazole, fluconazole, terbinafine, ketoconazole, clotrimazole and voriconazole). However, these cells became susceptible to nystatin, amphotericin B and the cell wall perturbing SDS. This gene deletion disturbs the complex calcium homeostasis mechanism, by which cells become increasingly prone to Ca^{2+} influx, leading to increased levels of uptake and thereby greater accumulation of cytosolic Ca^{2+} levels. Further investigations deduced no alterations to the rate of elimination of intracellular Ca^{2+} levels, and could be explained by significant increases in Ca^{2+} influx across the plasma membrane of Rch1p homozygous mutants (Jiang et al., 2012).

Furthermore, these increasing levels of intracellular Ca^{2+} resulted in the activation of the Ca^{2+}/calcineurin mechanism. This phenomenon was observed when a lack of RCH1 resulted in an increased expression of CaUTR2 (a downstream member of the signalling pathway), as a consequence of a greater accumulation of extracellular Ca^{2+}. However, upon addition of cyclosporin A (a specific inhibitor of calcineurin), Ca^{2+} hypersensitivity, caused by the deletion of the RCH1 gene, was entirely abolished. Similarly, addition of magnesium ions (Mg^{2+}) also reiterated this same response (Jiang et al., 2012).

Interestingly, Jiang and colleagues also showed both TM9 and TM10 domains of the Rch1p protein not to be essential to the subcellular localisation and salt/drug sensitivity of the transporter. TM8, however, was shown to be crucial in the stabilisation, cellular function and subcellular localisation of Rch1p, but required the support of other TMDs to achieve this successfully (Jiang et al., 2012).

It is very unlikely for Rch1p to be involved in Ca^{2+} efflux and instead has been suggested by Jiang and colleagues to play a critical role as a negative regulator of the Ca^{2+} influx mechanism, in which an efficient feedback inhibition of Ca^{2+}

influx is initiated upon increasing extracellular Ca^{2+} levels. This is the very first evidence for such a role from a homologous SLC10A7 protein, thereby expanding the functional attributes of the SLC10 carrier family (Jiang et al., 2012).

In *C. albicans*, there are a minimum of two Ca^{2+} influx systems: the low-affinity Ca^{2+} uptake system (LACS); and the high-affinity Ca^{2+} uptake system (HACS - involving the Mid1p and Cch1p Ca^{2+} channels). Alber and colleagues in 2013 (Alber et al., 2013) investigated in greater detail the possible functional associations of Rch1p with both these Ca^{2+} influx processes.

CaCch1p is a major member of the HACS process and therefore of interest to assess if any association with Rch1p is present/possible. Upon further research, it was identified that CaRch1p failed to associate with CaCch1p in a functional manner. Therefore, Alber and colleagues stated the CaRch1p to maintain Ca^{2+} homeostasis via an independent fashion to CaCch1p and hence, suggested CaRch1p to be a novel negative regulator of Ca^{2+} homeostasis, playing a key role in the LACS, rather than the HACS system (Alber et al., 2013).

Furthermore, in the LACS system of *S. cerevisiae*, Ca^{2+} influx was demonstrated to be abolished by the addition of low extracellular Mg^{2+} concentrations (Muller et al., 2001; Cui and Kaandorp, 2006). Since the Ca^{2+} hypersensitivity of *Candida albicans* cells lacking the RCH1 gene is also entirely reversed and revitalised when Mg^{2+} is introduced (Jiang et al., 2012), this outlined additional evidence supporting a significant correlation of Rch1p associating with the LACS system (Alber et al., 2013; Zhao et al., 2016).

An overview of both these systems in *C. albicans* and possible functional associations of Rch1p in maintaining Ca^{2+} homeostasis is illustrated in figure 1.29 below.

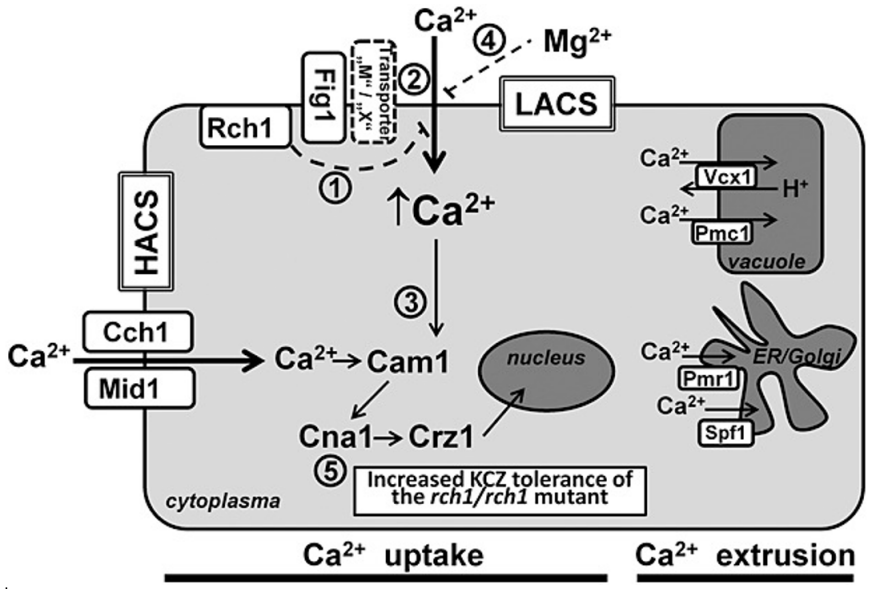


Figure 1.29: The proposed role of Rch1p in maintaining Ca^{2+} homeostasis in *C. albicans*.
1) Negative regulation of Ca^{2+} influx through the LACS system.
2) Deletion of Rch1p leads to greater uptake of Ca^{2+}.
3) The Ca^{2+}/calcineurin process is initiated upon cytosolic accumulation of Ca^{2+}.
4) Extracellular Mg^{2+} inhibits the influx of Ca^{2+} and can reverse the hypersensitivity to Ca^{2+} following deletion of the RCH1 gene.
5) rch1/rch1 homozygous double mutation leads to increased cytosolic Ca^{2+} influx, but also increases tolerance to well established antifungals and azoles.
Ca^{2+} extrusion is vital to cell survival in an over-abundnace of intercellular Ca^{2+} levels. Low Ca^{2+} concentrations are maintained via the Vcx1p Ca^{2+}/H^{+} exchanger and vacuolar Pmc1p Ca^{2+} pump in *C. albicans* cells; and the Spf1p and Pmr1p P-type Ca^{2+} ATPases in the ER and Golgi.
This figure was taken from Alber et al., 2013.

Evidence for a direct/indirect involvement of Rch1p on the regulation of calcium homeostasis and interaction with specific targets are unfortunately not yet identified, and as of present speculative. As such, an in-depth understanding of the molecular mechanism of action for this SLC10A7 homologue in yeast is not currently provided.

1.7.6 SLC10A7 is a Negative Regulator of Intracellular Calcium Signalling in Non-excitable Eukaryotic Cells

In eukaryotic cells, Ca^{2+} is regarded as being one of the most adaptable secondary messengers involved in a multitude of cellular systems, including vesicle exocytosis, gene expression, muscle contraction and cell growth/proliferation. Compared to excitable cells (where Ca^{2+} influx is exhibited by voltage-gated calcium channels), non-excitable cell Ca^{2+} intake is operated through a store-operated calcium entry (SOCE) mechanism, mediated by calcium release-activated channels of the cell membrane (Karakus et al., 2020).

Two well established SOCE members exist, including the canonical transient receptor potential (TRPC) and Orai1 channel proteins. The latter is involved in the activation and transport of stromal interaction molecules (STIM1), and shares a similar function to TRPC behaving in both a STIM1-independent and/or -dependent fashion. In reaction to a depletion of Ca^{2+} levels in the endoplasmic reticulum (ER), SOCE calcium channels are stimulated to open. This initiates the intracellular release of Ca^{2+} stores and activates STIM1 (the ER Ca^{2+} sensor), which forms a complex association with Orai1 to form a strong and stable selectivity for Ca^{2+} binding. This STIM-Orai complex is stabilised by the CRAC regulator 2A channel (CRACR2A); and inactivated by the SOCE-associated regulator factor (SARAF), inhibiting the intracellular intoxication of the cell as a result of too great an increase in Ca^{2+}. Finally, refilling of Ca^{2+} stores and a return to normal calcium homeostasis levels, results in a negative feedback regulation mechanism, which deactivates SOCE channel proteins (Karakus et al., 2020).

In order to regulate low resting cytosolic calcium levels, the sarcoplasmic/endoplasmic reticulum (SR/ER) ATPase (SERCA) is equipped in translocating Ca^{2+} from the cytosol to SR/ER lumen. The regulation of this transporter is exhibited by the extracellular signals and cellular needs of sarcolipin and phospholamban proteins, which when activated, lower the calcium affinity of SERCA (Karakus et al., 2020).

Through the generation of SLC10A7 overexpressing (HEKP7+tet), knockout (HAP1-KOP7) and wild-type control (HAP1) cell lines, Karakus and colleagues successfully discovered a negative correlation and suppression of this SOCE pathway with positive SLC10A7 expression. Cytosolic calcium signals, as well as intracellular calcium influx and uptake, were significantly decreased in the presence of an overexpression of SLC10A7, compared to control and SLC10A7 knockout cell lines. In contrast, deletion of this SLC10A7 gene, led to a greater total amount of intracellular calcium ions, when compared to control cells (Karakus et al., 2020).

Fascinatingly, the authors also uncovered strong individual colocalisation associations between SLC10A7 and SERCA2, STIM1 and Orai1, of which STIM1-

SLC10A7 showed the greatest intensity. In light of this, the consensus was that the overexpression of SLC10A7 hindered the SOCE mechanism, leading to disrupted uptake and removal of Ca^{2+}, via a possible deactivation and inhibition of Orai1, SERCA2 and/or STIM1. However, evidence for a direct/indirect association of SLC10A7 (and specific molecular mechanism of action) with the previously mentioned SOCE members remains unclear, and requires further experimental testing (Karakus et al., 2020).

Previously identified SLC10A7 mutations (from Dubail et al., 2018, Ashikov et al., 2018, and Laugel-Hauschalter et al., 2019), including the point mutations L74P, G112D, and G130; and the exon skipping variants 9, 9+10 and 10, expressed in HEK293 cells, failed to show any inhibitory action on the influx and uptake of calcium ions, compared to SLC10A7 overexpression, outlining a loss of function in the transporter as a result of severe genomic and protein disruptions.

Figure 1.30 below displays a possible molecular association of SLC10A7 on the SOCE mechanism, as described above.

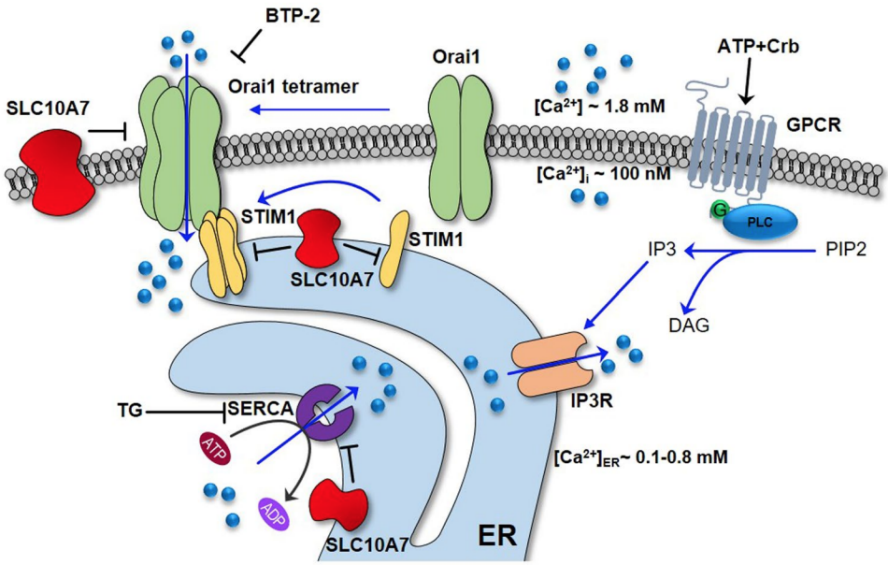


Figure 1.30: The proposed molecular mechanism of calcium signalling in non-excitable eukaryotic cells. The stimulation and activation of IP3 and G protein-coupled receptors (GPCR), initiates the IP3 receptor (IP3R), resulting in the depletion of ER Ca^{2+} stores. Depletion is also achieved via ionomycin and TG, albeit by separate mechanisms entirely. STIM1 senses the sudden decline in ER calcium levels and oligomerises to form a complex association with Orai1, initiating store-operated calcium entry (SOCE) channel opening. SERCA ATPase transporters pump Ca^{2+} back in the ER, thereby refilling the calcium stores. It is proposed SLC10A7 negatively regulates Orai1, SERCA2 and/or STIM1 through a direct protein-protein interaction, though this has not been experimentally observed.
TG: SERCA inactivator; BTP-2: SOCE inhibitor.
This figure was taken from Karakus et al., 2020.

1.7.7 Other Possible Functional Roles of the SLC10A7 Transporte in Humans

Inverted Insertional Translocation of Chromosomes in Hypergonadotropic Hypogonadism

Tzschach and colleagues reported on a chromosomal aberration in a 30 year old male patient, presenting with azoospermia (impaired spermatogenesis) and primary hypergonadotropic hypogonadism. This patient suffered from primary testicular dysfunction; elevated luteinizing (LH) and follicle-stimulating hormones (FSH), and low serum testosterone levels. Further investigation revealed a non-reciprocal inverted insertional (inv ins) translocation of chromosome 4q28.3q31.22 into chromosome 2p24. A breakpoint-spanning clone was additionally identified in chromosome 4q31.22 coding for the genes LSM6 and SLC10A7 (Tzschach et al., 2009).

Chromosome breakpoints can affect genes in one of two ways: they can either interfere with associated genes directly, or initiate the expression of neighbouring genes. This chromosome aberration therefore highlights a new gene locus responsible for primary hypogonadism in either 2p24 or 4q28-31, affecting three specific genes (PCDH10, LSM6 and SLC10A7), all of which still remain elusive in their functional roles and involvement in associated diseases (Tzschach et al., 2009).

Alternative Splicing Events in Colorectal Cancer

Recent research has established a strong link in aberrant alternative splicing (AS) events leading to the formation, pervasion and advancement of cancers. However, regarding colorectal carcinogenesis, these events remain evasive. In 2018, Liu and colleagues uncovered specific genes (SLC10A7, NOL8, PPAT, ALDH4A1 and PSMD2) associated in AS events, playing critical roles in the carcinogenesis and evolution of colorectal cancer (Liu et al., 2018).

Typically, AS genes lead to multiple transcripts, frequently entailing specific proteins with aberrant and sometimes opposite functional abilities. Under normal conditions, these AS events result in specific protein isoforms to be involved in complex biological mechanisms. However, if disturbed, specific tumour cells can produce disordered proteins with distinctly different functional abilities, resulting in tumourgenesis (Liu et al., 2018).

Regarding SLC10A7, Liu and colleagues identified an alternate terminator AS event, resulting in a different terminating exon to be encoded. It is unclear how this is involved in the development to colorectal cancer, but is assumed to lead to a disruption in the formation of the cytoskeleton and/or adhesion of cells. Most

commonly, AS events leading to these disruptions have previously been regarded to be essential hallmarks of cancers. Therefore, these novel aberrant AS events and their associations in colorectal cancer will provide critical insights into their complex initiation and development, as well as provide specialised therapeutic targets (Liu et al., 2018).

1.8 Aims and Objectives

The major aims and objectives of this PhD study are outlined in greater detail below:

- **Primary aim: To solve the 3D X-ray crystal structures of the ASBT and SLC10A7 bacterial orphan transporter homologues, in order to enhance our understanding of the human variants and help in the design of novel pharmaceutic drugs.**

  - **ASBT:** Previously crystallised bacterial homologues ($ASBT_{NM}$ (Hu et al., 2011) and $ASBT_{Yf}$ (Zhou et al., 2014)) have yielded monomeric ten TMD transporters, which differ significantly to both the amino acid sequence and alpha helical topology of the human ASBT membrane protein. A novel investigation of the ASBT *Leptospira biflexa* bacterial homologue however, has predicted a nine TMD topology (lacking TM1 helix and expected to orientate in an extracellular N-terminal and intracellular C-terminal fashion), which is in strong agreement to the human ASBT transporter, both of which are expected to form higher functional oligomeric assemblies (typically forming a dimer) to drive solute translocation (Noppes et al., 2019). As such, we aimed to crystallise this novel transporter to amplify or knowledge of the human homologue.

  - **SLC10A7:** As was previously seen in figure 1.11 above, the human SLC10A7 solute transporter depicts significant alterations to the conservation of amino acids categorically involved in the binding of Na^+ ions in the $ASBT_{NM}$ structure (Hu et al., 2011), which are essential to drive the translocation of substrates across the plasma membrane. We hypothesise the human and bacterial homologues of this transporter to alter their access mechanisms in a manner separate to the use of Na^+ ions, and with the presence of a positively charged Lysine residue at $ASBT_{NM}$ Q264, we suspect a possible transport operation to occur via the use of H^+ ions (protons). We therefore aimed to solve the crystal structures of the bacterial homologues of this fascinating membrane protein, in order to shed some much needed light on the conformational changes and workings on this unique SLC10 member.

- **Secondary aim: To characterise the functional roles of the ASBT and SLC10A7 bacterial orphan transporter homologues, to greatly enhance our knowledge of substrates/solutes involved in the alternating access mechanism of translocation and physiological functions they associate with.**

  - **ASBT:** The human ASBT transporter is fundamentally associated in the transport and re-uptake of bile acids from the intestines. Bacterial homologue transporters however, will certainly not share this physiological role. We therefore aimed to investigate alternative (potentially native) solutes which are transported across the cell membrane via the $ASBT_{Lb}$ membrane protein.

  - **SLC10A7:** It is proposed the human SLC10A7 member to be functionally involved in the negative regulation of calcium influx homeostasis. We believe this to also be reiterated in bacterial homologues. We therefore wished to investigate the binding/transport of potentially native substrates involved in the homeostasis of calcium as well as test the hypothesis of a hydrogen (proton) mediated secondary active transport mechanism.

Chapter 2

Biochemical and Biophysical Materials and Methods

Structural and functional characterisations of a membrane protein offer challenging prospectives, especially within the fields of molecular and structural biology, and specifically regarding in-vitro membrane protein over-expression, cell extraction, purification and crystallisation (Engel and Gaub, 2008; Cournia et al., 2015). However, these are fields that can generate substantially crucial information on the molecular assemblies and mechanisms of the function of membrane proteins, which can significantly help drive the design of vital drugs to target often devastating human physiological disorders (Balakrishnan and Polli, 2006; Claro da Silva et al., 2013).

As can be seen in figure 2.1 below, we designed an experimental strategy with the fundamental aim of characterising the bacterial homologues of the ASBT and SLC10A7 members, both structurally and functionally. Bacterial homologues were selected due to their tendency to present with far better overexpression levels when compared to eukaryotic membrane proteins. However, we acknowledge that key structural differences will arise, namely the loss of glycosylation sites and altered protein-lipid associations present in eukaryotic species, which can play significantly different alterations to the functionality of the membrane protein.

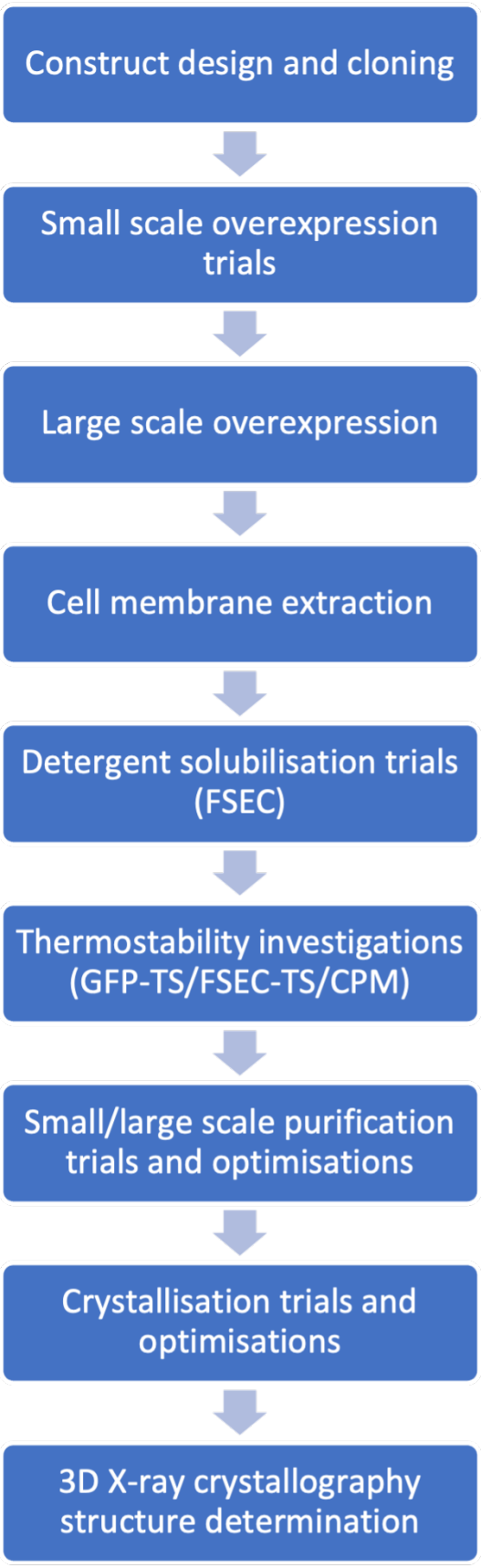


Figure 2.1: The PhD experimental strategy.

2.1 Expression and purification

2.1.1 The NEB5α and Lemo21(DE3) bacterial strains

NEB5α

Bacterial membrane protein DNA construct plasmids were cloned, replicated and extracted successfully with NEB5α chemically competent cells (New England BioLabs Inc.). This DH5α derived *Escherichia coli* cell line allowed for a high efficiency rate of DNA transformation (>1-3 $\times$ 10^9 cfu/µg pUC19 DNA) and was both T1 phage (fhuA2) resistant and endA1 deficient. This cell line was therefore used to produce high-quality DNA plasmid replication.

Lemo21(DE3)

Bacterial membrane protein overexpression was successfully achieved in Lemo21(DE3) chemically competent cells (New England BioLabs Inc.). This BL21(DE3) derived *Escherichia coli* strain was suited for tunable T7 expression of challenging membrane, toxic and insoluble proteins. Resistant to T1 phage(fhuA2), and proteases Lon and OmpT, efficient DNA transformation rates occured at >1-3 $\times$ 10^7 cfu/µg pUC19 DNA. In addition, the cell line had a p15A origin of replication, exhibiting resistance to chloramphenicol. This cell line was therefore used to yield optimum expression levels of membrane proteins.

2.1.2 Preparation of chemical competent bacterial cells

100 ml Luria Broth/Lysogeny Bertani (LB) (with appropriate antibiotics for selectivity (NEB5α: None; Lemo21 (DE3): 30 µg/ml Chloramphenicol)) was inoculated with a single colony of bacterial strain and incubated at 37 °C and shaken at 180-200 rpm, until an $OD_{600\,nm}$ 0.5 was reached. The bacterial growth culture was split into two 50 ml aliquots and incubated on ice for 10 min. Both aliquots were centrifuged at 2600 rcf at 4 °C for 10 min and pellets gently resuspended in 10 ml ice cold 0.1 M $CaCl_2$. Resuspended pellets were further incubated on ice for 10 min, before centrifuged and pelleted again at 2600 rcf at 4 °C for 10 min. Bacterial cell pellets were resuspended in 2 ml ice cold 0.1 M $CaCl_2$ with 20 % glycerol, and again further incubated on ice for 10 min. 50 µl aliquots of resuspended cells were evenly distributed into microtubes, immediately snap frozen in liquid nitrogen, before being stored at −80 °C.

2.1.3 Bacterial transformations and overnights

Bacterial transformations

50 µl chemically competent bacterial cell aliquots were thawed on ice for 10 min, before 0.5 µl of >100 ng/µl template DNA was added and mixed 5-10 times. Aliquots were incubated on ice for 30 min and heat shocked at 42 °C for 10 s (Lemo21(DE3)) or 30 s (NEB5α). After heat shock, cells were incubated on ice for a further 5 min before 950 µl Super Optimal broth (SOC media) with Catabolite repression, was added. The mixture was then incubated at 37 °C and 900 rpm, on an eppendorf Thermomixer comfort machine, for 1 h. 75-100 µl transformation culture was plated and streaked onto appropriate LB agar antibiotic selection plates (for NEB5α: 50 µg/ml Kanamycin; Lemo21(DE3): 50 µg/ml Kanamycin and 30 µg/ml Chloramphenicol) and incubated overnight at 37 °C.

Bacterial overnights

For bacterial overnight growths, a 5 ml aliquot of LB media, with appropriate antibiotic selections (50 µg/ml Kanamycin and/or 30 µg/ml Chloramphenicol) was inoculated with a single bacterial colony and incubated overnight for (15-20 h) at 37 °C and 180-200 rpm.

2.1.4 DNA cloning

Many of the analysed SLC10A7 constructs (including *Escherichia coli* full-length and R316 C-terminal truncated, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Pseudomonas putida*) had already previously been cloned in a pWaldo-GFPe vector, entailing a green fluorescent protein (GFP) fusion tag, 8x Histidine tag and TEV cleavable site. All these constructs were successfully re-cloned into a 3C cleavable tag entailing pWaldo-GFPe vector, yielding improved downstream purification procedures. Additional SLC10A7 constructs (including *Asticcacualis taihuensis*, *Cupriavidus necator*, and *Vibrio breoganii*), as well as the ASBT *Leptospira biflexa* construct (cloned by Oliver Huxley - MBio student 2019) had their g-Block Gene Fragments subsequently PCRed and also successfully cloned into a pWaldo-GFPe 3C cleavable vector. This vector is derived from the pET28(+) vector, with protein overexpression regulated by the transcriptional control of the T7 promoter, with kanamycin selectivity. The 3C protease cleavable recognition site is located at the interface of the membrane protein and GFP-His8 tag, allowing for its successful excision from the expressed protein. A much more detailed protocol for the cloning in the pWaldo 3C-GFP-His8 vector, and subsequent overexpression and purification procedures can be found in (Drew et al., 2006).

2.1.5 DNA agarose gel electrophoresis

DNA samples were separated into different molecular weight fragments on a 0.8 % (w/v) agarose gel. 0.8 g agarose powder was dissolved in 100 ml 1 × TAE buffer (40 mM Tris acetate, 1 mM EDTA), and microwaved until all agarose was fully dissolved. Upon cooling to 40-50 °C, 10 µl GelRed stain was added, mixed and poured into a gel setting kit with a comb in place. Once fully set, the gel was placed in a Geneflow gel tank and filled with 1 × TAE buffer. Quick-Load Purple 1 kb DNA Ladder and prepared DNA samples were loaded into the wells and run at 100 V for 45 min. Visualisation and imaging of the gel was performed with a Syngene G:Box gel illuminator analysis system, using ethidium bromide as the fluorescent illumination.

2.1.6 DNA extraction minipreps

DNA plasmids were successfully cloned and extracted from bacterial overnights via the Thermo Scientific GeneJET plasmid miniprep kit, following the manufacturer's manual instructions and guidelines.

2.1.7 DNA sequencing and plasmid construct validation

DNA plasmids, extracted via miniprep kits, were sent to eurofins genomics (previously known as GATC Biotech) for sequencing. 5 µl of (80-100 ng/µl) DNA template was mixed with 5 µl of 5 µM forward/reverse primers. Sequencing results were uploaded on a Warwick University database a few days later. The analysis and interpretation of results was conducted via SnapGene Viewer, whereby linear DNA sequence was extracted and converted to a protein sequence via the Expasy Translate portal. A BlastP search was finally conducted, comparing both the results from protein sequencing and known associated protein sequences. An overview of the DNA primers used in the sequencing of constructs is depicted below in table 2.1, and refer to the T7 forward primer, and GFP reverse primer.

Table 2.1: DNA sequencing primers used to quality check cloned constructs.

Primer Name	**Sequence**
T7 Promoter	TAATACGACTCACTATAGGG
pWaldo GFP end terminator	GAAAAGTTCTCCTCCTTTGCT

2.1.8 Bacterial expression media: PASM-5052 autoinductive media

Bacterial membrane protein overexpression was conducted in the PASM-5052 autoinductive media. All growth components of this media followed in detail the protocol from Lee et al., 2014, which is an updated and modified version of (Studier, 2005). The components briefly consist of the following:

20 × NPS (NPS = 100 mM PO_4, 25 mM SO_4, 50 mM NH_4, 100 mM Na, 50 mM K

Table 2.2: 20 × NPS solution recipe.

Reagent	Volume for 1 l	Concentration
Water	900 ml	/
Anhydrous $(NH_4)_2SO_4$	66 g	0.5 M
Anhydrous KH_2PO_4	136 g	1 M
Anhydrous Na_2HPO_4	142 g	1 M

Compounds were added in order, and stirred until fully dissolved. The 20 × NPS solution in sterile water gave a pH of roughly 6.2. The solution was autoclaved for 15 min at 121 °C and 15 psi.

50 × 5052 (5052 = 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose)

Table 2.3: 50 × 5052 solution recipe.

Reagent	Volume for 1 l
Glycerol	250 g
Water	730 ml
Glucose	25 g
α-lactose	100 g

Compounds were added in sequence, with the glycerol weighed first due to its highly viscose nature. The solution was stirred until all compounds were fully dissolved. The α-lactose took >2 h at room temperature to dissolve, but was quickened upon heating in a microwave. The solution was filter sterilised.

1 M $MgSO_4$

24.65 g $MgSO_4.7H_2O$ was dissolved in 100 ml sterile water and autoclaved for 15 min at 121 °C and 15 psi.

1000 × Trace metals

To 36 ml sterile water, the following reagents were added:

Table 2.4: 1000 × trace metals recipe.

Volume	**Compound**	**Anhydrous molecular weight (MWt)**	**1 × concentration for trace element**
50 ml	0.1 M $FeCl_3.6H_2O$ (Dissolved in 0.1 M HCl)	270.3	50 mM
2 ml	1 M $CaCl_2$	110.99	20 mM
1 ml	1 M $MnCl_2.4H_2O$	197.91	10 mM
1 ml	1 M $ZnSO_4.7H_2O$	287.56	10 mM
1 ml	0.2 M $CoCl_2.6H_2O$	237.95	2 mM
2 ml	0.1 M $CuCl_2.2H_2O$	170.49	2 mM
1 ml	0.2 M $NiCl_2.6H_2O$	237.72	2 mM
2 ml	0.1 M $Na_2MoO_4.5H_2O$	241.98	2 mM
2 ml	0.1 M $Na_2SeO_3.5H_2O$	263.03	2 mM
2 ml	0.1 M H_3BO_3	61.83	2 mM

Each metal solution was prepared individually and autoclaved for 15 min at 121 °C and 15 psi. Upon mixing, a precipitate forms on addition of $Na_2SeO_3.5H_2O$, but this clears with time and subsequent addition of remaining reagents. The solution was filter sterilised and stored at 4 °C.

25 mg/ml methionine

25 mg/ml Methionine was prepared in sterile water, and autoclaved for 15 min at 121 °C and 15 psi.

17aa solution (10 mg/ml each, minus (CYM))

To 90 ml of stirring sterile water, 1 g of the following aa were added in this specific order:

Table 2.5: 17aa solution recipe.
* Floats on surface. Increase stirring rate to dissolve fully.
** Stir until all has dissolved.
*** Turns solution brown, and some will remain undissolved

Amino Acid (aa)
1. Sodium Glutamate
2. Aspartate (Slow to dissolve, continue to next aa)
3. Lysine-HCl
4. Arginine-HCl
5. Histidine-HCl
6. Alanine
7. Proline
8. Glycine
9. Threonine
10. Serine
11. Glutamine
12. Asparagine
13. Valine *
14. Leucine *
15. Isoleucine *
16. Phenylalanine **
17. Tryptophan ***

All amino acids were L-amino acids. The solution was left to dissolve fully overnight. Filter sterilise the solution, and the brown colour almost disappears. 17aa solution was stored at 4 °C.

PASM-5052 1 l growth media preparation summary

Table 2.6: 1 l growth PASM-5052 AI media recipe.

Component	**Volume for 1 l PASM-5052 media**	**Final Compound Concentration**
Water	887 ml	/
1 M $MgSO_4$	2 ml	2 mM $MgSO_4$
1000 × Trace metals	200 µl	0.2 × Trace metals
50 ×5052	20 ml	0.5 % Glycerol
		0.05 % Glucose
		0.2 % α-Lactose
20 ×NPS	50 ml	50 mM Na_2HPO_4
		50 mM KH_2PO_4
		25 mM $(NH_4)_2SO_4$
100 µM Vitamin B12	1 ml	100 µM Vitamin B12
17aa (-CYM)	20 ml	200 µg/ml each aa
25 mg/ml methionine	20 ml	500 µg/ml methionine
Antibiotics	1 ml	50 µg/ml Kanamycin
		30 µg/ml Chloramphenicol

2.1.9 Sodium dodecyl sulphate - polyacrylamide gel electrophoresis (SDS-PAGE) and Laemmli sample buffer

SDS-PAGE gels

Based on the molecular weights for all SLC10A7 and ASBT constructs, protein gel analysis was possible on 14 % SDS-PAGE gels, yielding good molecular weight separation, forming distinct bands. The recipe to produce these gels is shown below:

Table 2.7: 14 % SDS-PAGE gel recipe.

Resolving Gel		
Reagent	**Volume for 2 gels**	**Final Concentration**
1.5 M Tris pH 8.8	2.82 ml	0.37 M
Water	3.02 ml	/
30 % acrylamide (29:1)	5.32 ml	14 %
10 % SDS	120 µl	0.1 %
10 % AMPS	120 µl	0.1 %
100 % N,N,N',N'-tetramethylethane-1,2-diamine (TEMED)	10 µl	0.1 %
Stacking Gel		
Reagent	**Volume for 2 gels**	**Final Concentration**
0.5 M Tris pH 6.8	1.26 ml	0.13 M
Water	2.77 ml	/
30 % acrylamide (29:1)	670 µl	4.2 %
10 % SDS	50 µl	0.1 %
10 % AMPS	50 µl	0.1 %
100 % N,N,N',N'-tetramethylethane-1,2-diamine (TEMED)	5 µl	0.1 %

2 × Laemmli sample buffer

All gel samples were prepared in 2 × Laemmli sample buffer before loading on a 14 % SDS-PAGE gel. The recipe for this 2 × sample buffer is shown below:

Table 2.8: 2 × Laemmli sample buffer recipe.

Reagent	**Volume for 10 ml**	**Final Concentration (2 ×)**
10 % SDS	4 ml	4 %
100 % Glycerol	2 ml	20 %
0.5 M Tris pH 6.8	2.4 ml	120 mM
Water	1.6 ml	/
Bromophenol blue	Small spatula amount	roughly 0.02 %
1 M Dithiothreitol (DTT)	0.1 ml	10 mM

2.1.10 Small-scale membrane protein overexpression trials

Bacterial membrane protein overnight growths were setup as previously outlined. A 1:50 dilution (100 µl) was added to 5 ml PASM-5052 media containing a serial L-rhamnose concentration range: 0 mM, 0.1 mM, 0.25 mM, 0.5 mM, 0.75 mM, and 1 mM. The varying L-rhamnose cell growth conditions were grown at 37 °C and 200 rpm until an $OD_{600\,nm}$ 0.5 was reached. Trials with and without the addition of 0.4 mM IPTG induction were investigated and incubated overnight for 15-20 h) at 25 °C. 1 ml cell growth aliquots were centrifuged at 16 000 rcf for 10 min and the pellet resuspended in 100 µl 1 × PBS buffer. The resuspension was transferred to a 96-well black Nunc plate and relative fluorescence GFP units (RFU) were measured on a CLARIOstar (BMG LABTECH) fluorescent plate reader. The emitted GFP was measured at 512 nm, using an excitation wavelength of 488 nm. With pure GFP standard, the plate reader was calibrated in order to accurately convert membrane protein fluorescence measurements to mg/ml quantities.

This small scale overexpression analysis was repeated for overexpression trials with the addition of L-rhamnose concentrations at $OD_{600\,nm}$ 0.5.

2.1.11 Large scale membrane protein overexpression

Large-scale membrane protein overexpression followed, in large part, the same experimental setup as small-scale overexpression, but was scaled up to 1 l baffled shaker flasks. Based on the outcomes from small-scale investigations, the optimal L-rhamnose/IPTG concentrations were added and grown in the exact same manner. After overnight induction at 25 °C, the cell cultures were harvested in centrifugation pots, spun at 5000 rcf for 15 min, and the pellets collected, flash frozen in liquid nitrogen and stored at −80 °C.

2.1.12 Cell membrane extraction and isolation

Harvested bacterial cell pellets were thawed and resuspended in 50 ml/l pellet, resuspension buffer (1 × PBS, 1 mM $MgCl_2$, small spatula of AEBSF (protease inhibitor) and DNase I) at 4 °C, stirring at 200 rpm for 1 h. Fully resuspended cell solution was filtered and lysed 3 × in a cell disruptor (Constant Systems) at 25 kpsi and 4 °C. Cell debris was removed by centrifugation at 24 000 rcf for 12 min in a JA 25.50 rotor at 4 °C. The supernatant was collected and centrifuged further at 41 000 rpm for 45 min in a 45Ti rotor using an ultracentrifuge, at 4 °C. The supernatant was discarded and the membrane pellets homogenized with resuspension buffer to a final volume of 100 ml. Homogenised cell membranes were then flash frozen in liquid nitrogen and stored at −80 °C.

2.1.13 Fluorescence size exclusion chromatography (FSEC)

In 1.5 ml Beckman Coulter Benchtop ultracentrifuge tubes, 100 µl crude bacterial cell membrane was mixed with 100 µl 10 % detergent (n-Dodecyl-β-D-Maltopyranoside (DDM), n-Undecyl-β-D-Maltopyranoside (UDM), n-Decyl-β-D-Maltopyranoside (DM), n-Nonyl-β-D-Maltopyranoside (NM), n-Octyl-β-D-Glucopyranoside (β-OG), Lauryldimethylamine N-oxide (LDAO), and Lauryl Maltose Neopentyl Glycol (LMNG)) and 800 µl solubilisation buffer (1 × PBS and 150 mM NaCl), and solubilised on a rotating platform for 2 h at 4 °C. After solubilisation, samples were centrifuged at 43 000 rpm in a benchtop ultracentrifuge for 1 h using the TLA 100.3 rotor at 4 °C. 500 µl of the supernatant was subsequently loaded onto a Superose 6 Increase 10/300 GL (GE-healthcare) gel filtration column pre-equilibrated with 20 mM Tris pH 7.5, 150 mM NaCl and 0.03 % DDM, running at a flow rate of 0.4 ml/min. The eluted GFP-fused protein fluorescence was measured in a 96-well NUNC black fluorescent plate using the CLARIOStar (BMG LABTECH) fluorescent reader (emission wavelength of 512 nm and excitation wavelength of 488 nm) and the subsequent GFP fluorescent readings and corresponding retention/elution volumes were plotted and analysed in GraphPad Prism (San Diego, CA).

2.1.14 $ASBT_{Lb}$ membrane solubilisation and purification

Solubilisation buffer: 1 × PBS pH 7.4, 150 mM NaCl, 10 mM Imidazole pH 7.4, 1 % DDM

Wash buffer 1: 1 × PBS pH 7.4, 150 mM NaCl, 10 mM Imidazole pH 7.4, 0.03 % DDM

Wash buffer 2: 1 × PBS pH 7.4, 150 mM NaCl, 20 mM Imidazole pH 7.4, 0.03 % DDM

Wash buffer 3: 20 mM Tris pH 7.5, 150 mM NaCl, 30 mM Imidazole pH 7.5, 0.03 % DDM

Elution buffer: 20 mM Tris pH 7.5, 150 mM NaCl, 250 mM Imidazole pH 7.5, 0.03 % DDM

SEC buffer: 20 mM Tris pH 7.5, 150 mM NaCl, 0.03 % DDM

6L (100 ml) ASBT *Leptospira biflexa* cell membranes were thawed and solubilised in 300 ml solubilisation buffer, stirring at 200 rpm at 4 °C for 3 h. Once fully solu-

bilised, the solution was ultracentrifuged at 41 000 rpm for 45 min in the 45Ti rotor at 4 °C. The ultracentrifuge supernatant was collected and bound to Ni-NTA resin pre-equilibrated in wash buffer 1, to form a slurry, constant stirring at 150 rpm at 4 °C, for roughly 4 h. 1 ml of resin typically binds to 1 mg of membrane protein. When <30 % unbound membrane protein remained or full binding saturation had occurred, the slurry was transferred to a glass Econo-column (BioRad), allowing for the elution of the flow-through (FT). Subsequent washing of the resin with 10 column volumes (CV) of wash buffer 1, 20CV wash buffer 2 and 15CV wash buffer 3, was performed. The 3C-GFP-His8 tag was then cleaved overnight by addition of equimolar amounts of 3C protease directly to the protein-bound resin. The following day, the cleaved protein FT was collected and chased through with 5CV wash buffer 3, collecting elutions in 5 ml fractions. Elution buffer was used to elute the bound GFP-His8 tag, for future gel analysis and purity check. Eluted membrane protein concentrations were measured with a nanodrop (see below) and relevant fractions pooled together. The pooled protein was then run on a 5 ml HisTrap column (GE Healthcare) (reverse immobilised metal affinity chromatography (rIMAC)), allowing for the removal of additional 3C contaminants. The post Histrap FT was then concentrated up to 8-10 mg/ml in a Vivaspin 100,000MW cut off concentrator and centrifuged at 3500 rpm in a benchtop centrifuge. Once concentrated, the protein was centrifuged further at 13 200 rpm in a microcentrifuge for 1 h prior to loading 500 µl onto a Superdex 200 Increase 10/300 GL gel filtration column pre-equilibrated in SEC buffer, for SEC analysis. SEC runs were typically used to buffer exchange into smaller micelled "harsher" detergents (such as NM), as well as removing imidazole and increasing membrane protein purity. Size exclusion fractions were pooled together based on their UV chromatogram trace profile, and concentrated further to desired protein concentrations (usually 10-15 mg/ml for vapour diffusion (sitting drops) and 25-30 mg/ml for lipidic cubic phase (LCP)) and crystallisation attempts were conducted.

2.1.15 SLC10A7 membrane solubilisation and purification

Solubilisation buffer: 20 mM Tris pH 7.5, 300 mM $NaHCO_3$ pH 7.5, 1 % DDM

Wash buffer 1: 20 mM Tris pH 7.5, 300 mM $NaHCO_3$ pH 7.5, 10 mM Imidazole pH 7.4, 0.03 % DDM

Wash buffer 2: 20 mM Tris pH 7.5, 300 mM $NaHCO_3$ pH 7.5, 20 mM Imidazole pH 7.4, 0.03 % DDM

Wash buffer 3: 20 mM Tris pH 7.5, 300 mM $NaHCO_3$ pH 7.5, 30 mM Imidazole pH 7.4, 0.03 % DDM

Elution buffer: 20 mM Tris pH 7.5, 300 mM $NaHCO_3$ pH 7.5, 250 mM Imidazole pH 7.4, 0.03 % DDM

SEC buffer: 20 mM Tris pH 7.5, 150 mM $NaHCO_3$ pH 7.5, 0.03 % DDM

6L (100 ml) SLC10A7 cell membranes were thawed and solubilised in 300 ml solubilisation buffer, stirring at 200 rpm at 4 °C for 3 h. Once fully solubilised, the solution was ultracentrifuged at 41 000 rpm for 45 min in the 45Ti rotor at 4 °C. The ultracentrifuge supernatant was collected and bound to Ni-NTA resin pre-equilibrated in wash buffer 1, to form a slurry, constant stirring at 150 rpm at 4 °C, for roughly 4 h. 1 ml of resin typically binds to 1 mg of membrane protein. When <30 % unbound membrane protein remained or full binding saturation had occurred, the slurry was transferred to a glass Econo-column (BioRad), allowing for the elution of the flow-through (FT). Subsequent washing of the resin with 10 column volumes (CV) of wash buffer 1, wash buffer 2 and wash buffer 3 was performed. The 3C-GFP-His8 tag was then cleaved overnight by addition of equimolar amounts of 3C protease directly to the protein-bound resin. The following day, the cleaved protein FT was collected and chased through with 5CV wash buffer 3, collecting elutions in 5 ml fractions. Elution buffer was used to elute the bound GFP-His8 tag, for future gel analysis and purity check. Eluted membrane protein concentrations were measured with a nanodrop (see below) and relevant fractions pooled together. The pooled protein was then run on a 5 ml His-Trap column (GE Healthcare) (reverse immobilised metal affinity chromatography (rIMAC)), allowing for the removal of additional 3C contaminants. In addition, a second run on a 5 ml MBP-trap column (GE Healthcare) was required for the purification of the SLC10A7 constructs in order to remove bound MBP, originating from the 3C pro-

tease. The post His-trap/MBP-trap FT was then concentrated up to 8-10 mg/ml in a Vivaspin 100,000MW cut off concentrator and centrifuged at 3500 rpm in a benchtop centrifuge. Once concentrated, the protein was centrifuged further at 13 200 rpm in a microcentrifuge for 1 h prior to loading 500 µl onto a Superdex 200 Increase 10/300 GL gel filtration column pre-equilibrated in SEC buffer, for SEC analysis. SEC runs were typically used to buffer exchange into smaller micelled "harsher" detergents (such as NM or LDAO), as well as removing imidazole and increasing membrane protein purity. Size exclusion fractions were pooled together based on their UV chromatogram trace profile, and concentrated further to desired protein concentrations (usually 10-15 mg/ml for vapour diffusion (sitting drops) and 25-30 mg/ml for lipidic cubic phase (LCP)) and crystallisation attempts were conducted. Across the 4 years of this PhD, many alterations to the SLC10A7 purification procedure were trialled concerning changes to the components of the solubilisation, wash 1, wash 2, wash 3 and sec buffers, as well as trials with a multitude of bacterial constructs. In particular, purification attempts in buffers with NaCl as well as no salt buffers were trialled, originating from thermostable outcomes from both GFP-TS and FSEC-TS assays. However, the best results in terms of purity and success at crystallisation arose from the above outlined purification strategy in $NaHCO_3$.

2.1.16 DNA/Protein nanodrop concentration measurements

All purified DNA and protein concentrations were quantified via a NanoDrop NS-1000 spectrophotometer, following manufacturer's guidelines. For both ASBT and SLC10A7protein constructs, the molecular extinction coefficients were required for reliable and accurate concentration measurements.

2.2 Functional studies: thermostability assays

2.2.1 GFP-TS assay

In 2 ml Eppendorf microtubes, 150 µl cell membranes were mixed with 150 µl 10 % DDM detergent and 1200 µl buffer containing (20 mM buffering ion (typically Tris pH 7.5), 150-300 mM salt, and ± 1 mM substrate). Solubilisation of the cell membranes, in various conditions, was performed for 1 h on a rotating platform at 4 °C. 150 µl solubilised membranes were aliquoted in microtubes containing 1 % β-OG detergent and incubated at 4 °C for 5 min, flicking/mixing the tubes 5-10 ×. Samples were then incubated at a range of temperatures (4 °C, 20 °C, 30 °C, 40 °C, 50 °C, 60 °C, 70 °C, 80 °C, 90 °C, and 100 °C) for 10 min, before being centrifuged in a microcentrifuge at 16 000 rcf for 30 min at 4 °C. 100 µl supernatant was transferred to a 96-well NUNC black fluorescent plate and subsequent GFP fluorescence was measured using the CLARIOStar (BMG LABTECH) fluorescent plate reader (emission wavelength of 512 nm and excitation wavelength of 488 nm). GFP RFUs were plotted, normalised and analysed in GraphPad Prism (San Diego, CA), and sigmoidal functions were conducted to generate the T_m melting temperature measurements for each condition. Two separate, independent replicates of every analysed condition was performed (as per the literary standard) (Nji et al., 2018), in order to yield accurate error bars and reliability in the data.

2.2.2 FSEC-TS assay

In 1.5 ml Beckman Coulter Benchtop ultracentrifuge tubes, 100 µl bacterial cell membrane was mixed with 100 µl of 10 % DDM detergent and 800 µl solubilisation buffer (typically 20 mM Tris pH 7.5, and 150-300 mM salt) and was solubilised on a rotating platform at 4 °C for 1 h. Samples were then ultracentrifuged in a TLA 100.3 rotor, at 43 000 rpm for 45 min. The supernatant was removed and 100 µl of 10 % β-OG detergent was added and further incubated at 4 °C for 10 min on a rotating platform. Samples were then heat treated at a range of temperatures (4 °C, 20 °C, 30 °C, 40 °C, 50 °C, and 60 °C) for 10 min, before being ultracentrifuged again at 43 000 rpm for 25 min. 500 µl of the ultracentrifuge supernatant was loaded onto a Superose 6 10/300 GL gel filtration column, pre equilibrated with 20 mM Tris pH 7.5, 150-300 mM salt and 0.03 % DDM. Elutions were collected in a 96-well NUNC black fluorescent plate and GFP fluorescence measurement established using a CLARIOStar (BMG LABTECH) fluorescent plate reader (emission wavelength of 512 nm and excitation wavelength of 488 nm). All GFP fluorescent readings were plotted and analysed in GraphPad Prism (San Diego, CA).

2.2.3 CPM-assay

The CPM thermostability assay was conducted as originally proposed by Alexandrov et al., 2008. 4 mg/ml stock concentrations of the 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM) fluorophore (Sigma) were dissolved in 100 % DMSO and aliquoted to 10 µl aliquots, flash frozen in liquid nitrogen and stored at −80 °C. 120 µl buffer solution containing 20 mM Tris pH 7.5, 150 mM NaCl, 0.03 % DDM and ± 1 mM substrate was mixed and 2 × 50 µl aliquots were dispensed to white 96-well qPCR plates, for two separate and independent biological replicates. An initial serial concentration range (1 µg, 2.5 µg, 5 µg, 10 µg, 15 µg, and 20 µg) of post-SEC purified protein was added to each qPCR well, before using 2.5 µg for future substrate investigations. An aliquot of CPM dye was thawed and diluted 1:40 to 0.1 mg/ml and 1.25 µl was added (in the dark) to each qPCR wells, yielding a final concentration of 2.5 µg/ml CPM dye. A clear plastic cover film was stuck to the plate and briefly spun in a benchtop centrifuge, before a pre-programmed CPM thermostability assay was run on a qPCR machine. CPM fluorescence values were recorded at an emission wavelength of 463 nm and excitation wavelength of 387 nm, with 71 measurements reported at a range of increasing temperatures (25-95 °C). The relative fluorescent units (RFU) generated were plotted and analysed in GraphPad Prism (San Diego, CA), and the T_m melting temperatures calculated by tracing sigmoidal curves.

2.3 Membrane protein crystallisation and 3D X-ray crystallographic structure determination

2.3.1 Crystallisation trials

A multitude of 96-well broad screen crystallisation condition screens, including MemGold, MemGold2, MemTrans, MemMeso, and MemGoldMeso, as well as broad additive screens including MemAdvantage and the Hampton Additive Screen, supplied from Molecular Dimensions and Hampton Research, were used in both vapour diffusion and LCP crystallisation trials, to increase the likelihood of successful membrane protein crystallisation. Correct usage and setup in crystallisation experiments followed the manufacturer's manual instructions and guidelines.

Sitting drop vapour diffusion crystallisation

Post-SEC protein (typically buffer exchanged into DDM or NM detergent buffer, with or without substrate) was concentrated to 10-15 mg/ml and centrifuged in a microcentrifuge at 16 000 rcf for 30 min and either flash frozen in liquid nitrogen for storage at −80 °C or used immediately for sitting drop vapour diffusion crystallisation setup. The above mentioned broad screen crystallisation plates were trialled. Using a Mosquito robot (TTP labtech), 100 nl protein sample was mixed with 100 nl of the 50 µl crystallisation buffer reservoir/precipitant, of a sitting drop crystallisation plate. A clear plastic cover film sealed the plate, before storing in a 20 °C incubator. Protein crystals formed within a couple of days to a week under the right crystallographic conditions.
Successful protein crystals were harvested and mounted in cryo-loops, and immediately flash frozen in liquid nitrogen before being stored in crystal pucks and sent to Diamond Light Source, Oxfordshire for X-ray analysis.

Lipidic cubic phase (LCP)

Post-SEC protein (typically buffer exchanged into DDM or NM detergent buffer, with or without substrate) was concentrated to 25-30 mg/ml and centrifuged in a microcentrifuge at 16 000 rcf for 30 min and either flash frozen in liquid nitrogen for stroage at −80 °C or used immediately for LCP crystallisation setup. Aliquots of frozen monoolein lipid were thawed and incubated at 40 °C.50 µl (for protein) and 100 µl (for monoolein lipid) gas-tight syringes were used for the correct and accurate production of the lipidic cubic phase state of the protein. A 2:3 protein:monoolein ratio was used to determine accurate quantities to be mixed, with subsequently amounts added to the relevant gas-tight syringes separately. A cou-

pler connected both syringes together, and the LCP was formed upon mixing both the protein and monoolein by pushing the syringes slowly from side to side, back and forth. The solution initially turned white/grey but quickly faded to a transparent state, indicative of a correctly formed stable LCP state. The homogenous LCP was transferred to the 50 µl gas-tight syringe, and the coupler disassembled, before a thin dispensing needle was assembled. The syringe was subsequently attached to a Mosquito robot (TTP labtech) machine following the manufacturer's manual instructions and guidelines.
The broad screen crystallisation plates mentioned above were again used in LCP crystallisation. Humidity levels were set to 50-75 % to prevent evaporation of the crystal drops. 50 nl LCP was mixed with 800 nl of the crystallisation buffer reservoir/precipitant on a glass plate, before being sandwiched with a thinner glass cover plate. The plates were subsequently incubated at 20 °C and left to form crystals for a week. Protein crystal observation was investigated using microscopes with polariser filters. The MemGold2 A8 and MemTrans A5 conditions yielded initial crystal hits which were further optimised in fine screens, varying crystallisation precipitant concentrations including salt, pH and PEG. In addition MemAdvantage and the Hampton Additive Screen were added to these fine screens to improve the formed protein crystals.
Successful protein crystals were harvested and mounted in the following manner: glass-cutting stones were used to open relevant crystal conditions by cutting the thin glass cover plate. LCP crystals were harvested using Mitogen crystal loops and immediately frozen in liquid nitrogen before being stored in crystal pucks and sent to Diamond Light Source, Oxfordshire for X-ray analysis.

2.3.2 3D protein structure determination

An initial 3.3 Å crystal dataset was collected at the I24 micro-focus beamline at Diamond Light Source, Oxfordshire. The intensities were integrated and scaled in DIALS (Winter et al., 2018) before initiating molecular replacement strategies in Phaser (McCoy et al., 2007) with the $ASBT_{NM}$ (PDB (Berman et al., 2000) entry 3ZUY). Further refinement, model building and generation of electron-density maps were conducted in Phenix (Adams et al., 2002, 2010) refinement and COOT (Emsley et al., 2010). This yielded a nine TMD solved $ASBT_{Lb}$ structure which was further used as a molecular replacement model for the collected 2.2 Å crystal dataset. Automated DIMPLE (Wojdyr et al., 2013) molecular replacement was conducted, followed by autobuilding and refinement strategies in Phenix (Adams et al., 2002, 2010). This yielded a finalised nine TMD structure.

Chapter 3

Results: Membrane Protein Bioinformatics, Overexpression, Extraction/Isolation and Solubilisation

3.1 Membrane Protein Bioinformatics

3.1.1 The ASBT$_{Lb}$ Bacterial Homologue

Via a bacterial BLAST (Altschul et al., 1990) search of the human ASBT transporter, we identified a novel *Leptospira biflexa* bacterial homologue, presenting with a predicted nine TMD topology. In order to investigate the functional and structural characteristics of this homologue further, successful cloning into a pWaldo plasmid, entailing a C-terminal 3C-GFP-His(8) tag, was established previously by Oliver Huxley (MBio student - 2019).

Comparisons of this novel ASBT bacterial homologue to the human NTCP, human ASBT, and previously crystallised ASBT *N. meningitidis*, *Y. frederiksanii* bacterial homologues are shown below in table 3.1, depicting core differences in amino acid sequence chain lengths, as well as amino acid sequence identity and similarity to both human NTCP and ASBT variants. Predicted amino acid sequence disruption plots generated in RONN (Yang et al., 2005), for all aforementioned members is presented in Appendix 2.

Table 3.1: The amino acid sequence identities and similarities of the human ASBT/NTCP and ASBT bacterial homologue transporters.
Percentage conservations were retrieved from Blastp analysis.

Organism / Transporter	**Total amino acid Residues**	**Human NTCP amino acid sequence identity / similarity**	**Human ASBT amino acid sequence identity / similarity**
Human NTCP	349	N/A	36 % / 56 %
Human ASBT	348	36 % / 56 %	N/A
$ASBT_{Yf}$	307	27 % / 47 %	31 % / 53 %
$ASBT_{NM}$	315	26 % / 49 %	33 % / 54 %
Leptospira biflexa	292	33 % / 56 %	35 % / 56 %

As can be clearly identified in table 3.1 above, the novel *Leptospira biflexa* bacterial homologue shows good sequence identity and similarity to both human NTCP and ASBT transporters.

In comparison to the previously crystallised bacterial ASBT homologues ($ASBT_{NM}$ and $ASBT_{Yf}$), $ASBT_{Lb}$ demonstrates slightly greater sequence identity and similarity to both human NTCP and ASBT. Furthermore, sequence identities and similarities of $ASBT_{Lb}$ to both $ASBT_{NM}$ and $ASBT_{Yf}$ also showed good conservation of amino acid residues with 28 %/55 % and 30 %/51 %, respectively.

In order to assess in greater detail the specific amino acid sequence differences among all five transporters listed in table 3.1 above, a multiple sequence alignment, shown below in figure 3.1, was generated.

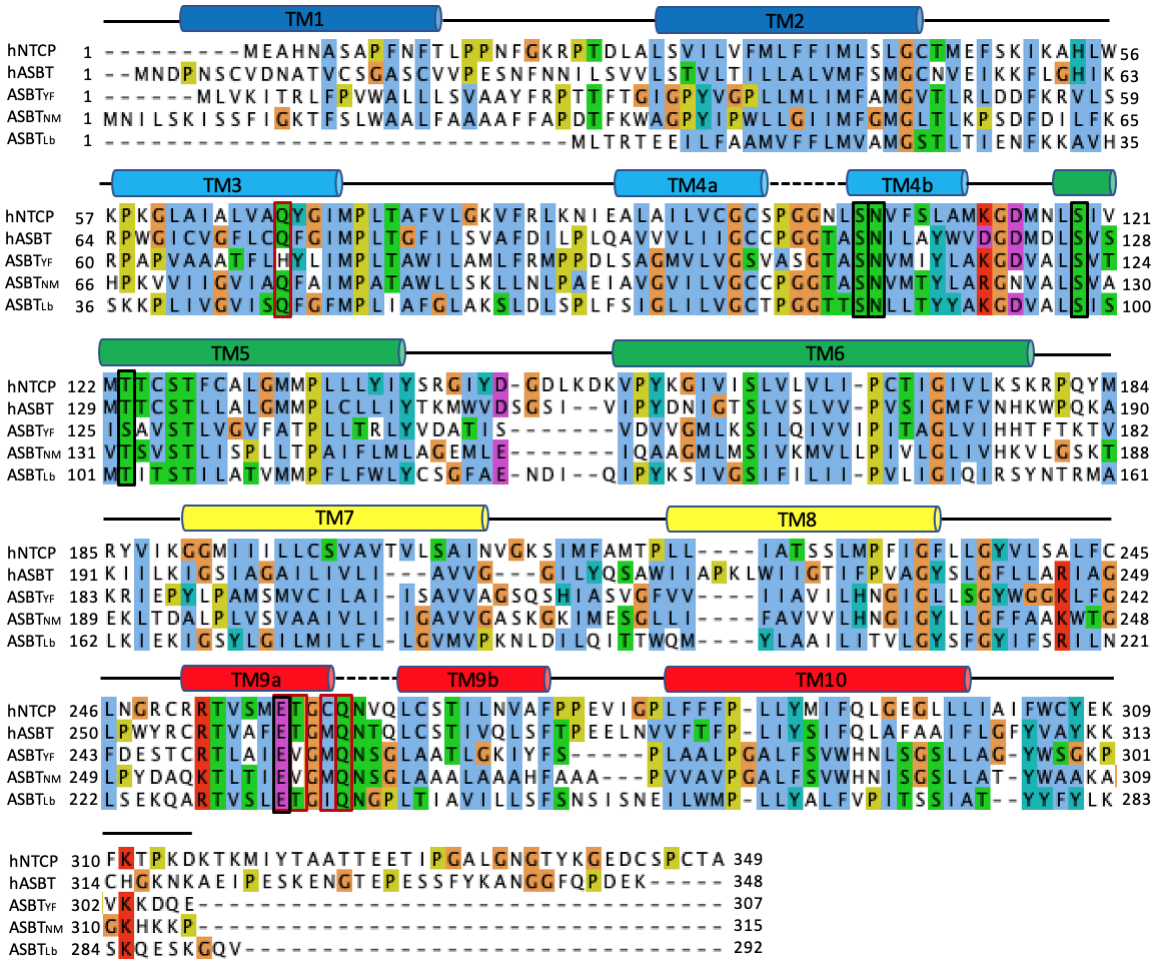


Figure 3.1: The Human NTCP, ASBT; and bacterial $ASBT_{Yf}$, $ASBT_{NM}$ and $ASBT_{Lb}$ homologues multiple sequence alignment. Respective amino acid sequences were analysed in Jalview (Waterhouse et al., 2009), and specifically aligned with Clustalx (Larkin et al., 2007) and MuscleWS (Edgar, 2004). The established $ASBT_{NM}$ transmembrane (TM) domains (as reported by Hu et al.,2011) are presented above the sequences in cylinders, depicting the structural TM1-TM10 domains from the N-terminal (blue) to C-terminal (red). Na^+ binding site amino acid residues implicated for Na1 and Na2, are depicted in red and black boxes, respectively. Although the alignment shows apparent helical association to the first residues of human NTCP and ASBT transporters, this is predicted not to be the case in reality.
Clustalx (Larkin et al., 2007) notation of residues: Blue: Hydrophobic, Red: Positive charge, Magenta: Negative charge, Green: Polar, Pink: Cysteines, Orange: Glycines, Yellow: Prolines, Cyan: Aromatic, and White: Unconserved.

As a whole, a very strong conservation of amino acid residues is depicted in the sequence alignment for all members analysed. As can be seen above, the established residues involved in Na^+ binding in $ASBT_{NM}$, are also unanimously conserved in the $ASBT_{Lb}$ construct, thereby suggesting both sodium ions to also be implicated in the translocation of its substrates. Furthermore, the TMD1 domain present in both $ASBT_{NM}$ and $ASBT_{Yf}$ is clearly missing in the $ASBT_{Lb}$ bacterial homologue, outlining a significant difference to its nine TMD topology, which was expected.

Taken together, due to its strong amino acid sequence conservation to human

NTCP and ASBT, it is strongly assumed $ASBT_{Lb}$ also reiterates similar structural and functional characteristics previously established from $ASBT_{NM}$ and $ASBT_{Yf}$.

3.1.2 The SLC10A7 Bacterial Homologues

Due to the wide taxonomical conservation of the human SLC10A7 orphan carrier transporter in both vertebrate and non-vertebrate species, including mammalian, yeast, plants and bacteria (Zou et al., 2005; Geyer et al., 2006; Godoy et al., 2007); the latter was selected as a suitable candidate organism for novel SLC10A7 biochemical, functional and structural research. Analogous to eukaryotic organisms, prokaryotes also fundamentally require regulated systems of intracellular Ca^{2+} signalling, for use in motility and transport, regulation of cell structure, sporulation and cell differentiation (Dominguez, 2004). Therefore, just as its conserved role in humans, mice, zebrafish and yeast cells (Jiang et al., 2012; Alber et al., 2013; Zhao et al., 2016; Ashikov et al., 2018; Dubail et al., 2018; Laugel-Haushalter et al., 2019; Karakus et al., 2020), it is presumed the SLC10A7 transporter to also be involved in Ca^{2+} homeostasis in bacterial species.

In light of this, specific SLC10A7 bacterial homologues were identified through the BLAST (Altschul et al., 1990) database, and selected to be suitable, based on their degree of amino acid sequence identity to human SLC10A7, as well as their sequence disorder prediction in RONN (Yang et al., 2005) (see Appendix 1). The chain length and sequence identity to human SLC10A7 of these bacterial homologues are displayed in table 3.2 below.

Table 3.2: The studied bacterial SLC10A7 homologues

Bacterial Organism	**Total Amino Acid Residues**	**Human SLC10A7 Sequence Identity**
Asticcacaulis taihuensis	326	36%
Cupriavidus necator	341	34%
Escherichia coli	332	28%
Escherichia coli Q322 C-terminal Truncation	322	28%
Escherichia coli Q320 C-terminal Truncation	320	28%
Escherichia coli R316 C-terminal Truncation	316	28%
Klebsiella pneumoniae	332	27%
Pseudomonas aeruginosa R318 C-terminal Truncation	318	29%
Pseudomonas putida	333	28%
Vibrio breoganii	326	34%

Overall, there was good sequence identity (27-36 %) to human SLC10A7, among the studied bacterial homologues. The decision to generate truncated mutants for SLC10A7 *Escherichia coli* and *Pseudomonas aeruginosa* was based entirely on their predicted amino acid RONN disorder plots (Yang et al., 2005), and aimed to reduce the flexible tail region of the C-terminal domain.

Specific amino acid differences among all these bacterial homologues, as well as the reported human homologues, are highlighted below in figure 3.2, in a multiple sequence alignment.

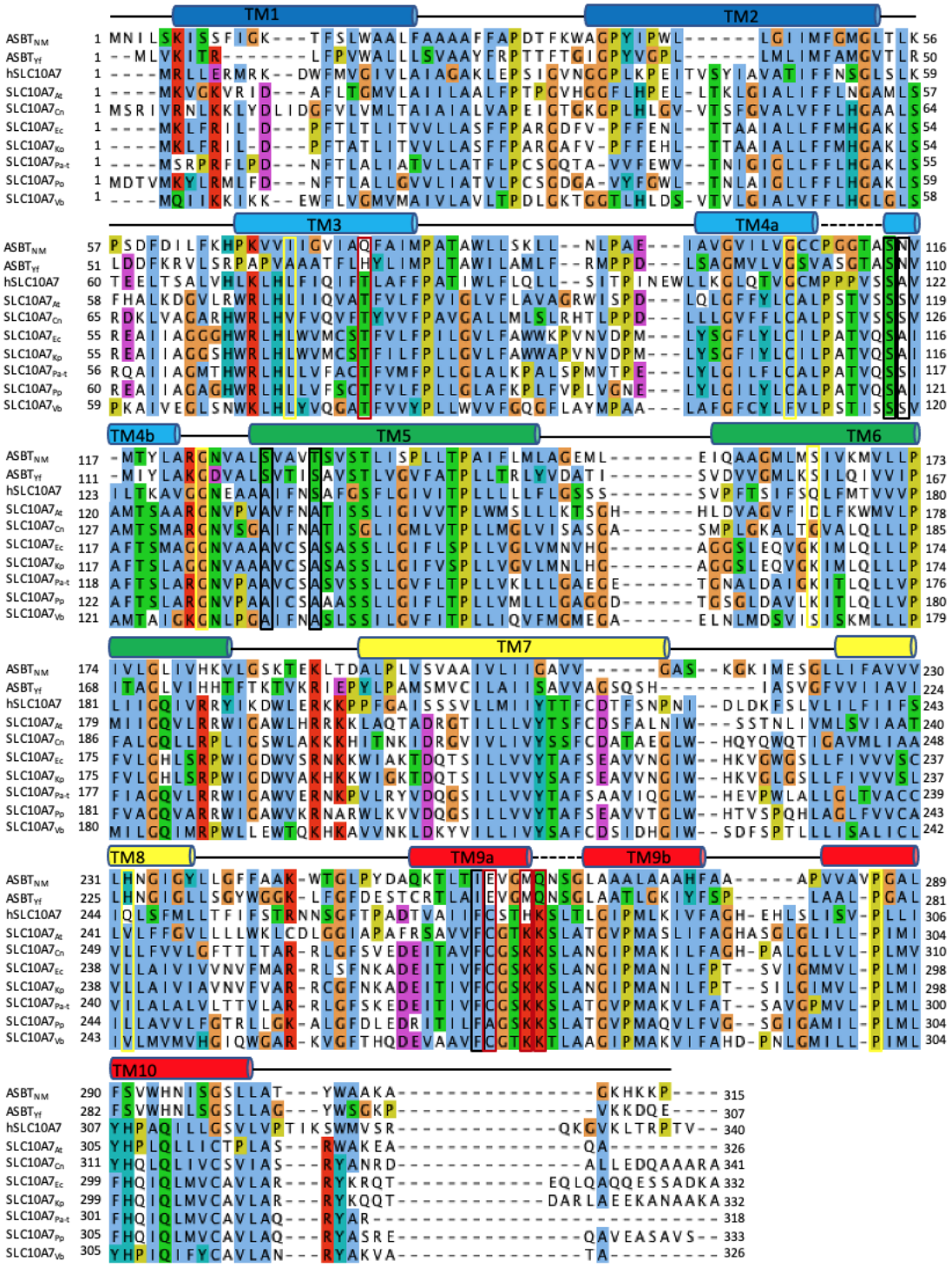


Figure 3.2: The SLC10A7 human and bacterial homologues; and ASBT$_{NM}$ and ASBT$_{Yf}$bacterial homologues multiple sequence alignment. Respective amino acid sequences were analysed in Jalview (Waterhouse et al., 2009), and specifically aligned with Clustalx (Larkin et al., 2007) and MuscleWS (Edgar, 2004). The established ASBT$_{NM}$ transmembrane (TM) domains (as reported by Hu et al.,2011) are presented above the sequences in cylinders, depicting the structural TM1-TM10 domains from the N-terminal (blue) to C-terminal (red). Na^+ binding site amino acid residues implicated for Na1 and Na2, are depicted in red and black boxes, respectively. Boxed yellow residues depict the reported SLC10A7 mutations in humans (from table 1.3.

Clustalx (Larkin et al., 2007) notation of residues: Blue: Hydrophobic, Red: Positive charge, Magenta: Negative charge, Green: Polar, Pink: Cysteines, Orange: Glycines, Yellow: Prolines, Cyan: Aromatic, and White: Unconserved.

As can be seen in figure 3.2 above, the analysed bacterial SLC10A7 amino acid sequences correlate extremely well with one another, yielding very strong amino acid conservations. However, amino acid comparison of these SLC10A7 transporters to $ASBT_{NM}$, shows a substantially weaker conservation of amino acid residues, especially at residues known to be implicated in driving sodium ions for solute translocation. This again reiterates the hypothesis that this SLC10A7 transporter operates instead perhaps via the co-transportation of protons.

Interestingly, all human SLC10A7 mutations (depicted in figure 3.2 by yellow boxes), with the exception of G112D, are localised to specific $ASBT_{NM}$ TMDs. As such, it can be assumed these disrupted residues to result in complete TMD deletions and aggregation of the SLC10A7 protein as a whole. Therefore, we suspect the severe clinical displays seen in affected individuals previously (Ashikov et al., 2018; Dubail et al., 2018; Laugel-Haushalter et al., 2019), to be due to an SLC10A7 loss-of-function, as a result of its structural alteration and degradation.

3.2 Small-scale Membrane Protein Overexpression

Understanding the unique behavioural characteristics of any novel membrane protein is imperative for further analyses to take place. These investigations can include downstream biochemical, molecular biological, functional and structural analyses, including but not limited to membrane protein extraction and detergent screening, purification, and crystallisation (Drew et al., 2006). Therefore, sufficient overexpression yields of membrane proteins, in their correctly localised, stable and folded states, must be successfully, accurately and efficiently investigated. As part of this PhD research, expression trials were conducted using the finely regulated and tunable Lemo21(DE3) *Escherichia coli* host strain, which allowed for the controlled monitoring of membrane protein overexpression.

The Lemo21(DE3) *E. coli* Host Platform

In the vast majority of cases, the natural abundance of membrane proteins is usually insufficient for downstream functional and structural research. Therefore, the overexpression of membrane proteins, via the "hijacking" of bacterial replication systems, is a fundamental process required to generate sufficient quantities for said investigations (Wagner et al., 2008).

Although in theory, membrane protein expression should be quite a simple process (crudely put, more mRNA produced = more expressed protein), in practice this is not often the case. In fact, the unregulated and non-optimised membrane protein overexpression is seen to be highly toxic to bacterial cells, resulting in altered levels of biomass formation; lowered total yields of protein; and the accummulation of cytoplasmic aggregates containing proteases, precursors of outer membrane proteins, inclusion bodies, chaperones, vital cytoplasmic proteins and overexpressed membrane protein (Wagner et al., 2008; Schlegel et al., 2012). These aggregate formations are believed to originate from a saturation of the Sec translocon, which maintains the trafficking of secretory proteins and the insertion of membrane proteins to the lipid bilayer. In scenarios where membrane proteins are overexpressed, too much protein is required to be processed by the Sec translocon, resulting in its saturation. The system subsequently becomes overwhelmed, resulting in the inability of membrane proteins to be embedded into the lipid bilayer correctly, leading to their aggregation and denaturation in the cytoplasm (Wagner et al., 2008; Schlegel et al., 2012). Therefore, tunable optimisation of protein overexpression is crucial, in order to suppress the saturation of the stress-induced Sec translocon, and minimise adverse and toxic side effects (Drew et al., 2006).

Fortunately, Lemo21(DE3) cells offer a highly regulated, tunable and optimised expression system, in which efficient recombinant membrane protein generation is possible. Here, a target protein's gene is cloned into a plasmid and transformed into Lemo21(DE3) cells. Inside is located the bacteriophage T7 RNA polymerase (T7RNAP), which itself is strongly regulated by the non-titratable, IPTG inducible *lac*UV5 promoter. This T7RNAP exclusively binds and activates the T7 promoter, thereby initiating membrane protein transcription at an eight-fold greater rate than *E. coli* RNAP, resulting in the overexpression of the target protein (Wagner et al., 2008; Lee et al., 2014a). However, as previously mentioned, a lack of regulation of this overexpression can lead to lowered proteins yields, with severe adverse cellular side effects. In light of this, Lemo21(DE3) cells behold a natural T7RNAP inhibitor (T7 lysozyme), which is itself strongly regulated by the extremely well-titratable and tuneable L-rhamnose inducible promoter rhaBAD (Wagner et al., 2008; Schlegel et al., 2012; Lee et al., 2014a). Via the addition of increasing concentrations of L-rhamnose, the functional activity of T7RNAP is "delayed", due to its inhibition by the strong activation of T7 lysozyme. This fundamental inhibitory effect is seen to govern the optimisation ratio of correctly folded and properly inserted membrane proteins to the cytoplasmic lipid bilayer; to aggregated, unfolded and non-inserted membrane proteins (Schlegel et al., 2012).

The proper regulation and optimisation of membrane protein overexpression, via this rhamnose-inhibitory process in Lemo21(DE3) cells, guarantees a maximal yield of membrane protein, as well as a minimal level of adverse cellular side effects, thereby ensuring a stable semicontinuous overexpression (Schlegel et al., 2012). The subsequently successfully expressed membrane protein can then be subjected to downstream biochemical, structural and functional studies.

A schematic summary of the regulation of protein overexpression in Lemo21(DE3) cells, via the fine tunability of L-rhamnose, is outlined in figure 3.3 below.

GFP-Fused Membrane Proteins can Accurately Assess Levels of Overexpression

Commonly, red or green fluorescent proteins (RFP or GFP) are fused to recombinant membrane protein targets, in order to aid molecular biological processes involving cloning, overexpression, membrane extraction, purification and crystallisation (Drew et al., 2006). GFP is most frequently used due to its fascinating tolerance to detergents; organic and chaotropic salts; alkaline pH; proteases; photobleaching and heat (Halan et al., 2014).

In light of this, the ASBT and SLC10A7 bacterial homologues were success-

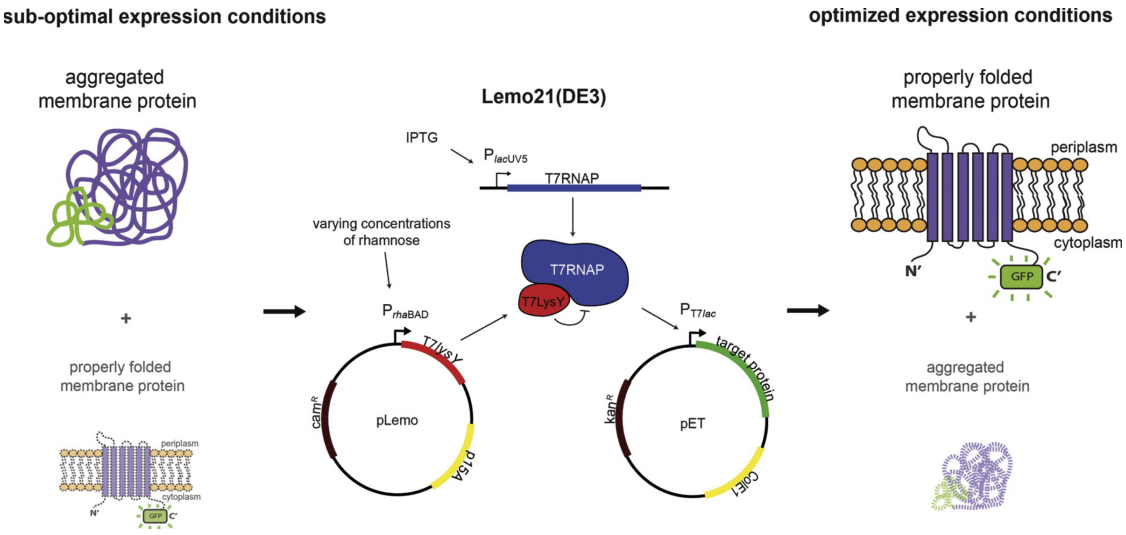


Figure 3.3: The Lemo21(DE3) optimised protein overexpression system. Under sub-optimal expression conditions, the majority of the culture will consist of aggregated expression states of the target membrane protein, due to saturation of the Sec translocon and production of adverse side effects. However, via fine optimisation of L-rhamnose, toxic bacterial side effects are depleted, resulting in almost homogeneous cultures of properly folded and localised membrane proteins. T7 RNAP is mediated by the non-titratable, IPTG-inducible *lac*UV5 promoter. T7 lysozyme (the T7 RNAP natural inhibitor) is monitored by the very well-titratable L-rhamnose inducible promoter rhaBAD, in the pLemo plasmid. The target membrane protein gene resides in a pET vector, and its expression is initiated at the T7 promoter. Addition of increasing amounts of L-rhamnose, delays the activity of T7 RNAP, due to inhibition by T7 lysozyme, which ultimately optimises the amounts of membrane protein generated, preventing the overexpression of unfolded and nonstable protein.
This figure was taken from Schlegel et al., 2012.

fully constructed and cloned into a C-terminal 3C-GFP-His(8) fusion expressing pET28(a) derived pWaldo-GFPe plasmid (Drew et al., 2001; Minhas et al., 2018). This incorporated GFP tag provided accurate monitoring and surveillance features of membrane protein overexpression in the cytoplasmic lipid bilayer, via both in-gel and whole-cell fluorescence techniques (Drew et al., 2006; Schlegel et al., 2012; Lee et al., 2014a).

If the GFP moiety fails to fold correctly, no GFP fluorescence is observed, thus indicating that the membrane protein-GFP fusion has failed to translocate to the cytoplasmic membrane, and has resulted in its denaturation and aggregation. Therefore, the GFP moiety will only fluoresce if it is correctly folded and localised within the cytoplasmic membrane, thus providing a convenient manner of assessing membrane protein overexpression levels which positively associate to the protein's functional characterisation (Wagner et al., 2008; Schlegel et al., 2012). However, it is worth noting that the GFP rate of folding remains slow ($t_{1/2}$ 30 minutes) (Waldo et al., 1999; Fukuda et al., 2000), suggesting that the bound membrane protein can misfold and become degraded, before the GFP tag has fully folded. Since GFP is an extremely stable protein, it is possible to form a correctly folded and fluorescent

GFP state, with an aggregated and denatured bound membrane protein, although this happens quite rarely. Typically, the stability of a target membrane protein remains proportional to the fluorescence of cytosolic GFP, whereby membrane protein GFP fluorescence overexpression can be accurately measured in whole-cell and isolated membranes, yielding well-established purified protein yields (Drew, 2005).

The PASM-5052 Auto-inductive Medium

For most structural and molecular biologists, luria/lysogeny broth (LB) remains the most mainstream available medium for membrane protein overexpression, and is typically accompanied by the induction of Isopropyl β-D-1-thiogalactopyranoside (IPTG) at the mid-log phase. However, auto-inductive media can provide additional features to the overexpression of membrane proteins, including the targeting of proteins for downstream crystallography and NMR techniques and induction of IPTG at the stationary phase (Lee et al., 2014a).

In 2014, Lee and colleagues proposed a novel membrane protein expression amplification recipe, using the Lemo21(DE3) host strain, IPTG and a PASM-5052 auto-inductive medium (Lee et al., 2014a). This medium had been seen to deplete levels of leakiness of the *lac*UV5 promoter, thereby reducing the toxic side effects of overexpressed membrane proteins. The addition of IPTG to autoinduction media, although in theory should not be necessary, in practice proves to be sometimes crucial on a protein-to-protein basis. Typically, this addition would lead to overproduced levels of the membrane protein, resulting in highly toxic overexpression levels, leading to cell death and overall reduced amounts of properly folded and localised target protein (Wagner et al., 2008; Schlegel et al., 2012). However, the opposite was witnessed from Lee and colleagues, whereby the overexpression of membrane proteins in the PASM-5052 auto-inductive media, with the addition of IPTG, produced high levels of biomass (which were comparable to those expressed in rich media) and resulted in greater quantity of protein being generated per host cell (Lee et al., 2014a). As can be seen for the ASBT_{NM} bacterial construct (figure 3.4), the addition of 0.4 mM IPTG was necessary to achieve maximal optimal overexpression yields of the protein, once again highlighting the protein-to-protein dependency on the application of IPTG in autoinduction media.

With an overall ten-fold greater overexpression of Lemo21(DE3)-expressing membrane proteins in PASM-5052 media, compared to LB (Lee et al., 2014a), we decided to use and recreate this medium and protocol, in order to fine tune, optimise and maximise properly folded levels of ASBT and SLC10A7 bacterial homologue overexpressions.

3.2.1 ASBT$_{Lb}$ and SLC10A7 Bacterial Homologue Small-scale Overexpression Trials

Optimal overexpression conditions for the ASBT *Leptospira biflexa* bacterial homologue, in Lemo21(DE3) with L-rhamnose tunable regulation and titratable IPTG addition, was previously investigated by Oliver Huxley (MBio student - 2019), and found to require the addition of 0.25 mM L-rhamnose and 0.4 mM IPTG at OD_{600nm} 0.5.

For the SLC10A7 bacterial homologues, small-scale (5 ml) Lemo21(DE3) protein expression trials in PASM-5052 autoinductive media were conducted, to assess the optimal growth conditions required (including bacterial strain, media, temperature, induction time etc.), to produce maximal stable membrane protein yields. The addition of 0-1 mM L-rhamnose was analysed, at both OD_{600nm} 0 and OD_{600nm} 0.5 growth cultures, in order to assess the optimal time of induction, yielding optimal overexpression. The cells were also induced with 0.4 mM IPTG at OD_{600nm} 0.5 growth culture, to assess the negative impact of glycerol (present in PASM-5052), in hindering auto-induced membrane protein overexpression, during high aeration levels (Lee et al., 2014a).

The ASBT$_{NM}$ bacterial homologue was additionally used in overexpression trials as a positive control, due to its extraordinarily high yield of protein expression, under already established L-rhamnose and IPTG concentrations, and Lemo21(DE3) and PASM-5052 optimised conditions. Figure 3.4 below depicts the ASBT$_{NM}$ overexpressed GFP fluorescence levels obtained in this setup, which were then used to compare expression levels to the SLC10A7 bacterial homologues.

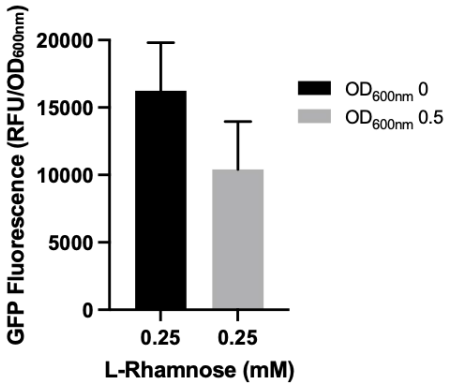
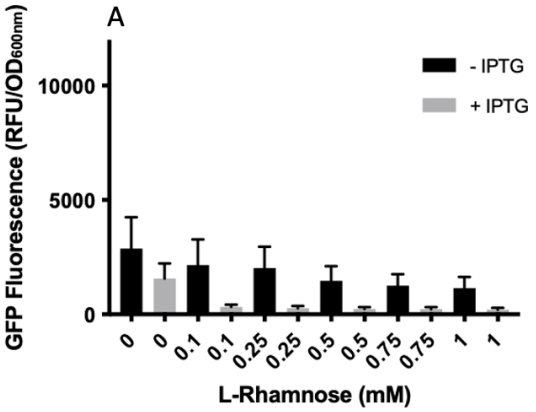


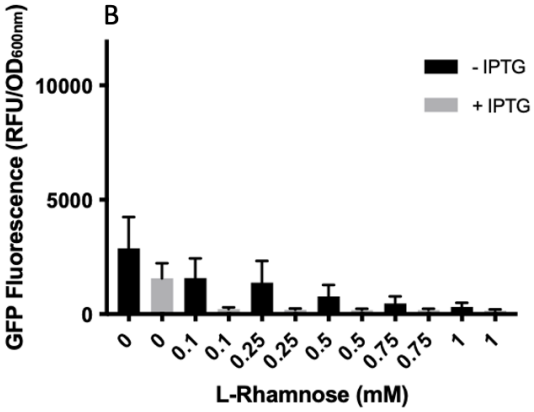
Figure 3.4: The optimal ASBT *Neisseria meningitidis* small-scale overexpression positive control. Maximal overexpression yields were attained with the addition of 0.25 mM L-rhamnose and 0.4 mM IPTG. The addition of L-rhamnose at OD_{600nm} 0 (black bar) and at OD_{600nm} 0.5 (grey bar) was also investigated. Error bars are shown as black "T" lines above the histograms and were calculated as the standard deviations from the mean of all expression trials (a minimum of three independent repeats). Data were analysed using GraphPad Prism.

As can be seen in figure 3.4, very high levels of protein overexpression were achieved: roughly 20,000 (OD_{600nm} 0) and 15,000 (OD_{600nm} 0.5) GFP counts/OD_{600nm}. These findings highlight $ASBT_{NM}$ to be very suitable for downstream analysis, but additionally can serve as a positive control, comparing membrane protein overexpressions in Lemo21(DE3) cells with PASM-5052 autoinductive medium.

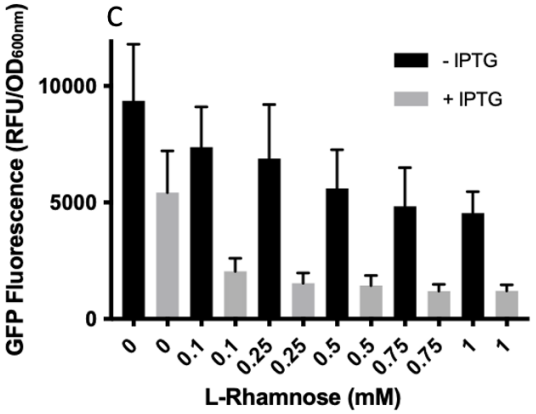
All generated SLC10A7 bacterial constructs, previously depicted in table 3.2, were subjected to the overexpression protocol mentioned above and yields analysed via measurement of relative GFP-fluorescence of whole-cells. Figure 3.5 below, highlights the optimal overexpression conditions for these bacterial homologues.



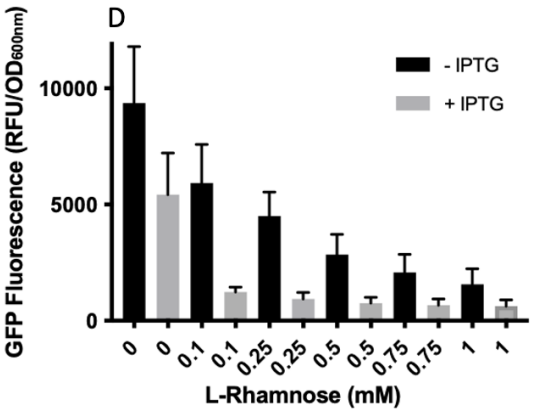
Asticcacaulis taihuensis (OD_{600nm} 0)



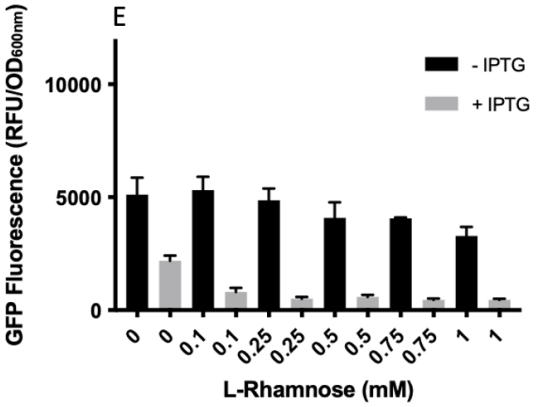
Asticcacaulis taihuensis (OD_{600nm} 0.5)



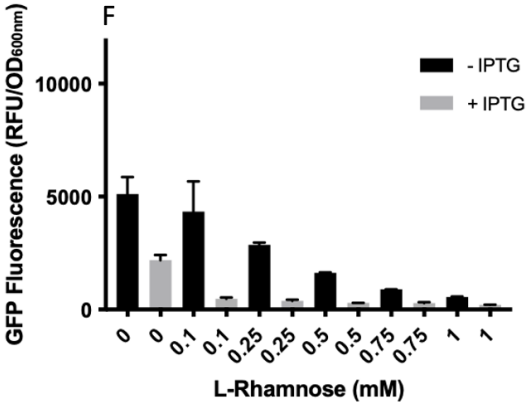
Cupriavidus necator (OD_{600nm} 0)



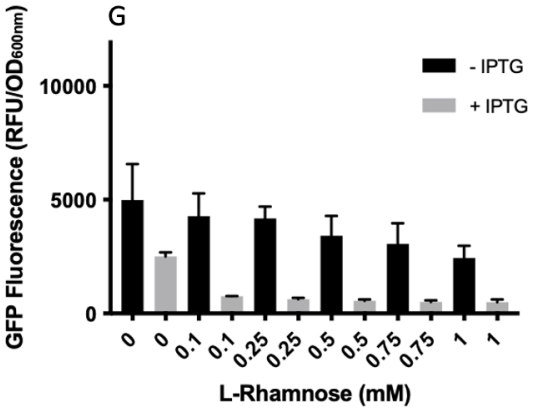
Cupriavidus necator (OD_{600nm} 0.5)



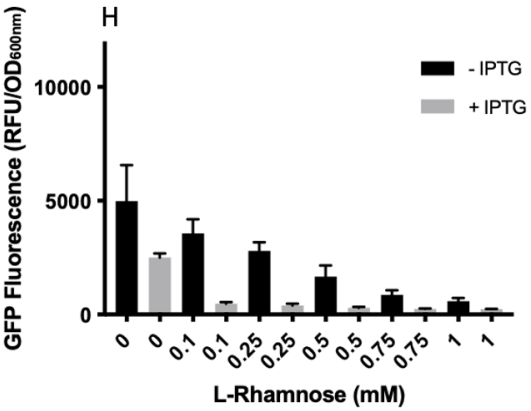
Escherichia coli (OD_{600nm} 0)



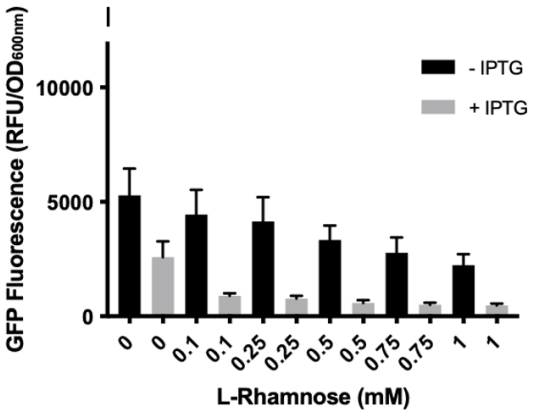
Escherichia coli (OD_{600nm} 0.5)



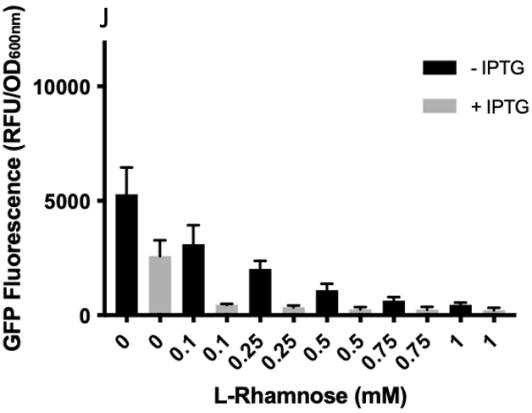
Escherichia coli R316 C-terminal Truncation (OD_{600nm} 0)



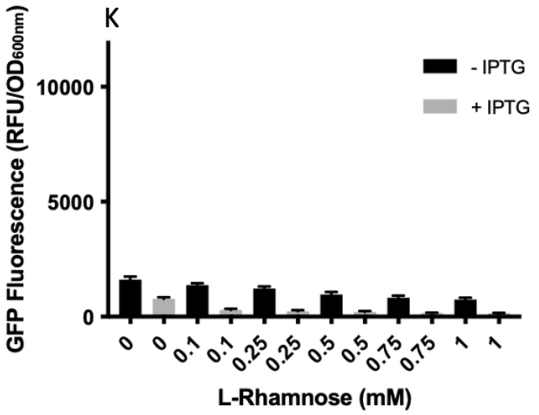
Escherichia coli R316 C-terminal Truncation (OD_{600nm} 0.5)



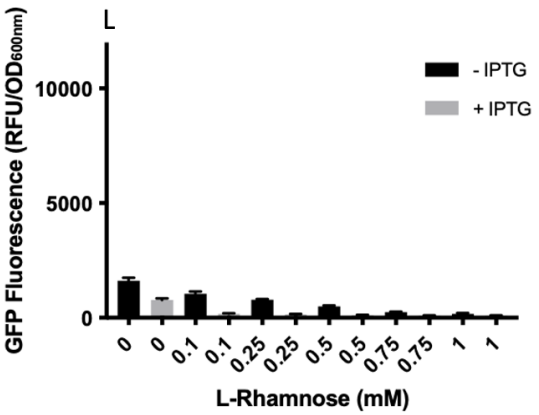
Klebsiella pneumoniae (OD_{600nm} 0)



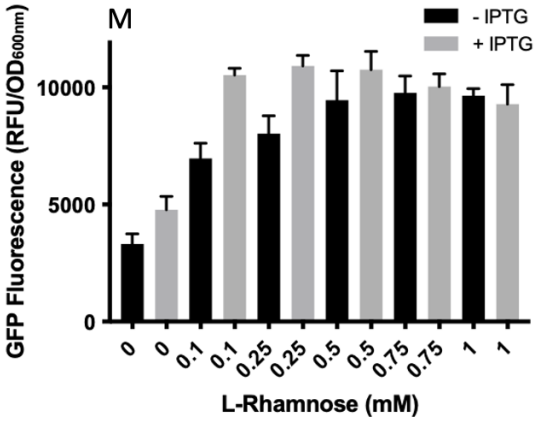
Klebsiella pneumoniae (OD_{600nm} 0.5)



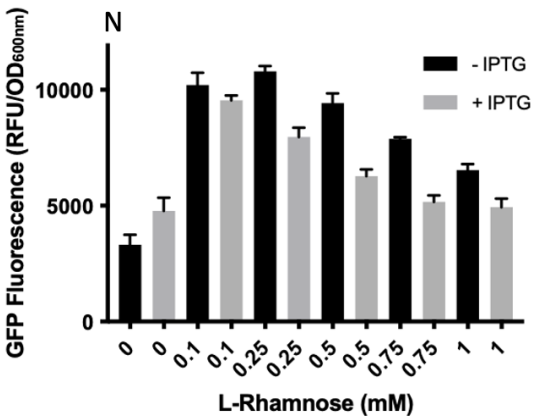
Pseudomonas aeruginosa R318 C-terminal Truncation (OD_{600nm} 0)



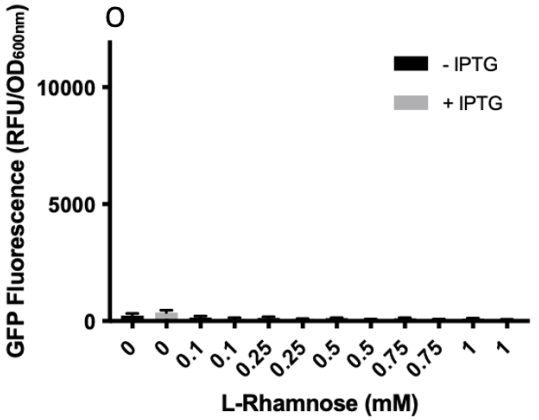
Pseudomonas aeruginosa R318 C-terminal Truncation (OD_{600nm} 0.5)



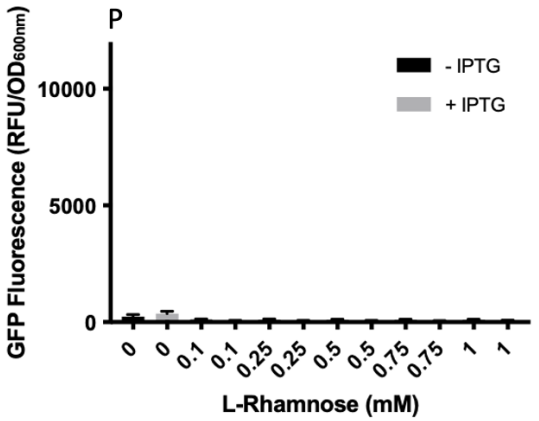
Pseudomonas putida (OD_{600nm} 0)



Pseudomonas putida (OD_{600nm} 0.5)



Vibrio breoganii (OD_{600nm} 0)



Vibrio breoganii (OD_{600nm} 0.5)

Figure 3.5: The optimal SLC10A7 *A. taihuensis*, *C. necator*, *E. coli* full-length and R316 C-terminal truncated, *K. pneumoniae*, *P. aeruginosa* R318 C-terminal truncated, *P. putida*, and *Vibrio breoganii* small-scale overexpression trials. Increasing concentrations of L-Rhamnose (0-1 mM) was added at OD_{600nm} 0 (A, C, E, G, I, K, M, and O) and OD_{600nm} 0.5 (B, D, F, H, J, L, N, and P) of the overexpression culture. Negative and positive inductions of 0.4 mM IPTG are also reported in black and grey bars, respectively. Error bars are shown as black "T" lines above the histograms and were calculated as the standard deviations from the mean of all expression trials (a minimum of three independent repeats). Data were analysed in GraphPad Prism.

The SLC10A7 bacterial overexpressions, as seen in figure 3.5.A-P above, overall yielded good expression levels. Both SLC10A7 *Cupriavidus necator* and *Pseudomonas putida* yielded the greatest overexpression at roughly 12,000 GFP counts/OD_{600nm} each, whilst SLC10A7 *Pseudomonas aeruginosa* R318 C-terminal truncated and *Vibrio breoganii*, in comparison, generated the lowest expression counts at roughly 1,750 and 480 GFP counts/OD_{600nm} respectively.

A summary of the maximal bacterial overexpression yields for all analysed constructs, and their optimised expression conditions, ranked from highest to lowest, are presented in table 3.3 below.

Table 3.3: Optimal SLC10A7 bacterial homologue small-scale overexpression yields

Rank	**Bacterial Organism**	**Maximal Overexpression Yield (GFP Fluorescence (RFU/OD_{600nm})**	**Optimised Expression Condition**
1	*Cupriavidus necator*	12,000	0 mM L-rhamnose, 0 mM IPTG
1	*Pseudomonas putida*	12,000	OD_{600nm} 0: (0.5 mM L-rhamnose, 0.4 mM IPTG) OD_{600nm} 0.5: (0.25 mM L-rhamnose, 0 mM IPTG)
2	*Escherichia coli* R316 C-terminal Truncation	6,800	0 mM L-rhamnose, 0 mM IPTG
3	*Klebsiella pneumoniae*	6,500	0 mM L-rhamnose, 0 mM IPTG
4	*Escherichia coli*	6,000	0 mM L-rhamnose, 0 mM IPTG
5	*Asticcacaulis taihuensis*	4,250	0 mM L-rhamnose, 0 mM IPTG
6	*Pseudomonas aeruginosa* R318 C-terminal Truncation	1,750	0 mM L-rhamnose, 0 mM IPTG
7	*Vibrio breoganii*	480	0 mM L-rhamnose, 0.4 mM IPTG

As can be seen in both figure 3.5.A-P and table 3.3 above, all SLC10A7 bacterial optimal expression yields, with the exception of both *Pseudomonas putida* and *Vibrio breoganii*, were achieved in the absence of both L-rhamnose and IPTG.

For SLC10A7 *P. putida* maximal GFP counts were achieved under varying concentrations of L-rhamnose: when the latter was added immediately to the starting culture at an OD_{600nm} 0, conditions of 0.5 mM L-rhamnose and 0.4 mM IPTG yielded maximal expression; whilst addition at OD_{600nm} 0.5 favoured 0.25 mM L-rhamnose and 0 mM IPTG conditions. These results indicate that for *Pseudomonas putida*, the requirement of IPTG, to yield optimal expression, is dependent on when L-rhamnose is added to the culture.

SLC10A7 *V. breoganii* on the other hand yielded optimal expression with the sole addition of 0.4 mM IPTG. However, as seen in figure 3.5.O-P, these overexpression yields were extremely low and thus deemed to be insufficient for downstream analytical studies including FSEC, Purification and Crystallisation. As such, this construct was excluded from these biochemical analyses.

Finally, in comparison to the exceptionally well-expressing $ASBT_{NM}$ membrane protein, the SLC10A7 bacterial homologues achieved significantly lower overexpression yields. However, since $ASBT_{NM}$ can be regarded as an exceptional and unique expressing membrane protein, we can conclude the SLC10A7 bacterial homologues (with the exception of *V. breoganii* yielded good levels of overexpression and were deemed to be more than sufficiently expressed for downstream structural and functional analyses.

3.3 Large-scale Membrane Protein Overexpression and Extraction

3.3.1 $ASBT_{Lb}$ Overexpression and Membrane Extraction

Following the identification of optimal overexpression conditions for the $ASBT_{Lb}$ membrane protein by Oliver Huxley (MBio student - 2019), large scale overexpression was conducted in Lemo21(DE3) cells with 0.25 mM L-rhamnose (added at OD_{600nm} 0), and 0.4 mM IPTG (added at OD_{600nm} 0.5). After 15-20 h growth, membrane extraction was established, following a standardised procedure and use of a cell disruptor (see methods section for details).

Figure 3.6 below demonstrates both in-gel fluorescence and SDS-PAGE gel analysis of this overexpression and membrane extraction, with GFP fluorescence providing good indication of the overexpression yields.

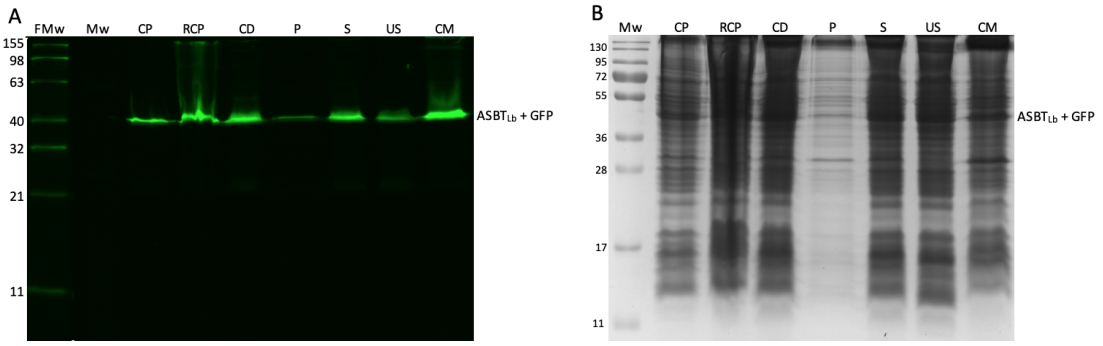


Figure 3.6: $ASBT_{Lb}$ protein overexpression and membrane extraction. In-gel fluorescence (A) and SDS-PAGE (B) gel images.
FMw: Fluorescent Molecular Weight ladder, Mw: Molecular Weight ladder, CP: Cell Pellet, RCP: Resuspended Cell Pellet, CD: Cell Disruptor, P: Pellet, S: Supernatant, US: Ultracentrifuge Supernatant, CM: Cell Membrane.
Gel fluorescence imaged at 1 s intensity.

As can be seen in figure 3.6 above, the fused protein:GFP tag was able to visually assess the intensity and quantity of $ASBT_{Lb}$ protein, in all collected samples, with the greatest fluorescence intensity present in the extracted cell membrane fraction. Overall, very good protein yields were established and were sufficiently expressed for downstream analysis in purification and crystallisation trials.

3.3.2 SLC10A7 *E. coli* Overexpression and Membrane Extraction

Following the identification of optimal overexpression conditions for the multiple SLC10A7 bacterial homologue membrane proteins (see above), large scale overexpression for the *Escherichia coli* construct was established in Lemo21(DE3) cells with 0 mM L-rhamnose, and 0 mM IPTG. After 15-20 h growth, membrane extraction was again conducted, following a standardised procedure and use of a cell disruptor (see methods section for details).

Figure 3.7 below demonstrates both in-gel fluorescence and SDS-PAGE gel analysis of this overexpression and membrane extraction, with GFP fluorescence providing good indication of the overexpression yields.

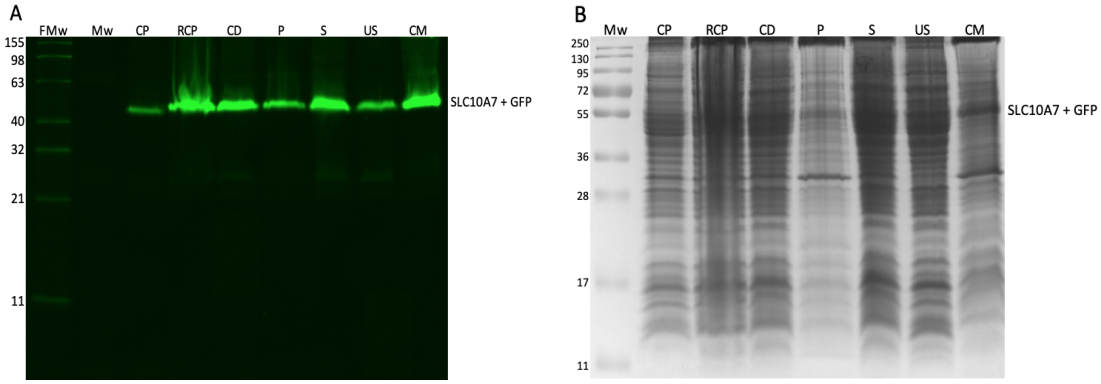


Figure 3.7: SLC10A7 *E. coli* protein overexpression and membrane extraction. (A) In-gel fluorescence and (B) SDS-PAGE gel images.
FMw: Fluorescent Molecular Weight ladder, Mw: Molecular Weight ladder, CP: Cell Pellet, RCP: Resuspended Cell Pellet, CD: Cell Disruptor, P: Pellet, S: Supernatant, US: Ultracentrifuge Supernatant, CM: Cell Membrane.
Gel fluorescence imaged at 1 s intensity.

As can be seen in figure 3.7 above, the use of a GFP fusion protein tag is an incredibly useful tool to be able to track visually the intensity of the membrane protein. Overall, the SLC10A7 *E. coli* overexpression and extracted membrane protein fluorescence yields represent a stable GFP fusion to the protein, with excellent yields for downstream biochemical and biophysical analysis.

3.3.3 SLC10A7 *E. coli* Q320 Truncated Overexpression and Membrane Extraction

Due to the predicted instability and flexibility of the C-terminal domain of the SLC10A7 *E. coli* transporter (as predicted by RONN disorder plots (Yang et al., 2005) (see Appendix 1)), further large-scale overexpression and membrane extraction investigations were conducted for the Q320 truncated bacterial homologue, with the fundamental aim of generating a greatly improved stabilised protein state (as was identified in GFP-TS studies, see chapter 4), to yield better downstream purity and increased chances of generating well-diffracting protein crystals (see chapter 5).

Again membrane protein overexpression and extraction followed the same protocols and procedures as outlined previously. Figure 3.8 below, highlights the protein intensities in in-gel fluorescent and SDS-PAGE gels.

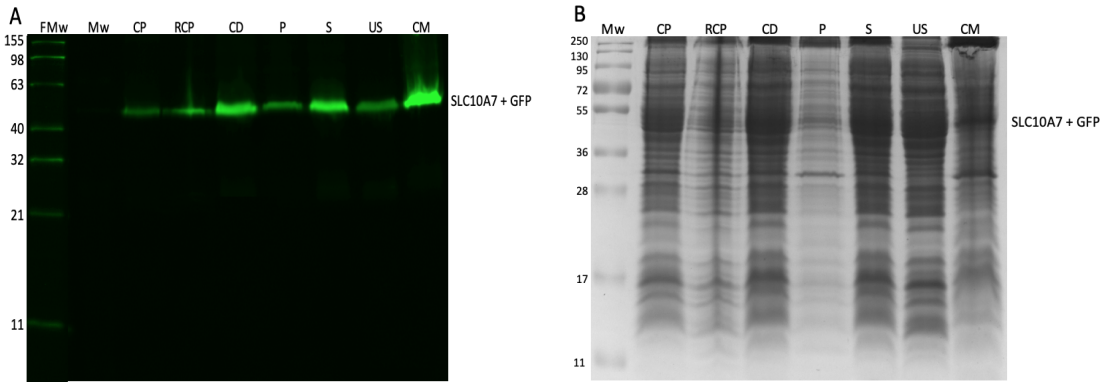


Figure 3.8: SLC10A7 *E. coli* Q320 truncated protein overexpression and membrane extraction. In-gel fluorescence (A) and SDS-PAGE (B) gel images.
FMw: Fluorescent Molecular Weight ladder, Mw: Molecular Weight ladder, CP: Cell Pellet, RCP: Resuspended Cell Pellet, CD: Cell Disruptor, P: Pellet, S: Supernatant, US: Ultracentrifuge Supernatant, CM: Cell Membrane.
Gel fluorescence imaged at 1 s intensity.

As can be seen in figure 3.8 above, similar to the full-length *E. coli* construct, very good GFP-protein fusion is presented, yielding high GFP intensities, indicative of good overexpression and membrane levels.

3.4 Membrane Protein Solubilisation - Fluorescence Size-Exclusion Chromatography (FSEC)

In order to study challenging membrane proteins in-vitro, their natural lipidic environment must be stripped away and replaced by artificial detergent micelles. This often leads to increased membrane protein flexibility and instability, crucially hindering their crystallisation ability (due to their increased tendency to form aggregate states), thereby severely impacting the determination of 3D crystal structures (Hattori et al., 2012; Halan et al., 2014). Typically, detergents forming the smallest micelles tend to produce the most well-ordered crystals, however, they require the utmost stability and monodisperity of encapsulated proteins (Michel, 1983; Ostermeier and Michel, 1997). These specific factors can be investigated efficiently, prior to performing resource and time consuming purification and crystallisation trials, via the use of fluorescence size-exclusion chromatography (FSEC) (Kawate and Gouaux, 2006; Halan et al., 2014).

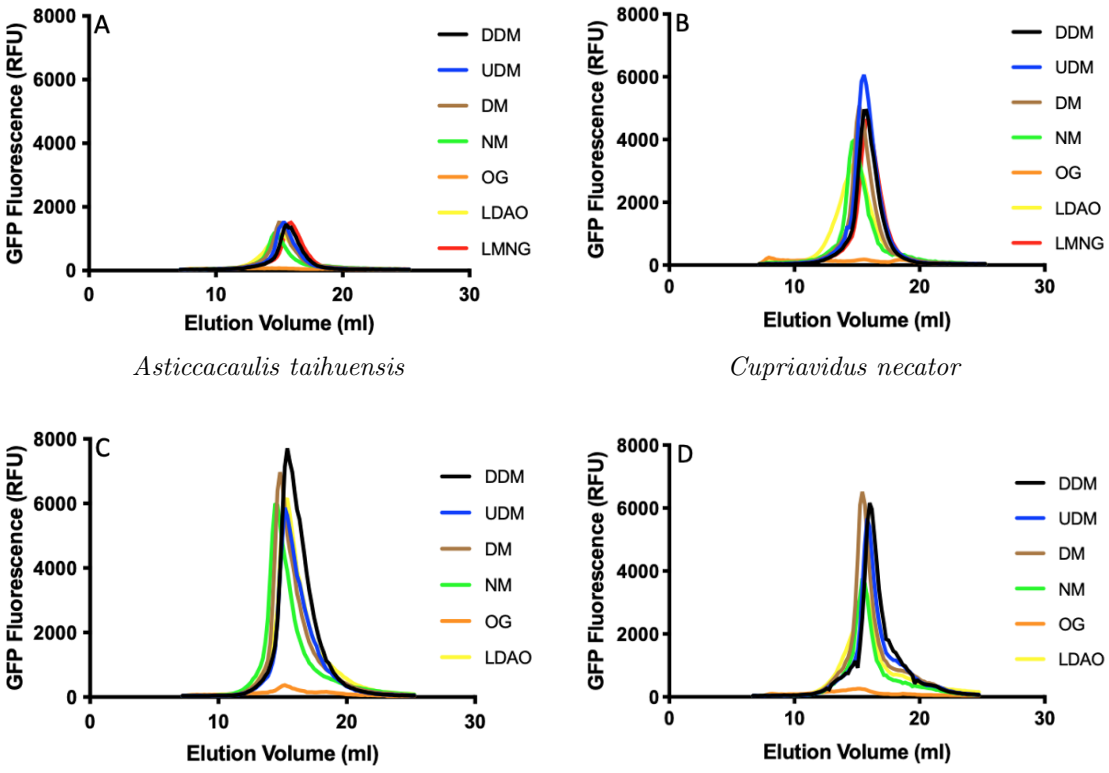
FSEC is an analytical technique used to fundamentally screen for precrystallisation attributes, such as assessing membrane protein stability; localisation; degree of monodispersity; structure-function relationships; level of expression and aggregation; protein-detergent complex associations; and approximate molecular mass (Kawate and Gouaux, 2006; Halan et al., 2014). The technique is rapid and reliable, and can generate high-throughput data, with the sole requirement being nanogram amounts of GFP-fused, non-purified solubilised cell membranes, or whole cell lysates (Kawate and Gouaux, 2006; Halan et al., 2014). Therefore the characterisation of GFP-tagged target proteins, without the need for purification, is possible and of extreme interest and importance.

Assessing the quality and functional ability for a detergent, to efficiently solubilise and stabilise a novel membrane protein, is vital for functional and structural investigations. Whether or not a detergent is suitable for crystallisation purposes, is essential for its success and production of high-resolution diffracting crystal data (Kawate and Gouaux, 2006). In light of this, FSEC was applied to the bacterial homologues of ASBT and SLC10A7, in order to screen for a variety of detergents, investigating their effect on membrane protein stability, and providing an imperative first step in increasing the likelihood their crystallisation (Kawate and Gouaux, 2006).

3.4.1 ASBT$_{Lb}$ and SLC10A7 Bacterial Homologue FSEC Trials

FSEC trials were once again previously investigated for the ASBT *Leptospira biflexa* bacterial homologue by Oliver Huxley (MBio student - 2019), and found to be significantly stable and monodisperse in a wide array of detergents including: n-dodecyl-β-D-maltopyranoside (DDM); n-undecyl-β-D-maltopyranoside (UDM); n-decyl-β-D-maltopyranoside (DM); n-Nonyl-β-D-maltopyranoside (NM); and Lauryldimethylamine N-oxide (LDAO). n-octyl-β-D-glucopyranoside (β-OG) was the only detergent investigated which proved to be destabilising and denaturing to the ASBT *L. biflexa* membrane protein.

For the multiple SLC10A7 bacterial homologues, FSEC trials were also conducted to assess the solubilising ability of a multitude of detergents. These results are outlined in greater detail in figure 3.9 below.

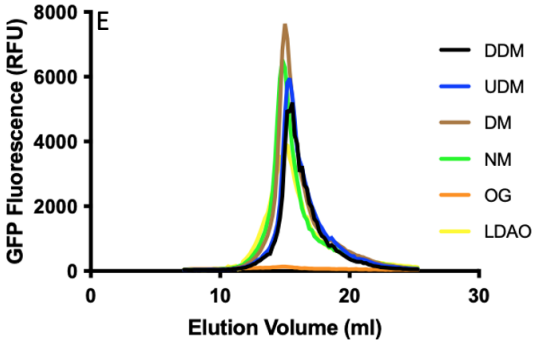


Asticcacaulis taihuensis

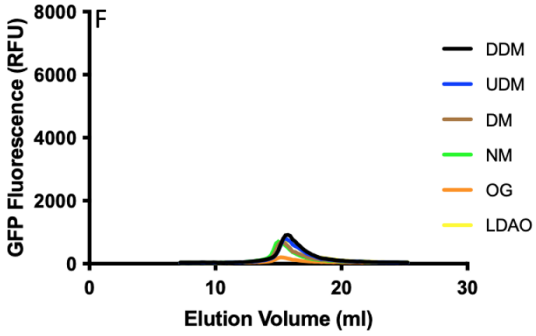
Cupriavidus necator

Escherichia coli

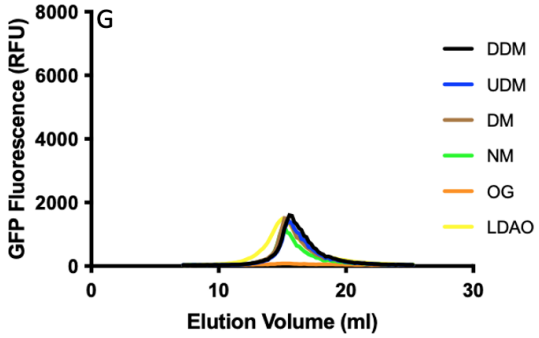
Escherichia coli R316 C-terminal Truncation



Klebsiella pneumoniae



Pseudomonas aeruginosai R318 C-terminal Truncation



Pseudomonas putida

Figure 3.9: SLC10A7 *A. taihuensis* (A), *C. necator* (B), *E. coli* full-length (C) and R316 C-terminal truncated (D), *K. pneumoniae* (E), *P. aeruginosa* R318 C-terminal truncated (F), and *P. putida* (G) FSEC detergent solubilisation trials.
DDM: n-dodecyl-β-D-maltopyranoside; UDM: n-undecyl-β-D-maltopyranoside; DM: n-decyl-β-D-maltopyranoside; NM: n-nonyl-β-D-maltopyranoside; OG: n-octyl-β-D-glucopyranoside; LDAO: lauryldimethylamine N-oxide; LMNG: lauryl Maltose neopentyl glycol.
Samples were run on a Superose 6 Increase 10/300 GL column and data analysed and processed in GraphPad Prism.

Overall the SLC10A7 bacterial homologue FSEC detergent solubilisation trials were well solubilised in a wide array of analysed detergents. In each case (with the exception of β-OG), a lack of both aggregation and lone GFP peaks were visible, indicating good, stable bacterial overexpression levels and tightly bound fusions with GFP.

The height of the individual fluorescent peak profiles provide crucial information on the expression and solubilisation levels of the SLC10A7 membrane protein in corresponding detergents. Whilst both SLC10A7 *E. coli* and *K. pneumoniae* were solubilised the most efficiently, with fluorescence levels reaching roughly 8,000 GFP counts each; SLC10A7 *A. taihuensis*, *Pseudomonas aeruginosa* R318 C-terminal truncation and *P. putida*, in contrast, yielded lower than expected solubilised GFP fluorescence levels at roughly 1,500, 1,000 and 1,700 GFP counts, respectively. SLC10A7 *P. putida* was especially surprising, considering it was a construct which overexpressed the best, suggesting a poor solubilising ability in the

detergents tested. Finally, both SLC10A7 *C. necator* and *E.coli* R316 C-terminal truncation showed moderate levels of solubility at roughly 6,000 and 6,500 GFP counts, respectively.

In all instances, the highest peaks (usually DDM) were eluted after roughly 15.5 ml, corresponding with a monomeric state of the protein. With gradual solubilisation in harsher detergents, protein peaks were seen to be eluted slightly earlier, highlighting a growing instability in the membrane protein. However, with the exception of β-OG, all detergents analysed, yielded sharp, symmetrical and monodisperse peaks, following a Gaussian distribution, indicative of a stabilised protein state.

Taken together, these FSEC results depict good solubilisation of all SLC10A7 bacterial constructs, whereby, DDM, UDM, DM, NM, LDAO and LMNG could serve as suitable detergent candidates for future crystallisation attempts. FSEC results from SLC10A7 *E.coli* full length and R316 C-terminal truncated, as well as *K. pneumoniae* and *C. necator*, are all promising for future downstream functional and structural investigations; whilst the poor solubility results of *P. aeruginosa* R318 C-terminal truncation in particular have excluded it from these analyses.

Chapter 4

Results: Membrane Protein Functional Thermostability

4.1 CPM Thermostability

Membrane protein structural and functional stability are both pivotal attributes which govern the successfulness of crystallisation and subsequent structure determination. For in-vitro analysis, their expulsion from the lipid bilayer requires the use of detergents, which replace the stable lateral pressures exerted from lipids of the cell membrane, with a weaker, freely coated "belt" of detergent monomers, resulting in the formation of protein-detergent complexes (PDC) (Alexandrov et al., 2008). Unsurprisingly, solubilisation in such detergents can lead to unfolded, destabilised and aggregated states of the protein. Therefore, it is imperative for the stability of a membrane protein to be fully optimised (as well as establishing good levels of solubility and homogeneity), in order to yield successful outcomes in downstream biochemical procedures, including both protein purification and crystallisation (Marsh, 1996; Alexandrov et al., 2008). Fortunately, the use of a thiol-specific fluorochrome N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM) offers the possibility of investigating, in a high-throughput manner, potential buffer solutions, detergent and ligand conditions, which can increase the thermostability and solubility of a membrane protein of interest (Alexandrov et al., 2008).

Exact specifications on how this CPM assay works, remains a little elusive. Alexandrov and colleagues originally proposed the CPM fluorochrome (which is naturally non-fluorescent in its unbound state (Ayers et al., 1986)) to bind specifically to embedded membrane protein cysteine residues (Alexandrov et al., 2008). At increasing temperatures, the membrane protein begins to unfold, allowing for buried cysteine residues in the hydrophobic core to become readily accessible to

the CPM probe, resulting in the formation of thioether bonds, and emitting high fluorescence intensities (Alexandrov et al., 2008). However, Wang and colleagues instead suggested cysteine residues (and other sulfhydryl residues) not to be essential as prerequisites for this assay, instead proposing the non-covalent binding of the CPM dye to other hydrophobic regions/patches such as lysine residues (Wang et al., 2015). Although significantly lower fluorescent measurements were obtained when compared to cysteine residues, binding to lysine residues did successfully yield sufficient fluorescent intensities to generate accurate melting temperature (T_m) measurements (Alexandrov et al., 2008; Wang et al., 2015).

Regardless of the mechanics of the assay, the result of CPM binding generates a two-state model, consisting of a low fluorescent, natively folded state, and an upper fluorescent, denatured state. Monitoring of the fluorescent changes between both these states elucidates the membrane protein's structural thermostability, allowing for the successful and reliable calculation of its T_m value) (Alexandrov et al., 2008; Wang et al., 2015).

Unfortunately, like many techniques, there are limitations to what can be successfully analysed. Primarily, the CPM dye has been seen to reach lower fluorescent intensities at pH's lower than 6.0, outlining a significant reduction in its reactive ability with cysteines/hydrophobic patches. Furthermore, stability investigations at pH levels <5.0, significantly altered the sigmoidal appearance of the melting curve, thereby highlighting the limitation of the dye for investigations of membrane protein stability at low pH's. Consequently, at pH levels above 8.0, the CPM dye was also seen to experience reduced chemical selectivity, also rendering membrane stability investigations in these conditions unsuitable (Alexandrov et al., 2008).

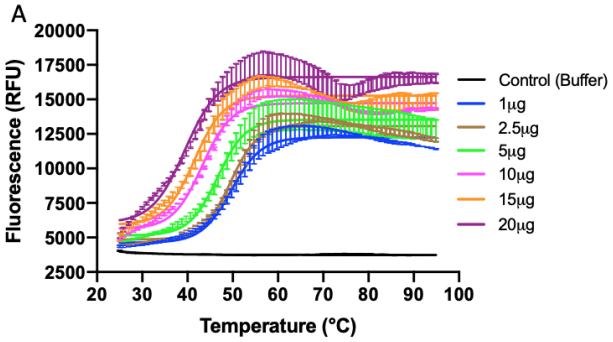
However, the CPM stability assay, requiring purified protein quantities of <10 µg, provides a highly sensitive, accurate, versatile and reliably reproducible assessment of a membrane protein's structural stability, in a wide array of conditions. Indeed, very good compatibility with a multitude of detergents and other ligands/compounds is possible, yielding high signal-to-noise ratios (Alexandrov et al., 2008). In comparison, the use of other thermofluor dyes, such as sypro-orange, to analyse membrane protein stability, produces high background signals, which significantly interfere with the unfolding transition of solubilised membrane proteins (Yeh et al., 2006; Alexandrov et al., 2008). Therefore, taken together, we deemed the application of the CPM assay to be a fundamental technique to investigate thermostability of the ASBT *L. biflexa* membrane protein, thereby investigating a wide array of different compounds, buffer conditions and potential substrates, which potentially increase the thermostability of the protein.

4.1.1 ASBT$_{Lb}$ CPM Functional Thermostability Investigations

The presence of cysteine aa residues in membrane proteins, confers protein conformational stability and oligomerisation, via the generation of intramolecular and intermolecular disulfide bonds and linkages, respectively. These covalent associations have been shown to be fundamentally coupled to a protein's function, stability and trafficking (Chothe et al., 2018).

In light of this, ASBT$_{Lb}$ thermostability was investigated via the CPM assay, with a concentration dependent series depicted below in figure 4.1, outlining the best protein concentration to use (i.e. yielding the most reliable sigmoidal curve), which can generate reliable comparisons for downstream analysis in a variation of buffering conditions and substrates.

A



B

Membrane Protein Concentration	Melting Temperature, Tm (°C)
1 µg	49.71 ± 0.24
2.5 µg	49.43 ± 0.25
5 µg	46.61 ± 0.36
10 µg	43.23 ± 0.30
15 µg	41.52 ± 0.39
20 µg	39.44 ± 0.48

Figure 4.1: ASBT$_{Lb}$ CPM thermostability concentration series investigation. (A) Thermostability traces in 20 mM Tris pH 7.5, 150 mM NaCl, 0.03 % DDM and 0-20 µg purified ASBT$_{Lb}$ protein, were investigated. (B) Calculated Tm values for these CPM investigations are displayed in a table.
Data was analysed and processed in GraphPad Prism.

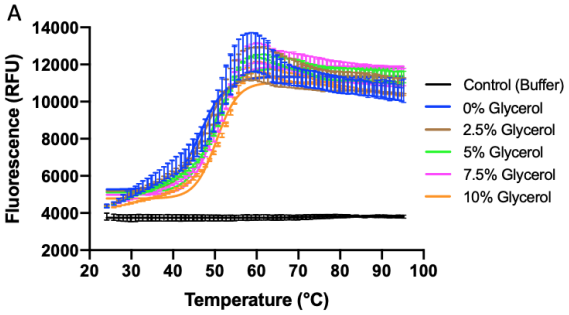
As can be seen in figure 4.1 above, a range of protein concentrations (from 0-20 µg) were investigated, in order to determine a suitable candidate for downstream Tm comparisons. As the protein concentration added increased, the overall thermostability and deduced melting temperature steadily declined, which was demonstrated by the plateau of the sigmoidal traces at earlier temperatures. However, we propose this not to be a decline in Tm, but rather an increase in sensitivity of the protein to the assay, due to increased concentration levels. Therefore, lower concentrations of the protein were deemed more suitable for downstream Tm comparisons, and the 2.5 µg concentration, yielding a reasonable stability of 49.4 °C, was selected for this purpose.

CPM Glycerol Stability Investigations

Glycerol has been shown to stabilise extracellular/intracellular loops of a protein, by reducing its conformational flexibility via the inhibition of local thermal backbone dynamics (Tsai et al., 2000; Weinkauf et al., 2001; Xia et al., 2004). As such, its addition guarantees the greatest native structural compact state of the protein (Alexandrov et al., 2008).

Thermostability investigations of the $ASBT_{Lb}$ protein with increasing concentrations of glycerol were investigated with the CPM assay. 2.5 µg $ASBT_{Lb}$ protein was used for the investigation in standardised buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.03 % DDM). Results from this investigation are displayed below in figure 4.2.

A



B

Glycerol Condition	Melting Temperature, Tm (°C)
0 % Glycerol (Control)	46.50 ± 0.66
2.5 % Glycerol	47.28 ± 0.43
5 % Glycerol	49.28 ± 0.29
7.5 % Glycerol	49.72 ± 0.30
10 % Glycerol	50.39 ± 0.29

Figure 4.2: $ASBT_{Lb}$ CPM thermostability glycerol investigation. (A) Thermostability traces in 20 mM Tris pH 7.5, 150 mM NaCl, 0.03 % DDM, 0-10 % glycerol, and 2.5 µg $ASBT_{Lb}$, were investigated. (B) Calculated Tm values for these CPM investigations are displayed in a table. Data was analysed and processed in GraphPad Prism.

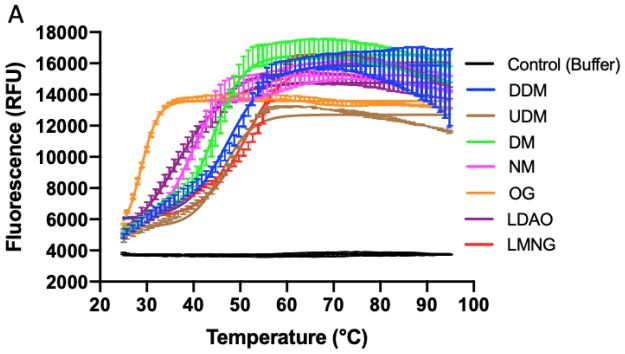
As can be observed in figure 4.2 above, gradual increase in glycerol concentration from 0-10 %, significantly increased the Tm of $ASBT_{Lb}$. Indeed, the greatest melting temperature was seen at 10 % glycerol, with a reported Tm of 50.4 °C. Compared to the 46.5 °C melting temperature in the absence of glycerol, this thermal shift of 3.9 °C is significant, and should be considered for future downstream biochemical/biomolecular processes.

CPM Detergent Stability Investigations

Since the chemical natures of detergent surfactants can vary significantly, the elucidation of their levels of solubility and stability on target membrane proteins are of the utmost importance (Odahara, 2004). These investigations are possible via CPM analysis.

In light of this, a range of detergents, with varying CMC, were selected, to investigate the thermostability properties of the $ASBT_{Lb}$ membrane protein. Results from these investigations are depicted below in figure 4.3.

A



B

Detergent Condition (3x CMC)	Melting Temperature, Tm (°C)
0.03 % DDM (Control)	47.29 ± 0.53
0.09 % UDM	46.15 ± 0.34
0.26 % DM	44.09 ± 0.31
0.84 % NM	39.41 ± 0.32
2.67 % β-OG	28.77 ± 0.25
0.07 % LDAO	36.32 ± 0.56
0.003 % LMNG	50.89 ± 0.44

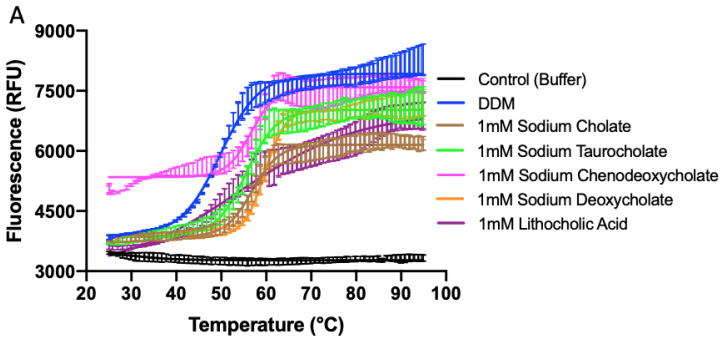
Figure 4.3: $ASBT_{Lb}$ CPM thermostability detergent investigation. (A) Thermostability traces in 20 mM Tris pH 7.5, 150 mM NaCl, 2.5 µg $ASBT_{Lb}$, and a range of detergents (0.03 % DDM, 0.09 % UDM, 0.26 % DM, 0.84 % NM, 2.67 % β-OG, 0.07 % LDAO, and 0.003 % LMNG), were investigated. (B) Calculated Tm values for these CPM investigations are displayed in a table. Data was analysed and successfully processed in GraphPad Prism.

As is depicted in figure 4.3 above, gradual incorporation of the ASB_{Lb} protein into 'harsher"/smaller micelled detergents led to decreased overall thermostability. Indeed, all tested detergents, apart from LMNG, were less stabilising than DDM, with the β-OG detergent exhibiting the weakest protein thermostability at 28.8 °C. This calculated Tm demonstrated a significant thermal shift of −18.5 °C from DDM incorporated $ASBT_{Lb}$. LMNG, on the other hand, was the only detergent to increase the melting temperature of the $ASBT_{Lb}$ protein when compared to DDM. Indeed, a thermal shift of 3.6 °C was observed, which is a significant result and must be considered for future purification strategies.

CPM Substrate Stability Investigations

Specific additives and ligands such as sugars, organic compounds, salts, cofactors and detergents have previously been seen to interact with proteins, affecting their biophysical properties. Assessing the potential stabilities and involvements of this diverse group of molecules on target proteins, is readily possible via the application of CPM assays (Alexandrov et al., 2008).

Since the human ASBT transporter is categorically involved in the reabsorption and transport of bile acids in the EHC, from the intestines to the liver, BA substrates were investigated via CPM thermostability assays for the $ASBT_{Lb}$ bacterial homologue. The results from these CPM investigations are portrayed below in figure 4.4.



B

Bile Acid Condition	Melting Temperature, Tm (°C)
20 mM Tris pH 7.5 + 150 mM NaCl + 0.03 % DDM (Control)	49.74 ± 0.26
1 mM Sodium Cholate	56.89 ± 0.24
1 mM Sodium Taurocholate	55.57 ± 0.27
1 mM Sodium Chenodeoxycholate	56.68 ± 0.41
1 mM Sodium Deoxycholate	59.64 ± 0.15
1 mM Lithocholic Acid	53.84 ± 1.46

Figure 4.4: $ASBT_{Lb}$ CPM thermostability bile acids investigation. (A) Thermostability traces in 20 mM Tris pH 7.5, 150 mM NaCl, 0.03 % DDM, 2.5 µg $ASBT_{Lb}$, and a range of bile acids (1 mM Sodium cholate, 1 mM Sodium taurocholate, 1 mM Sodium chenodeoxycholate, 1 mM Sodium deoxycholate, and 1 mM Lithocholic acid), were investigated. (B) Calculated Tm values for these CPM investigations are displayed in a table.
Data was analysed and successfully processed in GraphPad Prism.

As is depicted in figure 4.4 above, the addition of 1 mM BAs to the $ASBT_{Lb}$ membrane protein, had a significant effect on the generated protein Tm values. Indeed, every investigated BA increased the thermostability of the $ASBT_{Lb}$ protein, highlighting a significant interaction and thermal increase in stability. Most interestingly, addition of 1 mM Sodium deoxycholate (DCA), demonstrated the greatest increase in the melting temperature of the $ASBT_{Lb}$ protein with 59.6 °C, corresponding to a thermal shift of 9.9 °C, from DDM control conditions. In light of this finding, DCA was incorporated in downstream purification and crystallisation strategies, in order to increase the likelihood of successful 3D protein structure determination.

A wide array of other solutes and potential substrates were also investigated with the CPM assay, based on their availability in the laboratory and weak/minor associations to bile acids and the EHC. These substrates and their associated Tm values are portrayed in table 4.1 below.

Table 4.1: $ASBT_{Lb}$ substrate stability investigations with CPM assays.

Substrates	**Melting Temperature, Tm (°C)**
20 mM Tris pH 7.5 + 150 mM NaCl (Control)	**51.7 ± 1.1**
1 mM a, a, D-Trehalose Dihydrate	50.9 ± 0.3
1 mM 4-Aminobenzoic Acid (PABA)	49.3 ± 0.5
1 mM Aluminium Chloride	49.6 ± 0.6
1 mM Ammonium Bicarbonate	49.2 ± 1.5
1 mM Ammonium Chloride	48.9 ± 1.5
1 mM Ammonium Fluoride	48.3 ± 1.2
1 mM Ammonium Formate	48.6 ± 5.2
1 mM Ammonium Molybdate	46.1 ± 0.5
1 mM Ammonium Sulfate	49.6 ± 0.9
1 mM Anthranilic Acid	48.4 ± 1.5
1 mM L+Arabinose	49.6 ± 0.7
1 mM L-Ascorbic Acid (Vitamin C)	48.8 ± 1.2
1 mM Benzoic Acid	49.6 ± 1.2
1 mM Beta-Alanine	48.5 ± 1.2
1 mM Betaine Monohydrate	48.8 ± 1.2
1 mM Cadaverine	48.8 ± 2.0
1 mM Cadmium Chloride Hemi(pentahydrate)	47.0 ± 0.4
1 mM Caesium Chloride	48.6 ± 0.6
1 mM Calcium Chloride	48.7 ± 2.2
1 mM Chloranilic Acid	47.9 ± 0.6
1 mM Choline Chloride	51.2 ± 0.9

1 mM Chromic Chloride	46.6 ± 0.5
1 mM Cobalt Chloride	48.7 ± 0.3
1 mM Cobaltous Nitrate	47.9 ± 0.5
1 mM 3-4, Dihydroxyphenylacetic Acid	47.0 ± 0.3
1 mM EDTA	49.0 ± 0.3
1 mM EGTA	48.5 ± 0.5
1 mM Folic Acid	47.7 ± 0.9
1 mM D-Fructose	49.0 ± 0.2
1 mM D-Galactose pH 7.5	49.3 ± 0.5
1 mM D-Glucose	51.0 ± 0.4
1 mM Glucuronic Acid pH 7.5	48.8 ± 0.3
1 mM L-Glutamic Acid	48.9 ± 3.3
1 mM β-Glycerophosphate	48.9 ± 1.3
1 mM Glycine	49.2 ± 3.2
1 mM Glycylglycine	48.7 ± 0.5
1 mM Guanidine	49.2 ± 0.6
1 mM 1,6-Hexanediol	48.9 ± 2.6
1 mM Hexamine Cobalt Chloride	48.4 ± 0.9
1 mM Hydroxylamine	48.4 ± 1.6
1 mM α-Ketoglutaric Acid	49.7 ± 4.6
1 mM α-Lactose	48.5 ± 1.4
1 mM Lithium Chloride	46.9 ± 0.3
1 mM Lithium Nitrate	46.9 ± 3.0
1 mM Lithium Sulphate	48.0 ± 2.5

1 mM Magnesium Acetate	48.8 ± 1.8
1 mM Magnesium Chloride	48.5 ± 0.4
1 mM Magnesium Sulphate	49.9 ± 1.1
1 mM L-Malic Acid	49.9 ± 2.4
1 mM Malonic Acid	47.6 ± 2.0
1 mM Maltose	49.0 ± 1.8
1 mM Manganese (II) Chloride	49.5 ± 3.7
1 mM D-Mannitol	48.3 ± 1.4
1 mM Methyl-B-Cyclodexin	49.1 ± 0.3
1 mM Myo-Inositol	51.2 ± 1.8
1 mM N-Acetyl-L-Glutamic Acid	49.8 ± 0.4
1 mM N-Bromo-Succinimide	41.2 ± 4.6
1 mM N-Hydroxy Succinimide	46.5 ± 1.5
1 mM Niacinimide	49.1 ± 0.1
1 mM Nicotinic Acid	49.4 ± 1.0
1 mM N-Lauryl Sarcosine	55.7 ± 0.1
1 mM N-Methyl-D-Glucamine	48.4 ± 0.7
1 mM O-Phenylene-Diamine	49.4 ± 0.5
1 mM Oxalic Acid	48.2 ± 1.5
1 mM Poly(acrylic acid sodium salt)	50.0 ± 0.2
1 mM Potassium Bicarbonate	49.2 ± 0.1
1 mM Potassium Bromide	48.4 ± 0.3
1 mM Potassium Carbonate	48.4 ± 0.2

1 mM Potassium Chloride	48.5 ± 0.1
1 mM Potassium Fluoride	48.6 ± 0.3
1 mM Potassium-Sodium Tartrate Tetrahydrate	50.7 ± 0.7
1 mM Pyridoxine Monohydrochloride	48.3 ± 0.4
1 mM Rubidium Chloride	47.4 ± 0.3
1 mM Sodium Benzoate	50.9 ± 2.4
1 mM Sodium Bicarbonate	51.4 ± 0.7
1 mM Sodium Borate	49.4 ± 0.9
1 mM Sodium Citrate	50.6 ± 1.0
1 mM Sodium Fluoride	51.7 ± 0.8
1 mM Sodium Formate	51.2 ± 1.1
1 mM (+) Sodium L-Ascorbate	50.2 ± 0.7
1 mM Sodium Molybdate	50.2 ± 2.2
1 mM Sodium Pantoate	49.0 ± 0.6
1 mM Sodium Pyrophosphate Decahydrate	50.5 ± 2.6
1 mM Sodium Pyruvate	49.8 ± 0.3
1 mM Sodium Selenite	51.0 ± 2.2
1 mM Sodium Sulfate Decahydrate	52.2 ± 0.8
1 mM Sodium Thiocyanate	52.4 ± 4.1
1 mM D-Sorbitol	49.8 ± 1.1
1 mM Sucrose	50.3 ± 0.2
1 mM (-) D-Tartaric Acid	49.9 ± 0.9
1 mM (+) L-Tartaric Acid	51.9 ± 1.3
1 mM Taurine	50.1 ± 0.6

1 mM Triethanolamine Hydrochloride	50.7 ± 0.5
1 mM Trimethylamine N-Oxide Dihydrate	49.7 ± 1.4
1 mM Xylitol	48.7 ± 0.3
1 mM Zinc Sulfate	47.0 ± 0.6

As is shown in table 4.1 above, the vast majority of investigated substrates failed to substantially increase the melting temperature of the $ASBT_{Lb}$ protein. Typically, a Tm value of >3-5 °C is deemed to be a significant result. Only a few substrates showed increased thermostability, including sodium sulfate decahydrate (thermal shift 0.5 °C), sodium thiocyanate (thermal shift 0.7 °C) and (+) L-tartaric acid (thermal shift 0.2 °C), however, these were so minimal, were deemed insignificant and eliminated as possible $ASBT_{Lb}$ solutes. N-lauryl sarcosine (thermal shift 4.0 °C) was the only substrate to provide a reasonably significant result. However, since this is an un-natural, industrialised, anionic surfactant, most likely increasing the thermostability of the transporter by forming mixed micelles with the DDM detergent, this compound was also disregarded as being a potential $ASBT_{Lb}$ substrate.

Taken together, these CPM assays provided extremely reliable thermostability results for the $ASBT_{Lb}$ transporter, yielding extremely interesting findings. The significantly increased thermal shift (9.9 °C) caused by the addition of 1 mM DCA, prompted its incorporation in future downstream purification procedures, as is explored in greater detail in chapter 5. The addition of 10 % glycerol (thermal shift 3.9 °C) was also considered and can provide an alternative option for future crystallisation attempts.

4.2 FSEC-TS and GFP-TS Thermostability

In comparison to the CPM assay (which requires the use of well-purified protein samples), the presence of a GFP-fusion can greatly facilitate the monitoring of a membrane protein's stability, via thermostability techniques such as FSEC-TS and GFP-TS. Here, non-purified µg amounts of membrane protein GFP samples can be effectively and reliably investigated, as is explored in greater detail below.

Fluorescence Size-Exclusion Chromatography Thermostability (FSEC-TS)

Fluorescence detection size-exclusion chromatography thermostability (FSEC-TS), is a simple, efficient and reliably fast technique, used to yield crucial, mid-throughput novel information on the thermostability of ng-µg target protein, under varying conditions (Hattori et al., 2012). The covalently bound GFP tagged membrane protein can be used to screen for a wide array of specific additives, solubilised in detergent buffers, such as buffering capacitance, salts, pH, ions (anions and cations), substrates, ligands, lipids, etc (Hattori et al., 2012; Nji et al., 2018). The samples are heat treated at (4-100 °C), and analysed by FSEC, as described previously (Kawate and Gouaux, 2006). The results generate a temperature-induced denaturation of a target protein, elucidating its T_m (the temperature at which half the intrinsic fluorescence is lost), which can then be used as reference to compare investigations of future additive thermostabilisation studies (Hattori et al., 2012).

In specific situations were membrane protein solutes/substrates are unknown (as is the case for both ASBT and SLC10A7 bacterial homologues) and/or target proteins are known to be unstable, FSEC-TS offers the ability to investigate potential compounds and conditions which can significantly increase protein stability. Indeed, the successful isolation of small compounds and molecules, which are shown to increase membrane protein stability, leads to a direct increase in the likelihood of efficient purification and crystallisation procedures, of novel membrane proteins (Hattori et al., 2012).

Green Fluorescent Protein Thermostability (GFP-TS)

Following the same characteristic principles of FSEC-TS, Nji and colleagues, developed a simpler, more efficient GFP-TS thermal-shift assay, allowing for the generation of high-throughput additive, lipid and ligand screening interactions, with GFP-tagged membrane proteins (Nji et al., 2018). The core difference between techniques was the replacement of the limiting SEC stage in FSEC-TS (used to separate aggregates from protein), with a centrifugation step in GFP-TS.

Initial replacement of this SEC step, led to significant overestimations of Tm for analysed target proteins. This was believed to be largely due to inadequate exclusion of fluorescent aggregates in the centrifugated sedimentation step. The addition of β-OG to samples, prior to heat treatment, rectified this issue, thereby readily precipitating the aforementioned fluorescent aggregates. Through this optimised protocol, comparative and accurate Tm estimates for GFP-TS and FSEC-TS were then successfully reported, outlining the fundamentality of β-OG addition. Furthermore, comparisons to the N-[4-(7-diethylamino-4-methyl-3-coumarinyl) phenyl] maleimide (CPM) thermal-shift assay, with purified protein, also outlined a strong positive correlation in the generation of Tm values obtained from GFP-TS (Nji et al., 2018).

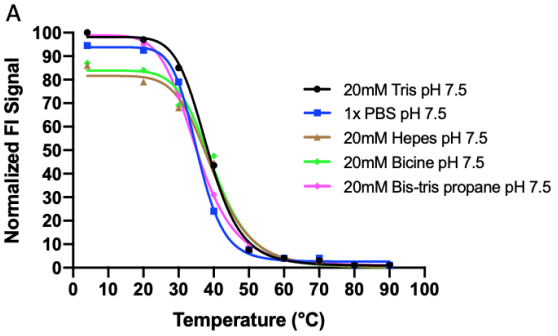
4.2.1 SLC10A7 *Cupriavidus necator* GFP-TS and FSEC-TS Functional Thermostability Investigations

SLC10A7 thermostability was investigated, in order to identify potential substrates, molecules and/or buffer conditions which increase Tm values. The *Cupriavidus necator* bacterial homologue was specifically selected for these studies, due to its significantly greater expression and solubilisation levels; as well as displaying the greatest sequence identity to human SLC10A7.

Buffer Ion Stability (GFP-TS)

A range of buffering ions (Tris, 1x PBS, Hepes, Bicine, and Bis-tris propane) at pH 7.5, were investigated, in order to investigate their buffering capacitance abilities and stability effect on the transporter. These ions were concentrated to 20 mM at pH 7.5, in the presence of 150 mM NaCl and 1 % DDM detergent. These GFP-TS thermostability traces and calculated Tm values are depicted in figure 4.5 below.

A



B

Buffer Condition	Melting Temperature, Tm (°C)
20 mM Tris pH 7.5 + 150 mM NaCl	38.42 ± 0.67
1 × PBS pH 7.5 + 150 mM NaCl	35.49 ± 0.44
20 mM Hepes pH 7.5 + 150 mM NaCl	37.32 ± 1.69
20 mM Bicine pH 7.5 + 150 mM NaCl	37.83 ± 1.95
20 mM Bis-tris propane pH 7.5 + 150 mM NaCl	35.15 ± 0.32

Figure 4.5: SLC10A7 *C. necator* buffering ion stability GFP-TS optimisation studies. (A) Thermostability traces in Tris, 1x PBS, Hepes, Bicine and Bis-tris propane, at 20 mM concentrations and pH 7.5, in the presence of 150 mM NaCl, and 1 % DDM, were investigated. Calculated Tm values (B) are also displayed in a table.
Data were analysed and processed in GraphPad Prism.

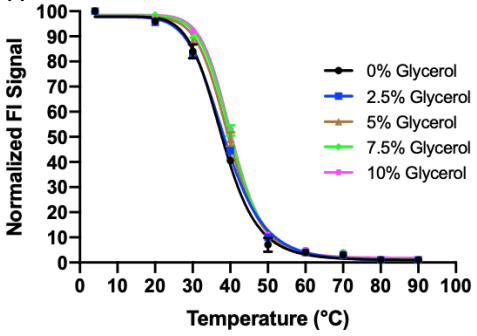
As can be seen in figure 4.5 above, SLC10A7 *C. necator* thermostability, is well conserved with minor Tm differences. Overall, Tris buffer seemed to stabilise the protein the best, yielding a Tm value of 38.4 ± 0.7℃. Hepes and Bicine seemed to yield slightly weaker melting temperatures at 37.3 ± 1.7℃, and 37.8 ± 2.0℃, respectively, however, these buffering ions inflicted a slightly reduced level of protein solubilisation. Finally, 1x PBS and Bis-tris propane yielded significantly lower SLC10A7 protein melting temperatures at 35.5 ± 0.4℃, and 35.2 ± 0.3℃, respectively.

Taken together, only minor Tm differences are reported from the analysed buffering ions, with Tris ions representing the greatest thermostability and solubilisation of the SLC10A7 protein. As such, Tris ions were used as the main buffering agent for future stability investigations.

Glycerol Stability (GFP-TS)

Glycerol-induced membrane protein stability was also analysed for the SLC10A7 *C. necator* transporter. As previously mentioned (see above in CPM section), the addition of glycerol stabilises the flexible extracellular/intracellular loops of a membrane protein, via the inhibition of the local thermal backbone dynamics (Tsai et al., 2000; Weinkauf et al., 2001; Xia et al., 2004). As such, buffered conditions consisting of 20 mM Tris pH 7.5, 150 mM NaCl, 1 % DDM detergent, and 0-10 % glycerol were investigated. These thermostability traces, as well as calculated Tm units, are portrayed in figure 4.6 below.

A



B

Glycerol Condition	Melting Temperature, Tm (°C)
20 mM Tris pH 7.5 + 150 mM NaCl	37.88 ± 0.71
20 mM Tris pH 7.5 + 150 mM NaCl + 2.5 % Glycerol	38.54 ± 0.68
20 mM Tris pH 7.5 + 150 mM NaCl + 5 % Glycerol	39.62 ± 0.64
20 mM Tris pH 7.5 + 150 mM NaCl + 7.5 % Glycerol	40.20 ± 0.61
20 mM Tris pH 7.5 + 150 mM NaCl + 10 % Glycerol	40.35 ± 0.39

Figure 4.6: SLC10A7 *C. necator* glycerol stability GFP-TS optimisation studies. (A) Thermostability traces in 20 mM Tris pH 7.5, 150 mM NaCl, 1 % DDM, and 0-10 % glycerol were investigated. Calculated Tm values (B) are also displayed in a table.
Data were analysed and processed in GraphPad Prism.

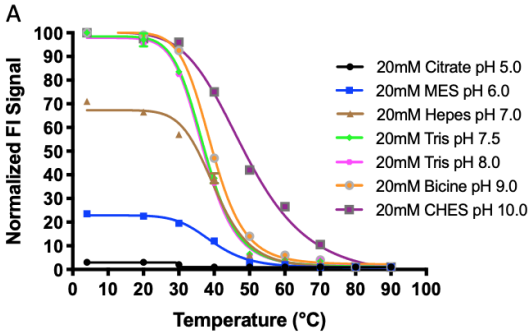
As can be observed in figure 4.6 above, the addition of increasing concentrations (0-10 %) of glycerol, has only a very negligible effect on the thermostability of SLC10A7 *C. necator*. Indeed, increasing amounts of glycerol concentrations, in the buffered system, had a direct positive correlation to an increase in membrane protein thermostability. However, this increase was very small, whereby

comparison of Tm measurements at 0 % and 10 % glycerol only yielded a thermal shift change of 2.5 °C. Since this addition was of such a minimal thermal shift to the stability of the membrane protein, we decided to continue analysing other thermostability associations of the target protein, without the addition of glycerol.

pH Stability (GFP-TS)

Analysing the stability thresholds of the SLC10A7 *C. necator* transporter at various pH levels was also investigated. Our initial hypothesis for the alternating access mechanism of the SLC10A7 transporter, was the use of coupled proton ions (H^+), due to the lack of conserved aa residues associated with the Na^+ ion sites seen previously in $ASBT_{NM}$. We believed that by altering the pH in solutions, the concentration of H^+ ions would change, thus inflicting important effects and alterations on the stability of the transporter. To test this hypothesis, GFP-TS analysis with a range of pH buffering agents were analysed, including 20 mM concentrations of citrate pH 5.0, MES pH 6.0, Hepes pH 7.0, Tris pH 7.5, Tris pH 8.0, Bicine pH 9.0, and CHES pH 10.0, in buffer conditions entailing 150 mM NaCl, and 1 % DDM detergent. Again, thermostability traces, and analysed Tm measurements, are outlined in figure 4.7 below.

A



B

Buffer pH Condition	Melting Temperature, Tm (°C)
20 mM Citrate pH 5.0 + 150 mM NaCl	0 - N/A
20 mM MES pH 6.0 + 150 mM NaCl	0 - N/A
20 mM Hepes pH 7.0 + 150 mM NaCl	34.82 ± 1.9
20 mM Tris pH 7.5 + 150 mM NaCl	37.42 ± 0.66
20 mM Tris pH 8.0 + 150 mM NaCl	37.08 ± 0.65
20 mM Bicine pH 9.0 + 150 mM NaCl	39.37 ± 0.40
20 mM CHES pH 10.0 + 150 mM NaCl	49.02 ± 1.48

Figure 4.7: SLC10A7 *C. necator* pH stability GFP-TS optimisation studies. (A) Thermostability traces in citrate pH 5.0, MES pH 6.0, Hepes pH 7.0, Tris pH 7.5, Tris pH 8.0, Bicine pH 9.0 and CHES pH 10.0, at 20 mM concentrations, with 150 mM NaCl, and 1 % DDM, were investigated. Calculated Tm values (B) are also analysed in a table.
Data were analysed and processed in GraphPad Prism.

The general trend in stability of this SLC10A7 *C. necator* transporter, solubilised at varying pH'ed buffers (as seen above in figure 4.7), showed the membrane protein to be more stable at higher (more alkaline) pH's in comparison to lower (more acidic) pH.

This is certainly the case for the buffers pH'ed from 7.5-10 , where a gradual increase in the stability and Tm of the transporter is observed. Most significantly, solubilisation and stability of SLC10A7 in CHES buffer at pH 10.0, yielded the greatest increase at a Tm of 49.0 ± 1.5°C, which, is a thermal shift of 11.6 °C when

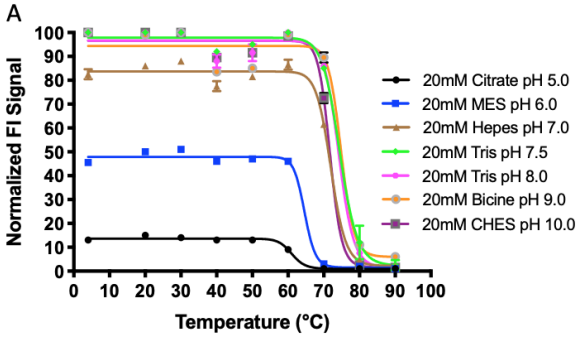
compared to Tris buffer at pH 7.5 (Tm of 37.4 ± 0.7℃).

In comparison, lower (more acidic) pH'ed buffers yielded significant instability in the membrane protein. Indeed citrate pH 5.0 and MES pH 6.0 buffers, both denatured the solubilised protein, by up to 80-100 %, at the initial solubilisation temperature of 4 °C. Due to such detrimental protein instability, melting temperatures of 0 °C were attributed.

Solubilisation of the transporter in Hepes pH 7.0 buffer depicted slightly better solubilisation and stability compared to citrate and MES, although roughly 30 % was still degraded at 4 °C. A Tm value of 34.8 ± 1.9 °C was established, presenting a substantially greater level of instability when compared to pH buffers at 7.5-10

In order to ensure that these generated Tm values for the SLC10A7 transporter are genuine and representative of the membrane protein (and not a GFP-induced stabilisation effect), pH GFP-TS investigations were conducted with pure standardised GFP, which is analysed and plotted below in figure 4.8.

A



B

Buffer pH Condition	Melting Temperature, Tm (°C)
20 mM Citrate pH 5.0 + 150 mM NaCl	0 - N/A
20 mM MES pH 6.0 + 150 mM NaCl	0 - N/A
20 mM Hepes pH 7.0 + 150 mM NaCl	72.01 ± 1.52
20 mM Tris pH 7.5 + 150 mM NaCl	74.45 ± 1.51
20 mM Tris pH 8.0 + 150 mM NaCl	74.20 ± 1.77
20 mM Bicine pH 9.0 + 150 mM NaCl	74.86 ± 1.11
20 mM CHES pH 10.0 + 150 mM NaCl	71.74 ± 1.74

Figure 4.8: Pure GFP pH stability GFP-TS optimisation studies. (A) Thermostability traces in citrate pH 5.0, MES pH 6.0, Hepes pH 7.0, Tris pH 7.5, Tris pH 8.0, Bicine pH 9.0 and CHES pH 10.0, at 20 mM concentrations, with 150 mM NaCl, and 1 % DDM, were investigated. Calculated Tm values (B) are also analysed in a table.
Data were analysed and processed in GraphPad Prism.

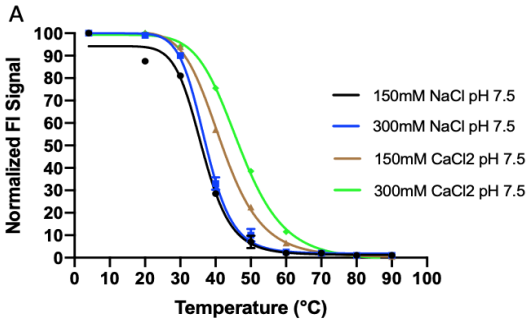
As can be seen for the pure GFP pH investigation in figure 4.8 depicted above, a very similar thermostability profile is seen for the varying pH buffers investigated for SLC10A7 *C. necator*. Again, at lower pH's, including citrate pH 5.0 and MES pH 6.0, very strong destabilisation of pure GFP can be observed. In comparison, at higher pH (including Hepes pH 7.0, Tris pH 7.5, Tris pH 8.0, Bicine pH 9.0, and CHES pH 10.0), much greater protein stability is observed, with calculated Tm measurements ranging from 71.7-74.9 °C. These high melting temperatures are as expected for pure GFP, and similar to its reported Tm of 78 °C (Ward et al., 1982; Tsien, 1998).

Taken together, the SLC10A7 *C. necator* transporter's instability, witnessed in figure 4.7 at lower pH'ed buffers, appears to be an effect of the instability of pure GFP and is not necessarily due to the instability of the SLC10A7 membrane protein. Since this technique measures the fluorescence of GFP to deduce Tm

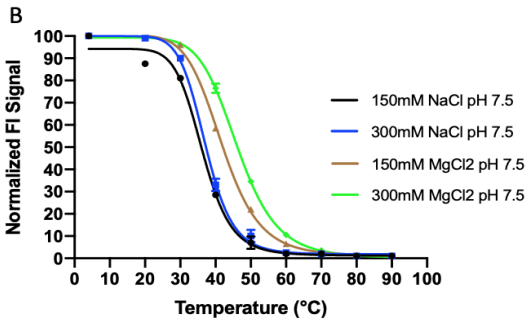
measurements of membrane proteins, the denaturation of GFP at low pH, deems the GFP-TS assay to be unsuitable for analysis of membrane protein stability at low pH. As such we were unable to prove or disprove our initial hypothesis of the transporter requiring H^+ for solute translocation. However, since pure GFP remains very stable at high pHs, we can confidently conclude that the SLC10A7 *C. necator* membrane protein transporter exhibits increased stability at higher (more alkaline) pH's.

Salt Stability (GFP-TS)

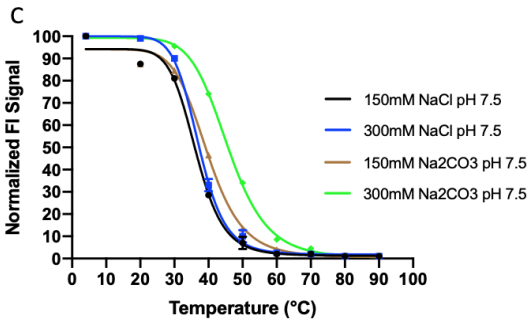
An in-depth analysis was conducted for a plethora of salts, by GFP-TS assays, in detergent buffered solution (20 mM Tris pH 7.5, 1% DDM), at 150 mM and 300 mM concentrations, with respective T_m values calculated. Although greater pH was seen to stabilise the target membrane protein better (see above), a salt pH 7.5 was selected, due to the adherence to standardised purification protocols for the protein. Further salt and buffer pH combination optimisations should be investigated in the future, to establish the optimal buffer-salt pH ratios which yield the greatest protein stability. Figure 4.9 below, highlights just a few of the many different salts studied.



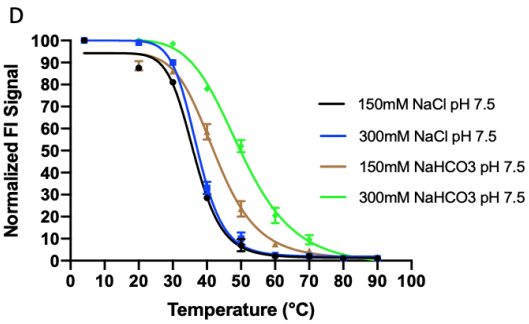
Calcium Chloride pH 7.5



Magnesium Chloride pH 7.5



Sodium Carbonate pH 7.5



Sodium Bicarbonate pH 7.5

E

Buffer Salt Condition	Melting Temperature, Tm (°C)
20 mM Tris pH 7.5 + 150 mM NaCl	36.37 ± 1.09
20 mM Tris pH 7.5 + 300 mM NaCl	37.15 ± 0.52
20 mM Tris pH 7.5 + 150 mM $CaCl_2$	41.74 ± 0.37
20 mM Tris pH 7.5 + 300 mM $CaCl_2$	46.89 ± 0.51
20 mM Tris pH 7.5 + 150 mM $MgCl_2$	41.93 ± 0.45
20 mM Tris pH 7.5 + 300 mM $MgCl_2$	46.27 ± 0.31
20 mM Tris pH 7.5 + 150 mM Na_2CO_3	39.72 ± 1.17
20 mM Tris pH 7.5 + 300 mM Na_2CO_3	45.81 ± 0.38
20 mM Tris pH 7.5 + 150 mM $NaHCO_3$	42.61 ± 1.37
20 mM Tris pH 7.5 + 300 mM $NaHCO_3$	50.09 ± 1.10

Salt Melting Temperatures

Figure 4.9: SLC10A7 *C. necator* salt stability GFP-TS optimisation studies. Thermostability in sodium chloride (NaCl, (A-D)), calcium chloride ($CaCl_2$, (A)), magnesium chloride ($MgCl_2$, (B)), sodium carbonate (Na_2CO_3, (C)) and sodium bicarbonate ($NaHCO_3$, (D)) at 150 mM and 300 mM concentrations were investigated. Bar chart (E) represents the Tm values obtained from these investigations.
Data were analysed and processed in GraphPad Prism.

As can be identified from figure 4.9.A-E, varying salt components and their concentrations in buffer, have significant effects on Tm values. The general trend for all salts analysed, was that increasing salt concentrations from 150 mM to 300 mM, positively correlated with an increase in Tm protein thermostability. Typically, the greater the proportion of charged surface residues in membrane proteins, the better its solubility, stability and prevention against aggregation (Lindman et al., 2006). Therefore, the increasing salt concentrations witnessed in figure 4.9.A-E, most likely result from a reduction in the repulsive forces of the surface charged residues, thereby resulting in a membrane protein with greater surface charge and increased stability (Lindman et al., 2006).

Interestingly, and unexpectedly, sodium chloride (NaCl) presented with the lowest generated Tm values at 36.4 ± 1.1 °C and 37.2 ± 0.5℃, at 150 mM and 300 mM concentrations, respectively. These values are both respectively low, suggesting a greater degree of membrane protein instability, and thus reduced likelihood of successful purification and crystallisation.

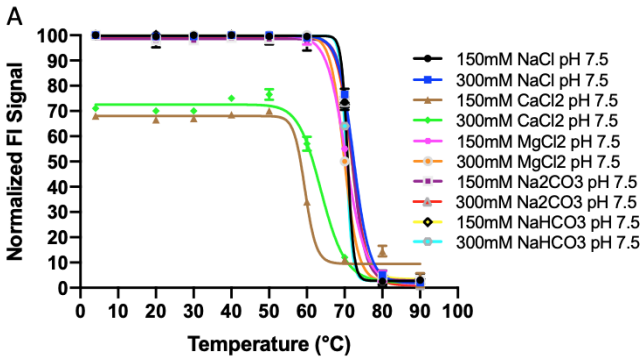
In comparison, calcium chloride ($CaCl_2$), magnesium chloride ($MgCl_2$), sodium carbonate (Na_2CO_3) and sodium bicarbonate ($NaHCO_3$), all yielded significantly greater protein stabilisation, at pH 7.5. For both $CaCl_2$ and $MgCl_2$, at 300 mM concentrations, a thermal shift of Tm 9.7 °C and 9.1 °C were achieved, respectively, in reference to 300 mM NaCl. This is extremely encouraging, and depicts an almost 10 °C greater thermal shift, strongly highlighting a more stabilised state of the SLC10A7 carrier protein.

Similarly, both Na_2CO_3 and $NaHCO_3$ also significantly increased Tm values, when compared to NaCl. Here, at 300 mM concentrations, thermal shifts of Tm 8.7 °C and 12.9 °C were observed, respectively, in comparison to NaCl. Sodium bicarbonate especially presented with the greatest SLC10A7 membrane protein stability, significantly increasing its chances of yielding efficient protein crystals. In light of this, sodium bicarbonate was selected for downstream protein purification and crystallisation trials.

Other SLC10A7 bacterial homologue constructs were also trialled in thermostable investigations with this bicarbonate salt. These all depicted very similar Tm values to the ones reported here, however, interestingly the *E. coli* Q320 C-terminal truncated protein construct generated even higher melting temperatures, across all analysed bacterial constructs. This finding seems to suggest the C-terminal domain of the full-length *E.coli* protein to be rather flexible and an unstable domain/region of the transporter. As such, its subsequent truncation and solubilisation in bicarbonate buffer led to the generation of the most stable SLC10A7 protein. Downstream purification and crystallisation studies for this truncated construct, were characterised further (see below).

To ensure these results were accurately depicting SLC10A7 stabilisation and not merely an influenced GFP effect, pure GFP Tm thermostability (positive control) was assessed in the same buffers mentioned above. Figure 4.10 below, outlines these findings.

A



B

Buffer Salt Condition	Melting Temperature, Tm (°C)
20 mM Tris pH 7.5 + 150 mM NaCl	70.80 ± 0.63
20 mM Tris pH 7.5 + 300 mM NaCl	72.42 ± 0.4
20 mM Tris pH 7.5 + 150 mM $CaCl_2$	59.52 ± 1.54
20 mM Tris pH 7.5 + 300 mM $CaCl_2$	63.97 ± 1.11
20 mM Tris pH 7.5 + 150 mM $MgCl_2$	70.47 ± 0.21
20 mM Tris pH 7.5 + 300 mM $MgCl_2$	69.92 ± 0.05
20 mM Tris pH 7.5 + 150 mM Na_2CO_3	71.98 ± 0.93
20 mM Tris pH 7.5 + 300 mM Na_2CO_3	72.02 ± 0.91
20 mM Tris pH 7.5 + 150 mM $NaHCO_3$	71.76 ± 0.46
20 mM Tris pH 7.5 + 300 mM $NaHCO_3$	70.43 ± 0.82

Figure 4.10: Pure GFP salt stability GFP-TS optimisation studies. (A) Thermostability traces in sodium chloride (NaCl), calcium chloride ($CaCl_2$), magnesium chloride ($MgCl_2$), sodium carbonate (Na_2CO_3) and sodium bicarbonate ($NaHCO_3$), at 150 mM and 300 mM concentrations, with 20 mM Tris pH 7.5 and 1 % DDM, were investigated. Calculated Tm values (B) are also analysed in a table.
Data was analysed and successfully processed in GraphPad Prism.

As can be seen in figure 4.10 above, with the exception of $CaCl_2$, pure GFP, in the presence of the above mentioned salts, yielded significantly higher Tm values of 71-73 °C. These values are to be expected from GFP, with it being a very thermostable protein, correlating well to its expected and reported melting tem-

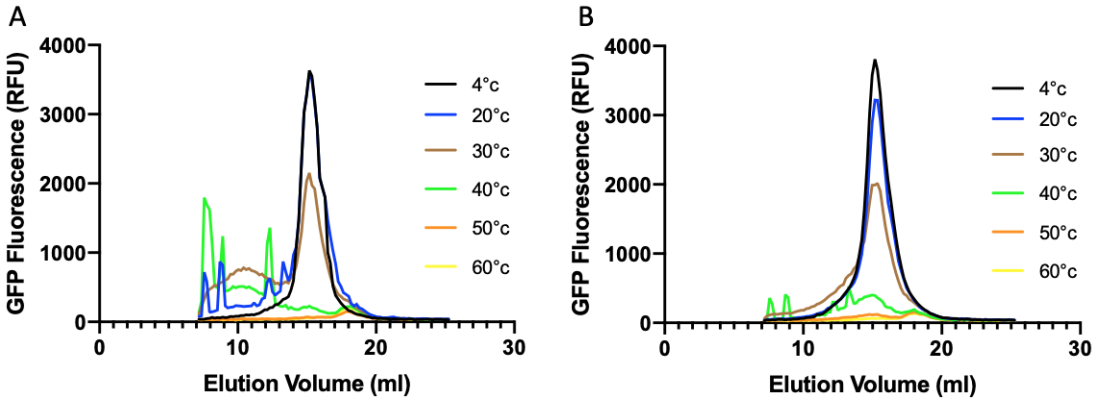
perature Tm 78 °C (Ward et al., 1982; Tsien, 1998).

Interestingly, solubilisation of pure GFP in detergent buffer with 150 mM and 300 mM $CaCl_2$ pH 7.5 concentrations, yielded drastically reduced melting temperatures (Tm of 60 °C and 64 °C, respectively), albeit still significantly greater than SLC10A7 Tm values. This salt thereby hinders the stability of pure GFP, and its involvement in future GFP-associated stability assays should be strongly considered.

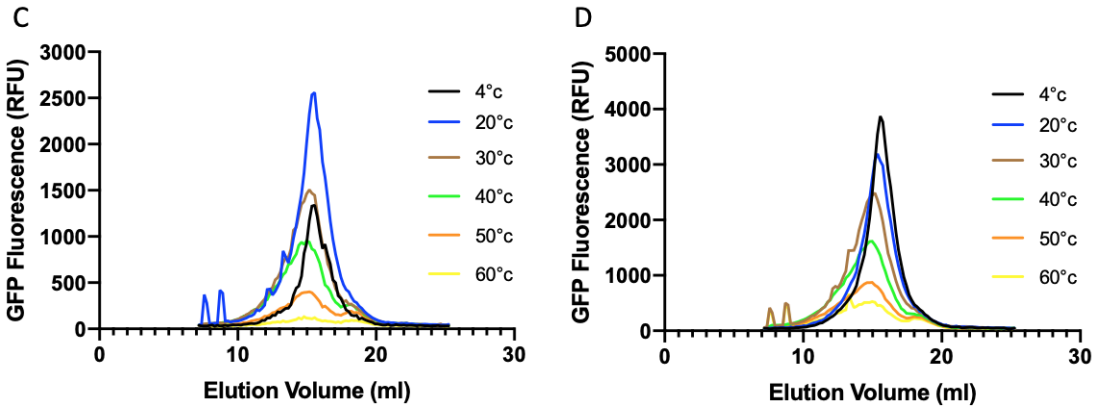
Taken together, based on the findings of stability from pure GFP, Tm values calculated for SLC10A7 *C. necator* (with the exception of $CaCl_2$, as seen above, are reliable and accurate, and depict true melting temperatures. Sodium bicarbonate provides an extraordinary finding, yielding the greatest thermal shift and stability in the membrane protein. As such, downstream purification and crystallisation trials with this salt were investigated further.

Salt Stability (FSEC-TS)

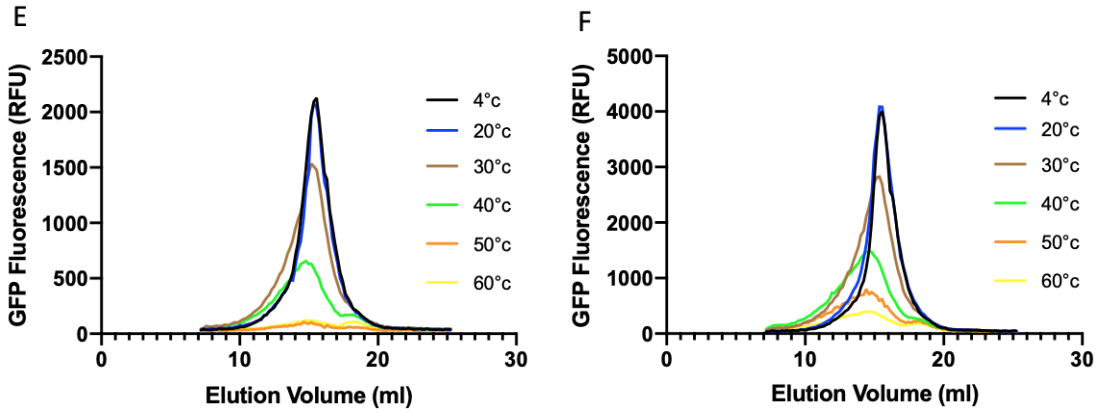
In order to further characterise the SLC10A7 *C. necator* thermostability, FSEC-TS investigations were conducted for the same salts mentioned above, using the same GFP-TS experimental setup. The results are displayed below in figure 4.11.



150mM Sodium Chloride pH 7.5 300mM Sodium Chloride pH 7.5

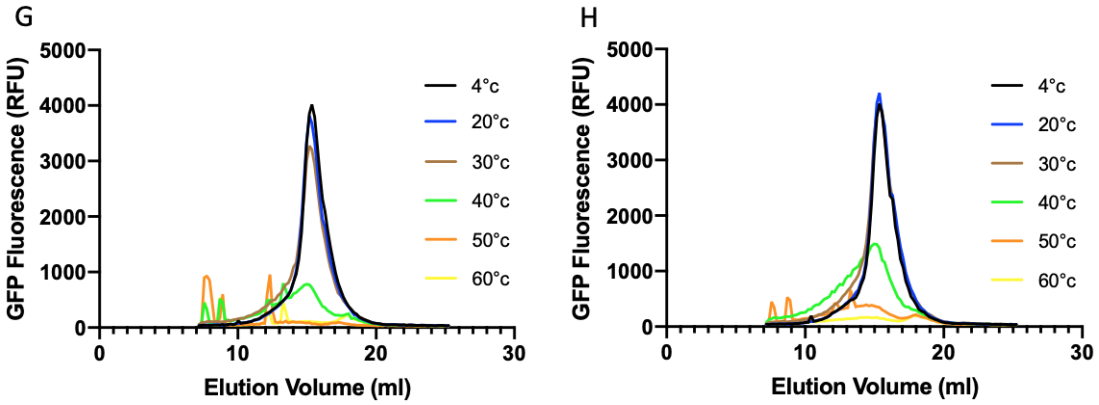


150mM Calcium Chloride pH 7.5 300mM Calcium Chloride pH 7.5



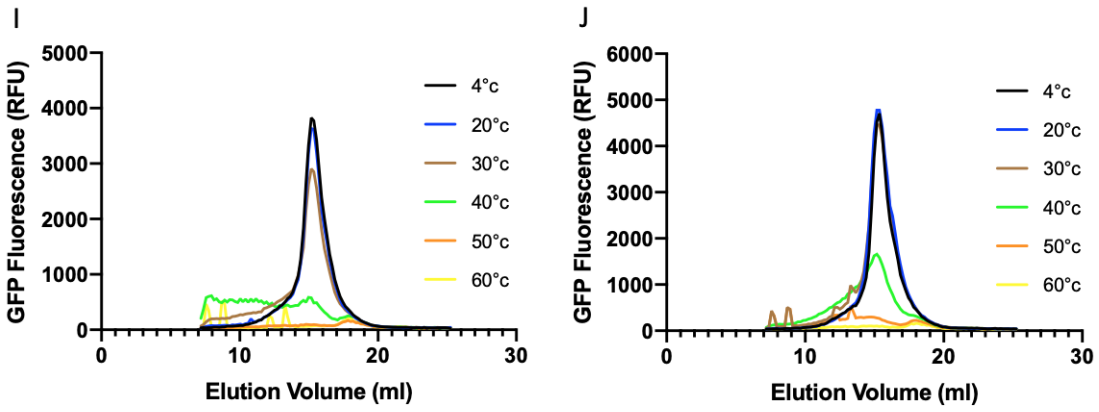
150mM Magnesium Chloride pH 7.5

300mM Magnesium Chloride pH 7.5



150mM Sodium Carbonate pH 7.5

300mM Sodium Carbonate pH 7.5



150mM Sodium Bicarbonate pH 7.5

300mM Sodium Bicarbonate pH 7.5

Figure 4.11: SLC10A7 *C. necator* salt stability FSEC-TS optimisation studies. Thermostability in sodium chloride(NaCl, (A-B)), calcium chloride ($CaCl_2$, (C-D)), magnesium chloride ($MgCl_2$, (E-F)), sodium carbonate (Na_2CO_3, (G-H)) and sodium bicarbonate ($NaHCO_3$, (I-J)) at 150 mM and 300 mM concentrations were investigated.
Data was analysed and successfully processed in GraphPad Prism.

In general, the SLC10A7 *C. necator* salt optimisation FSEC-TS profiles, depicted in figure 4.11.A-J, yielded sharp monomeric and stable solubilised fluorescent peaks. A similarity in trends to GFP-TS studies, regarding the increase in membrane protein stability with an increase in salt concentrations from 150-

300 mM was also observed. This is particularly emphasised in $CaCl_2$, $MgCl_2$ and $NaHCO_3$, whereby increased differences in fluorescence values (and hence sharpness of profiles) between these salt concentrations were 1,316, 1,958 and 952 GFP RFU, respectively.

Comparable to GFP-TS studies, NaCl demonstrated the weakest solubilisation and fluorescence profile, albeit producing sharp elution peaks. At the 150 mM concentration, significant aggregation/void peaks are noticeable, strongly implying a greater tendency for the SLC10A7 protein to form denatured states. However, increase in concentration to 300 mM removes the large majority of this aggregation, albeit a minimal increase in total protein solubilisation and fluorescence.

Regarding $CaCl_2$, although a minimal aggregation peak is depicted, the elution peaks are considerably broader in appearance, when compared to all other salts investigated. Although an increase in the total protein solubilisation and hence maximal fluorescence peak is displayed in the different salt concentrations (as explained above), it is important to note that the maximal fluorescent peaks are eluted at slightly earlier volumes, upon increase in thermal temperatures. Similar plots are also seen in $MgCl_2$ investigations, albeit slightly increased peak sharpness.

The greatest symmetrical gausian profile peaks, outlining the best protein solubility and monodisperity, are displayed in both Na_2CO_3 and $NaHCO_3$ salt studies. In particular, $NaHCO_3$ presented with the greatest peak sharpness, with minimal aggregation, and yielded the greatest solubilisation and fluorescence intensity at both 150 mM and 300 mM concentrations. Based on these findings, $NaHCO_3$ was distinguished to be more than suitable for incorporation in downstream purification and crystallisation trials.

As observed above in figure 4.11.A-J, the results from FSEC-TS correlate extremely well with the generated outcomes of GFP-TS studies, strongly highlighting the accurate reliability and powerful generations of both techniques.

Substrate Stability (GFP-TS)

It is thought the binding of associated substrates to a transporter will alter its structural conformation/architechture to a ligand-bound state, thereby changing its flexibility characteristics and enhancing its overall stability and Tm (Errasti-Murugarren et al., 2021). As such, a wide array of potential SLC10A7 substrates were investigated via GFP-TS and presented in tables 4.2, 4.3, 4.4, and 4.5 below.

Table 4.2: SLC10A7 substrate stability investigation

Substrates	**Melting Temperature, Tm (°C)**
20 mM Tris pH 7.5 + 150 mM NaCl (Control)	37.9 ± 0.4
1 mM Ammonium Molybdate	27.6 ± 1.9
1 mM Anthranilic Acid	36.8 ± 0.6
1 mM Betaine Monohydrate	37.6 ± 0.9
1 mM β-Glycerophosphate	37.4 ± 1.0
1 mM EDTA	37.7 ± 0.6
1 mM Ferric Chloride	38.1 ± 0.9
1 mM Glucuronic Acid	36.8 ± 0.4
1 mM Glycine	37.1 ± 1.9
1 mM Guanidine	35.9 ± 1.5
1 mM Methyl-β-Cyclodextrin	37.4 ± 0.8
1 mM Nicotinic Acid	36.0 ± 1.5
1 mM N-Methyl-D-Glucamine	38.5 ± 0.8
1 mM Oxalic Acid	36.9 ± 0.6
1 mM Quercetin Dihydrate	36.7 ± 0.5
1 mM Sodium Citrate	33.5 ± 1.1
1 mM Sodium Fluoride	35.6 ± 1.3
1 mM Sodium L-Ascorbate	37.0 ± 0.3
1 mM Sodium Pantoate	37.7 ± 0.7
1 mM Sodium Pyrophosphate	33.4 ± 0.9
1 mM Sodium Pyruvate	36.6 ± 1.2
1 mM Taurine	36.8 ± 0.8

As can be seen in table 4.2, although a wide array of potential solutes were investigated, no significant increase in Tm, in comparison to control studies, were reported. Indeed, only 1 mM ferric chloride (thermal shift 0.7 °C) and 1 mM N-methyl-D-glucamine (thermal shift 1.0 °C) seemed to impose slightly greater thermostability measurements, however these values are so minimal (when compared to control studies), that these were deemed insignificant.

Since the human SLC10A1 (NTCP) and SLC10A2 (ASBT) members are categorically associated in the transport of bile acids and regulation of the EHC, specific bile acid GFP-TS investigation studies in the SLC10A7 *C. necator* bacterial protein carrier were also conducted. Since bacteria have obviously no association or requirement for bile acids, we did not expect the Tm of the transporter to change. These thermostable results are displayed in table 4.3 below.

Table 4.3: SLC10A7 bile acid substrate stability investigation

Bile Acid Substrates	**Melting Temperature, Tm (°C)**
20 mM Tris pH 7.5 + 150 mM NaCl (Control)	37.9 ± 0.4
1 mM Sodium Cholate	37.1 ± 0.5
1 mM Choline Chloride	36.9 ± 1.5
1 mM Sodium Taurocholate	38.0 ± 0.7
1 mM Sodium Deoxycholate	37.3 ± 1.0
1 mM Sodium Chenodeoxycholate	37.2 ± 1.5

As expected, as can be seen above in table 4.3, no significant increase in Tm was observed, thereby strongly suggesting a lack of involvement of this bacterial transporter (and most likely its human homologue) in the regulation of bile acid homeostasis. Although 1 mM Sodium Taurocholate showed an ever so slight increase in protein stability (thermal shift 0.5 °C), this is extremely negligible and again insignificant.

Having been well established as a negative regulator of calcium homeostasis in humans, rats, mice, zebrafish and yeast; calcium-linked stability substrates, including vitamin D derivatives and calcium signalling agents, were additionally investigated for the *C. necator* bacterial homologue, with Tm values presented in table 4.4 below.

Table 4.4: SLC10A7 calcium-associated substrate stability investigation

Calcium-Associated Substrates	**Melting Temperature, Tm (°C)**
20 mM Tris pH 7.5 + 150 mM NaCl (Control)	37.9 ± 0.4
0.5 mM Calcitriol	36.5 ± 0.6
1 mM Cholecalciferol	37.0 ± 0.7
0.5 mM Ercalcitriol	37.9 ± 0.8
1 mM Ergocalciferol	37.5 ± 0.6
0.5 mM Inositol Trisphosphate (IP3)	35.5 ± 0.4
0.5 mM Inositol Hexaphosphate (IP6)	36.5 ± 0.7

A 0.5-1 mM substrate concentration was deemed sufficient to invoke a thermostable Tm change in the target membrane protein. As can be seen in table 4.4 above, a lack of any thermal shift was presented, for all investigated calcium-linked substrates.

Finally, specific sugars were also investigated, and their stability in the SLC10A7 *C. necator* bacterial homologue are displayed in table 4.5 below.

Table 4.5: SLC10A7 sugar substrate stability investigation

Sugar Substrates	**Melting Temperature, Tm (°C)**
20 mM Tris pH 7.5 + 150 mM NaCl (Control)	37.9 ± 0.4
1 mM Alpha-Lactose	38.0 ± 1.0
1 mM D-Fructose	35.7 ± 1.2
1 mM D-Galactose	36.1 ± 0.4
1 mM D-Glucose	36.1 ± 0.8
1 mM Maltose	36.8 ± 0.9
1 mM D-Mannitol	36.9 ± 1.0
1 mM Myo-Inositol	36.5 ± 0.8
1 mM D-Sorbitol	36.0 ± 1.1
1 mM Sucrose	36.1 ± 1.1
1 mM D-Trehalose	36.8 ± 1.1
1 mM Xylitol	36.5 ± 0.5

Here again, a lack of a significant thermal shift was observed. Although 1 mM Alpha-Lactose caused a very slight increase in Tm (thermal shift 0.8 °C), this was deemed insignificant and not indicative of an associated SLC10A7 solute.

Taken together, although many potential SLC10A7 substrates were trialled in GFP-TS investigations, none were able to increase the melting temperature of the SLC10A7 *C. necator* membrane protein significantly. As such, the drive for the discovery of this transporter's associated solute continues.

Chapter 5

Results: Membrane Protein Purification, Crystallisation and Structural Determination

5.1 Membrane Protein Purification

The presence of a 3C cleavable tag, with a GFP fusion and polyhisitidine C-terminal tail, as part of the GFP pWaldo fusion vector (Drew et al., 2006; Minhas et al., 2018), provides a fundamental ability to optimally characterise and purify membrane proteins successfully, whilst taking full beneficial advantage of the stable fluorescence of GFP (as seen above) (Kawate and Gouaux, 2006).

In light of this, purification procedures for both the $ASBT_{Lb}$ and SLC10A7 bacterial homologues, followed standardised protocols involving Ni-NTA, IMAC, and SEC, with a multitude of buffers, at set ranges of imidazole and salt concentrations. Due to the findings from GFP-TS studies, SLC10A7 bacterial homologues were subjected to buffers entailing sodium bicarbonate ($NaHCO_3$) salt instead of the usual sodium chloride (NaCl). Details of their purification, overall purity and stability are displayed in greater detail below.

5.1.1 ASBT$_{Lb}$ Purification

In-gel fluorescence and SDS-PAGE gel analysis for the purification of the ASBT$_{Lb}$ bacterial homologue is displayed below in figure 5.1, providing good assessments of the level of protein bound and purified by Ni-NTA resin, as well as the successful removal of the 3C cleavable tag fused to GFP.

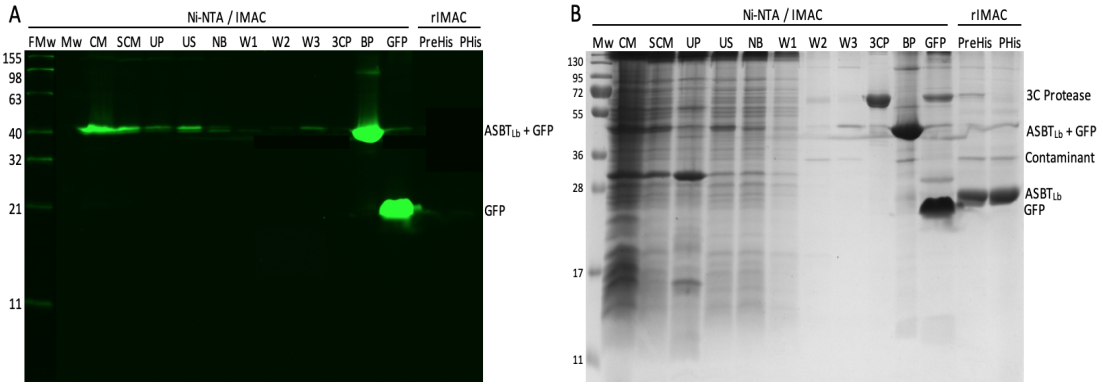


Figure 5.1: ASBT$_{Lb}$ protein Ni-NTA, and reverse IMAC (rIMAC) purification. In-gel fluorescence (A) and SDS-PAGE (B) gel images.
FMw: Fluorescent Molecular Weight ladder, Mw: Molecular Weight ladder, CM: Cell Membrane, SCM: Solubilised Cell Membrane, UP: Ultracentrifuge Pellet, US: Ultracentrifuge Supernatant, NB: Nickel Binding Flowthrough, W1: Wash1 (250 ml), W2: Wash2 (500 ml), W3: Wash3 (350 ml), 3CP: 3C Protease, BP: Pre-3C Cleaved Bound Protein, GFP: Post 3C Cleaved GFP Elution (50 ml), PreHis: Pooled Pre-His trap protein, PHis: Post His-trap protein.
Gel fluorescence imaged at 1 s intensity.

As is presented in figure 5.1 above, very good levels of ASBT$_{Lb}$ purity are established, via the binding to Ni-NTA resin, and subsequent cleavage by 3C protease. Purity levels increased after rIMAC (to remove excess 3C protease contaminants), and yielded sufficient quantities of purified protein for SEC analysis and crystallisation studies. Interestingly, an additional contaminant was also expressed at roughly 34kDa, but was not seen to affect the stability and characterisation of the protein.

Figure 5.2 below, outlines the ASBT$_{Lb}$ SEC purification UV chromatogram profiles in DDM, DDM + 1mM DCA, and NM, with their purity levels assessed in an SDS-PAGE gel.

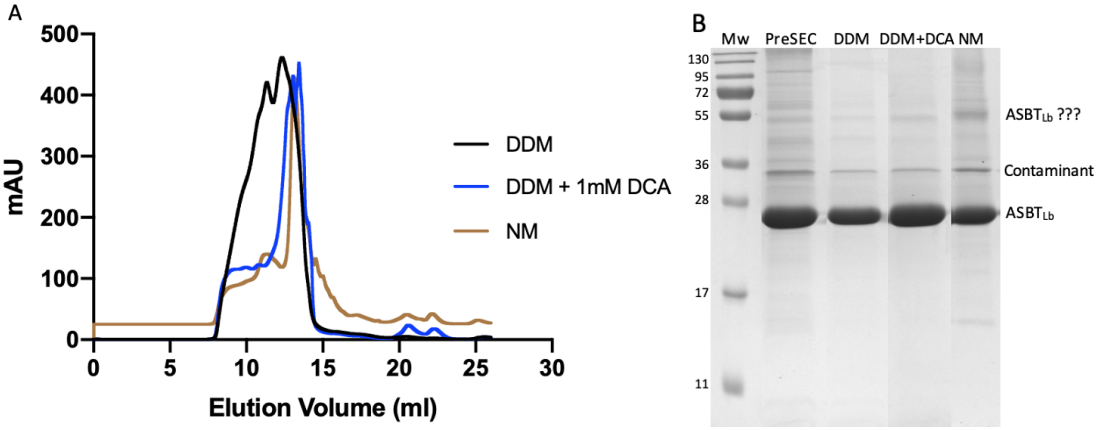


Figure 5.2: $ASBT_{Lb}$ protein size-exclusion chromatography (SEC) and purity display on SDS-PAGE. (A) SEC UV traces (detected at absorbance 280 nm) were analysed in different conditions including 0.03 % DDM, 0.03 % DDM + 1 mM DCA, and 0.84 % NM, with (B) their respective levels of purity assessed in SDS-PAGE.
Samples were run on a Superdex 200 Increase 10/300 GL column.

As can be seen in figure 5.2 above, the SEC purified $ASBT_{Lb}$ protein yields stable, monomeric UV chromatograms for all analysed conditions. Analysis by SDS-PAGE highlights significantly increased purity levels from the SEC runs, although the aforementioned 34kDa contaminant is still present. A band corresponding to a potential dimeric state of the $ASBT_{Lb}$ protein is also seen at roughly 53kDa.

Taken together, the $ASBT_{Lb}$ membrane protein is seen to express and purify extremely well, yielding more than sufficient quantities post-SEC for attempts at protein crystallisation.

5.1.2 SLC10A7 *E. coli* Purification

Based on the very interesting sodium bicarbonate salt stability findings elucidated from GFP-TS and FSEC-TS, *E. coli* SLC10A7 purification studies were carried out in buffers containing 150-300 mM $NaHCO_3$, 20 mM Tris pH 7.5, and 1 % DDM. These successful purification attempts are displayed in both in-gel fluorescent and SDS-PAGE gels, below in figure 5.3.

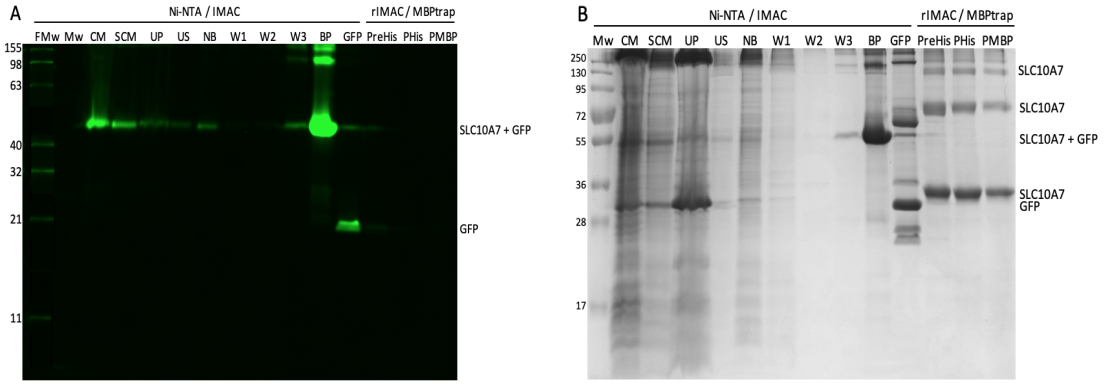


Figure 5.3: SLC10A7 *E. coli* protein Ni-NTA, reverse IMAC (rIMAC) and MBP-binding purification. In-gel fluorescence (A) and SDS-PAGE (B) gel images.
FMw: Fluorescent Molecular Weight ladder, Mw: Molecular Weight ladder, CM: Cell Membrane, SCM: Solubilised Cell Membrane, UP: Ultracentrifuge Pellet, US: Ultracentrifuge Supernatant, NB: Nickel Binding Flowthrough, W1: Wash1 (250 ml), W2: Wash2 (250 ml), W3: Wash3 (250 ml), BP: Pre-3C Cleaved Bound Protein, GFP: Post 3C Cleaved GFP Elution (50 ml), PreHis: Pooled Pre-His trap protein, PHis: Post His-trap protein, PMBP: Post MBP-trap protein.
Gel fluorescence imaged at 1 s intensity.

As is shown in figure 5.3 above, the SLC10A7 *E. coli* membrane protein demonstrates good levels of purity. The binding to Ni-NTA resin, and subsequent cleavage with 3C protease was successfully demonstrated, yielding satisfactory amounts of purified protein for downstream size-exclusion chromatography (SEC) investigations, as is shown below in figure 5.4, assessing SEC analysis in DDM, NM and LDAO detergent buffers.

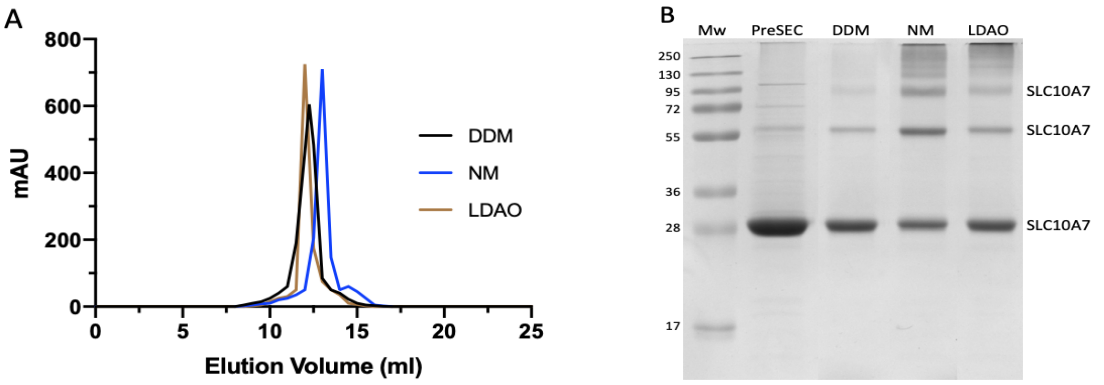


Figure 5.4: SLC10A7 *E. coli* protein size-exclusion chromatography (SEC) and purity display on SDS-PAGE. (A) SEC UV traces were analysed in DDM, NM and LDAO detergents, and (B) their respective levels of purity assessed in SDS-PAGE.
Samples were run on a Superdex 200 Increase 10/300 GL column.

As is shown in 5.4 above, the purified SLC10A7 protein runs on a SEC column as stable, monomeric UV chromatograms, indicating extremely good levels of purity on an SDS-PAGE gel, in DDM, NM and LDAO.

5.1.3 SLC10A7 *E. coli* Q320 Truncated Purification

Finally, purification trials were repeated for the Q320 C-terminal truncated bacterial construct, with aim to increase the chances of forming well-diffracting protein crystals. Alternative purification strategies in the other Q322 and R316 truncated *E. coli* bacterial homologue constructs, failed to be cleaved from the Ni-NTA resin upon addition of 3C protease. Only the Q320 C-terminal truncated construct was able to be successfully purified, and as such is demonstrated in the reported findings below:

Again, purification studies were investigated in sodium bicarbonate salt buffers containing 150-300 mM $NaHCO_3$, 20 mM Tris pH 7.5 and 1 % DDM, with their depictions in in-gel fluorescent and SDS-PAGE gels shown below in figure 5.5.

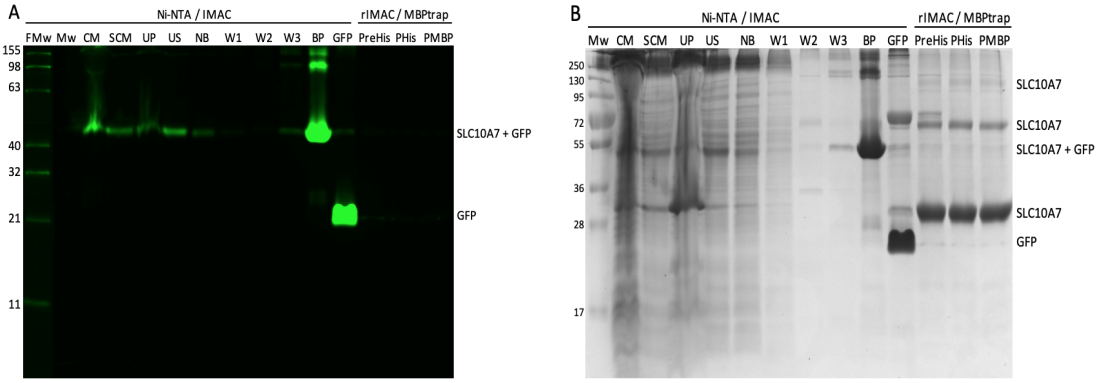


Figure 5.5: SLC10A7 *E. coli* Q320 truncated protein Ni-NTA, reverse IMAC (rIMAC) and MBP-binding purification. In-gel fluorescence (A) and SDS-PAGE (B) gel images.
FMw: Fluorescent Molecular Weight ladder, Mw: Molecular Weight ladder, CM: Cell Membrane, SCM: Solubilised Cell Membrane, UP: Ultracentrifuge Pellet, US: Ultracentrifuge Supernatant, NB: Nickel Binding Flowthrough, W1: Wash1 (250 ml), W2: Wash2 (250 ml), W3: Wash3 (250 ml), BP: Pre-3C Cleaved Bound Protein, GFP: Post 3C Cleaved GFP Elution (50 ml), PreHis: Pooled Pre-His trap protein, PHis: Post His-trap protein, PMBP: Post MBP-trap protein.
Gel fluorescence imaged at 1 s intensity.

As seen in figure 5.5, good levels of protein purity are once again readily achieved, again indicating successful binding to Ni-NTA resin and cleavage with 3C protease. SEC analysis in different salts (NaCl, $MgCl_2$, and $NaHCO_3$) were performed, to investigate the homogeneity of the samples, and cross referenced on an SDS-PAGE gel, as is depicted in figure 5.6.

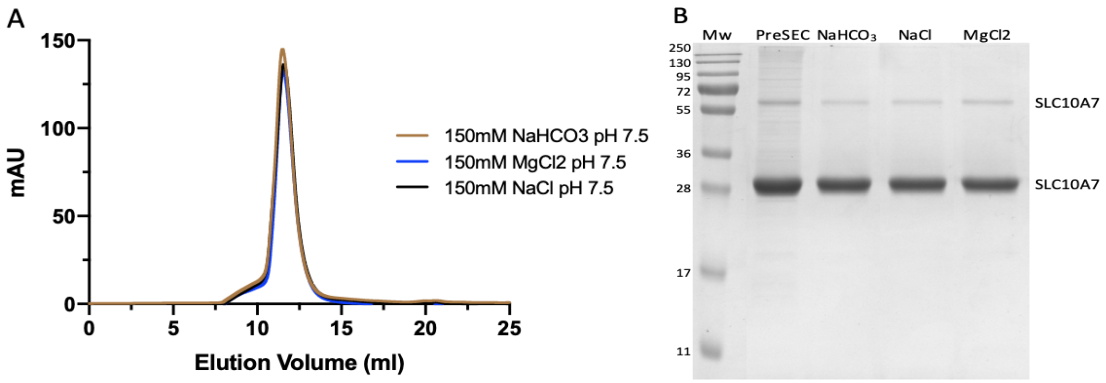


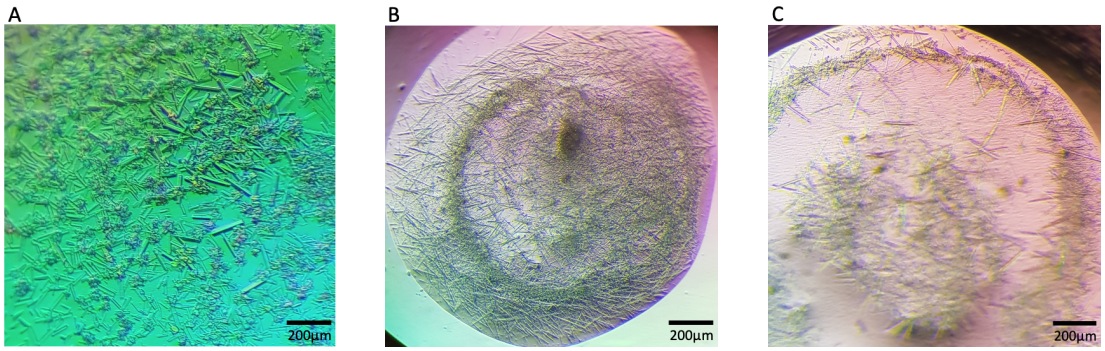
Figure 5.6: SLC10A7 *E. coli* Q320 truncated protein size-exclusion chromatography (SEC) and purity display on SDS-PAGE. (A) SEC UV traces were analysed in different salts including NaCl, $MgCl_2$, and $NaHCO_3$ at 150 mM concentrations, and (B) their respective levels of purity assessed in SDS-PAGE.
Samples were run on a Superdex 200 Increase 10/300 GL column.

As can be seen in figure 5.6, SEC profiles for the different salts analysed yielded very sharp, monomeric peaks, indicating a properly folded and stable protein state, with a lack of any early aggregated profile peak. The purity levels on SDS-PAGE were good, showing clear separation of monomeric and dimeric protein assemblies.

5.2 Membrane Protein Crystallisation

5.2.1 ASBT$_{Lb}$ Crystallisation

Once adequate levels of ASBT$_{Lb}$ purity, with and without the addition of 1 mM DCA, (via Ni-NTA IMAC, rIMAC and SEC) were achieved, crystallisation trials in both vapour diffusion (sitting drops) and lipidic cubic phase (LCP) were trialled. The most successful vapour diffusion (sitting drop) crystal hits are displayed below in figure 5.7, with their corresponding crystal conditions from broad screens attached.



15 mg/ml ASBT$_{Lb}$, 1 mM DCA, 10 mM Zinc acetate dihydrate, 100 mM MES pH 6.0, 40 % PEG 200.

15 mg/ml ASBT$_{Lb}$, 0.2 M Calcium acetate hydrate, 0.1 M Sodium acetate pH 5.0, 38 % PEG 400.

15 mg/ml ASBT$_{Lb}$, 0.08 M Calcium acetate hydrate, 0.05 M Sodium chloride, 0.1 M MOPS pH 7.0, 28 % PEG 400.

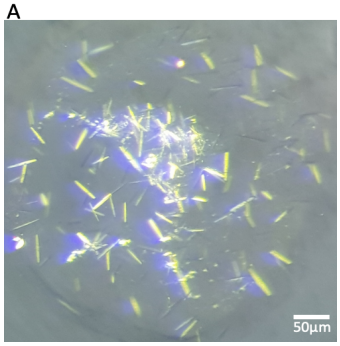
Figure 5.7: ASBT$_{Lb}$ post-SEC DDM sitting drop vapour diffusion crystallisation studies. A multitude of broad crystallisation screens were investigated with (A) MemMeso A8, (B) MemGold2 E6 and (C) MemTrans F5 conditions yielding crystals.
All crystal images were taken and processed using a Samsung Galaxy S8 mobile phone.

As can be identified above, well-forming, stable ASBT$_{Lb}$ protein crystals were obtained from sitting drop vapour diffusion, in variable crystal conditions, in the presence and absence of 1 mM DCA. These crystals however diffracted poorly, prompting further investigations with LCP instead.

Growing membrane protein crystals in the meso/LCP has a tendency to generate crystal structures at higher resolution (due to their lower solvent composition), when compared to vapour-diffusion crystallisation (Nji et al., 2018). As such, many challenging membrane proteins, including secondary active transporters, have previously been successfully crystallised via the use of this technique (Parker and Newstead, 2016; Li and Caffrey, 2020). Since our aim was to produce well-diffracting high resolution crystal structures of bacterial homologues of the SLC10 family, LCP crystallisation studies were trialled for the ASBT$_{Lb}$ transporter in the presence and absence of 1 mM DCA, where DDM purified protein was concentrated to roughly 30 mg/ml and mixed with monoolein lipid at a 2:3 protein-to-lipid ratio,

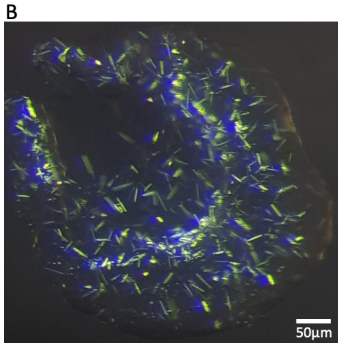
as previously demonstrated by (Caffrey and Cherezov, 2009). The most successful crystal "hits" are displayed below in figure 5.8.

A



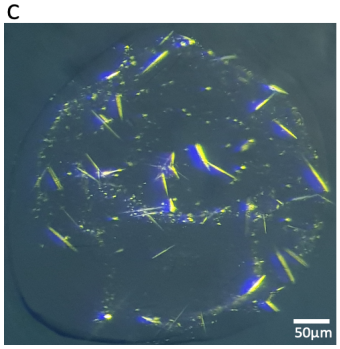
31 mg/ml ASBT$_{Lb}$, 2.5 % Glycerol, 1 mM DCA, 0.1 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 36 % PEG 400.

B



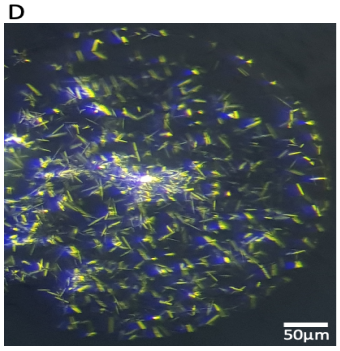
30 mg/ml ASBT$_{Lb}$, 1 mM DCA, 0.1 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 34 % PEG 300.

C



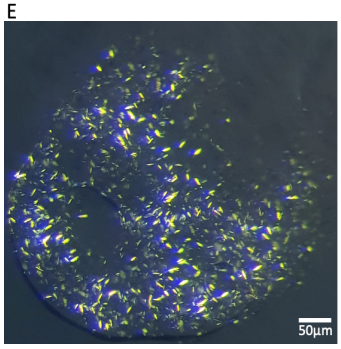
* 25 mg/ml ASBT$_{Lb}$, 1 mM DCA, 0.2 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 42 % PEG 400.

D



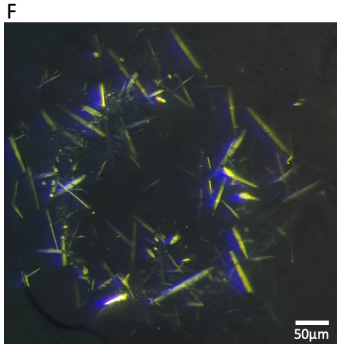
30 mg/ml ASBT$_{Lb}$, 0.1 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 34 % PEG 400.

E



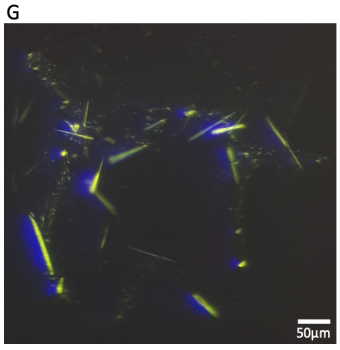
30 mg/ml ASBT$_{Lb}$, 0.1 M Cadmium chloride hemi (pentahydrate), 0.1 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 32 % PEG 200.

F



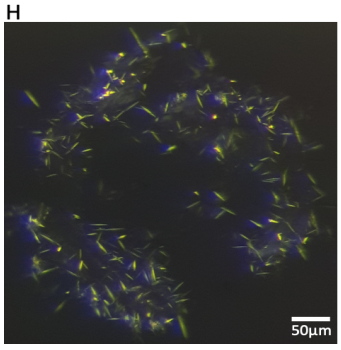
30 mg/ml ASBT$_{Lb}$, 0.1 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 38 % PEG 200.

G



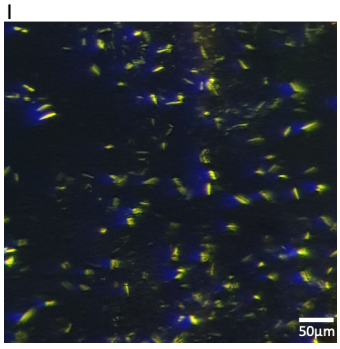
30 mg/ml ASBT$_{Lb}$, 1 mM DCA, 0.1 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 30 % PEG 400.

H



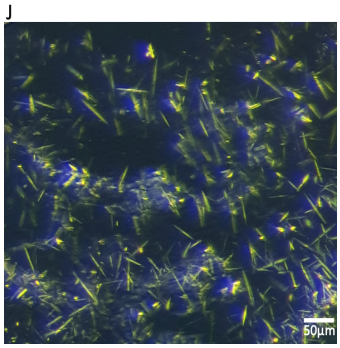
30 mg/ml ASBT$_{Lb}$, 2.5 % Glycerol, 1 mM DCA, 0.1 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 38 % PEG 350MME.

I



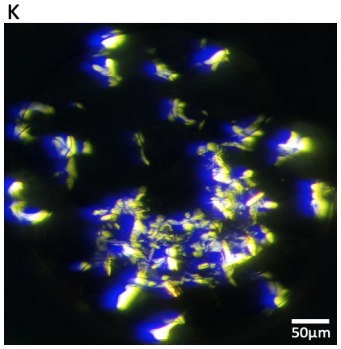
30 mg/ml ASBT$_{Lb}$, 1 mM DCA, 0.05 M Sodium chloride, 0.05 M MES pH 6.5, 34 % PEG 200.

J



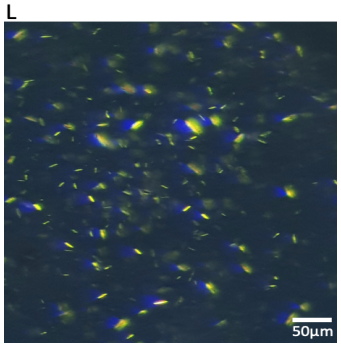
30 mg/ml ASBT$_{Lb}$, 1 mM DCA, 0.2 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 0.5 % MERPOL® HCS surfactant, 42 % PEG 400.

K



** 30 mg/ml ASBT$_{Lb}$, 1 mM DCA, 0.12 M Magnesium formate dihydrate, 0.1 M Sodium chloride, 0.1 M Tris pH 8.5, 33 % PEG 600.

L



28 mg/ml ASBT$_{Lb}$, 0.1 M Cadmium chloride hemi (pentahydrate),0.1 M Lithium chloride, 0.1 M Sodium acetate pH 4.5, 26 % PEG 200.

Figure 5.8: ASBT$_{Lb}$ post-SEC LCP crystallisation studies. The MemGold2 A8 and MemTrans A5 conditions yielded initial crystallisation hits, which were then further optimised in fine screens and displayed above.
* This condition yielded a crystal diffraction dataset at 3.3 Å resolution.
** This condition yielded a crystal diffraction dataset at 2.2 Å resolution.
All crystal images were viewed on a microscope with a cross-polarizer enabled, and imaged using a Samsung Galaxy S8 mobile phone.

As can be seen in figure 5.8A-L above, the ASBT$_{Lb}$ protein produces well-formed, stable LCP crystals, in a multitude of different crystal and buffering conditions, both with and without 1 mM DCA. These crystal conditions were harvested and their respective resolution investigated at the I24 micro-focus beamline at the Diamond Light Source, Oxfordshire. Two of the crystal figures, 5.8.C and 5.8.K yielded very well-diffracting X-ray datasets at 3.3 Å and 2.2 Å resolutions respectively. The structural determination, processing and analysis of this 2.2 Å dataset is investigated in greater detail below.

5.2.2 SLC10A7 *E. coli* full-length and Q320 Truncated Crystallisation

After SLC10A7 bacterial homologue purification trials were well-achieved, crystallisation trials were attempted, in both vapour diffusion (sitting drops) and LCP techniques. Figure 5.9 below details some of the crystal-like "objects" observed from both the SLC10A7 *E. coli* Wt full-length and Q320 truncated bacterial homologue constructs.

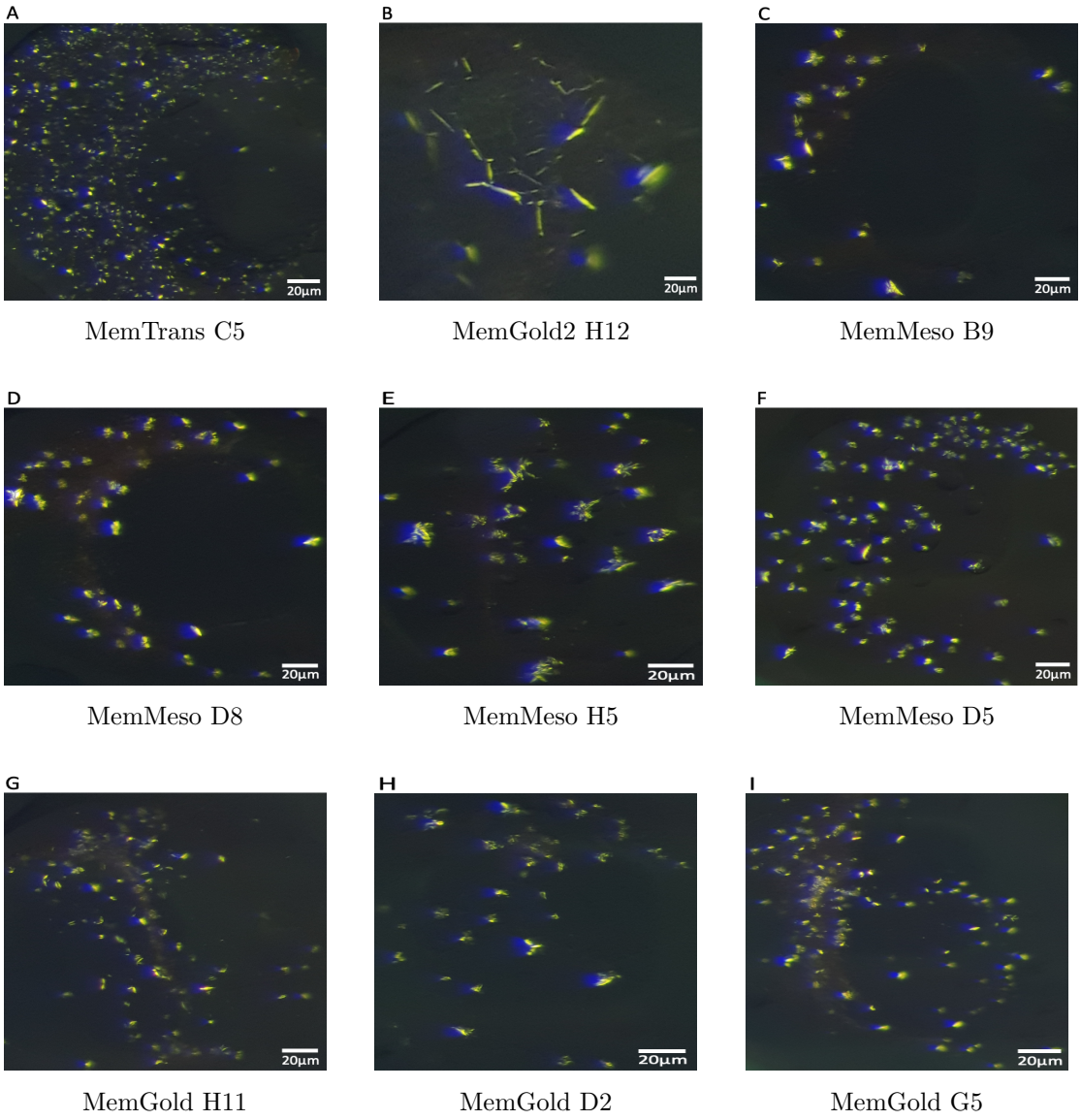
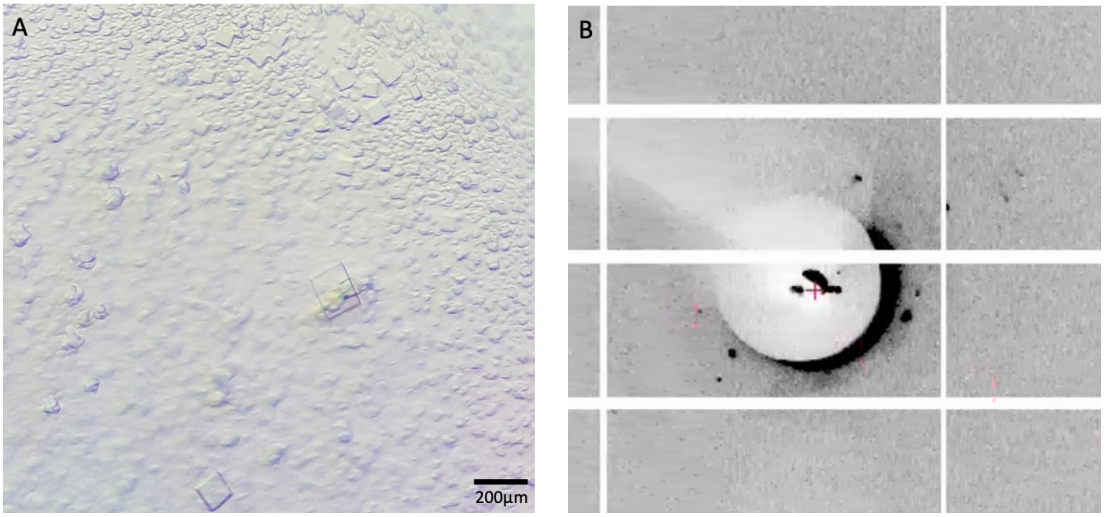


Figure 5.9: SLC10A7 *E. coli* Wt full-length and Q320 truncated crystallisation studies. Only MemGold, MemGold2, MemTrans, and MemMeso broad screens yielded crystal-like looking LCP "objects".
All crystal images were viewed on a microscope with a cross-polarizer enabled, and imaged using a Samsung Galaxy S8 mobile phone.

As seen in figure 5.9A-I, LCP crystallisation investigations yielded several broad screen "hits" presenting with crystal-like looking "objects". These "crys-

tals" were harvested and processed at the I24 micro-focus beamline at Diamond Light Source, but unfortunately failed to show any diffraction of protein nature.

Sitting drop vapour diffusion crystals were also obtained for the very first time in the *E. coli* Q320 C-terminal truncated construct, purified in sodium bicarbonate buffer. The crystal image and respective X-ray diffraction is portrayed below, in figure 5.10.



MemGold2 D5

Figure 5.10: SLC10A7 *E. coli* Q320 C-terminal truncated sitting drop crystal and protein diffraction pattern. (A) Crystal image of the protein grown in MemGold2 D5 (0.2 M Choline chloride, 0.1 M Tris pH 7.5, 14 % w/v PEG 2000 MME). (B) X-ray diffraction pattern for the harvested and analysed crystal.

As can be seen above in figure 5.10, crystal like objects were obtained in MemGold2 D5 broad screen. The crystals were harvested and processed at the Diamond Light Source, Oxfordhire, with the X-ray diffraction pattern displayed. Although very low resolution is shown, the X-ray spots are indicative of protein origin which is an exciting and encouraging development for this project.

5.3 3D X-ray Structure Determination

5.3.1 ASBT$_{Lb}$ 2.2 Å 3D Crystal Structure Analysis

As mentioned previously, both crystal figures, 5.8.C and 5.8.K yielded well-diffracting X-ray datasets at 3.3 Å and 2.2 Å resolutions respectively. The reflections from the 3.3 Å dataset were indexed, scaled and merged via DIALS (Winter et al., 2018), before molecular replacement in Phaser (McCoy et al., 2007) with the whole ASBT$_{NM}$ (PDB (Berman et al., 2000) entry 3ZUY) was initiated. Generated structural coordinates were then edited and model built in Coot (Emsley et al., 2010) and further processed and refined in both PHENIX (Adams et al., 2002, 2010) and REFMAC (Murshudov et al., 2011). A novel nine TMD structure was solved for this 3.3 Å dataset and used in molecular replacement to model the 2.2 Å ASBT$_{Lb}$ dataset in DIMPLE (Wojdyr et al., 2013). Further refinement procedures in PHENIX (Adams et al., 2002, 2010), with TLS and NCS restrictions produced substantially better electron density maps and structural coordinates, with reported R-free and R-work values 0.2483 and 0.2007, respectively. Representation of the 2.2 Å dataset diffraction pattern is seen below in figure 5.11.

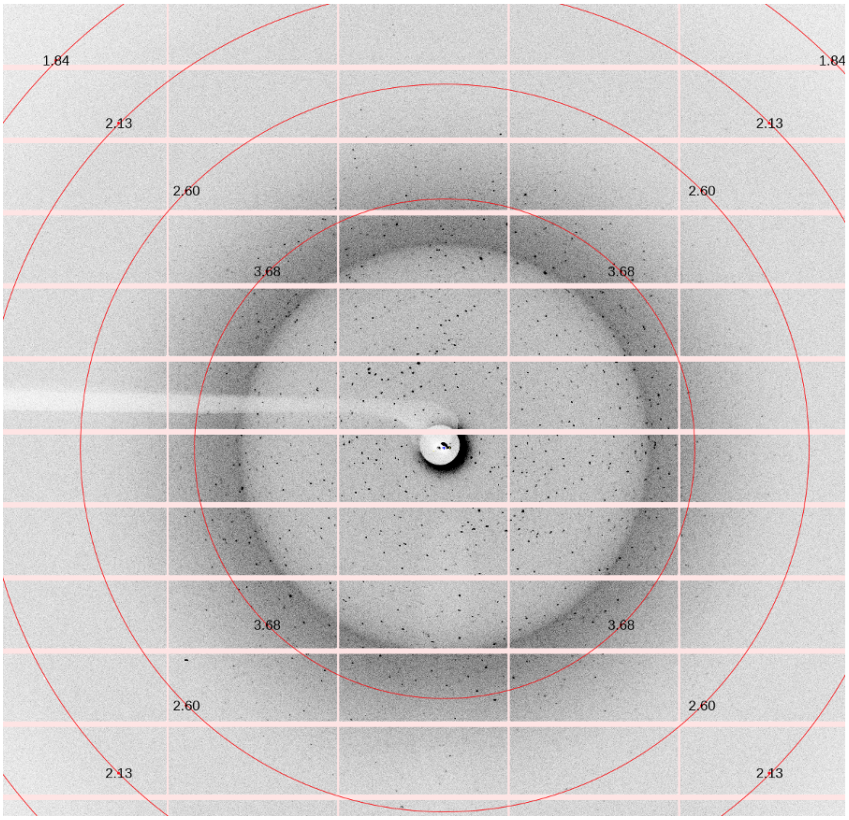


Figure 5.11: The 2.2 Å ASBT$_{Lb}$ dataset X-ray diffraction pattern. Ten images from the data collection were compiled and analysed in dials.image.viewer, highlighting high resolution diffraction spots.

As can be seen above in figure 5.11, nicely shaped, high resolution spots, extending further than 2.6 Å, are clearly visible. However, due to a lack of regularity in the pattern of diffraction (i.e. lack of a lattice display), this diffraction is most likely a representation of multiple crystals. Regardless, a novel transporter entailing nine TMDs, with an extracellular N-terminal and intracellular C-terminal assembly was generated, with in-depth crystal and refinement statistics displayed in tables 5.1 and 5.2 below.

Table 5.1: $ASBT_{Lb}$ 2.2 Å Diamond DLS data collection statistics.
Values in parentheses () represent data in the highest resolution shell.
Statistics were generated in Phenix.

Crystal Statistics	**$ASBT_{Lb}$**
Beamline	DLS/I24
Wavelength (Å)	0.9999
Oscillation range (°)	0.10
Exposure time (s)	0.015
Transmission (%)	10.04
Beam size (µm)	20 ×20
Space group	P 2_1 2_1 2_1
Resolution (Å)	53.95-2.212 (2.291-2.212)
Cell dimensions (Å)	**a**=65.5831, **b**=94.902, **c**=107.602, α=90, β=90, γ=90
Number of measured reflections	213121 (17669)
Number of unique reflections	34240 (3340)
Completeness (%)	99.79 (98.93)
Redundancy/Multiplicity	6.2 (5.3)
I/σ(I)	7.52 (1.01)
Rpim	0.1069 (0.7583)
CC1/2	0.466 (0.427)

As can be seen in table 5.1 above, the crystal dataset was analysed at the Diamond light source synchrotron on beamline I24. A wavelength of approximately 1 Å was selected, roughly corresponding to the radius/width of an atom and the scale of covalent chemical bonds (Suryanarayana and Norton, 2013). The oscillation range, exposure time and transmission of X-rays, were all carefully selected, to ensure the production of high quality resolution data, with the lowest crystal

radiation damage, was possible. The beam size was set based on the size of the protein crystal to ensure an accurate diffraction pattern was possible, and sufficient reflections were detected. A space group P 2_1 2_1 2_1 was deduced for the crystal dataset, representing an orthorhombic translational and rotational symmetry in real space, with a rectangular base and prism, presenting an orthogonal lattice (Grenier and Ballou, 2012). A high resolution up to 2.2 Å was generated, which presents the level at which no information is lost upon truncation of the data, and the shortest distance between planes of the crystal lattice, deduced by the reflections of the diffraction pattern (Evans and Murshudov, 2013). At this proposed resolution, good clear identification of the membrane protein backbone and sidechains is visible (Karplus and Diederichs, 2015; Dubach and Guskov, 2020). The cell dimensions display the smallest possible unit cell which governs the spacing of spots on a diffraction pattern (Evans and Murshudov, 2013). This also gives indication on the orthorhombic symmetry and space grouping of the crystal as is designated by the α=90, β=90, and γ=90 angles. The completeness percentage of the data is a good indication of its quality, as this represent a ratio of the amount of crystallographic reflections collected to the unique amount theoretically possible for the symmetry of a crystal (Arkhipova et al., 2017). Typically, for accurate protein model building and refinement >90 % is desirable (Arkhipova et al., 2017). Here, 99.79 %/98.93 % (highest resolution shell) completeness shows extremely good quality of the crystal data. The redundancy/multiplicity statistic provides an assessment of the total independent detections of every reflection in the crystallographic data set (Rhodes). Here, each independent structure factor was identified 6.2 and 5.3 (highest resolution shell) times on average, portraying good quality of the data. I/σ(I) provides an assessment of the signal to noise ratio of the crystallographic data, providing an important analysis of the quality of the data (Evans and Murshudov, 2013). Here, a highest resolution shell of 1.01 is reported, presenting poor signal to noise data. Typically, 2.0 is considered the limit for adequate data quality, suggesting a resolution cutoff to be perhaps necessary for this 2.2 Å dataset. However, because the data was a collection of reflections from multiple crystals, it is possible the sigma merging statistics have been incorrectly/inefficiently measured, which can result in the disregard of well diffracting reflections. As such, in this situation, I/σ(I) may not provide the best identification of data quality, and must be interpreted further and compared to other resolution determining statistics (such as Rpim and CC1/2), in order to accurately interpret the quality of the real data. The Rpim R factor statistic provides an analysis of the precision of the averaged intensities/amplitudes of merged crystallographic data (Evans and Murshudov, 2013). Taking into account the redundancy of the data, this statistic serves as a good indicator of data quality and

can provide vital information on the resolution cutoff. With a high resolution bin statistic of 0.7583, adequate data quality is depicted. Finally, the CC1/2 statistic further characterises the quality of the merged crystallographic data, by assessing the correlation coefficient between random estimates of intensity of half data sets (Karplus and Diederichs, 2015; Evans and Murshudov, 2013). This CC1/2 statistic is typically regarded as being the primary indicator for the high resolution cutoff of merged crystallographic data. Using the Xia2 (Winter, 2010) default data processing database, a cutoff of the high resolution bin data at CC1/2 of 0.427 was selected, which resulted in the generation of outstanding density maps, correlating well with a resolution of 2.2 Å. Depiction of these fantastic density maps at flat bulky amino acid residues (including TRP118, TYR133 and PHE212) are displayed below in figure 5.12 and can clearly identify the high quality of this crystallographic data.

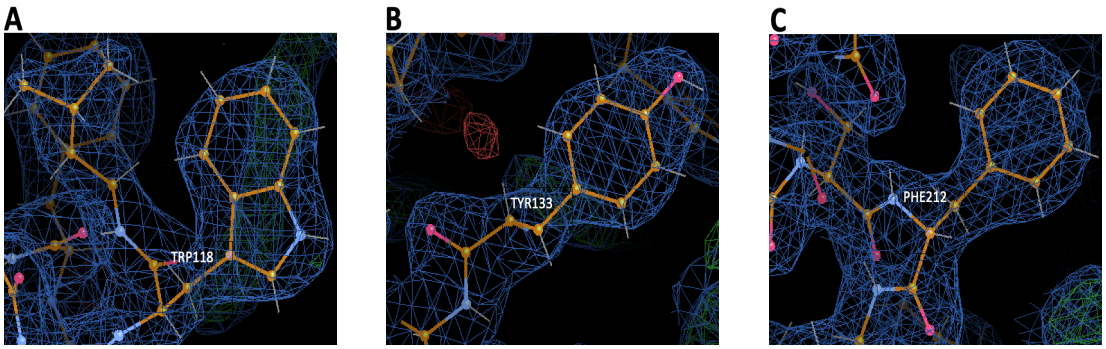


Figure 5.12: Electron density maps of $ASBT_{Lb}$ bulky amino acid residues: (A) TRP118, (B) TYR133, and (C) PHE212. Well defined blue meshed 2Fo-Fc electron density maps are a depiction of high resolution data/quality.
These images were generated in Coot (Emsley et al., 2010).

Taken together, the determination of the resolution of crystallographic data is reliant upon a multitude of statistics, including but not limited to I/σ(I), Rpim, and CC1/2. These statistics for the high resolution bin suggest the data should be cut back and re-examined before the structure can be fully deposited and published. However, although this is necessary, the generation of fantastic density maps correlating to a 2.2 Å resolution, as seen above in figure 5.12, is a significantly encouraging result and one which can be further optimised.

Table 5.2: $ASBT_{Lb}$ 2.2 Å refinement statistics.
Values in parentheses () represent data in the highest resolution shell.

Refinement	**$ASBT_{Lb}$**
Number of reflections used	34186 (3319)
Number of atoms	4550
R-factor/R-work	0.2007 (0.2478)
R-free	0.2483 (0.2803)
R.m.s.d from ideal values	
Bond lengths (Å)	0.015
Bond angles (°)	1.59
Ramachandran outliers (%)	0
Ramachandran favoured (%)	97.55

The R-factor and R-free statistics provide a measure of the refinement quality, by assessing how well the experimental diffraction pattern fits to a simulated diffraction pattern (i.e. fit of the data to the model) (Wang, 2015). As such, these statistics provide an important evaluation on the quality of the model. A statistic of 0.20 is typically acceptable. Here, outer shell R-factor (0.2478) and R-free (0.2803) values portray adequate matching, indicating good model quality. However, because refinement programs introduce bias to these R-values, these may not be the best method to assess overall quality (Wang, 2015). Ramachandran outliers highlight amino acid residues presenting with un-favourable dihedral angles, which typically arise from errors in model building and refinement (Sobolev et al., 2020). Here no outliers were identified which is a representation of good data. Ramachandran favoured in comparison, are fully allowed regions of the aa residues in the data set (Sobolev et al., 2020). Here, a good percentage (97.55 %) is favoured, which again is indicative of good quality data.

This structural dataset generated a nine TMD membrane protein, with an extracellular N-terminal and intracellular C-terminal conformation. Figure 5.13 below, outlines the structural topology and architechture for this exciting novel $ASBT_{Lb}$ membrane transporter.

Figure 5.13: The ASBT$_{Lb}$ 3D X-ray crystal structure, solved to 2.2 Å. A total of nine TM domains, with an extracellular N-terminal and intracellular C-terminal orientation is depicted. TM domains were numbered from 1-9 and coloured from blue at the N-terminal to red at the C-terminal. TMDs 3 and 8 both form disrupted helical arrangements, and as such have been categorised into TMD 3a, TMD 3b, TMD 8a and TMD 8b. Additionally present in the structure are two proposed sodium ions (Na1 and Na2) and a co-crystallised formate molecule, displayed as magenta spheres, and sticks, respectively.
This image was generated in PyMol (Schrödinger, 2015).

This ASBT$_{Lb}$ structure can be divided into two functional domains comprising of a panel domain (TMDs 1, 5 and 6) and a core domain (TMDs 2, 3a, 3b, 4, 7, 8a, 8b, and 9), with proposed sodium ions and bound formate, localised in close proximity to the TMD3-TMD8 discontinuous helical crossover region (see in more detail below). Figure 5.14 below highlights the location and topological orientation of these TMDs accordingly.

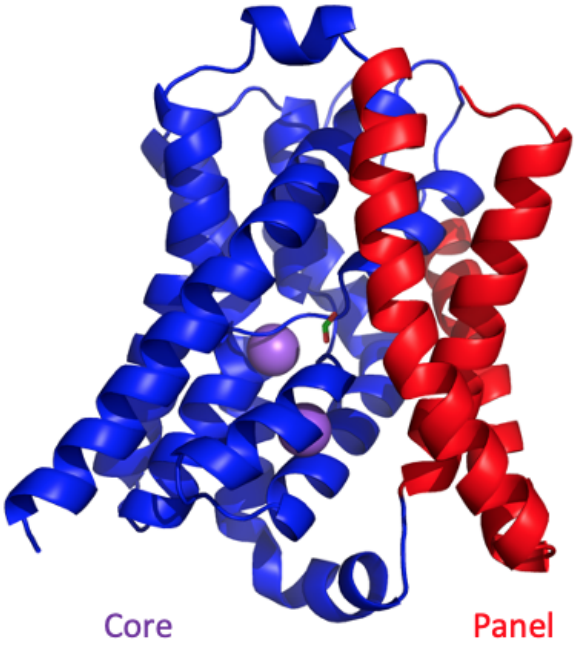


Figure 5.14: The respective core and panel domains of the ASBT$_{Lb}$ crystal structure. The core domain is composed of TMDs 2-4 and 7-9, whilst the panel domain is encompassed by TMDs 1, 5 and 6. For clarity, the panel domain has been coloured in red, and the core domain in blue. Proposed sodium binding sites (Na1 and Na2) are displayed as magenta coloured spheres, and formate in stick mode.
This image was generated in PyMol (Schrödinger, 2015).

An assessment of the charged residues present in the ASBT$_{Lb}$ structure is seen in figure 5.15 below, highlighting a surface depiction of the protein with respective positively and negatively charged residues.

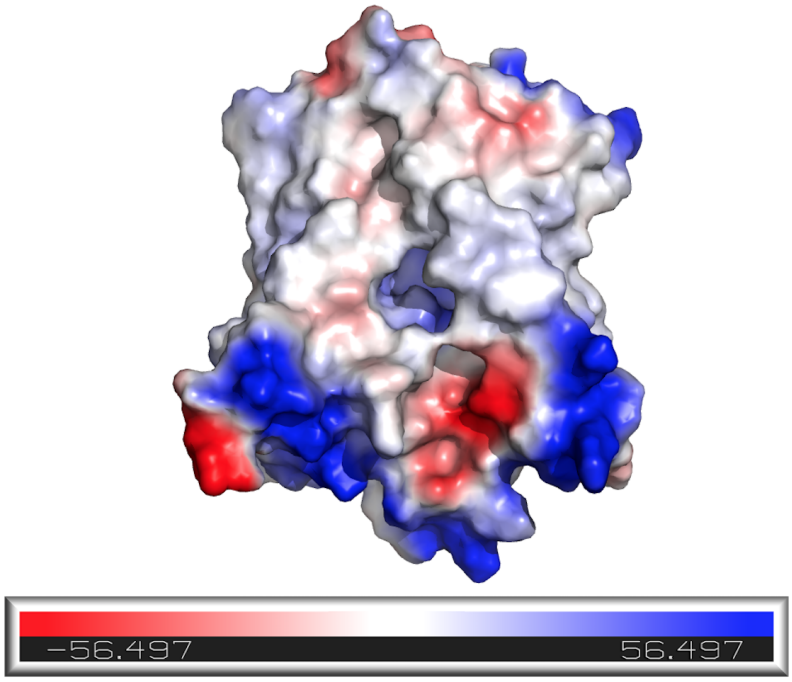


Figure 5.15: Electrostatic charged surface of the ASBT$_{Lb}$ structure. Areas coloured red show negative charge, and blue areas show positive charge. White areas show neutral charge.
This image was generated in PyMol (Schrödinger, 2015).

As can be seen in the $ASBT_{Lb}$ electrostatic surface representation above, a large proportion of the aa residues residing in or near the outer surface, are positively and negatively charged. In comparison, the large majority of the central binding cavity of the transporter exhibits fairly neutral potentials.

Thermal motion of the $ASBT_{Lb}$ structure were also investigated by portraying the conformational architecture with its corresponding B-factor flexibility. Figure 5.16 below, highlights the flexible regions of the membrane protein in both surface and putty representations.

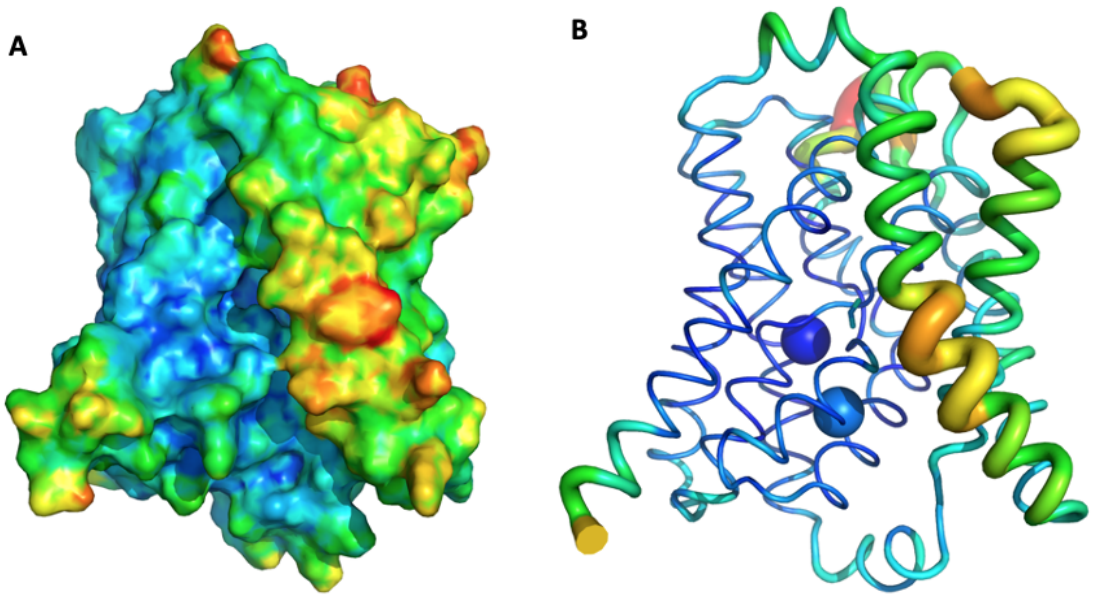


Figure 5.16: B-factor representation of the $ASBT_{Lb}$ structure. (A) surface and (B) putty representations of the flexible regions of the structure as termed by the B-factors.
This image was generated in PyMol (Schrödinger, 2015).

As can quite clearly be identified in figure 5.16 above, the panel domain in particular exhibits greater levels of flexibility with regards to the core domain of the protein, which remains fairly static. As would be expected, extracellular and intracellular loops also show some degree of thermal flexibility.

Further analysis of the assigned sodium ions and formate molecule are given below in figures 5.17, 5.18, 5.19 and 5.20.

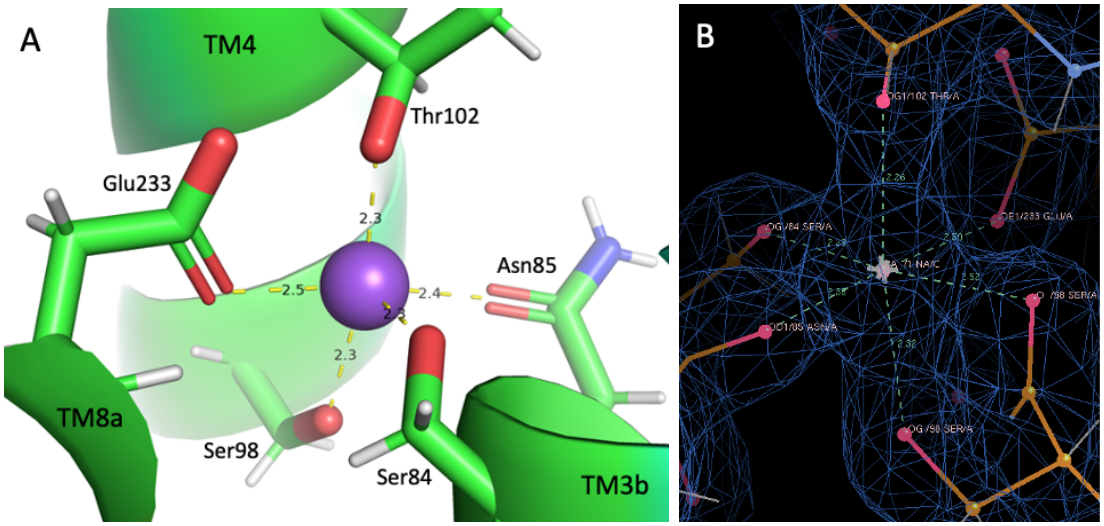


Figure 5.17: The proposed sodium 1 (Na1) binding site of the ASBT$_{Lb}$ transporter. (A) Cartoon representation of the strong molecular interactions between the Na1 ion and neighbouring TMD3b, TMD4 and TMD8a helices. The proposed sodium ion is depicted as a sphere coloured in magenta, with interacting aa residues displayed as sticks. Key distances to nearby residues are shown as dashed lines. (B) The 2Fo-Fc blue electron density map representation of the same Na1-ASBT$_{Lb}$ molecular interactions.
(A) was generated in PyMol (Schrödinger, 2015). (B) was generated in Coot (Emsley et al., 2010).

As can be identified in figure 5.17 above, the proposed Na1 ion is coordinated by the backbone carbonyls of Ser84, Asn85, Ser98, Thr102 and Glu233, all located near the TMD3-TMD8 discontinuous helical crossover region.

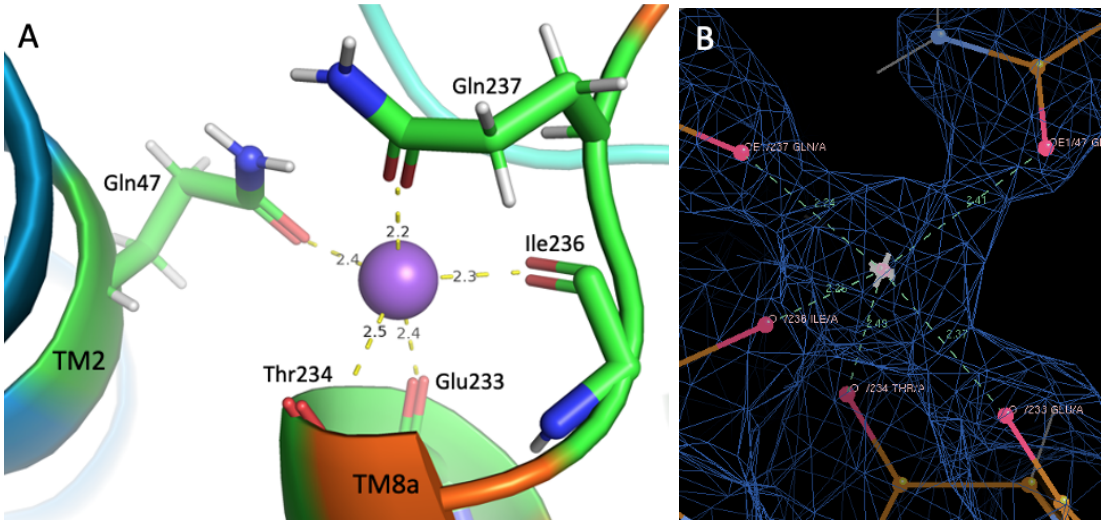
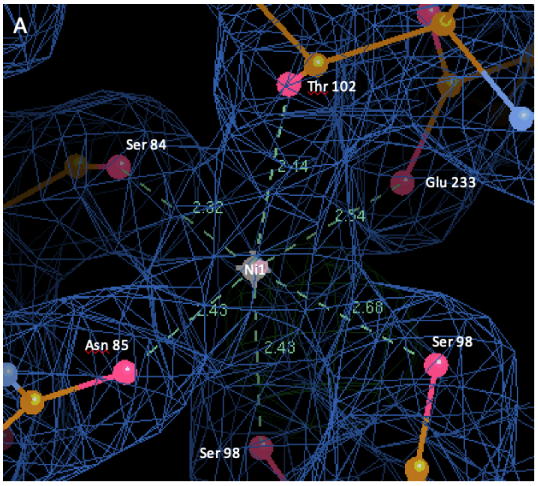


Figure 5.18: The proposed sodium 2 (Na2) binding site of the ASBT$_{Lb}$ transporter. (A) Cartoon representation of the strong molecular interactions between the Na2 ion and neighbouring TMD2, and TMD8a helices. The proposed sodium ion is depicted as a sphere coloured in magenta, with interacting aa residues displayed as sticks. Key distances to nearby residues are shown as dashed lines. (B) The 2Fo-Fc blue electron density map representation of the same Na2-ASBT$_{Lb}$ molecular interactions.
(A) was generated in PyMol (Schrödinger, 2015). (B) was generated in Coot (Emsley et al., 2010).

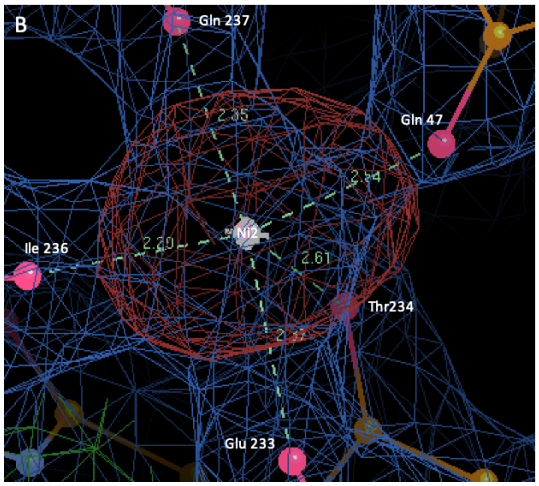
As can be identified in figure 5.18 above, the proposed Na2 ion is coordinated by

the backbone carbonyls of Gln47, Glu233, Thr234, Ile236 and Gln237, all located near the TMD3-TMD8 discontinuous helical crossover region.

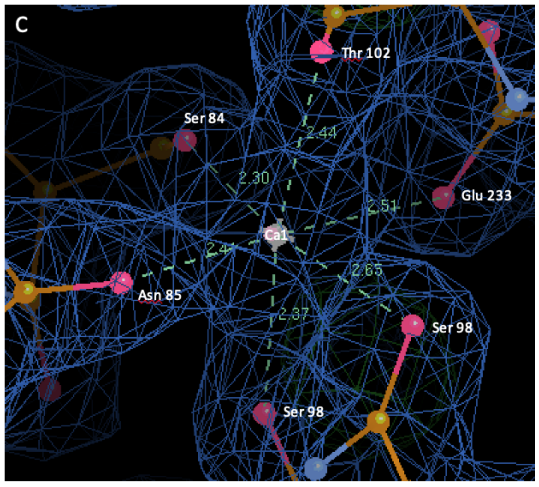
Since the incorporation of 2 Na^+ ions into the ASBT structure is crucial to drive solute translocation in both *N. meningitidis* (Hu et al., 2011) and *Y. frederiksanii* (Zhou et al., 2014) bacterial homologue structures, we also wished to identify the necessity of these proposed Na^+ ions in the $ASBT_{Lb}$ bacterial homologue. In order to ensure the ions observed in the electron density of this novel structure are correctly and safely assigned as sodium (and not another metal ion entirely), we modelled and refined the $ASBT_{Lb}$ crystal structure with other ions, including Ni^{2+}, Ca^{2+}, K^+, and Mg^+, substituted at both conserved Na^+ ion binding sites. These alternative ions were selected based on their addition in both overexpression and purification procedures. Figure 5.19 below portrays the electron density maps for these additional ion refinement results, with aim to provide stronger evidence for allocating the bound co-transporting ions at the binding sites as sodium.



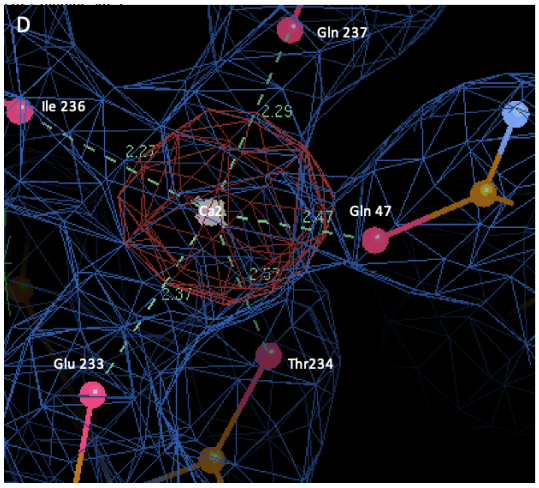
Ni^{2+} Binding Site 1



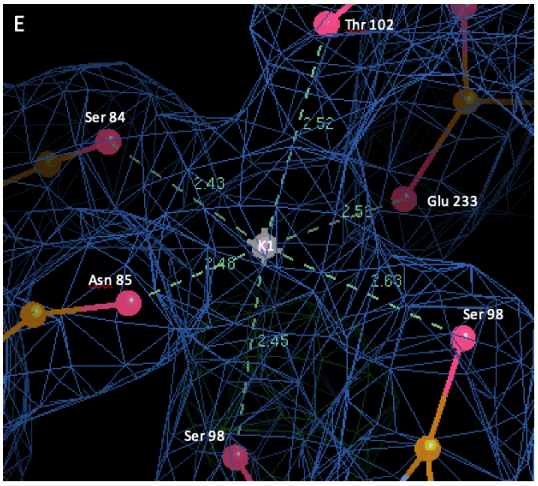
Ni^{2+} Binding Site 2



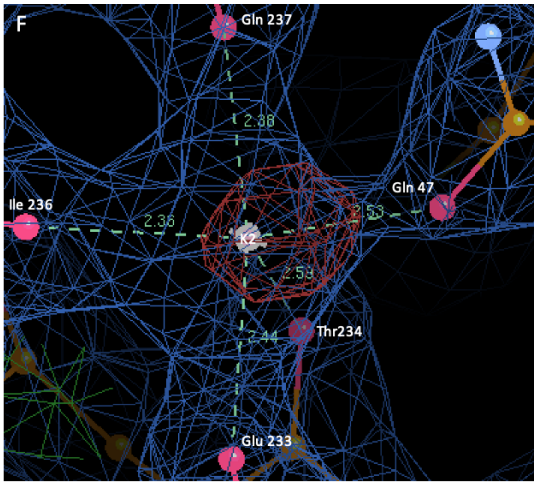
Ca^{2+} Binding Site 1



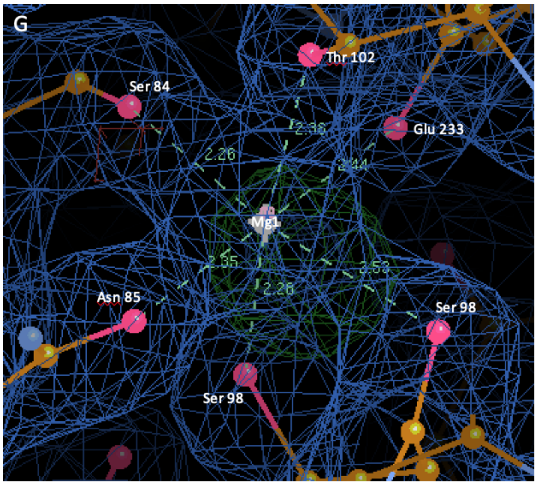
Ca^{2+} Binding Site 2



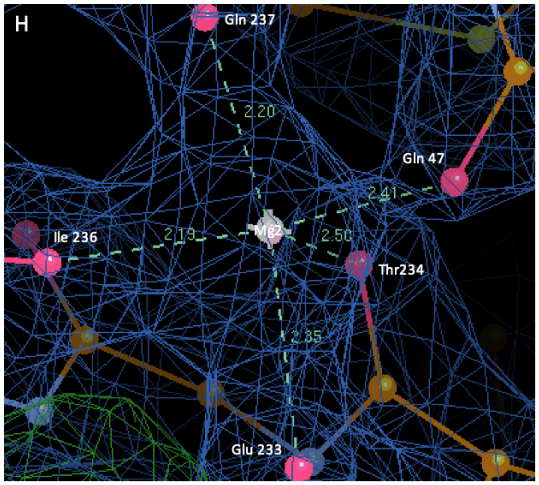
K^+ Binding Site 1



K^+ Binding Site 2



Mg^{2+} Binding Site 1



Mg^{2+} Binding Site 2

Figure 5.19: The electron density maps of the $ASBT_{Lb}$ modelled Ni^{2+} (A-B), Ca^{2+} (C-D), K^+ (E-F), and Mg^+ (G-H) ion binding sites. Specific ions were modelled into the binding sites 1 and 2 with Coot (Emsley et al., 2010), and refined in PHENIX (Adams et al., 2002, 2010). The 2Fo-Fc blue maps were displayed in Coot (Emsley et al., 2010) along with the returned structural coordinates from PHENIX (Adams et al., 2002, 2010). In addition the Fo-Fc red and green difference map was overlayed, in order to identify areas of the structure with negative electron density. Key distances to nearby residues are shown as dashed lines.
Images were generated in Coot (Emsley et al., 2010).

As can be well identified in the figure 5.19A-H above, overall mapping of the investigated alternative ions to the binding sites of the $ASBT_{Lb}$ membrane protein (with the exception of Mg^{2+} ions), was very poor, especially at the binding site 2. Indeed, all ions trialled (with the exception of Mg^{2+} ions) displayed red negative electron density difference maps at the location of the binding site with both Ni^{2+} and Ca^{2+} presenting with the largest negative density difference maps. These results were unsurprising, considering the greater size of the atomic numbers of both these ions, in comparison to the Na^+ ions, and thus, these refinements have allowed us to exclude the possibility of bound Nickel or Calcium ions at the

$ASBT_{Lb}$ ion binding site 2. K^+ ions also presented with negative electron density at the binding site 2, albeit smaller, due to smaller atomic number and size of the ion. As such, we were able to rule out a possible binding site for all three of these ion species (Ni^{2+}, Ca^{2+}, and K^+). Mg^{2+} ions, in contrast, were modelled well into the $ASBT_{Lb}$ binding sites, showing no negative electron density. Being extremely similar in size to Na^+ ions (entailing one extra electron in the atomic number), we were unable to distinguish or rule out the possible binding of Mg^{2+} ions in the $ASBT_{Lb}$ transporter.

In addition, fascinating research conducted by Marjorie Harding, has provided accurate and novel insights into the expected/predicted distances between protein aa residues and bound metallic ion ligands, in 3D X-ray crystal structures (Harding, 2001, 2002, 2006). Here, for Na^+ ions modelled into a protein structure, an expected mean distance of 2.41 Å has been proposed (Harding, 2006), with an accepted range of 2.01-2.70 Å (Harding, 2002). With reference to figures 5.17 and 5.18 seen above, the distances of aa residues to the modelled bound Na^+ sites lie well within this range, thus providing encouraging evidence for the bound ions being Na^+, as has been the case for previously crystallised $ASBT_{NM}$ (Hu et al., 2011) and $ASBT_{Yf}$ (Zhou et al., 2014) bacterial homologues. For modelled Mg^{2+} ions in a protein structure, an expected mean distance of 2.07 Å, with a standard deviation of <0.05 Å, has been reliably and accurately proposed (Harding, 2006). As can be seen in figure 5.19 above, the exhibited $ASBT_{Lb}$ aa residue-Mg^{2+} distances are far greater than expected, highlighting the incorrect modelling of these magnesium ions in the binding pockets.

Taken together, upon thorough analysis of both electron density difference maps and refined metal ion distances, we can confidently conclude the assigned ions at both binding sites of the novel $ASBT_{Lb}$ transporter to be sodium.

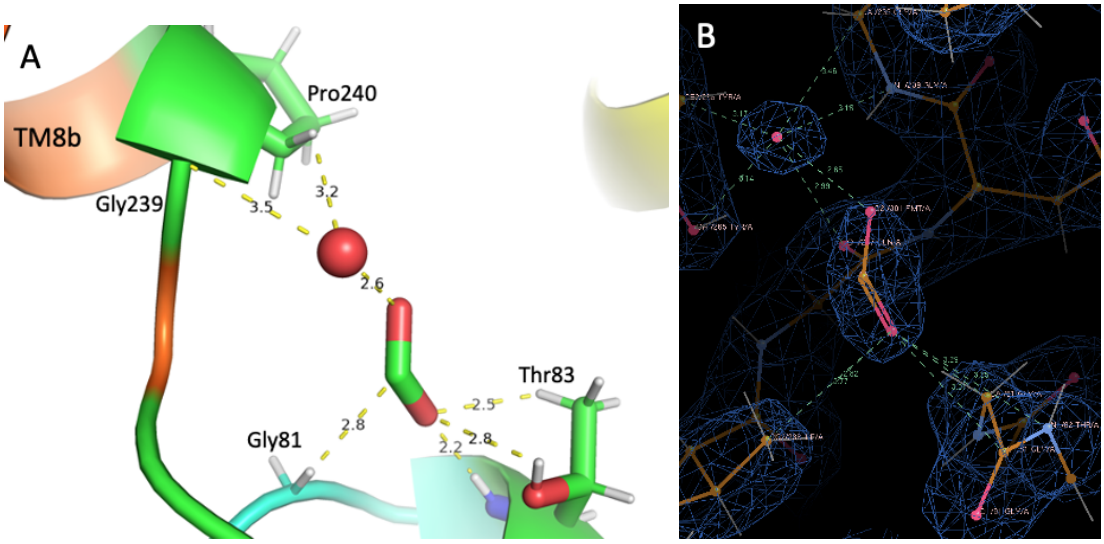
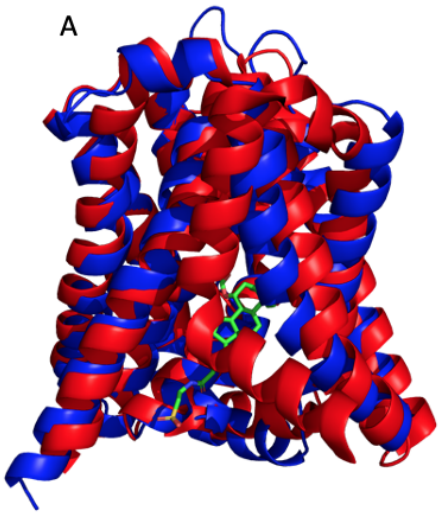


Figure 5.20: The formate binding site of the ASBT$_{Lb}$ transporter. (A) cartoon representation of the strong molecular interactions between the formate ligand and neighbouring TMD3b, and TMD8b helices. The formate molecule and interacting amino acid residues are depicted as sticks, with an additional water molecule displayed as a red sphere. Key distances to nearby residues are shown as dashed lines. (B) The 2Fo-Fc blue electron density map representation of the same formate-water-ASBT$_{Lb}$ molecular interactions.
(A) was generated in PyMol (Schrödinger, 2015). (B) was generated in Coot (Emsley et al., 2010).

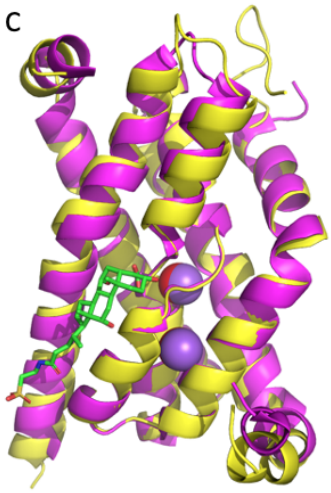
As can be identified in figure 5.20 above, the formate molecule bound to the central cavity of the ASBT$_{Lb}$ transporter, interacts with Gly81, Thr83, and a water molecule, which further interacts with Gly239 and Pro240, near the TMD3-TMD8 discontinuous helical crossover region. Interestingly, although formate was trialled in CPM thermostability assays, no increase in Tm was identified.

As mentioned previously both ASBT$_{NM}$ and ASBT$_{Yf}$ exhibit a topology of ten TMDs. The ASBT$_{Lb}$ as seen above, has presented with just nine TMDs, lacking TMD1 present in both aforementioned bacterial species. In order to examine further the structural effect of this lack of TMD1, superposition alignments of the ASBT$_{Lb}$ with both ASBT$_{NM}$ and ASBT$_{Yf}$ structures were conducted, in order to compare structural similarities and differences in both core and panel domain regions, as well as conserved sodium and substrate binding sites. Figures 5.21 and 5.22 below highlight the superpositions of this novel ASBT$_{Lb}$ structure with ASBT$_{NM}$ and ASBT$_{Yf}$ respectively.

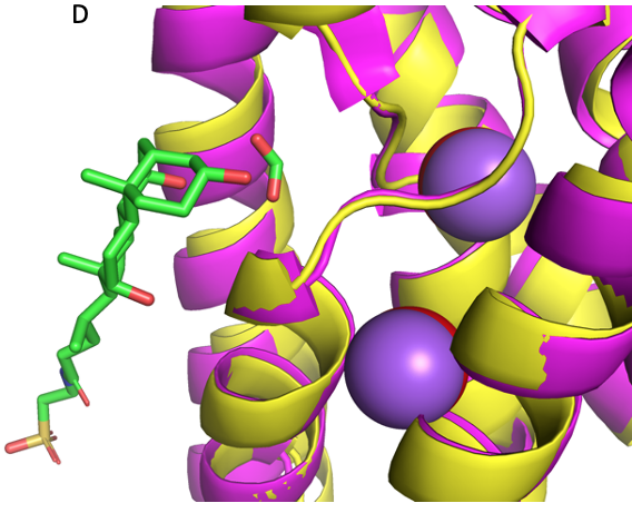


$ASBT_{Lb}$-$ASBT_{NM}$ whole structure

$ASBT_{Lb}$-$ASBT_{NM}$ panel domain



$ASBT_{Lb}$-$ASBT_{NM}$ core domain

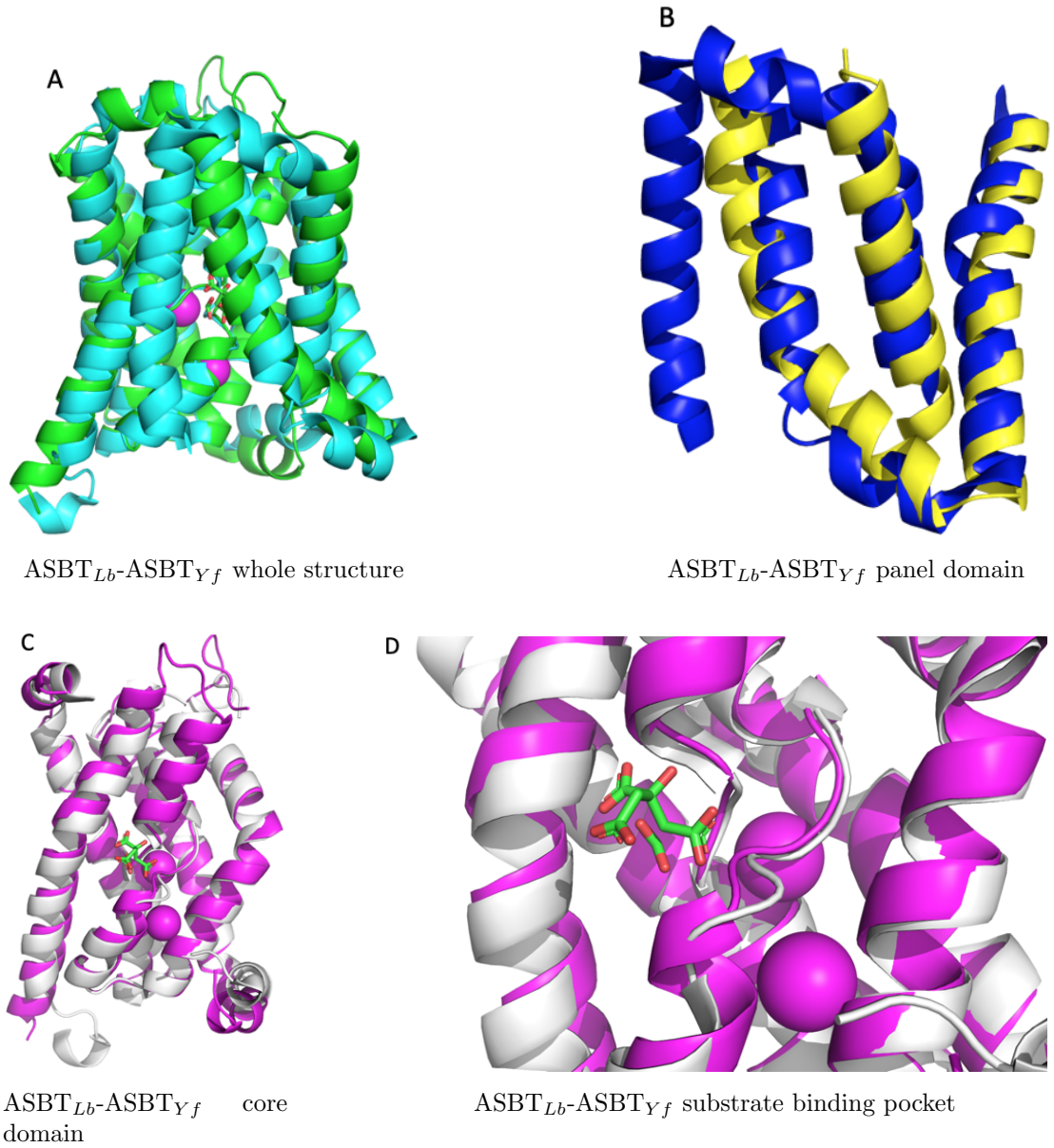


$ASBT_{Lb}$-$ASBT_{NM}$ substrate binding pocket

Figure 5.21: Superposition of $ASBT_{Lb}$ and $ASBT_{NM}$. (A) whole structure superposition with $ASBT_{Lb}$ coloured blue, and $ASBT_{NM}$ coloured red. Both formate and taurocholate substrates are shown as sticks in the substrate binding pocket. (B) panel domain superposition with $ASBT_{Lb}$ coloured red, and $ASBT_{NM}$ coloured green. (C-D) core domain superposition with $ASBT_{Lb}$ coloured magenta, and $ASBT_{NM}$ coloured green. TCA and formate are displayed as sticks and sodium ions are presented as red/magenta spheres.
All images were generated in PyMol (Schrödinger, 2015).

As can be seen in figure 5.21 above, the $ASBT_{Lb}$ structure fits moderately well to the $ASBT_{NM}$ (PDB entry: 3ZUY) structure, with a calculated RMSD of 1.365. However, since $ASBT_{Lb}$ lacks the TMD1 present in $ASBT_{NM}$, separation into panel and core domains is required to outline similarities and differences in the two overlaid structures. Superimposing both panel domains of the structures led to an increase in the RMSD value to 2.620, highlighting key differences in the overlay of both structures. In comparison, the superposition of both core

domains portrayed the lowest RMSD value of 0.905, demonstrating significant conservation of this region. Most interestingly, the binding sites for the sodium ions and substrates in the core domain, for both overlaid structures showed very strong conservation.



$ASBT_{Lb}$-$ASBT_{Yf}$ whole structure

$ASBT_{Lb}$-$ASBT_{Yf}$ panel domain

$ASBT_{Lb}$-$ASBT_{Yf}$ core domain

$ASBT_{Lb}$-$ASBT_{Yf}$ substrate binding pocket

Figure 5.22: Superposition of $ASBT_{Lb}$ and $ASBT_{Yf}$. (A) whole structure superposition with $ASBT_{Lb}$ coloured green, and $ASBT_{Yf}$ coloured grey. Both formate and citrate substrates are shown as sticks in the substrate binding pocket, with sodium ions shown as magenta spheres. (B) panel domain superposition with $ASBT_{Lb}$ coloured green, and $ASBT_{Yf}$ coloured blue. (C-D) core domain superposition with $ASBT_{Lb}$ coloured magenta, and $ASBT_{Yf}$ coloured grey. Citrate and formate are displayed as sticks and sodium ions are presented as magenta spheres. All images were generated in PyMol (Schrödinger, 2015).

Similar to the structural overlay with $ASBT_{NM}$, superposition of the $ASBT_{Lb}$ structure to $ASBT_{Yf}$ (PDB entry: 4N7W) portrays moderate fitting, with a whole structure RMSD value of 1.681. Superposition alignment of panel domains again

indicate substantial structural differences, with an increased RMSD value of 2.771. Again, the core domain superpositions of both structures are very well conserved, with a reported RMSD value of 1.069. Since there were no electron density for the sodium binding sites in the $ASBT_{Yf}$, only sodium ions from the $ASBT_{Lb}$ structure are displayed. However, both formate and citrate substrates are seen to be bound in very close proximities to one another, again highlighting the conserved localisation of the binding site of the ASBT transporter.

Chapter 6

Discussion

The fundamental aims of this PhD study were to investigate in greater detail the novel functional and structural classifications of the SLC10 family members. We specifically targeted bacterial homologues of the ASBT and SLC10A7 solute transporters, and primarily aimed to solve their 3D X-ray crystal structures. As part of this study, functional investigations were additionally undertaken, with specific interest targeted to the characterisation of the solutes involved in the translocation across the cell membrane for both "orphan transporters".

Although both $ASBT_{NM}$ and $ASBT_{Yf}$ have previously been well crystallised and provided a well proposed mechanism of solute translocation through established outward and inward-facing structures, their structural topology consist of ten TMDs with intracellular N- and C-terminal domains. The human ASBT transporter however is known to exhibit just nine TMDs and as such will mediate substrate translocation significantly differently, especially as it is also suspected to form dimeric state conformations. This novel $ASBT_{Lb}$ transporter however also exhibits nine TMDs with an extracellular N-terminal and intracellular C-terminal domain, exhibiting loss of TMD1 present in both $ASBT_{NM}$ and $ASBT_{Yf}$. We aimed to solve the 3D X-ray crystal structure of this transporter, potentially in a dimeric assembly, in order to yield a better understanding of the potential topological architecture of the human ASBT transporter.

As was previously explored in the introduction (see chapter 1), the human SLC10A7 transporter is the only membrane of the SLC10 family to encode ten TMDs, with all remaining members predicted to encode nine TMDs (Claro da Silva et al., 2013). Being the most taxonomically conserved member of the family (with homologous proteins expressed in a multitude of vertebrate and non-vertebrate species), SLC10A7 has been strongly associated to play crucial roles in negatively regulating the influx of calcium ions into cells (Dubail et al., 2018; Ashikov et al., 2018; Laugel-Haushalter et al., 2019; Karakus et al., 2020). This novel functional role is unexpected, considering other members of the SLC10 family have been

shown to instead transport bile acids and drugs (including NTCP and ASBT), as well as steroidal sulfates/hormones (SOAT) (Geyer et al., 2006). Furthermore, previously crystallised bacterial homologues of the ASBT transporter ($ASBT_{NM}$ (Hu et al., 2011) and $ASBT_{Yf}$ (Zhou et al., 2014)) have revealed conserved residues in all SLC10 members (with the exception of SLC10A7), which are crucial for the binding of co-transported Na1 and Na2 sodium ions. Lack of these residues in the SLC10A7 transporter, highlight a altered hypothesised mechanism for its solute translocation, perhaps instead operating via the co-transport of protons (H^+), since a lysine residue is reported to be present at the highly conserved $ASBT_{NM}$ Q264 residue, which is fundamentally involved in driving Na^+ and BA translocation. Interestingly, the proton antiporter NapA exhibits structural similarities to both the $ASBT_{NM}$ and $ASBT_{Yf}$ transporters, albeit expressing a distinctly separate non-homologous aa sequences (Lee et al., 2013, 2014b). Fascinatingly, the H^+ binding site in NapA, corresponding to the Na1 binding site of $ASBT_{Yf}$, has replaced the conserved Gln258 and His71 residues with positively charged Arg331 and Lys305 aa residues near the helical TM4a/TM11a crossover region, in order to drive proton binding and translocation (Lee et al., 2013, 2014b). In light of this, it is possible to suggest SLC10A7 can operate in a similar fashion to NapA, thereby potentially driving proton co-transportation via positively charged Lysine residues in the typical sodium binding sites of $ASBT_{NM}$ and $ASBT_{Yf}$. These abnormal and contrasting, but highly interesting findings therefore question the accuracy and reliability in assigning this calcium-associated transporter to the SLC10 family, considering only >20 % aa sequence identity with other family members is required for its classification.

An in depth discussion of the key findings obtained as part of this PhD study, for both SLC10 projects, is presented in the sections below.

6.1 Membrane Protein Overexpression, Extraction and Fluorescence Size-Exclusion Chromatography (FSEC)

6.1.1 Membrane Protein Small/Large-scale Overexpression and Cell Extraction

The $ASBT_{Lb}$ bacterial homologue

The $ASBT_{Lb}$ bacterial homologue membrane transporter had been previously investigated by Oliver Huxley (MBio student - 2019), to yield optimal overexpression yields, when the addition of 0.25 mM L-rhamnose (added at OD_{600nm} 0) and 0.4 mM IPTG (added at OD_{600nm} 0.5) was attributed to Lemo21(DE3) cells grown in PASM-5052 autoinductive media (Wagner et al., 2008; Schlegel et al., 2012; Lee et al., 2014a). These deduced overexpression conditions translated exceptionally well to large-scale overexpression trials, yielding very well overexpressed, fluorescent GFP-tagged protein, which was further extracted from the cell membrane in a stabilised and well-folded state, with the final cell membrane fraction generating the greatest GFP fluorescence intensity, with only a marginal loss of GFP-fused protein seen in pelleted fractions. Ultimately, this provided encouraging insights for downstream biochemical, biomolecular and biophysical investigations.

The SLC10A7 bacterial homologues

Regarding the SLC10A7 bacterial homologues, analysis of small-scale membrane protein overexpression, identifying the optimal growth conditions for all analysed bacterial constructs (*Asticcacaulis taihuensis*, *Cupriavidus necator*, *Escherichia. coli*, *Escherichia coli* R316 C-terminal truncation, *Escherichia coli* Q320 C-terminal truncation, *Escherichia coli* Q322 C-terminal truncation, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* R318 C-terminal truncated, *Pseudomonas putida* and *Vibrio breoganii*), was achieved.

These aforementioned constructs (with the exception of *P. putida* and *V. breoganii*), produced optimal overexpressed protein yields at 0 mM L-rhamnose and 0 mM IPTG induction. These findings therefore suggest a mediated operation of the overexpression machinery of the Lemo21(DE3) host cells, without the necessity of additional inhibitory reagents, all while ensuring saturated levels of the Sec translocon are not reached (Wagner et al., 2008; Schlegel et al., 2012). Furthermore, increasing the concentration levels of the L-rhamnose reagent, past its optimal inhibitory dose, did not increase the overexpression levels, and in fact

caused the complete opposite, producing significantly lower quantities of overexpressed SLC10A7 proteins, due to the increased "delay" in initiating expression, via the inhibition of T7 RNAP (Schlegel et al., 2012).

The addition of L-rhamnose at OD_{600nm} 0.5 produced significantly lower yields of SLC10A7 protein, when compared to addition at OD_{600nm} 0; which was similarly reiterated with the induction of 0.4 mM IPTG at OD_{600nm} 0.5 also significantly negatively effecting the total yield of protein, when compared to growth in its complete absence.

Regarding the SLC10A7 *P. putida* and *V. breoganii* bacterial homologues, significantly altered optimal overexpression conditions, with variable amounts of L-rhamnose and IPTG, are required, evidently displaying dependence on the time of addition of both reagents to the growth mixture.

Whilst the SLC10A7 *C. necator* protein presented with the greatest fluorescent overexpression yields; *V. breoganii* and *Pseudomonas aeruginosa* R318 C-terminal truncation displayed with the lowest levels overall protein overexpression. Plausible explanations for this underproduction of SLC10A7 in the latter constructs could arise from possible alterations in the functional roles of the membrane transporter, which may generate adverse toxic effects on the viability of the host bacterial cell (Schlegel et al., 2012).

Interestingly, the overexpression of the SLC10A7 *Escherichia coli* full-length Wt and R316 C-terminal truncation constructs saw different quantitative levels of protein being produced, with the latter yielding slightly increased levels under the same optimised conditions. This difference might be explained by the disorder predictions from the RONN database (Yang et al., 2005) (see Appendix 1), which shows an increase in the level of de-stabilisation of the full-length *E. coli* SLC10A7 bacterial construct towards the C-terminal domain. As such, we suspect the deletion of 16 aa residues from this C-terminal tail, to produce the R316 C-terminal truncated construct, increased the stability of the protein, thereby increasing the overexpression levels witnessed in the results.

Taken together, with the exceptions of *Vibrio breoganii* and *Pseudomonas aeruginosa* R318 C-terminal truncation, these SLC10A7 bacterial homologues overexpressed very well, and translated successfully to large-scale overexpression trials. This yielded very well overexpressed, fluorescent GFP-tagged protein, which was further extracted from the cell membrane in a stabilised and well-folded state, with the final cell membrane fraction generating the greatest GFP fluorescence intensity, with only a marginal loss of GFP-fused protein seen in pelleted fractions. Ultimately, this provided encouraging insights for downstream biochemical, biomolecular and biophysical investigations.

6.1.2 Membrane Protein FSEC Trials

The $ASBT_{Lb}$ bacterial homologue

Similarly to the small-scale overexpression trials, FSEC trials had also previously been conducted by Oliver Huxley (MBio student - 2019) for this $ASBT_{Lb}$ bacterial homologue. Excellent solubilisation, stability and monodisperity of the membrane protein was shown in DDM, UDM, DM, NM, and LDAO detergents, with only β-OG exhibiting a denatured and aggregated state of the transporter. These findings from Oliver Huxley were crucially applied in downstream detergent buffer exchanges for optimised purification procedures, in order to enhance success in crystallisation attempts.

The SLC10A7 bacterial homologues

FSEC solubilisation trials were conducted for the bacterial homologues of the SLC10A7 orphan transporter, in order to identify optimal detergents for cell membrane extraction, and formation of stable and functional membrane protein states, with little-to-no precipitation/aggregation (Kawate and Gouaux, 2006; Hattori et al., 2012; Halan et al., 2014).

Overall, the larger-forming micelle detergents, including DDM, UDM, DM, and LMNG, all produced substantially promising solubilisation and stabilisation levels for all the SLC10A7 bacterial membrane proteins. In addition, although both NM and LDAO are perceived to be harsh smaller-micelled detergents, they also provided a respectable stabilisation of the SLC10A7 protein, albeit weaker GFP fluorescent solubilisation yields were achieved. Large, sharp, monomeric and symmetrical Gaussian peaks, indicative of highly stable, monodisperse, correctly-folded and well-solubilised membrane protein assemblies, were witnessed in the above mentioned detergents, for all analysed SLC10A7 bacterial homologues. As such, these detergents were deemed to be more than suitable for downstream solubilisation, purification and crystallisation trials.

Undoubtedly and unsurprisingly, the solubilisation of all SLC10A7 bacterial homologues in β-OG (known to be one of the harshest detergents, resulting in the formation of very small micelles), depicted the weakest GFP fluorescent measurements, presenting with extremely poor membrane protein solubilisation yields. The solubilisation in β-OG resulted in a complete aggregation and denaturation state of the membrane protein, in all analysed bacterial homologues, and as such was classified as a poor choice of detergent for use in future solubilisation, purification and crystallisation studies.

Whilst the SLC10A7 *E. coli* and *K. pneumoniae* bacterial constructs yielded

the greatest levels of solubilised GFP-fused protein, both SLC10A7 *P. putida* and *P. aeruginosa* R318 C-terminal truncated produced the lowest overall membrane protein solubilisation. The FSEC results of the *P. putida* construct are unfortunate, as this bacterial homologue exhibited the greatest overexpression yields in Lemo21(DE3) cells. Its solubilisation, however, in the investigated detergents, produced significantly smaller peaks (indicating weaker solubilisation levels), albeit presenting as sharp and monomeric. This finding evidently demonstrates the lack of guarantee of a membrane protein's high overexpression levels translating to well solubilisation yields. Clearly, the alteration of the lipidic environment, which the transporter is accustomed to, instead into an artificial, non-natural solubilised detergent micelle, proved to be too challenging and stressful on both the *P. aeruginosa* R318 C-terminal truncated and *P. putida* constructs (Nji et al., 2018).

A possible solution to rectify these low solubilisation yields, however, is to redesign the bacterial constructs, to entail the GFP fusion tag at the N-terminal, rather than the C-terminal domain of the protein (Kawate and Gouaux, 2006). Indeed, in certain cases this redesign significantly improved the overexpression and detergent solubilisation profiles, yielding more symmetrical FSEC peaks (Kawate and Gouaux, 2006).

Taken together, the SLC10A7 bacterial homologues were solubilised efficiently in a wide array of detergents (including DDM, UDM, DM, NM, LDAO and LMNG), and presented as monodisperse, stable and homogenous species. As such, their investigations in purification and crystallisation trials can be initiated.

6.2 Membrane Protein Functional Thermostability

6.2.1 Membrane Protein CPM Functional Stability

The $ASBT_{Lb}$ bacterial homologue

CPM thermostability assays (Alexandrov et al., 2008) were generated in order to assess the membrane protein melting temperature, Tm differences under varying buffer and substrate conditions. Investigations were initially trialled in both $ASBT_{Lb}$ and SLC10A7 bacterial homologue transporters, with only the former producing reliably accurate sigmoidal Tm curves, which could be explained by the lowered stability of the SLC10A7 transporter with regards to the $ASBT_{Lb}$ transporter. In light of this SLC10A7 thermostability assays were instead conducted using both GFP-TS and FSEC-TS assays (see below).

An $ASBT_{Lb}$ concentration series was first investigated, in order to establish a low enough concentration to yield accurate and reliable results. As the concentration of the transporter increased, a decline in the thermostability and Tm of the membrane protein was depicted. However, we instead proposed this not to be a decline in Tm, but rather an increase in the sensitivity of the protein to the assay, due to increased concentration levels. This was concluded from the increase in fluorescence (RFU) observed at both the base and threshold levels, indicating greater sensitivity to the conducted experiment. Therefore, we concluded, that at 2.5 µg concentrations, the transporter showed good thermostability levels, with a reported melting temperature of 49.43 °C, in the standardised buffer condition (20 mM Tris pH 7.5, 150 mM NaCl, 0.03 % DDM). In light of this, this 2.5 µg $ASBT_{Lb}$ concentration was selected for future thermostability investigations.

Increase in glycerol conditions (from 0-10 %) yielded gradual increases in the thermostability of the $ASBT_{Lb}$ transporter. These thermal shifts, although low (<5 °C), did provide a significant increase in the stabilisation of the protein, and should be considered for downstream purification and crystallisation strategies. Interestingly, these slight increases in thermostability correlated well with similar glycerol-induced protein stability previously reported in FSEC-TS and GFP-TS studies of the SLC10A7 transporters.

Detergent CPM $ASBT_{Lb}$ thermostability investigations generated expected results and findings. As a transporter becomes solubilised into smaller "harsher" detergents, with greatly increasing CMCs, we would expect a decrease in the thermostability and hence lower Tm measurements (Kawate and Gouaux, 2006; Alexandrov et al., 2008; Hattori et al., 2012; Halan et al., 2014). This was indeed

seen to be the case for the $ASBT_{Lb}$ transporter, as the harshest analysed detergent (β-OG), reported a significantly decreased and unstable membrane protein state, at a reported Tm of just 28.77 °C. In comparison, the LMNG detergent, forming the greatest micellular solubilisation of the $ASBT_{Lb}$ transporter, generated the best thermostability of the transporter at 50.89 °C. Similar transporter-detergent thermostability trends and traits, in many other detergent-screening techniques such as FSEC (Kawate and Gouaux, 2006), FSEC-TS (Hattori et al., 2012) and GFP-TS (Nji et al., 2018), have also been conclusively portrayed for the SLC10A7 bacterial homologues. Taken together, the detergent CPM assay investigations for the $ASBT_{Lb}$ transporter provided accurate, reliable and expected thermostable results, correlating exactly with the degrees of detergent harshness, and thus providing a very good control for investigation.

Since both the human NTCP and ASBT variants are categorically involved in the regulated transport and circulation of BAs in the EHC of the liver and intestine, respectively; BA CPM studies were conducted for the $ASBT_{Lb}$ transporter. Interestingly, all investigated BAs showed significant increases to the Tm of $ASBT_{Lb}$, with the secondary BA, DCA, reportedly demonstrating the greatest thermal shift of 9.9 °C, when compared to the DDM positive control. Therefore, we can state that all the investigated BAs, but DCA in particular, are interacting in a direct manner with the transporter, most likely establishing strong binding to increase protein stability. Unfortunately, radioactive ligand binding and transport studies were unable to be conducted, due to time limiting constraints, and so transport of investigated BAs cannot be conclusively stated without further experimental analysis. Furthermore, although an increase in $ASBT_{Lb}$ thermostability had been reported for these BAs, these definitely do not represent the native substrates of the transporter, since the polypeptide originates from the *Leptospira biflexa* bacterium, which has no need for BAs in their biological homeostatic regulations. As such, another, as of yet unidentified solute, must be specifically associated to the $ASBT_{Lb}$ translocation pathway.

Finally, a wide array of potential substrates, ligands and solutes were additionally investigated using the CPM $ASBT_{Lb}$ thermostability assays. The vast majority failed to increase the Tm values of the protein, with only a selected few (N-lauryl sarcosine (thermal shift 4.02 °C), sodium sulfate decahydrate (thermal shift 0.5 °C), sodium thiocyanate (thermal shift 0.66 °C), and (+) L-tartaric acid (thermal shift 0.16 °C)), showing slight, but insignificant changes in Tm values. N-lauryl sarcosine demonstrated the greatest thermal shift, however was not considered a potential $ASBT_{Lb}$ substrate, due to this compound presenting as an un-natural, industrialised, anionic surfactant, which most likely increases the thermostability of the transporter by forming a mixed micelle with the DDM

detergent.

6.2.2 Membrane Protein GFP-TS and FSEC-TS

The SLC10A7 *Cupriavidus necator* bacterial homologue

GFP-TS (Nji et al., 2018) and FSEC-TS (Hattori et al., 2012) functional stability studies were conducted for the SLC10A7 *C. necator* bacterial homologue, as this construct yielded both high overexpression and solubilisation levels, and presented with one of the greatest sequence identity (34 %) to human SLC10A7.

Modification of buffer ion capacitances, presented with minor stabilisation differences among 20 mM Tris pH 7.5, 1 × PBS pH 7.5, 20 mM Hepes pH 7.5, 20 mM Bicine pH 7.5, and 20 mM Bis-tris propane pH 7.5, in detergent buffer. Fundamentally, Tris pH 7.5 buffer displayed the greatest membrane protein stabilisation Tm at 38.42 °C, with the remaining analysed buffers presenting with reduced but negligible Tm differences. In light of this, 20 mM Tris pH 7.5, with 150 mM NaCl and 1 % DDM, was used as a control reference, for comparison of Tm values to all future thermostability studies.

SLC10A7 stability investigations in detergent buffer with increasing concentrations of glycerol, were also analysed. Overall a positive correlation between glycerol and Tm was observed: as the concentration of glycerol increased from (0-10 %), so did the generated Tm measurements. Detergent buffer with 10 % glycerol yielded the greatest membrane protein stability, with a melting temperature of 40.35 °C. However, this yielded a thermal shift of only 2.47 °C, when compared to the control reference in the absence of glycerol (Tm of 37.88 °C). Taken together, we can state that although glycerol seems to impose slightly increased stabilisation levels to the SLC10A7 protein, these unfortunately remain marginally low and insignificant to be considered for downstream purification and crystallisation trials.

Altered pH buffered stability investigations were also assessed for the SLC10A7 *C. necator* construct, ranging from pH 5.0-10.0 . Stability of the carrier transporter at lower (acidic) pH's (including 20 mM Citrate pH 5.0, and 20 mM MES pH 6.0), lead to a complete unfolded state of the protein, with roughly 80-100 % degradation at 4 °C. It was later concluded from pure GFP pH investigations, that this instability, resulted from the fused GFP which was also significantly disrupted. In comparison, stabilisation of the SLC10A7 membrane protein carrier at higher pH's (ranging from pH 7.5-10.0), saw a significant increase in stabilisation with 20 mM CHES pH 10.0 presenting with a Tm of 49.02 °C, which is a phenomenal thermal shift of 11.6 °C, from the 20 mM Tris pH 7.5, reference control. When analysing again whether this could be due to a GFP-induced effect, it was highlighted that

this is indeed a reliable depiction of the stabilisation of the SLC10A7 *C. necator* bacterial homologue. Taken together, although we are unable to state the levels of stability of this membrane protein at low (acidic) pH's (due to the instability of the GFP tag at these pH's), incredibly, there is conclusive evidence to suggest this SLC10A7 protein operates with greater functional stability at higher (alkaline) pH's.

Varying salt stabilisation studies of the SLC10A7 *C. necator* bacterial homologue were investigated in sodium chloride (NaCl), calcium chloride ($CaCl_2$), magnesium chloride ($MgCl_2$), sodium carbonate (Na_2CO_3) and sodium bicarbonate ($NaHCO_2$), at both 150 mM and 300 mM concentrations. All analysed salts displayed the same positive correlation: increasing salt concentrations from 150-300 mM, increased the generated Tm measurements for the SLC10A7 *C. necator* membrane protein. Fascinatingly, the NaCl salt, at both 150 mM and 300 mM, showed detrimentally negative effects on the stability of the protein, in comparison to all other analysed salts. Furthermore, significantly large Tm thermal shifts were reported in all the salts: $CaCl_2$, $MgCl_2$, Na_2CO_3, and $NaHCO_2$, with the latter presenting the greatest Tm measurement of 50.09 °C, corresponding to a thermal shift of 12.94 °C compared to the 300 mM NaCl reference control. To ensure these reported thermal shifts were a direct representation of the SLC10A7 transporter and not due to influential effects of the GFP tag, pure GFP salt investigations were also conducted. Only $CaCl_2$, at both 150 mM and 300 mM concentrations, showed disruption of the pure GFP protein of up to 30 % at 4 °C. This resulted in the generation of lower than expected Tm values at 59.52 °C and 63.97 °C, respectively. Due to this unexpected effect on the stability of pure GFP, it was determined $CaCl_2$ was an unsuitable salt for stabilisation studies, however all remaining salts analysed provided encouraging prospects for this SLC10A7 bacterial homologue. Taken together, the salt stabilisation GFP-TS results, generated for the SLC10A7 *C. necator* construct, with the exception of $CaCl_2$, are genuine representations of its stability, with $NaHCO_3$ providing the strongest potential to yield well-diffracting protein crystals.

In order to verify and cross-check these GFP-TS salt investigations further, FSEC-TS studies were also performed, yielding some very interesting outcomes. In general, good membrane protein solubilisation post heat-treatment is achieved, yielding sharp monomeric peak profiles in the majority of salts analysed. Similar to GFP-TS studies, the fluorescent profile peaks presented in FSEC-TS became larger and more intense as the salt concentrations were increased from 150 mM to 300 mM, indicative of a greater stabilisation of the protein. The fluorescent peaks from NaCl presented with the weakest protein stabilisation, with the presence of large aggregation peaks visible at the 150 mM concentration. Interestingly

however, although more prone to aggregation, the eluted protein-GFP fluorescent peaks were much sharper in comparison to the peaks portrayed in both $CaCl_2$ and $MgCl_2$, at both 150 mM and 300 mM concentrations. Indeed, both these salts also exhibited the lowest solubilised fluorescent counts at 150 mM concentrations, albeit with little-to-no aggregation present. The broadness in the display of the $CaCl_2$ fluorescent peaks may be explained by the reduced Tm values witnessed from pure GFP investigations in the GFP-TS studies. This would again suggest a direct effect of the GFP fused tag on the stability of the SLC10A7 transporter, and therefore is not a genuine display of the *C. necator* protein stability. Finally, both Na_2CO_3 and $NaHCO_3$ salts, presented with the greatest fluorescent profile peaks, which yielded the greatest protein solubilisation levels with sharp, monomeric profiles. Of particular interest was $NaHCO_3$, which presented with the lowest levels of aggregation and clearly depicted the most stable state of the protein, of all the salts analysed. Taken together, the FSEC-TS profile distributions correlate extremely well with the outcomes deduced from GFP-TS studies, whereby both techniques have come to the same conclusion: the solubilisation of the SLC10A7 *C. necator* protein in $NaHCO_3$, provides the greatest level of protein stability, thereby strongly increasing the likelihood of the successful generation of protein crystals for 3D X-ray crystallisation.

It is impossible to state whether this SLC10A7 is a transporter of bicarbonate molecules, without having established radioactive ligand investigations. Since a substantially increased Tm value is reported for the SLC10A7 transporter upon addition of bicarbonate, one can strongly believe this molecule forms strong molecular binding to the transporter. Therefore, although this molecule might not represent the substrate of this transporter, its increase in overall stability is encouraging for future experimental analyses.

Finally, since the SLC10A7 solute carrier membrane protein remains to this day an orphan transporter, potential GFP-TS substrate investigations in the *C. necator* bacterial homologue were investigated. A wide array of varying inorganic and organic compounds were tested, with the ultimate aim to identify novel compounds which increase the melting temperature of the membrane protein, thereby potentially uncovering a novel association. 0.5-1 mM compound concentrations were used as these were more than sufficient to witness, if any, a fundamental thermal shift. Since other members (including SLC10A1/NTCP and SLC10A2/ASBT) of the SLC10 family are categorised as well-established transporters in the regulation of BA homeostasis and the EHC in humans and mammals (Claro da Silva et al., 2013), we decided to investigate whether SLC10A7 could also play a vital role in this mechanism. Several BA substrates were tested including sodium cholate, sodium taurocholate, sodium deoxycholate, and sodium chenodeoxycholate. A

lack of T_m increase was reported for all analysed bile acid substrates, which was not surprising, considering the previous functional transport findings reported by Godoy et al., 2007. Having wide functional implications in the negative regulation of intracellular calcium homeostasis (Jiang et al., 2012; Alber et al., 2013; Zhao et al., 2016; Dubail et al., 2018; Ashikov et al., 2018; Laugel-Haushalter et al., 2019; Karakus et al., 2020), calcium associated substrates, including the vitamin D derivatives (calcitriol, cholecalciferol, ercalcitriol, and ergocalciferol), as well as the calcium signalling agent, inositol trisphosphate (IP3), were also examined with the SLC10A7 *C. necator* homologue. Unfortunately, no increase in melting temperature was reported. Finally, sugar compounds and other organic/inorganic molecules were additionally trialled with no successes.

6.3 Membrane Protein Purification, Crystallisation and Structural Determination

6.3.1 Membrane Protein Purification

The $ASBT_{Lb}$ bacterial homologue

$ASBT_{Lb}$ purification studies followed Ni-NTA resin binding and 3C protease tag cleavage, producing large quantities of purified protein, which was easily identified by large protein bands at roughly 26kDa. 3C protease contaminants were effectively removed by passing purified fractions through a 5 ml His-trap IMAC column. However, interestingly, an additional contaminant, present at roughly 34kDa, was also visible. Oliver Huxley (MBio student - 2019) previously excised this band and sent it for mass spectrometry analysis at St. Andrews University. The results classified this contaminant as being cytochrome o ubiquinol oxidase, originating from the *E. coli* Lemo21(DE3) host strain. Although established and optimised washing steps (i.e. increased washing volumes of higher imidazole concentration buffers (20-30 mM)) were conducted, with the aim of enhancing the removal of this contaminant, unfortunately cytochrome o ubiquinol oxidase remained present in purified fractions, albeit at significantly weaker protein intensities, in comparison to $ASBT_{Lb}$ purified protein. This therefore suggests the contaminant to form strong binding interactions directly with the protein, prohibiting its removal. However, subsequent downstream purification and crystallisation strategies have provided fruitful and accurate results for the $ASBT_{Lb}$ membrane protein, outlining a lack of interference from the contaminant, with respect to the $ASBT_{Lb}$ functional and structural characterisation.

SEC purification was also conducted in DDM, DDM + 1 mM DCA, and NM conditions, all of which yielded monomeric, and stable protein profiles. Comparing pre- and post SEC fractions, highlighted a significantly improved level of purity in the latter, praising the essentiality of this technique prior to attempting crystallisation studies. As was expected, due to the smaller micelle and hence lower molecular weight, the NM SEC profile of $ASBT_{Lb}$ protein was eluted later in comparison to the DDM profile peak. Interestingly, however, the addition of 1 mM DCA to the DDM buffer, significantly altered its SEC profile, and strongly resembled the UV chromatogram of the NM peak. Indeed, this DCA induced profile peak eluted protein at a very similar elution volume of NM, and yielded the sharpest profile for $ASBT_{Lb}$ protein, generating the greatest level of purity. Since BAs can be regarded as detergents, and behave as such, it is assumed DCA embeds itself into the DDM micelle to form a mixed micelle, thereby yielding tighter and

stronger molecular interactions with the protein, resulting in increased stability and purity levels.

Interestingly, the SEC purity levels of both DDM + DCA, and NM conditions, yielded moderately intense protein bands at roughly 53kDa. This molecular weight would correspond with a dimeric state of the $ASBT_{Lb}$ protein, however, no indication has yet been identified as this to be the case. Therefore, mass spectrometry analysis of this molecular weight band should be investigated further in the future development of this project. Furthermore, the possibility of an oligomeric induced SDS-PAGE effect cannot be exclusively exempt.

Oliver Huxley (MBio student - 2019) had also previously shown the $ASBT_{Lb}$ membrane protein to be very stable and monomeric in SEC studies with LDAO detergent buffer. Due to time constraints, this was not investigated further, but provides a future optimisation pathway for potential crystallographic protein characterisation.

Taken together, all purified post-SEC $ASBT_{Lb}$ elutions yielded significant quantities of well-purified protein, which were used further in both vapour diffusion (sitting drops) and LCP crystallisation trials.

The SLC10A7 *Escherichia coli* Wt and Q320 C-terminal truncated bacterial homologues

Regarding purification attempts, both SLC10A7 *E. coli* constructs produced eluted fractions of high purity and quality, generating more than sufficient yields for downstream crystallisation investigations. These outcomes highlighted the functional importance of addition of the 3C protease, which was able to successfully cleave the GFP and polyhistidine tag. Further purification strategies including the administration on SEC, yielded extremely pure samples with different oligomeric states of the protein (although this may be due to an SDS-PAGE effect). As the protein was encapsulated in smaller micelled detergents (including NM and LDAO), greater intensities of molecular weight bands at supposed dimeric and trimeric oligomeric states were observed, and confirmed to be SLC10A7 by mass spectrophotometry at the University of St. Andrews.

Taken together, purification trials were highly successful, generating adequate amounts of the membrane protein at the purest of levels (post SEC). The generated protein in different detergents and/or salts would then be subjected to crystallisation attempts in both vapour diffusion and LCP techniques.

6.3.2 Membrane Protein Crystallisation

The $ASBT_{Lb}$ bacterial homologue

$ASBT_{Lb}$ crystallisation studies in both LCP and vapour diffusion (sitting-drop) were very successful, yielding stable and well-diffracting protein crystals, in a wide array of conditions.

Sitting drop crystals were trialled in DDM detergent, in the presence and absence of DCA. Although crystals did form, their harvesting and subsequent analysis at Diamond Light Source, Oxfordshire, classified these to be of too low and insignificant resolution. Due to time restrictions, vapour diffusion optimisation attempts in smaller-micelle forming detergents, such as NM and/or LDAO were not trialled. These however would most likely generate better diffracting crystals, due to the formation of closer molecular interactions arising from smaller protein-solubilised micelles. These investigations therefore provide a future optimisation pathway for obtaining a better structural classification of the $ASBT_{Lb}$ transporter.

For LCP, the $ASBT_{Lb}$ protein was concentrated to roughly 25-30 mg/ml and mixed with monoolein lipid at a 2:3 protein-to-lipid ratio. Although pre- and post-SEC protein fractions produced a multitude of LCP crystals in very similar conditions, the latter yielded significantly better diffraction. Furthermore, crystals were also obtained in similar conditions with the presence and absence of 1 mM DCA, with the former producing better diffraction. Therefore, the best LCP crystals were obtained from post-SEC DDM purified protein, in the presence of 1 mM DCA. Since the addition of DCA to the protein was seen to significantly increase its stability and melting temperature (T_m) from CPM assay investigations, it was not surprising to see better crystal diffraction upon its addition to post-SEC purified protein as well.

Post-SEC NM purified protein LCP crystallisation studies were also conducted, but yielded significantly smaller crystals in comparison to protein purified with DDM. Solubilisation in NM detergent, although significantly crucial for vapour diffusion crystallisation, is perhaps not vital for LCP crystallisation, as the detergent is "replaced"/"renatured" by the monoolein lipid to form a more stable interaction with the protein. Therefore, solubilisation in a harsher NM detergent and renaturation with monoolein may be less productive in comparison to solubilisation in a more stabilising DDM detergent, prior to replacement with monoloein.

The SLC10A7 bacterial homologues

Multiple crystallisation trials were attempted for the vast majority of the SLC10A7 bacterial homologues. Although LCP crystallisation generated what originally looked like potential protein crystals, upon analysis and further processing at the X-ray synchrotron at Diamond Light Source, these unfortunately proved not to be of protein origin and instead were strongly suspected to be presentations of the breakdown of the LCP setup.

Crystallisation success has been recently shown for the very first time in vapour diffusion sitting drop crystallisation of the SLC10A7 *E. coli* Q320 C-terminal truncated protein. This construct was purified in sodium bicarbonate entailing buffer and formed protein crystals at 15 mg/mL. Although the X-ray diffraction pattern presents spots of very low resolution, this was expected since crystals were obtained from DDM purified protein. However, the confirmation that these crystal-like objects are indeed of protein origin, is hugely encouraging and a step closer towards the structural characterisation of this transporter. Therefore, with subsequent solubilisation of this construct in smaller micelled and "harsher" detergents, will provide the greatest likelihood of yielding better diffraction vapour diffusion protein crystals.

6.3.3 Membrane Protein Structure Determination

The $ASBT_{Lb}$ bacterial homologue

The nine TMD predicted topology of this novel *L. biflexa* bile acid transporter is interesting, since the human ASBT transporter has also been predicted to categorically contain just nine TMDs. Whilst previously crystallised ASBT bacterial homologues from *Neisseria meningitidis* and *Yersinia frederiksenii* have depicted ten TMD topologies, this novel *L. biflexa* bacterial homologue, was predicted to be lacking the TMD1 of $ASBT_{NM}$ and $ASBT_{Yf}$, thereby potentially reiterating similar structurally and functionally conserved attributes of the human ASBT transporter. Of specific interest is the potential formation of a functionally active dimeric assembly, which has previously not been observed in either $ASBT_{Nm}$ and $ASBT_{Yf}$ transporters, but which is a known possible oligomeric state of the human variant.

Through the above mentioned crystallographic optimisation procedures in LCP, incredibly, a 2.2 Å $ASBT_{Lb}$ crystal dataset was produced. Previously established structural folds from $ASBT_{NM}$ (Hu et al., 2011), and $ASBT_{Yf}$ (Zhou et al., 2014), were able to classify the panel domain of $ASBT_{Lb}$ to entail TMDs 1, 5, and 6, and the core domain to contain TMDs 2-4 and 7-9. The specific positioning of the panel domain with regards to the core domain of the structure, demonstrated the conformation of the transporter to be in an open-inward facing state as part of the alternating access mechanism. Furthermore, representation of the transporters thermal movement flexibility, by portraying associated transporter B-factors, elucidated increased flexible movement of the panel domain with regards to the core domain, which remains relatively static. Indeed, the core domain exhibits low B-factors, whereas the panel domain entails high B-factors. Therefore, this strongly suggests the panel domain moves relative to the core domain, in order to drive solute translocation, and correlates significant well with the reported structural mechanism initially suggested by (Hu et al., 2011). (Zhou et al., 2014), on the other hand, seemed to suggest the core domain to move relative to the panel domain, which is very contradictory to the findings reported here for $ASBT_{Lb}$. To conclude, the specifications of the solute translocation alternating access mechanism is not clear for the ASBT transporter, and needs to be further investigated.

Zhou et al., (2014), and Wang et al., (2021a,b), proposed an elevator-style alternating access mechanism for $ASBT_{Yf}$ solute translocation across the cell membrane. Since one domain is seen to move relative to an immobile domain, the authors suggested the core domain of the transporter to be the mobile component in solute translocation, whereby a rigid-body motion in the TM4/TM9 crossover region drives a conformational alteration to drive transport. In contrast, Hu et al.,

(2011), instead suggested the panel domain to be the moving component, due to a strong conservation of the core domain and greater flexibility and variation of the panel domain. Since crystal structures only provide a snapshot of a transporter mid-translocation, multiple conformations, including intermediate occluded states, must be generated to conclusively determine the solute translocation mechanism.

Electrostatic charged surface representation of the transporter highlighted large areas of the outer aa residues to hold positive and negative charge, with the central hydrophobic TMDs region to exhibit mostly neutral aa residue charge.

Consistent with previous classifications of ASBT translocation to require two sodium ions (Na^+) to drive the energy for transport of one substrate, $ASBT_{Lb}$ also had two proposed Na^+ binding sites in the structure. Indeed, Na1 formed strong molecular interactions (2.3-2.5 Å) with associating aa residues including: Asn85 and Ser84 from TMD3b, Ser98 and Thr102 from TMD4, and Glu233 from TMD8a. These residues correlated exactly to the same interactions seen for $ASBT_{NM}$. The Na2 binding site also formed very strong molecular interactions (2.2-2.5 Å) with the following associated aa residues: Gln47 from TMD2, Glu233, Thr234, Ile236 and Gln237 from TMD8a and the linking loop to TMD8b. Again, these residues were conserved and correlated exactly to the same aa residues involved in the Na2 binding site in the $ASBT_{NM}$ transporter. The electron density plots for both these binding sites displayed fantastic maps, demonstrating very well the conservation of the residues of the transporter.

Although DCA was added to the purification and crystallisation procedures for the $ASBT_{Lb}$ transporter (as thermostability CPM assays established a significant increase in the stability and melting temperature of the protein), the crystallised structure instead showed a bound formate molecule to the binding cavity of the transporter, originating from the co-crystallisation buffering condition (MemGoldMeso G11). This formate molecule formed significant molecular interactions (2.2-3.5 Å) with associated aa residues including: Gly81 and Thr83 from TMD3b, and Gly239 and Pro240 from TMD8b. This binding is not unexpected, since the $ASBT_{Lb}$ transporter is expressed in an operon corresponding to the production of long and short chain fatty acid co-A synthases. Since formate can be classified as the shortest fatty acid chain, its binding to the transporter is deemed likely. As part of the binding site, a water molecule is also present, interacting with the formate and associated aa residues. We speculate this water can be replaced to fit longer chained fatty acids in the binding cavity of the transporter.

Due to the lack of bound DCA in the solved structure, it is more than likely DCA does not represent the substrate of the transporter, but rather we speculate it to bind to the outside of the transporter, forming strong interactions to produce

a stable state, which subsequently initiates the subsequent binding of formate and other chained fatty acids. This would not be unexpected, since bacteria have no use for bile acids, as they do not undergo EHC and as such BAs are regarded not to be native molecules/substrates. However, since the addition of 1 mM formate in $ASBT_{Lb}$ CPM investigations (see above) failed to show an increase in the T_m of the transporter, we cannot categorically state formate to represent a possible $ASBT_{Lb}$ substrate. Therefore, further experimental analysis of the formate molecule and its association with the $ASBT_{Lb}$ transporter must be conducted by means of radioactive ligand binding and transport assays.

Overlay of the $ASBT_{Lb}$ structure to both $ASBT_{NM}$ and $ASBT_{Yf}$ yielded significantly interesting and comparable outcomes. In both structural overlays, increase in the RMSD values was seen when superposition of the panel domain was conducted. This can be explained by the flexible tendency of this domain to drive solute translocation. In comparison both $ASBT_{NM}$ and $ASBT_{Yf}$ showed the lowest RMSD values when their core domain was superimposed independently to $ASBT_{Lb}$. This would correlate with the conservation of this domain representing a more static localisation with regards to the panel domain. Furthermore, the sodium ions from $ASBT_{NM}$ and $ASBT_{Lb}$ superimposed almost perfectly with one another, whilst the formate, taurocholate and citrate substrates are in very close proximity to one another, and are all positioned at the interface of the discontinuous helices, highlighting the significantly conserved substrate binding pocket of these ASBT homologous transporters, as was first proposed by (Wang et al., 2001, 2021a).

Although we did not see any binding for deoxycholate in the $ASBT_{Lb}$ structure, we did see density for a co-crystallised formate molecule, near the TM4/TM9 crossover region of the transporter. The binding of this formate molecule lies in the acidic tail site of the suggested solvent accessible binding cavity proposed by Wang et al., (2021a,b), and overlays extremely well with substrates previously crystallised in both $ASBT_{NM}$ and $ASBT_{Yf}$ structures. This suggests that the substrate binding cavity is accommodating to a wide array of differently sized solutes, which will always bind in the density in the same fashion, with acidic tails facing the core domain, and hydrophobic regions orientating towards the panel domain of the ASBT transporter. These findings propose a conservation in the binding site for substrates in the interface of the core and panel domain, proposing a potential solute translocation pathway for the human ASBT transporter.

Taken together, this novel $ASBT_{Lb}$ structure strongly resembles the conformational architecture of both $ASBT_{NM}$ and $ASBT_{Yf}$ transporters, presenting with a significantly conserved core domain, with similar localisation of sodium and substrate binding sites. The panel domain, however, is significantly different, although

not unexpected due to the aforementioned flexibility.

As such, this novel $ASBT_{Lb}$ structure provides an accurate portrayal of a homologous human ASBT transporter, with residues involved in the sodium and substrate binding sites being strongly conserved among a wide array of species. It is therefore strongly suspected the human ASBT transporter to also entail nine transmembrane domains, with a similar conservation of the core domain, and significantly altered panel domain, especially if the formation of a dimeric state is possible.

Chapter 7

Conclusion and Future Outlooks

As was seen previously, both human ASBT and SLC10A7 transporters have shown to be implicated in the regulatory roles of complex physiological homeostatic processes, including the regulation of the enterohepatic circulation of BAs in the liver and intestine, as well as maintenance of calcium homeostasis. When levels of these regulations are significantly altered, severe clinical phenotypes arise, such as primary bile acid malabsorption (Oelkers et al., 1997) in ASBT inhibited individuals; and amelogenesis imperfecta with skeletal dysplasias (Dubail et al., 2018; Ashikov et al., 2018; Laugel-Haushalter et al., 2019) in SLC10A7 deficient individuals. Therefore, the crystallisation and subsequent structural characterisation of these transporters are imperative in understanding their molecular and functional associations. Once their mechanism of action is understood, tailored pharmaceutical interventions can be applied, in order to provide novel treatments for a multitude of severe physiological conditions. Therefore, the structural investigations performed on the bacterial homologues of these human transporters are crucial to advance the development for novel treatments.

In conclusion, this novel PhD research has outlined significant relevance and advancement to the human SLC10 protein family. Indeed, investigations of bacterial homologues of both ASBT and SLC10A7 transporters, have provided in-depth analysis of their overexpression, solubilisation, thermostability, purification, crystallisation and structural determination attributes; yielding novel progression in their functional and structural characterisations.

The ultimate aim of this research was to solve the 3D X-ray crystal structures for both projects, in order to better understand the molecular mechanical systems of the alternating access mechanism. Although, unfortunately, this was not achieved for the SLC10A7 transporter; amazingly, the structural architecture of the $ASBT_{Lb}$ bacterial homologue was successfully classified at 2.2 Å resolution. In addition, a wide array of functional features for both projects were also successfully generated, providing novel understandings of the SLC10 family.

An additional aim of the research was to identify novel solutes involved in the translocation across the cell membrane in both "orphan transporters". Although extremely challenging, we used thermostability assays to investigate a plethora of compounds and molecules, yielding interesting outcomes. The $ASBT_{Lb}$ bacterial homologue showed increased Tm measurements upon binding of BAs, although these do not represent the native substrate. However, the observation of a bound formate in the crystal structure could associate this molecule to be a native substrate of the transporter, although this remains speculative without additional radioactive uptake/transport investigations. For the SLC10A7 bacterial homologues, Tm measurements were significantly increased upon addition of sodium bicarbonate. This unexpected finding has to be investigated further in future, with radioactive studies, in order to assess whether this molecule can be classified as an SLC10A7 solute.

Future outlooks and progressions for the $ASBT_{Lb}$ project involve the generation of novel mutated constructs at conserved Na^+ binding sites, in order to investigate in greater detail the fundamentality of these aa residues and role they possess in the alternating access mechanism. In addition, radioactive transport studies should be conducted, to investigate further the involvement of BA binding/translocation, as well as the novel bound formate molecule in the 2.2 Å structure. With the generation of these additional experiments, a strong possibility of solute identification is possible.

For the SLC10A7 bacterial homologues, future outlooks and progressions target the continue identification of calcium-associated solutes; as well as radioactive transport analysis of the fascinating sodium bicarbonate findings. Thermostability, purification and crystallisation optimisations should be enhanced, with aim to advance the structural classification of the novel transporter in LCP/vapour diffusion crystallisation. Finally, if continued challenges are experienced with X-ray crystallography techniques, electron microscopy could additionally provide an alternative structural characterisation pathway for the novel classification of these bacterial transporters.

Bibliography

Takuya Abe, Yoshitomi Kanemitu, Masateru Nakasone, Ichiro Kawahata, Tohru Yamakuni, Akira Nakajima, Naoto Suzuki, Masazumi Nishikawa, Takanori Hishinuma, and Yoshihisa Tomioka. SLC10A4 is a protease-activated transporter that transports bile acids. *The Journal of Biochemistry*, 154(1): 93–101, July 2013. ISSN 0021-924X. doi: 10.1093/jb/mvt031. URL `https://academic.oup.com/jb/article/154/1/93/790177`. Publisher: Oxford Academic.

Paul D. Adams, Ralf W. Grosse-Kunstleve, Li Wei Hung, Thomas R. Ioerger, Airlie J. McCoy, Nigel W. Moriarty, Randy J. Read, James C. Sacchettini, Nicholas K. Sauter, and Thomas C. Terwilliger. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallographica. Section D, Biological Crystallography*, 58(Pt 11):1948–1954, November 2002. ISSN 0907-4449. doi: 10.1107/s0907444902016657.

Paul D. Adams, Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, Li-Wei Hung, Gary J. Kapral, Ralf W. Grosse-Kunstleve, Airlie J. McCoy, Nigel W. Moriarty, Robert Oeffner, Randy J. Read, David C. Richardson, Jane S. Richardson, Thomas C. Terwilliger, and Peter H. Zwart. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 2):213–221, February 2010. ISSN 1399-0047. doi: 10.1107/S0907444909052925.

Ivan A. Adzhubei, Steffen Schmidt, Leonid Peshkin, Vasily E. Ramensky, Anna Gerasimova, Peer Bork, Alexey S. Kondrashov, and Shamil R. Sunyaev. A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4):248–249, April 2010. ISSN 1548-7105. doi: 10.1038/nmeth0410-248. URL `https://www.nature.com/articles/nmeth0410-248`. Number: 4 Publisher: Nature Publishing Group.

Joerg Alber, Linghuo Jiang, and Joachim Geyer. CaRch1p does not functionally interact with the high-affinity Ca2+ influx system (HACS) of Candida albicans.

Yeast, 30(11):449–457, 2013. ISSN 1097-0061. doi: 10.1002/yea.2981. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/yea.2981. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/yea.2981.

Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular biology of the cell*. Garland Science, 5 edition, 2008.

R. Aldini, M. Montagnani, A. Roda, S. Hrelia, P. L. Biagi, and E. Roda. Intestinal absorption of bile acids in the rabbit: different transport rates in jejunum and ileum. *Gastroenterology*, 110(2):459–468, February 1996. ISSN 0016-5085. doi: 10.1053/gast.1996.v110.pm8566593.

Alexander I. Alexandrov, Mauro Mileni, Ellen Y. T. Chien, Michael A. Hanson, and Raymond C. Stevens. Microscale Fluorescent Thermal Stability Assay for Membrane Proteins. *Structure*, 16(3):351–359, March 2008. ISSN 0969-2126. doi: 10.1016/j.str.2008.02.004. URL http://www.sciencedirect.com/science/article/pii/S0969212608000609.

Yilmaz Alguel, Sotiris Amillis, James Leung, George Lambrinidis, Stefano Capaldi, Nicola J. Scull, Gregory Craven, So Iwata, Alan Armstrong, Emmanuel Mikros, George Diallinas, Alexander D. Cameron, and Bernadette Byrne. Structure of eukaryotic purine/H+ symporter UapA suggests a role for homodimerization in transport activity. *Nature Communications*, 7(1):11336, April 2016. ISSN 2041-1723. doi: 10.1038/ncomms11336. URL https://www.nature.com/articles/ncomms11336. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Carrier proteins;Molecular biophysics;Structural biology Subject_term_id: carrier-proteins;molecular-biophysics;structural-biology.

G Alpini, Ss Glaser, R Rodgers, Jl Phinizy, We Robertson, J Lasater, A Caligiuri, Z Tretjak, and Gd LeSage. Functional expression of the apical Na+-dependent bile acid transporter in large but not small rat cholangiocytes. *Gastroenterology*, 113(5):1734–1740, November 1997. ISSN 00165085. doi: 10.1053/gast.1997.v113.pm9352879. URL https://linkinghub.elsevier.com/retrieve/pii/S001650859700543X.

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, October 1990. ISSN 0022-2836. doi: 10.1016/S0022-2836(05)80360-2.

Bo Angelin, Kenneth S. Hershon, and John D. Brunzell. Bile Acid Metabolism in Hereditary Formes of Hypertriglyceridemia: Evidence for an Increased Synthesis Rate in Monogenic Familial Hypertriglyceridemia. *Proceedings of the National Academy of Sciences of the United States of America*, 84(15):5434–5438, 1987. ISSN 0027-8424. URL https://www.jstor.org/stable/30052. Publisher: National Academy of Sciences.

Takatoshi Arakawa, Takami Kobayashi-Yurugi, Yilmaz Alguel, Hiroko Iwanari, Hinako Hatae, Momi Iwata, Yoshito Abe, Tomoya Hino, Chiyo Ikeda-Suno, Hiroyuki Kuma, Dongchon Kang, Takeshi Murata, Takao Hamakubo, Alexander D. Cameron, Takuya Kobayashi, Naotaka Hamasaki, and So Iwata. Crystal structure of the anion exchanger domain of human erythrocyte band 3. *Science*, 350(6261):680–684, November 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa4335. URL https://science.sciencemag.org/content/350/6261/680. Publisher: American Association for the Advancement of Science Section: Report.

Valentina Arkhipova, Albert Guskov, and Dirk-Jan Slotboom. Analysis of the quality of crystallographic data and the limitations of structural models. *Journal of General Physiology*, 149(12):1091–1103, October 2017. ISSN 0022-1295. doi: 10.1085/jgp.201711852. URL https://doi.org/10.1085/jgp.201711852.

Angel Ashikov, Nurulamin Abu Bakar, Xiao-Yan Wen, Marco Niemeijer, Glentino Rodrigues Pinto Osorio, Koroboshka Brand-Arzamendi, Linda Hasadsri, Hana Hansikova, Kimiyo Raymond, Dorothée Vicogne, Nina Ondruskova, Marleen E. H. Simon, Rolph Pfundt, Sharita Timal, Roel Beumers, Christophe Biot, Roel Smeets, Marjan Kersten, Karin Huijben, CDG group, Peter T. A. Linders, Geert van den Bogaart, Sacha A. F. T. van Hijum, Richard Rodenburg, Lambertus P. van den Heuvel, Francjan van Spronsen, Tomas Honzik, Francois Foulquier, Monique van Scherpenzeel, Dirk J. Lefeber, CDG group, Wamelink Mirjam, Brunner Han, Mundy Helen, Michelakakis Helen, van Hasselt Peter, van de Kamp Jiddeke, Martinelli Diego, Morkrid Lars, Brocke Holmefjord Katja, Hertecant Jozef, Alfadhel Majid, Carpenter Kevin, and Te Water Naude Johann. Integrating glycomics and genomics uncovers SLC10A7 as essential factor for bone mineralization by regulating post-Golgi protein transport and glycosylation. *Human Molecular Genetics*, 27(17):3029–3045, 2018. ISSN 1460-2083. doi: 10.1093/hmg/ddy213.

F. C. Ayers, G. L. Warner, K. L. Smith, and D. A. Lawrence. Fluorometric quantitation of cellular and nonprotein thiols. *Analytical Biochemistry*, 154(1): 186–193, April 1986. ISSN 0003-2697. doi: 10.1016/0003-2697(86)90513-0.

Leonardo Baiocchi, Tianhao Zhou, Suthat Liangpunsakul, Ilaria Lenci, Francesco Santopaolo, Fanyin Meng, Lindsey Kennedy, Shannon Glaser, Heather Francis, and Gianfranco Alpini. Dual Role of Bile Acids on the Biliary Epithelium: Friend or Foe? *International Journal of Molecular Sciences*, 20(8), April 2019. ISSN 1422-0067. doi: 10.3390/ijms20081869. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6514722/.

Anand Balakrishnan and James E. Polli. Apical Sodium Dependent Bile Acid Transporter (ASBT, SLC10A2): A Potential Prodrug Target. *Molecular pharmaceutics*, 3(3):223–230, 2006. ISSN 1543-8384. doi: 10.1021/mp060022d. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2796132/.

Nazzareno Ballatori. Biology of a Novel Organic Solute and Steroid Transporter, OST-OST. *Experimental Biology and Medicine*, 230(10):689–698, November 2005. ISSN 1535-3702. doi: 10.1177/153537020523001001. URL https://doi.org/10.1177/153537020523001001. Publisher: SAGE Publications.

Nazzareno Ballatori, Whitney V. Christian, Jin Young Lee, Paul A. Dawson, Carol J. Soroka, James L. Boyer, Michael S. Madejczyk, and Na Li. OST-OST: A major basolateral bile acid and steroid transporter in human intestinal, renal, and biliary epithelia. *Hepatology*, 42(6):1270–1279, 2005. ISSN 1527-3350. doi: https://doi.org/10.1002/hep.20961. URL https://aasldpubs.onlinelibrary.wiley.com/doi/abs/10.1002/hep.20961. _eprint: https://aasldpubs.onlinelibrary.wiley.com/doi/pdf/10.1002/hep.20961.

Diane C. Bassham and Natasha V. Raikhel. Transport proteins in the plasma membrane and the secretory system. *Trends in Plant Science*, 1(1):15–20, January 1996. ISSN 1360-1385. doi: 10.1016/S1360-1385(96)80018-5. URL https://www.sciencedirect.com/science/article/pii/S1360138596800185.

Michele A. Battle, Benjamin J. Bondow, Moriah A. Iverson, Scott J. Adams, Ronald J. Jandacek, Patrick Tso, and Stephen A. Duncan. GATA4 Is Essential for Jejunal Function in Mice. *Gastroenterology*, 135(5):1676–1686.e1, November 2008. ISSN 00165085. doi: 10.1053/j.gastro.2008.07.074. URL https://linkinghub.elsevier.com/retrieve/pii/S0016508508014170.

Oliver Beckstein and Fiona Naughton. General Principles of Secondary Active Transporter Function. *arXiv:1912.06275 [q-bio]*, February 2021. URL http://arxiv.org/abs/1912.06275. arXiv: 1912.06275.

Jeremy M. Berg, John L. Tymoczko, and Lubert Stryer. *Biochemistry*. W. H. Freeman and Company Palgrave Macmillan, 7 edition, 2012.

H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, January 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.235.

Jorge Bernardino de la Serna, Gerhard J. Schütz, Christian Eggeling, and Marek Cebecauer. There Is No Simple Model of the Plasma Membrane Organization. *Frontiers in Cell and Developmental Biology*, 4: 106, 2016. ISSN 2296-634X. doi: 10.3389/fcell.2016.00106. URL https://www.frontiersin.org/article/10.3389/fcell.2016.00106.

Deborah M. Boes, Albert Godoy-Hernandez, and Duncan G. G. McMillan. Peripheral Membrane Proteins: Promising Therapeutic Targets across Domains of Life. *Membranes*, 11(5):346, May 2021. ISSN 2077-0375. doi: 10.3390/membranes11050346. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8151925/.

Tjalling Bosse, Christina M. Piaseckyj, Ellen Burghard, John J. Fialkovich, Satish Rajagopal, William T. Pu, and Stephen D. Krasinski. Gata4 Is Essential for the Maintenance of Jejunal-Ileal Identities in the Adult Mouse Small Intestine. *Molecular and Cellular Biology*, 26(23):9060–9070, December 2006. ISSN 0270-7306, 1098-5549. doi: 10.1128/MCB.00124-06. URL https://mcb.asm.org/content/26/23/9060.

Patrick D. Bosshart and Dimitrios Fotiadis. Secondary Active Transporters. *Sub-Cellular Biochemistry*, 92:275–299, 2019. ISSN 0306-0225. doi: 10.1007/978-3-030-18768-2_9.

Olga Boudker and Grégory Verdon. Structural perspectives on secondary active transporters. *Trends in pharmacological sciences*, 31(9):418–426, September 2010. ISSN 0165-6147. doi: 10.1016/j.tips.2010.06.004. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2933288/.

M. A. Brink, J. F. Slors, Y. C. Keulemans, K. S. Mok, D. R. De Waart, M. C. Carey, A. K. Groen, and G. N. Tytgat. Enterohepatic cycling of bilirubin: a putative mechanism for pigment gallstone formation in ileal Crohn's disease. *Gastroenterology*, 116(6):1420–1427, June 1999. ISSN 0016-5085. doi: 10.1016/s0016-5085(99)70507-x.

Ma Brink, N Mendez-Sanchez, and Mc Carey. Bilirubin cycles enterohepatically after ileal resection in the rat. *Gastroenterology*, 110(6):1945–1957, June 1996. ISSN 00165085. doi: 10.1053/gast.1996.v110.pm8964422. URL https://linkinghub.elsevier.com/retrieve/pii/S0016508596002855.

Robert Brommage, Jeff Liu, Gwenn M Hansen, Laura L Kirkpatrick, David G Potter, Arthur T Sands, Brian Zambrowicz, David R Powell, and Peter Vogel. High-throughput screening of mouse gene knockouts identifies established and novel skeletal phenotypes. *Bone Research*, 2(1):14034, December 2014. ISSN 2095-6231. doi: 10.1038/boneres.2014.34. URL http://www.nature.com/articles/boneres201434.

Martin Caffrey and Vadim Cherezov. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols*, 4(5):706–731, May 2009. ISSN 1754-2189, 1750-2799. doi: 10.1038/nprot.2009.31. URL http://www.nature.com/articles/nprot.2009.31.

M. Camilleri, A. Nadeau, W. J. Tremaine, J. Lamsam, D. Burton, S. Odunsi, S. Sweetser, and R. Singh. Measurement of serum 7-hydroxy-4-cholesten-3-one (or 7C4), a surrogate test for bile acid malabsorption in health, ileal disease and irritable bowel syndrome using liquid chromatography-tandem mass spectrometry. *Neurogastroenterology & Motility*, 21(7):734–e43, 2009. ISSN 1365-2982. doi: https://doi.org/10.1111/j.1365-2982.2009.01288.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2982.2009.01288.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2982.2009.01288.x.

Deepti Chaturvedi and Radhakrishnan Mahalakshmi. Transmembrane -barrels: Evolution, folding and energetics. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1859(12):2467–2482, December 2017. ISSN 0005-2736. doi: 10.1016/j.bbamem.2017.09.020. URL https://www.sciencedirect.com/science/article/pii/S0005273617303024.

Isaac Chen and Forshing Lui. Physiology, Active Transport. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2021. URL http://www.ncbi.nlm.nih.gov/books/NBK547718/.

Nicolas Chignard, Martine Mergey, Danielle Veissière, Rolland Parc, Jacqueline Capeau, Raoul Poupon, Annick Paul, and Chantal Housset. Bile acid transport and regulating functions in the human biliary epithelium. *Hepatology*, 33(3):496–503, 2001. ISSN 1527-3350. doi: https://doi.org/10.1053/jhep.2001.22345. URL https://aasldpubs.onlinelibrary.wiley.com/doi/abs/10.1053/jhep.2001.22345. _eprint: https://aasldpubs.onlinelibrary.wiley.com/doi/pdf/10.1053/jhep.2001.22345.

Paresh P. Chothe, Lindsay C. Czuba, Robyn H. Moore, and Peter W. Swaan. Human bile acid transporter ASBT (SLC10A2) forms functional non-covalent homodimers

and higher order oligomers. *Biochimica Et Biophysica Acta. Biomembranes*, 1860 (3):645–653, March 2018. ISSN 0005-2736. doi: 10.1016/j.bbamem.2017.11.016.

D. M. Christie, P. A. Dawson, S. Thevananther, and B. L. Shneider. Comparative analysis of the ontogeny of a sodium-dependent bile acid transporter in rat kidney and ileum. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 271(2):G377–G385, August 1996. ISSN 0193-1857. doi: 10.1152/ajpgi.1996.271.2.G377. URL https://journals.physiology.org/doi/abs/10.1152/ajpgi.1996.271.2.G377. Publisher: American Physiological Society.

Tatiana Claro da Silva, James E. Polli, and Peter W. Swaan. The solute carrier family 10 (SLC10): Beyond bile acid transport. *Molecular Aspects of Medicine*, 34(2):252–269, April 2013. ISSN 0098-2997. doi: 10.1016/j.mam.2012.07.004. URL http://www.sciencedirect.com/science/article/pii/S0098299712000878.

Zoe Cournia, Toby W. Allen, Ioan Andricioaei, Bruno Antonny, Daniel Baum, Grace Brannigan, Nicolae-Viorel Buchete, Jason T. Deckman, Lucie Delemotte, Coral del Val, Ran Friedman, Paraskevi Gkeka, Hans-Christian Hege, Jérôme Hénin, Marina A. Kasimova, Antonios Kolocouris, Michael L. Klein, Syma Khalid, M. Joanne Lemieux, Norbert Lindow, Mahua Roy, Jana Selent, Mounir Tarek, Florentina Tofoleanu, Stefano Vanni, Sinisa Urban, David J. Wales, Jeremy C. Smith, and Ana-Nicoleta Bondar. Membrane Protein Structure, Function and Dynamics: A Perspective from Experiments and Theory. *The Journal of membrane biology*, 248 (4):611–640, August 2015. ISSN 0022-2631. doi: 10.1007/s00232-015-9802-0. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4515176/.

Ann L. Craddock, Martha W. Love, Rebecca W. Daniel, Lyndon C. Kirby, Holly C. Walters, Melissa H. Wong, and Paul A. Dawson. Expression and transport properties of the human ileal and renal sodium-dependent bile acid transporter. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 274(1):G157–G169, January 1998. ISSN 0193-1857, 1522-1547. doi: 10.1152/ajpgi.1998.274.1.G157. URL https://www.physiology.org/doi/10.1152/ajpgi.1998.274.1.G157.

Thomas J. Crisman, Shaogang Qu, Baruch I. Kanner, and Lucy R. Forrest. Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. *Proceedings of the National Academy of Sciences*, 106(49):20752–20757, December 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0908570106. URL https://www.pnas.org/content/106/49/20752. Publisher: National Academy of Sciences Section: Biological Sciences.

Jiangjun Cui and Jaap A. Kaandorp. Mathematical modeling of calcium homeostasis in yeast cells. *Cell Calcium*, 39(4):337–348, April 2006. ISSN 0143-4160. doi: 10.1016/j.ceca.2005.12.001. URL https://www.sciencedirect.com/science/article/pii/S0143416005002411.

Florian Cymer, Gunnar von Heijne, and Stephen H. White. Mechanisms of integral membrane protein insertion and folding. *Journal of molecular biology*, 427(5): 999–1022, March 2015. ISSN 0022-2836. doi: 10.1016/j.jmb.2014.09.014. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4339636/.

Anthony P. Davis. Bile Acid Scaffolds in Supramolecular Chemistry: The Interplay of Design and Synthesis. *Molecules*, 12(9):2106–2122, September 2007. doi: 10.3390/12082106. URL https://www.mdpi.com/1420-3049/12/9/2106. Number: 9 Publisher: Molecular Diversity Preservation International.

Paul A. Dawson. Role of the intestinal bile acid transporters in bile acid and drug disposition. *Handbook of Experimental Pharmacology*, (201):169–203, 2011. ISSN 0171-2004. doi: 10.1007/978-3-642-14541-4_4.

Paul A. Dawson. Chapter 53 - Bile Formation and the Enterohepatic Circulation. In Leonard R. Johnson, Fayez K. Ghishan, Jonathan D. Kaunitz, Juanita L. Merchant, Hamid M. Said, and Jackie D. Wood, editors, *Physiology of the Gastrointestinal Tract (Fifth Edition)*, pages 1461–1484. Academic Press, Boston, January 2012. ISBN 978-0-12-382026-6. doi: 10.1016/B978-0-12-382026-6.00053-1. URL https://www.sciencedirect.com/science/article/pii/B9780123820266000531.

Paul A. Dawson, Melissa Hubbert, Jamie Haywood, Ann L. Craddock, Noa Zerangue, Whitney V. Christian, and Nazzareno Ballatori. The Heteromeric Organic Solute Transporter -, Ost-Ost, Is an Ileal Basolateral Bile Acid Transporter. *Journal of Biological Chemistry*, 280(8):6960–6968, February 2005. ISSN 00219258. doi: 10.1074/jbc.M412752200. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925819628139.

Paul A. Dawson, Tian Lan, and Anuradha Rao. Bile acid transporters. *Journal of Lipid Research*, 50(12):2340–2357, December 2009. ISSN 0022-2275. doi: 10.1194/jlr.R900012-JLR200. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2781307/.

J. M. Dietschy. Mechanisms for the intestinal absorption of bile acids. *Journal of Lipid Research*, 9(3):297–309, May 1968. ISSN 0022-2275.

J. M. Dietschy, S. D. Turley, and D. K. Spady. Role of liver in the maintenance of cholesterol and low density lipoprotein homeostasis in different animal species,

including humans. *Journal of Lipid Research*, 34(10):1637–1659, October 1993. ISSN 0022-2275.

John M. Dietschy and Stephen D. Turley. Control of Cholesterol Turnover in the Mouse. *Journal of Biological Chemistry*, 277(6):3801–3804, February 2002. ISSN 00219258. doi: 10.1074/jbc.R100057200. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925820874766.

A. J. Doig, C. D. Andrew, D. A. Cochran, E. Hughes, S. Penel, J. K. Sun, B. J. Stapley, D. T. Clarke, and G. R. Jones. Structure, stability and folding of the alpha-helix. *Biochemical Society Symposium*, (68):95–110, 2001. ISSN 0067-8694. doi: 10.1042/bss0680095.

Delfina C. Dominguez. Calcium signalling in bacteria. *Molecular Microbiology*, 54(2):291–297, 2004. ISSN 1365-2958. doi: https://doi.org/10.1111/j.1365-2958.2004.04276.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2004.04276.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2958.2004.04276.x.

David Drew. *GFP as a tool to monitor membrane protein topology and overexpression in Escherichia coli*. PhD thesis, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, 2005. ISBN: 9789171551603 OCLC: 185155189.

David Drew, Mirjam Lerch, Edmund Kunji, Dirk-Jan Slotboom, and Jan-Willem de Gier. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods*, 3(4):303–313, April 2006. ISSN 1548-7091. doi: 10.1038/nmeth0406-303.

David E Drew, Gunnar von Heijne, Pär Nordlund, and Jan-Willem L de Gier. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Letters*, 507(2):220–224, October 2001. ISSN 0014-5793. doi: 10.1016/S0014-5793(01)02980-5. URL https://www.sciencedirect.com/science/article/pii/S0014579301029805.

W. C. Duane, L. A. Hartich, A. E. Bartman, and S. B. Ho. Diminished gene expression of ileal apical sodium bile acid transporter explains impaired absorption of bile acid in patients with hypertriglyceridemia. *Journal of Lipid Research*, 41 (9):1384–1389, September 2000. ISSN 0022-2275.

Victor R. A. Dubach and Albert Guskov. The Resolution in X-ray Crystallography and Single-Particle Cryogenic Electron Microscopy. *Crystals*, 10(7):580, July 2020. doi: 10.3390/cryst10070580. URL

https://www.mdpi.com/2073-4352/10/7/580. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

Johanne Dubail, Céline Huber, Sandrine Chantepie, Stephan Sonntag, Beyhan Tüysüz, Ercan Mihci, Christopher T. Gordon, Elisabeth Steichen-Gersdorf, Jeanne Amiel, Banu Nur, Irene Stolte-Dijkstra, Albertien M. van Eerde, Koen L. van Gassen, Corstiaan C. Breugem, Alexander Stegmann, Caroline Lekszas, Reza Maroofian, Ehsan Ghayoor Karimiani, Arnaud Bruneel, Nathalie Seta, Arnold Munnich, Dulce Papy-Garcia, Muriel De La Dure-Molla, and Valérie Cormier-Daire. SLC10A7 mutations cause a skeletal dysplasia with amelogenesis imperfecta mediated by GAG biosynthesis defects. *Nature Communications*, 9(1): 1–15, August 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-05191-8. URL https://www.nature.com/articles/s41467-018-05191-8. Number: 1 Publisher: Nature Publishing Group.

Barbara Döring, Thomas Lütteke, Joachim Geyer, and Ernst Petzinger. Chapter Four - The SLC10 Carrier Family: Transport Functions and Molecular Structure. In Mark O. Bevensee, editor, *Current Topics in Membranes*, volume 70 of *Co-Transport Systems*, pages 105–168. Academic Press, January 2012. doi: 10.1016/B978-0-12-394316-3.00004-1. URL http://www.sciencedirect.com/science/article/pii/B9780123943163000041.

Robert C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, 2004. ISSN 0305-1048. doi: 10.1093/nar/gkh340. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/.

David Eisenberg. The discovery of the -helix and -sheet, the principal structural features of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 100(20):11207–11210, September 2003. ISSN 0027-8424. doi: 10.1073/pnas.2034522100. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC208735/.

P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of *Coot*. *Acta Crystallographica Section D Biological Crystallography*, 66 (4):486–501, April 2010. ISSN 0907-4449. doi: 10.1107/S0907444910007493. URL http://scripts.iucr.org/cgi-bin/paper?S0907444910007493.

Andreas Engel and Hermann E. Gaub. Structure and Mechanics of Membrane Proteins. *Annual Review of Biochemistry*, 77(1):127–148, 2008. doi: 10.1146/annurev.biochem.77.062706.154450. URL

https://doi.org/10.1146/annurev.biochem.77.062706.154450. _eprint: https://doi.org/10.1146/annurev.biochem.77.062706.154450.

Ekaitz Errasti-Murugarren, Paola Bartoccioni, and Manuel Palacín. Membrane Protein Stabilization Strategies for Structural and Functional Studies. *Membranes*, 11(2):155, February 2021. ISSN 2077-0375. doi: 10.3390/membranes11020155.

P. R. Evans and G. N. Murshudov. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69(7): 1204–1214, July 2013. ISSN 0907-4449. doi: 10.1107/S0907444913000061. URL //scripts.iucr.org/cgi-bin/paper?ba5190. Publisher: International Union of Crystallography.

James W. Fairman, Nicholas Noinaj, and Susan K. Buchanan. The structural biology of -barrel membrane proteins: a summary of recent reports. *Current opinion in structural biology*, 21(4):523–531, August 2011. ISSN 0959-440X. doi: 10.1016/j.sbi.2011.05.005. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3164749/.

Sunzhenhe Fang, Xiaowei Huang, Xue Zhang, Minhua Zhang, Yahui Hao, Hui Guo, Lu-Ning Liu, Fang Yu, and Peng Zhang. Molecular mechanism underlying transport and allosteric inhibition of bicarbonate transporter SbtA. *Proceedings of the National Academy of Sciences*, 118(22), June 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2101632118. URL https://www.pnas.org/content/118/22/e2101632118. Publisher: National Academy of Sciences Section: Biological Sciences.

Joseph Feher. 2.5 - Passive Transport and Facilitated Diffusion. In Joseph Feher, editor, *Quantitative Human Physiology*, pages 126–133. Academic Press, Boston, January 2012. ISBN 978-0-12-382163-8. doi: 10.1016/B978-0-12-382163-8.00015-3. URL https://www.sciencedirect.com/science/article/pii/B9780123821638000153.

Carla F. Fernandes, José R. Godoy, Barbara Döring, Marcia C. O. Cavalcanti, Martin Bergmann, Ernst Petzinger, and Joachim Geyer. The novel putative bile acid transporter SLC10A5 is highly expressed in liver and kidney. *Biochemical and Biophysical Research Communications*, 361(1):26–32, September 2007. ISSN 0006-291X. doi: 10.1016/j.bbrc.2007.06.160.

Lucy R. Forrest, Reinhard Krämer, and Christine Ziegler. The structural basis of secondary active transport mechanisms. *Biochimica*

et Biophysica Acta (BBA) - Bioenergetics, 1807(2):167–188, February 2011. ISSN 0005-2728. doi: 10.1016/j.bbabio.2010.10.014. URL https://www.sciencedirect.com/science/article/pii/S000527281000722X.

H. Fukuda, M. Arai, and K. Kuwajima. Folding of green fluorescent protein and the cycle3 mutant. *Biochemistry*, 39(39):12025–12032, October 2000. ISSN 0006-2960. doi: 10.1021/bi000543l.

J. Geyer, T. Wilke, and E. Petzinger. The solute carrier family SLC10: more than a family of bile acid transporters regarding function and phylogenetic relationships. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 372(6):413–431, March 2006. ISSN 0028-1298. doi: 10.1007/s00210-006-0043-8.

J. Geyer, C. F. Fernandes, B. Döring, S. Burger, J. R. Godoy, S. Rafalzik, T. Hübschle, R. Gerstberger, and E. Petzinger. Cloning and molecular characterization of the orphan carrier protein Slc10a4: expression in cholinergic neurons of the rat central nervous system. *Neuroscience*, 152(4):990–1005, April 2008. ISSN 0306-4522. doi: 10.1016/j.neuroscience.2008.01.049.

Joachim Geyer, José R. Godoy, and Ernst Petzinger. Identification of a sodium-dependent organic anion transporter from rat adrenal gland. *Biochemical and Biophysical Research Communications*, 316(2):300–306, April 2004. ISSN 0006-291X. doi: 10.1016/j.bbrc.2004.02.048. URL http://www.sciencedirect.com/science/article/pii/S0006291X0400292X.

Joachim Geyer, Barbara Döring, Kerstin Meerkamp, Bernhard Ugele, Nadiya Bakhiya, Carla F. Fernandes, José R. Godoy, Hansruedi Glatt, and Ernst Petzinger. Cloning and Functional Characterization of Human Sodium-dependent Organic Anion Transporter (SLC10A6). *Journal of Biological Chemistry*, 282(27):19728–19741, July 2007. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M702663200. URL http://www.jbc.org/lookup/doi/10.1074/jbc.M702663200.

Nikhil Ranjan Ghosh. Active Transport and Transfer of Energy. *Nature*, 200 (4902):175–176, October 1963. ISSN 1476-4687. doi: 10.1038/200175a0. URL https://www.nature.com/articles/200175a0. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4902 Primary_atype: Research Publisher: Nature Publishing Group.

José R. Godoy, Carla Fernandes, Barbara Döring, Knut Beuerlein, Ernst Petzinger, and Joachim Geyer. Molecular and phylogenetic characterization

of a novel putative membrane transporter (SLC10A7), conserved in vertebrates and bacteria. *European Journal of Cell Biology*, 86(8):445–460, August 2007. ISSN 0171-9335. doi: 10.1016/j.ejcb.2007.06.001. URL http://www.sciencedirect.com/science/article/pii/S0171933507000787.

Félix M. Goñi. The basic structure and dynamics of cell membranes: An update of the Singer–Nicolson model. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838(6):1467–1476, June 2014. ISSN 0005-2736. doi: 10.1016/j.bbamem.2014.01.006. URL https://www.sciencedirect.com/science/article/pii/S000527361400008X.

Béatrice Grenier and Rafik Ballou. Crystallography: Symmetry groups and group representations. In E. D. P. Sciences, editor, *Contribution of Symmetries in Condensed Matter*, EPJ Web of Conferences, page 00006. EPJ Web of Conferences, 2012. doi: 10.1051/epjconf/20122200006. URL https://hal.archives-ouvertes.fr/hal-00963925.

B. Hagenbuch and P. J. Meier. Molecular cloning, chromosomal localization, and functional characterization of a human liver Na+/bile acid cotransporter. *The Journal of Clinical Investigation*, 93(3):1326–1331, March 1994. ISSN 0021-9738. doi: 10.1172/JCI117091.

Bruno Hagenbuch and Paul Dawson. The sodium bile salt cotransport family SLC10. *Pflugers Archiv: European Journal of Physiology*, 447(5):566–570, February 2004. ISSN 0031-6768. doi: 10.1007/s00424-003-1130-z.

Vivek Halan, Sunit Maity, Anurag S. Rathore, and Ira Krull. Fluorescence-Dependent Size-Exclusion Chromatography - An Analytical Technique with Multiple Applications. *LCGC North America*, 32(12):922–926, December 2014. URL http://www.chromatographyonline.com/fluorescence-detection-size-exclusion-chroma

S. Hara, J. Higaki, K. Higashino, M. Iwai, N. Takasu, K. Miyata, K. Tonda, K. Nagata, Y. Goh, and T. Mizui. S-8921, an ileal Na+/bile acid cotransporter inhibitor decreases serum cholesterol in hamsters. *Life Sciences*, 60(24):PL 365–370, 1997. ISSN 0024-3205. doi: 10.1016/s0024-3205(97)00242-7.

Takeshi Harayama and Howard Riezman. Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology*, 19 (5):281–296, May 2018. ISSN 1471-0080. doi: 10.1038/nrm.2017.138. URL https://www.nature.com/articles/nrm.2017.138. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 5 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Membrane

biophysics;Membrane lipids;Membrane proteins;Membrane structure and assembly;Membranes Subject_term_id: membrane-biophysics;membrane-lipids;membrane-proteins;membrane-structure-and-assembly;membranes.

M. M. Harding. Geometry of metal-ligand interactions in proteins. *Acta Crystallographica. Section D, Biological Crystallography*, 57(Pt 3):401–411, March 2001. ISSN 0907-4449. doi: 10.1107/s0907444900019168.

Marjorie M. Harding. Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallographica. Section D, Biological Crystallography*, 58(Pt 5):872–874, May 2002. ISSN 0907-4449. doi: 10.1107/s0907444902003712.

Marjorie M. Harding. Small revisions to predicted distances around metal sites in proteins. *Acta Crystallographica. Section D, Biological Crystallography*, 62(Pt 6): 678–682, June 2006. ISSN 0907-4449. doi: 10.1107/S0907444906014594.

K. A. Hassan, I. T. Paulsen, L. D. H Elbourne, Q. Ren, A. D. Cameron, and P. J. F Henderson. Microbial Solute Transporters. In *Reference Module in Biomedical Sciences*. Elsevier, January 2014. ISBN 978-0-12-801238-3. doi: 10.1016/B978-0-12-801238-3.02464-8. URL `http://www.sciencedirect.com/science/article/pii/B9780128012383024648`.

Motoyuki Hattori, Ryan E. Hibbs, and Eric Gouaux. A Fluorescence-Detection Size-Exclusion Chromatography-Based Thermostability Assay for Membrane Protein Precrystallization Screening. *Structure*, 20(8):1293–1299, August 2012. ISSN 09692126. doi: 10.1016/j.str.2012.06.009. URL `https://linkinghub.elsevier.com/retrieve/pii/S0969212612002316`.

Matthias A. Hediger, Michael F. Romero, Ji-Bin Peng, Andreas Rolfs, Hitomi Takanaga, and Elspeth A. Bruford. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Archiv*, 447(5):465–468, February 2004. ISSN 1432-2013. doi: 10.1007/s00424-003-1192-y. URL `https://doi.org/10.1007/s00424-003-1192-y`.

G. Von Heijne. The membrane protein universe: what's out there and why bother? *Journal of Internal Medicine*, 261(6):543–557, 2007. ISSN 1365-2796. doi: 10.1111/j.1365-2796.2007.01792.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2796.2007.01792.x`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2796.2007.01792.x.

James E. Heubi, Roger D. Soloway, and William F. Balistreri. Biliary lipid composition in healthy and diseased infants, children, and young adults. *Gastroenterology*, 82(6):1295–1299, June 1982. ISSN 00165085. doi: 10.1016/0016-5085(82)90061-0. URL https://linkinghub.elsevier.com/retrieve/pii/0016508582900610.

J. Higaki, S. Hara, N. Takasu, K. Tonda, K. Miyata, T. Shike, K. Nagata, and T. Mizui. Inhibition of ileal Na+/bile acid cotransporter by S-8921 reduces serum cholesterol and prevents atherosclerosis in rabbits. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 18(8):1304–1311, August 1998. ISSN 1079-5642. doi: 10.1161/01.atv.18.8.1304.

Richard H. Ho, Brenda F. Leake, Richard L. Roberts, Wooin Lee, and Richard B. Kim. Ethnicity-dependent Polymorphism in Na+-taurocholate Cotransporting Polypeptide (SLC10A1) Reveals a Domain Critical for Bile Acid Substrate Recognition. *Journal of Biological Chemistry*, 279(8):7213–7222, February 2004. ISSN 00219258. doi: 10.1074/jbc.M305782200. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925818446868.

Richard H. Ho, Rommel G. Tirona, Brenda F. Leake, Hartmut Glaeser, Wooin Lee, Christopher J. Lemke, Yi Wang, and Richard B. Kim. Drug and Bile Acid Transporters in Rosuvastatin Hepatic Uptake: Function, Expression, and Pharmacogenetics. *Gastroenterology*, 130(6):1793–1806, May 2006. ISSN 00165085. doi: 10.1053/j.gastro.2006.02.034. URL https://linkinghub.elsevier.com/retrieve/pii/S0016508506003908.

A. F. Hofmann and L. R. Hagey. Bile Acids: Chemistry, Pathochemistry, Biology, Pathobiology, and Therapeutics. *Cellular and Molecular Life Sciences*, 65(16):2461–2483, August 2008. ISSN 1420-682X, 1420-9071. doi: 10.1007/s00018-008-7568-6. URL http://link.springer.com/10.1007/s00018-008-7568-6.

A F Hofmann, G Molino, M Milanese, and G Belforte. Description and simulation of a physiological pharmacokinetic model for the metabolism and enterohepatic circulation of bile acids in man. Cholic acid in healthy man. *Journal of Clinical Investigation*, 71(4):1003–1022, April 1983. ISSN 0021-9738. doi: 10.1172/JCI110828. URL http://www.jci.org/articles/view/110828.

Alan F. Hofmann. Bile acids: Trying to understand their chemistry and biology with the hope of helping patients. *Hepatology*, 49(5):1403–1418, 2009. ISSN 1527-3350. doi: 10.1002/hep.22789. URL https://aasldpubs.onlinelibrary.wiley.com/doi/abs/10.1002/hep.22789. _eprint: https://aasldpubs.onlinelibrary.wiley.com/doi/pdf/10.1002/hep.22789.

Arthur Holzer, Simone Harsch, Olga Renner, André Strohmeyer, Silke Schimmel, Jan Wehkamp, Peter Fritz, and Eduard F. Stange. Diminished Expression of Apical Sodium-Dependent Bile Acid Transporter in Gallstone Disease Is Independent of Ileal Inflammation. *Digestion*, 78(1):52–59, 2008. ISSN 0012-2823, 1421-9867. doi: 10.1159/000159379. URL https://www.karger.com/Article/FullText/159379. Publisher: Karger Publishers.

Nien-Jen Hu, So Iwata, Alexander D. Cameron, and David Drew. Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. *Nature*, 478(7369):408–411, October 2011. ISSN 1476-4687. doi: 10.1038/nature10450. URL https://www.nature.com/articles/nature10450. Number: 7369 Publisher: Nature Publishing Group.

C. V. Hulzebos, L. Renfurm, R. H. Bandsma, H. J. Verkade, T. Boer, R. Boverhof, H. Tanaka, I. Mierau, P. J. Sauer, F. Kuipers, and F. Stellaard. Measurement of parameters of cholic acid kinetics in plasma using a microscale stable isotope dilution technique: application to rodents and humans. *Journal of Lipid Research*, 42(11):1923–1929, November 2001. ISSN 0022-2275.

Oleg Jardetzky. Simple Allosteric Model for Membrane Pumps. *Nature*, 211 (5052):969–970, August 1966. ISSN 1476-4687. doi: 10.1038/211969a0. URL https://www.nature.com/articles/211969a0. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 5052 Primary_atype: Research Publisher: Nature Publishing Group.

Linghuo Jiang, Joerg Alber, Jihong Wang, Wei Du, Xuexue Yang, Xichuan Li, Dominique Sanglard, and Joachim Geyer. The Candida albicans plasma membrane protein Rch1p, a member of the vertebrate SLC10 carrier family, is a novel regulator of cytosolic Ca2+ homoeostasis. *Biochemical Journal*, 444(3):497–502, June 2012. ISSN 0264-6021. doi: 10.1042/BJ20112166. URL /biochemj/article/444/3/497/46320/The-Candida-albicans-plasma-membrane-protein- Publisher: Portland Press.

Emre Karakus, Marie Wannowius, Simon Franz Müller, Silke Leiting, Regina Leidolf, Saskia Noppes, Stefan Oswald, Martin Diener, and Joachim Geyer. The orphan solute carrier SLC10A7 is a novel negative regulator of intracellular calcium signaling. *Scientific Reports*, 10(1):7248, April 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-64006-3.

P Andrew Karplus and Kay Diederichs. Assessing and maximizing data quality in macromolecular crystallography. *Current opinion in structural biology*, 34:

60–68, October 2015. ISSN 0959-440X. doi: 10.1016/j.sbi.2015.07.003. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4684713/.

Toshimitsu Kawate and Eric Gouaux. Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins. *Structure*, 14(4):673–681, April 2006. ISSN 09692126. doi: 10.1016/j.str.2006.01.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0969212606001092.

Ken Kitayama, Daisuke Nakai, Keita Kono, Arthur Gerritsen van der Hoop, Hitoshi Kurata, Elly C. de Wit, Louis H. Cohen, Toshimori Inaba, and Takafumi Kohama. Novel non-systemic inhibitor of ileal apical Na+-dependent bile acid transporter reduces serum cholesterol levels in hamsters and monkeys. *European Journal of Pharmacology*, 539(1):89–98, June 2006. ISSN 0014-2999. doi: 10.1016/j.ejphar.2006.04.005. URL https://www.sciencedirect.com/science/article/pii/S0014299906003979.

Edward D. Korn. Structure and Function of the Plasma Membrane. *The Journal of General Physiology*, 52(1):257–278, July 1968. ISSN 0022-1295. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2225798/.

Einar Krag and Bent Krag. Regional Ileitis (Crohn's Disease). *Scandinavian Journal of Gastroenterology*, 11(5):481–486, August 1976. ISSN 0036-5521. doi: 10.1080/00365521.1976.12097137. URL https://doi.org/10.1080/00365521.1976.12097137. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00365521.1976.12097137.

W. Kramer, F. Girbig, U. Gutjahr, S. Kowalewski, K. Jouvenal, G. Müller, D. Tripier, and G. Wess. Intestinal bile acid absorption. Na(+)-dependent bile acid transport activity in rabbit small intestine correlates with the coexpression of an integral 93-kDa and a peripheral 14-kDa bile acid-binding membrane protein along the duodenum-ileum axis. *Journal of Biological Chemistry*, 268(24):18035–18046, August 1993. ISSN 00219258. doi: 10.1016/S0021-9258(17)46808-6. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925817468086.

W. Kramer, S. Stengelin, K. H. Baringhaus, A. Enhsen, H. Heuer, W. Becker, D. Corsiero, F. Girbig, R. Noll, and C. Weyland. Substrate specificity of the ileal and the hepatic Na(+)/bile acid cotransporters of the rabbit. I. Transport studies with membrane vesicles and cell lines expressing the cloned transporters. *Journal of Lipid Research*, 40(9):1604–1617, September 1999. ISSN 0022-2275.

Werner Kramer and Heiner Glombik. Bile acid reabsorption inhibitors (BARI): novel hypolipidemic drugs. *Current Medicinal Chemistry*, 13(9):997–1016, 2006. ISSN 0929-8673. doi: 10.2174/092986706776361003.

Werner Kramer, Günther Wess, Ulrike Bewersdorf, Daniel Corsiero, Frank Girbig, Claudia Weyland, Siegfried Stengelin, Alfons Enhsen, Klaus Bock, Horst Kleine, Marie-Anne Le Dreau, and Hans-Ludwig Schäfer. Topological Photoaffinity Labeling of the Rabbit Ileal Na+/Bile-Salt-Cotransport System. *European Journal of Biochemistry*, 249(2):456–464, 1997. ISSN 1432-1033. doi: https://doi.org/10.1111/j.1432-1033.1997.00456.x. URL https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1432-1033.1997.00456.x. _eprint: https://febs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1432-1033.1997.00456.x.

Werner Kramer, Frank Girbig, Heiner Glombik, Daniel Corsiero, Siegfried Stengelin, and Claudia Weyland. Identification of a Ligand-binding Site in the Na+/Bile Acid Cotransporting Protein from Rabbit Ileum. *Journal of Biological Chemistry*, 276(38):36020–36027, September 2001. ISSN 00219258. doi: 10.1074/jbc.M104665200. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925819588767.

Diego Krapf. Compartmentalization of the plasma membrane. *Current Opinion in Cell Biology*, 53:15–21, August 2018. ISSN 1879-0410. doi: 10.1016/j.ceb.2018.04.002.

Harini Krishnamurthy and Eric Gouaux. X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature*, 481(7382): 469–474, January 2012. ISSN 1476-4687. doi: 10.1038/nature10737. URL https://www.nature.com/articles/nature10737. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7382 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Molecular conformation;Neurotransmitters;Structural biology;Transporters in the nervous system Subject_term_id: molecular-conformation;neurotransmitters;structural-biology;transporters-in-the-nervous-system.

Harini Krishnamurthy, Chayne L. Piscitelli, and Eric Gouaux. Unlocking the molecular secrets of sodium-coupled transporters. *Nature*, 459(7245): 347–355, May 2009. ISSN 1476-4687. doi: 10.1038/nature08143. URL https://www.nature.com/articles/nature08143. Number: 7245 Publisher: Nature Publishing Group.

A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580, January 2001. ISSN 0022-2836. doi: 10.1006/jmbi.2000.4315.

K. Kuhajda, J. Kandrac, S. Kevresan, M. Mikov, and J. P. Fawcett. Structure and origin of bile acids: An overview. *European Journal of Drug Metabolism and Pharmacokinetics*, 31(3):135–143, September 2006. ISSN 0378-7966, 2107-0180. doi: 10.1007/BF03190710. URL `http://link.springer.com/10.1007/BF03190710`.

G. A. Kullak-Ublick, B. Stieger, B. Hagenbuch, and P. J. Meier. Hepatic transport of bile salts. *Seminars in Liver Disease*, 20(3):273–292, 2000. ISSN 0272-8087. doi: 10.1055/s-2000-9426.

Gerd A. Kullak-Ublick, Bruno Stieger, and Peter J. Meier. Enterohepatic bile salt transporters in normal physiology and liver disease. *Gastroenterology*, 126(1): 322–342, January 2004. ISSN 0016-5085. doi: 10.1053/j.gastro.2003.06.005.

Leon Lack. Properties and Biological Significance of the leal Bile Salt Transport System. *Environmental Health Perspectives*, page 11, 1979.

Arthur Laganowsky, Eamonn Reading, Timothy M. Allison, Martin B. Ulmschneider, Matteo T. Degiacomi, Andrew J. Baldwin, and Carol V. Robinson. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature*, 510(7503):172–175, June 2014. ISSN 1476-4687. doi: 10.1038/nature13419. URL `https://www.nature.com/articles/nature13419`. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7503 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Lipids;Membrane proteins;X-ray crystallography Subject_term_id: lipids;membrane-proteins;x-ray-crystallography.

M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21):2947–2948, November 2007. ISSN 1367-4811. doi: 10.1093/bioinformatics/btm404.

Virginie Laugel-Haushalter, Séverine Bär, Elise Schaefer, Corinne Stoetzel, Véronique Geoffroy, Yves Alembik, Naji Kharouf, Mathilde Huckert, Pauline Hamm, Joseph Hemmerlé, Marie-Cécile Manière, Sylvie Friant, Hélène Dollfus, and Agnès Bloch-Zupan. A New SLC10A7 Homozygous Missense Mutation Responsible for a Milder Phenotype of Skeletal Dysplasia With Amelogenesis Imperfecta. *Frontiers in Genetics*, 10, 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.00504. URL `https://www.frontiersin.org/articles/10.3389/fgene.2019.00504/full`. Publisher: Frontiers.

K N Lazaridis, L Pham, P Tietz, R A Marinelli, P C deGroen, S Levine, P A Dawson, and N F LaRusso. Rat cholangiocytes absorb bile acids at their apical domain via

the ileal sodium-dependent bile acid transporter. *Journal of Clinical Investigation*, 100(11):2714–2721, December 1997. ISSN 0021-9738. doi: 10.1172/JCI119816. URL http://www.jci.org/articles/view/119816.

Chiara Lee, Hae Joo Kang, Christoph von Ballmoos, Simon Newstead, Povilas Uzdavinys, David L. Dotson, So Iwata, Oliver Beckstein, Alexander D. Cameron, and David Drew. A two-domain elevator mechanism for sodium/proton antiport. *Nature*, 501(7468):573–577, September 2013. ISSN 1476-4687. doi: 10.1038/nature12484. URL https://www.nature.com/articles/nature12484. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7468 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Membrane proteins;Permeation and transport;X-ray crystallography Subject_term_id: membrane-proteins;permeation-and-transport;x-ray-crystallography.

Chiara Lee, Hae Joo Kang, Anna Hjelm, Abdul Aziz Qureshi, Emmanuel Nji, Hassanul Choudhury, Konstantinos Beis, Jan-Willem de Gier, and David Drew. MemStar: A one-shot Escherichia coli-based approach for high-level bacterial membrane protein production. *FEBS Letters*, 588(20): 3761–3769, 2014a. ISSN 1873-3468. doi: 10.1016/j.febslet.2014.08.025. URL https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/j.febslet.2014.08.025. _eprint: https://febs.onlinelibrary.wiley.com/doi/pdf/10.1016/j.febslet.2014.08.025.

Chiara Lee, Shoko Yashiro, David L. Dotson, Povilas Uzdavinys, So Iwata, Mark S.P. Sansom, Christoph von Ballmoos, Oliver Beckstein, David Drew, and Alexander D. Cameron. Crystal structure of the sodium–proton antiporter NhaA dimer and new mechanistic insights. *Journal of General Physiology*, 144(6): 529–544, November 2014b. ISSN 0022-1295. doi: 10.1085/jgp.201411219. URL https://doi.org/10.1085/jgp.201411219.

M. C. Lewis and C. Root. In vivo transport kinetics and distribution of taurocholate by rat ileum and jejunum. *The American Journal of Physiology*, 259(2 Pt 1): G233–238, August 1990. ISSN 0002-9513. doi: 10.1152/ajpgi.1990.259.2.G233.

M. C. Lewis, L. E. Brieaddy, and C. Root. Effects of 2164U90 on ileal bile acid absorption and serum cholesterol in rats and mice. *Journal of Lipid Research*, 36 (5):1098–1105, May 1995. ISSN 0022-2275.

Dianfan Li and Martin Caffrey. Structure and Functional Characterization of Membrane Integral Proteins in the Lipid Cubic Phase. *Journal of Molecular Biology*, 432(18):5104–5123, August 2020. ISSN 0022-2836. doi: 10.1016/j.jmb.2020.02.024. URL https://www.sciencedirect.com/science/article/pii/S0022283620302011.

M.C. Lin, W. Kramer, and F.A. Wilson. Identification of cytosolic and microsomal bile acid-binding proteins in rat ileal enterocytes. *Journal of Biological Chemistry*, 265(25):14986–14995, September 1990. ISSN 00219258. doi: 10.1016/S0021-9258(18)77213-X. URL https://linkinghub.elsevier.com/retrieve/pii/S002192581877213X.

Stina Lindman, Wei-Feng Xue, Olga Szczepankiewicz, Mikael C. Bauer, Hanna Nilsson, and Sara Linse. Salting the Charged Surface: pH and Salt Dependence of Protein G B1 Stability. *Biophysical Journal*, 90(8):2911–2921, April 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.071050. URL https://www.sciencedirect.com/science/article/pii/S0006349506724717.

Jingwei Liu, Hao Li, Shixuan Shen, Liping Sun, Yuan Yuan, and Chengzhong Xing. Alternative splicing events implicated in carcinogenesis and prognosis of colorectal cancer. *Journal of Cancer*, 9(10):1754–1764, 2018. ISSN 1837-9664. doi: 10.7150/jca.24569.

J. H. Lorent, K. R. Levental, L. Ganesan, G. Rivera-Longsworth, E. Sezgin, M. Doktorova, E. Lyman, and I. Levental. Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nature Chemical Biology*, 16(6):644–652, June 2020. ISSN 1552-4469. doi: 10.1038/s41589-020-0529-6. URL https://www.nature.com/articles/s41589-020-0529-6. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Biophysical chemistry;Lipidomics;Membrane trafficking;Membranes Subject_term_id: biophysical-chemistry;lipidomics;membrane-trafficking;membranes.

Dmitry A. Los and Norio Murata. Membrane fluidity and its roles in the perception of environmental signals. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1666(1):142–157, November 2004. ISSN 0005-2736. doi: 10.1016/j.bbamem.2004.08.002. URL https://www.sciencedirect.com/science/article/pii/S0005273604002032.

Feiran Lu, Shuo Li, Yang Jiang, Jing Jiang, He Fan, Guifeng Lu, Dong Deng, Shangyu Dang, Xu Zhang, Jiawei Wang, and Nieng Yan. Structure and mechanism of the uracil transporter UraA. *Nature*, 472(7342):243–246, April 2011. ISSN 1476-4687. doi: 10.1038/nature09885. URL https://www.nature.com/articles/nature09885. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7342 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Carrier proteins;Structural biology Subject_term_id: carrier-proteins;structural-biology.

S. N. Marcus, C. D. Schteingart, M. L. Marquez, A. F. Hofmann, Y. Xia, J. H. Steinbach, H. T. Ton-Nu, J. Lillienau, M. A. Angellotti, and A. Schmassmann. Active absorption of conjugated bile acids in vivo. Kinetic parameters and molecular specificity of the ileal transport system in the rat. *Gastroenterology*, 100(1): 212–221, January 1991. ISSN 0016-5085. doi: 10.1016/0016-5085(91)90603-i.

D. Marsh. Lateral pressure in membranes. *Biochimica Et Biophysica Acta*, 1286(3): 183–223, October 1996. ISSN 0006-3002. doi: 10.1016/s0304-4157(96)00009-3.

A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4):658–674, August 2007. ISSN 0021-8898. doi: 10.1107/S0021889807021206. URL `http://scripts.iucr.org/cgi-bin/paper?he5368`. Number: 4 Publisher: International Union of Crystallography.

Walter E. Meihoff and Fred Kern. Bile salt malabsorption in regional ileitis, ileal resection, and mannitol-induced diarrhea. *Journal of Clinical Investigation*, 47 (2):261–267, February 1968. ISSN 0021-9738. doi: 10.1172/JCI105722. URL `http://www.jci.org/articles/view/105722`.

Erica J. Melief, Jeffrey T. Gibbs, Xianwu Li, R. Garrett Morgan, C. Dirk Keene, Thomas J. Montine, Richard D. Palmiter, and Martin Darvas. Characterization of cognitive impairments and neurotransmitter changes in a novel transgenic mouse lacking Slc10a4. *Neuroscience*, 324:399–406, June 2016. ISSN 0306-4522. doi: 10.1016/j.neuroscience.2016.03.037. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4838488/`.

Hartmut Michel. Crystallization of membrane proteins. *Trends in Biochemical Sciences*, 8(2):56–59, February 1983. ISSN 0968-0004. doi: 10.1016/0968-0004(83)90390-0. URL `https://www.sciencedirect.com/science/article/pii/0968000483903900`.

Gurdeep S Minhas, Daniel Bawdon, Reyme Herman, Michelle Rudden, Andrew P Stone, A Gordon James, Gavin H Thomas, and Simon Newstead. Structural basis of malodour precursor transport in the human axilla. *eLife*, 7:e34995, July 2018. ISSN 2050-084X. doi: 10.7554/eLife.34995. URL `https://doi.org/10.7554/eLife.34995`. Publisher: eLife Sciences Publications, Ltd.

E M Muller, E G Locke, and K W Cunningham. Differential regulation of two Ca(2+) influx systems by pheromone signaling in Saccharomyces cerevisiae. *Genetics*, 159(4):1527–1538, December 2001. ISSN 0016-6731. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1461924/`.

G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, and A. A. Vagin. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):355–367, April 2011. ISSN 0907-4449. doi: 10.1107/S0907444911001314. URL http://scripts.iucr.org/cgi-bin/paper?ba5152. Number: 4 Publisher: International Union of Crystallography.

Emmanuel Nji, Yurie Chatzikyriakidou, Michael Landreh, and David Drew. An engineered thermal-shift screen reveals specific lipid preferences of eukaryotic and prokaryotic membrane proteins. *Nature Communications*, 9(1):1–12, October 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06702-3. URL https://www.nature.com/articles/s41467-018-06702-3. Number: 1 Publisher: Nature Publishing Group.

Saskia Noppes, Simon Franz Müller, Josefine Bennien, Matthias Holtemeyer, Massimo Palatini, Regina Leidolf, Jörg Alber, and Joachim Geyer. Homo- and heterodimerization is a common feature of the solute carrier family SLC10 members. *Biological Chemistry*, 400(10):1371–1384, October 2019. ISSN 1437-4315, 1431-6730. doi: 10.1515/hsz-2019-0148. URL https://www.degruyter.com/view/journals/bchm/400/10/article-p1371.xml. Publisher: De Gruyter Section: Biological Chemistry.

H Nyhlin, M V Merrick, and M A Eastwood. Bile acid malabsorption in Crohn's disease and indications for its assessment using SeHCAT. *Gut*, 35 (1):90–93, January 1994. ISSN 0017-5749. doi: 10.1136/gut.35.1.90. URL https://gut.bmj.com/lookup/doi/10.1136/gut.35.1.90.

Takayuki Odahara. Stability and solubility of integral membrane proteins from photosynthetic bacteria solubilized in different detergents. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1660(1-2):80–92, January 2004. ISSN 00052736. doi: 10.1016/j.bbamem.2003.11.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0005273603003523.

P Oelkers, L C Kirby, J E Heubi, and P A Dawson. Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2). *Journal of Clinical Investigation*, 99(8):1880–1887, April 1997. ISSN 0021-9738. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC508012/.

Peter Oelkers and Paul A. Dawson. Cloning and chromosomal localization of the human ileal lipid-binding protein. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, 1257(2):199–202, July

1995. ISSN 0005-2760. doi: 10.1016/0005-2760(95)00098-W. URL https://www.sciencedirect.com/science/article/pii/000527609500098W.

C. Ostermeier and H. Michel. Crystallization of membrane proteins. *Current Opinion in Structural Biology*, 7(5):697–701, October 1997. ISSN 0959-440X. doi: 10.1016/s0959-440x(97)80080-2.

Arthur B. Pardee. Biochemical Studies on Active Transport. *The Journal of General Physiology*, 52(1):279–295, July 1968. ISSN 0022-1295. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2225799/.

Joanne L. Parker and Simon Newstead. Membrane Protein Crystallisation: Current Trends and Future Perspectives. *Advances in Experimental Medicine and Biology*, 922:61–72, April 2016. ISSN 0065-2598. doi: 10.1007/978-3-319-35072-1_5.URLhttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5033070/.

Russell Paterson. Transport of Ions against their Concentration Gradients in an Analogue System for Active Transport using Charged Membranes. *Nature*, 217 (5128):545–547, February 1968. ISSN 1476-4687. doi: 10.1038/217545a0. URL https://www.nature.com/articles/217545a0. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 5128 Primary_atype: Research Publisher: Nature Publishing Group.

Emelie Perland and Robert Fredriksson. Classification Systems of Secondary Active Transporters. *Trends in Pharmacological Sciences*, 38(3):305–315, March 2017. ISSN 0165-6147. doi: 10.1016/j.tips.2016.11.008. URL https://www.cell.com/trends/pharmacological-sciences/abstract/S0165-6147(16)301 Publisher: Elsevier.

Svetlana N. Popova and Irina Alafuzoff. Distribution of SLC10A4, a Synaptic Vesicle Protein in the Human Brain, and the Association of this Protein with Alzheimer's Disease-Related Neuronal Degeneration. *Journal of Alzheimer's Disease*, 37 (3):603–610, January 2013. ISSN 1387-2877. doi: 10.3233/JAD-130548. URL https://content.iospress.com/articles/journal-of-alzheimers-disease/jad130548. Publisher: IOS Press.

Matthias Quick and Jonathan A. Javitch. Monitoring the function of membrane transport proteins in detergent-solubilized form. *Proceedings of the National Academy of Sciences*, 104(9):3603–3608, February 2007. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0609573104. URL https://www.pnas.org/content/104/9/3603. Publisher: National Academy of Sciences Section: Biological Sciences.

Gale Rhodes. Quality of Macromolecular Models. URL https://spdbv.vital-it.ch/TheMolecularLevel/ModQual/Redundancy.

Carolyn Root, Chari D. Smith, Scott S. Sundseth, Heather M. Pink, Joan G. Wilson, and Michael C. Lewis. Ileal bile acid transporter inhibition, CYP7A1 induction, and antilipemic action of 264W94. *Journal of Lipid Research*, 43(8):1320–1330, August 2002. ISSN 0022-2275.

J. E. Rothman and J. Lenard. Membrane asymmetry. *Science (New York, N.Y.)*, 195(4280):743–753, February 1977. ISSN 0036-8075. doi: 10.1126/science.402030.

Aiman Sadaf, Kyung Ho Cho, Bernadette Byrne, and Pil Seok Chae. Chapter Four - Amphipathic Agents for Membrane Protein Study. In Arun K. Shukla, editor, *Methods in Enzymology*, volume 557 of *Membrane Proteins—Engineering, Purification and Crystallization*, pages 57–94. Academic Press, January 2015. doi: 10.1016/bs.mie.2014.12.021. URL https://www.sciencedirect.com/science/article/pii/S0076687914001359.

Lawrence R. Schiller, Reed B. Hogan, Stephen G. Morawski, Carol A. Santa Ana, M.Jon Bern, Richard P. Norgaard, George W. Bo-Linn, and John S. Fordtran. Studies of the prevalence and significance of radiolabeled bile acid malabsorption in a group of patients with idiopathic chronic diarrhea. *Gastroenterology*, 92(1): 151–160, January 1987. ISSN 00165085. doi: 10.1016/0016-5085(87)90852-3. URL https://linkinghub.elsevier.com/retrieve/pii/0016508587908523.

Susan Schlegel, John Löfblom, Chiara Lee, Anna Hjelm, Mirjam Klepsch, Marc Strous, David Drew, Dirk Jan Slotboom, and Jan-Willem de Gier. Optimizing Membrane Protein Overexpression in the Escherichia coli strain Lemo21(DE3). *Journal of Molecular Biology*, 423(4):648–659, November 2012. ISSN 0022-2836. doi: 10.1016/j.jmb.2012.07.019. URL http://www.sciencedirect.com/science/article/pii/S0022283612005864.

Stephanie Schmidt, Marcela Moncada, Simone Burger, and Joachim Geyer. Expression, sorting and transport studies for the orphan carrier SLC10A4 in neuronal and non-neuronal cell lines and in Xenopus laevis oocytes. *BMC neuroscience*, 16: 35, June 2015. ISSN 1471-2202. doi: 10.1186/s12868-015-0174-2.

LLC Schrödinger. The PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC. 2015.

Yigong Shi. Common Folds and Transport Mechanisms of Secondary Active Transporters. *Annual Review of Biophysics*, 42(1): 51–72, 2013. doi: 10.1146/annurev-biophys-083012-130429. URL

https://doi.org/10.1146/annurev-biophys-083012-130429. _eprint: https://doi.org/10.1146/annurev-biophys-083012-130429.

J. C. Skou. Enzymatic Basis for Active Transport of Na+ and K+ Across Cell Membrane. *Physiological Reviews*, 45(3):596–618, July 1965. ISSN 0031-9333. doi: 10.1152/physrev.1965.45.3.596. URL https://journals.physiology.org/doi/abs/10.1152/physrev.1965.45.3.596. Publisher: American Physiological Society.

Davor Slijepcevic and Stan F.J. van de Graaf. Bile Acid Uptake Transporters as Targets for Therapy. *Digestive Diseases (Basel, Switzerland)*, 35 (3):251–258, March 2017. ISSN 0257-2753. doi: 10.1159/000450983. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5516419/.

Oleg V. Sobolev, Pavel V. Afonine, Nigel W. Moriarty, Maarten L. Hekkelman, Robbie P. Joosten, Anastassis Perrakis, and Paul D. Adams. A Global Ramachandran Score Identifies Protein Structures with Unlikely Stereochemistry. *Structure*, 28(11):1249–1258.e2, November 2020. ISSN 0969-2126. doi: 10.1016/j.str.2020.08.005. URL https://www.cell.com/structure/abstract/S0969-2126(20)30287-2. Publisher: Elsevier.

Wilfred D. Stein and Thomas Litman. Chapter 6 - Primary Active Transport Systems. In Wilfred D. Stein and Thomas Litman, editors, *Channels, Carriers, and Pumps (Second Edition)*, pages 247–328. Elsevier, London, January 2015. ISBN 978-0-12-416579-3. doi: 10.1016/B978-0-12-416579-3.00006-X. URL https://www.sciencedirect.com/science/article/pii/B978012416579300006X.

William Stillwell. Membrane Transport. *An Introduction to Biological Membranes*, pages 423–451, 2016. doi: 10.1016/B978-0-444-63772-7.00019-1. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7182109/.

F. William Studier. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1):207–234, May 2005. ISSN 1046-5928. doi: 10.1016/j.pep.2005.01.016. URL http://www.sciencedirect.com/science/article/pii/S1046592805000264.

Linfeng Sun, Xin Zeng, Chuangye Yan, Xiuyun Sun, Xinqi Gong, Yu Rao, and Nieng Yan. Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. *Nature*, 490(7420):361–366, October 2012. ISSN 1476-4687. doi: 10.1038/nature11524. URL https://www.nature.com/articles/nature11524.

Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7420 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Structural biology;Transporters Subject_term_id: structural-biology;transporters.

C. Suryanarayana and M. Grant Norton. *X-Ray Diffraction: A Practical Approach*. Springer Science & Business Media, June 2013. ISBN 978-1-4899-0148-4. Google-Books-ID: RRfrBwAAQBAJ.

C. Tanford. Translocation pathway in the catalysis of active transport. *Proceedings of the National Academy of Sciences*, 80(12):3701–3705, June 1983. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.80.12.3701. URL http://www.pnas.org/cgi/doi/10.1073/pnas.80.12.3701.

Sanna Tolle-Sander, Kimberley A. Lentz, Dean Y. Maeda, Andrew Coop, and James E. Polli. Increased Acyclovir Oral Bioavailability via a Bile Acid Conjugate. *Molecular Pharmaceutics*, 1(1):40–48, January 2004. ISSN 1543-8384, 1543-8392. doi: 10.1021/mp034010t. URL https://pubs.acs.org/doi/10.1021/mp034010t.

L. Tougaard, B. Giese, B. Højlund Pedersen, and V. Binder. Bile Acid Metabolism in Patients with Crohn's Disease in Terminal Ileum. *Scandinavian Journal of Gastroenterology*, 21(5):627–633, January 1986. ISSN 0036-5521. doi: 10.3109/00365528609003110. URL https://doi.org/10.3109/00365528609003110. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/00365528609003110.

Amos M. Tsai, Dan A. Neumann, and Leonard N. Bell. Molecular Dynamics of Solid-State Lysozyme as Affected by Glycerol and Water: A Neutron Scattering Study. *Biophysical Journal*, 79(5):2728–2732, November 2000. ISSN 00063495. doi: 10.1016/S0006-3495(00)76511-8. URL https://linkinghub.elsevier.com/retrieve/pii/S0006349500765118.

Roger Y. Tsien. The green fluorescent protein. *Annual Review of Biochemistry*, 67(1):509–544, June 1998. ISSN 0066-4154. doi: 10.1146/annurev.biochem.67.1.509. URL https://www.annualreviews.org/doi/10.1146/annurev.biochem.67.1.509. Publisher: Annual Reviews.

A. Tzschach, C. Ramel, A. Kron, B. Seipel, C. Wüster, U. Cordes, T. Liehr, M. Hoeltzenbein, C. Menzel, H.-H. Ropers, R. Ullmann, V. Kalscheuer, J. Decker, and D. Steinberger. Hypergonadotropic hypogonadism in a

patient with inv ins (2;4). *International Journal of Andrology*, 32(3):226–230, 2009. ISSN 1365-2605. doi: 10.1111/j.1365-2605.2007.00839.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2605.2007.00839.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2605.2007.00839.x.

Gerrit van Meer, Dennis R. Voelker, and Gerald W. Feigenson. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology*, 9(2):112–124, February 2008. ISSN 1471-0080. doi: 10.1038/nrm2330. URL https://www.nature.com/articles/nrm2330. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 2 Primary_atype: Reviews Publisher: Nature Publishing Group.

Robert Vaser, Swarnaseetha Adusumalli, Sim Ngak Leng, Mile Sikic, and Pauline C Ng. SIFT missense predictions for genomes. *Nature Protocols*, 11(1):1–9, January 2016. ISSN 1754-2189, 1750-2799. doi: 10.1038/nprot.2015.123. URL http://www.nature.com/articles/nprot.2015.123.

L. Vítek and M. C. Carey. Enterohepatic cycling of bilirubin as a cause of 'black' pigment gallstones in adult life. *European Journal of Clinical Investigation*, 33(9):799–810, 2003. ISSN 1365-2362. doi: https://doi.org/10.1046/j.1365-2362.2003.01214.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2362.2003.01214.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2362.2003.01214.x.

Samuel Wagner, Mirjam M. Klepsch, Susan Schlegel, Ansgar Appel, Roger Draheim, Michael Tarry, Martin Högbom, Klaas J. van Wijk, Dirk J. Slotboom, Jan O. Persson, and Jan-Willem de Gier. Tuning Escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences*, 105(38):14371–14376, September 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0804090105. URL https://www.pnas.org/content/105/38/14371. Publisher: National Academy of Sciences Section: Biological Sciences.

G. S. Waldo, B. M. Standish, J. Berendzen, and T. C. Terwilliger. Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology*, 17(7):691–695, July 1999. ISSN 1087-0156. doi: 10.1038/10904.

E. Wallin and G. von Heijne. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science: A Publi-*

cation of the Protein Society, 7(4):1029–1038, April 1998. ISSN 0961-8368. doi: 10.1002/pro.5560070420.

Jimin Wang. Estimation of the quality of refined protein crystal structures. *Protein Science : A Publication of the Protein Society*, 24(5): 661–669, May 2015. ISSN 0961-8368. doi: 10.1002/pro.2639. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4420517/.

Wei Wang, Shanyan Xue, Sue Ann Ingles, Qiuxiong Chen, Anh T. Diep, Harold D. Frankl, Andrew Stolz, and Robert W. Haile. An Association between Genetic Polymorphisms in the Ileal Sodium-dependent Bile Acid Transporter Gene and the Risk of Colorectal Adenomas. *Cancer Epidemiology and Prevention Biomarkers*, 10(9):931–936, September 2001. ISSN 1055-9965, 1538-7755. URL https://cebp.aacrjournals.org/content/10/9/931. Publisher: American Association for Cancer Research Section: Research Articles.

X. Wang, Y. Lyu, Y. Ji, Z. Sun, and X. Zhou. An engineered disulfide bridge traps and validates an outward-facing conformation in a bile acid transporter. *Acta Crystallographica Section D: Structural Biology*, 77(1):108–116, January 2021a. ISSN 2059-7983. doi: 10.1107/S205979832001517X. URL https://journals.iucr.org/d/issues/2021/01/00/dw5214/. Number: 1 Publisher: International Union of Crystallography.

X. Wang, Y. Lyu, Y. Ji, Z. Sun, and X. Zhou. Substrate binding in the bile acid transporter ASBTYf from Yersinia frederiksenii. *Acta Crystallographica Section D: Structural Biology*, 77(1):117–125, January 2021b. ISSN 2059-7983. doi: 10.1107/S2059798320015004. URL https://journals.iucr.org/d/issues/2021/01/00/qh5068/. Number: 1 Publisher: International Union of Crystallography.

Zhaoshuai Wang, Cui Ye, Xinyi Zhang, and Yinan Wei. Cysteine residue is not essential for CPM protein thermal-stability assay. *Analytical and Bioanalytical Chemistry*, 407(13):3683–3691, May 2015. ISSN 1618-2650. doi: 10.1007/s00216-015-8587-4. URL https://doi.org/10.1007/s00216-015-8587-4.

William W. Ward, Hugh J. Prentice, Amy F. Roth, Chris W. Cody, and Sue C. Reeves. Spectral Perturbations of the Aequorea Green-Fluorescent Protein. *Photochemistry and Photobiology*, 35(6):803–808, 1982. ISSN 1751-1097. doi: 10.1111/j.1751-1097.1982.tb02651.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-1097.1982.tb02651.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-1097.1982.tb02651.x.

Andrew M. Waterhouse, James B. Procter, David M. A. Martin, Michèle Clamp, and Geoffrey J. Barton. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191, May 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp033. URL https://doi.org/10.1093/bioinformatics/btp033.

S. Weinkauf, J. F. Hunt, J. Scheuring, L. Henry, J. Fak, D. B. Oliver, and J. Deisenhofer. Conformational stabilization and crystallization of the SecA translocation ATPase from Bacillus subtilis. *Acta Crystallographica Section D: Biological Crystallography*, 57(4):559–565, April 2001. ISSN 0907-4449. doi: 10.1107/S0907444901001202. URL http://scripts.iucr.org/cgi-bin/paper?gr2107. Number: 4 Publisher: International Union of Crystallography.

S. A. Weinman. Electrogenicity of Na(+)-coupled bile acid transporters. *The Yale Journal of Biology and Medicine*, 70(4):331–340, August 1997. ISSN 0044-0086.

Steven A. Weinman, Michael W. Carruth, and Paul A. Dawson. Bile Acid Uptake via the Human Apical Sodium-Bile Acid Cotransporter Is Electrogenic. *Journal of Biological Chemistry*, 273(52):34691–34695, December 1998. ISSN 00219258. doi: 10.1074/jbc.273.52.34691. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925818371321.

A. M. Whited and A. Johs. The interactions of peripheral membrane proteins with biological membranes. *Chemistry and Physics of Lipids*, 192:51–59, November 2015. ISSN 1873-2941. doi: 10.1016/j.chemphyslip.2015.07.015.

Julian P. Whitelegge. Integral Membrane Proteins and Bilayer Proteomics. *Analytical chemistry*, 85(5):2558–2568, March 2013. ISSN 0003-2700. doi: 10.1021/ac303064a. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3664232/.

Frederick A. Wilson, Gerhard Burckhardt, Heini Murer, Gerhard Rumrich, and Karl J. Ullrich. Sodium-coupled Taurocholate Transport in the Proximal Convolution of the Rat Kidney In Vivo and In Vitro. *Journal of Clinical Investigation*, 67(4):1141–1150, April 1981. ISSN 0021-9738. doi: 10.1172/JCI110128. URL http://www.jci.org/articles/view/110128.

G. Winter. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, 43(1):186–190, February 2010. ISSN 0021-8898. doi: 10.1107/S0021889809045701. URL http://scripts.iucr.org/cgi-bin/paper?ea5113. Number: 1 Publisher: International Union of Crystallography.

G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter, and G. Evans. DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D: Structural Biology*, 74(2):85–97, February 2018. ISSN 2059-7983. doi: 10.1107/S2059798317017235. URL http://scripts.iucr.org/cgi-bin/paper?di5011. Number: 2 Publisher: International Union of Crystallography.

Marcin Wojdyr, Ronan Keegan, Graeme Winter, and Alun Ashton. DIMPLE - a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands. *Acta Crystallographica Section A Foundations of Crystallography*, 69:s299–s299, August 2013. doi: 10.1107/S0108767313097419.

Melissa H. Wong, Peter Oelkers, and Paul A. Dawson. Identification of a Mutation in the Ileal Sodium-dependent Bile Acid Transporter Gene That Abolishes Transport Activity. *Journal of Biological Chemistry*, 270(45):27228–27234, October 1995. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.270.45.27228. URL http://www.jbc.org/content/270/45/27228. Publisher: American Society for Biochemistry and Molecular Biology.

M.H. Wong, P. Oelkers, A.L. Craddock, and P.A. Dawson. Expression cloning and characterization of the hamster ileal sodium-dependent bile acid transporter. *Journal of Biological Chemistry*, 269(2):1340–1347, January 1994. ISSN 00219258. doi: 10.1016/S0021-9258(17)42263-0. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925817422630.

Xiaobing Xia, Gang Wang, and Hongqing Fang. Role of hydration in the conformational transitions between unliganded and liganded forms of loop 13 of the Na+/glucose cotransporter 1. *Biochemical and Biophysical Research Communications*, 315(4):1018–1024, March 2004. ISSN 0006291X. doi: 10.1016/j.bbrc.2004.02.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0006291X04002372.

Ling Xiao and Guoyu Pan. An important intestinal transporter that regulates the enterohepatic circulation of bile acids and cholesterol homeostasis: The apical sodium-dependent bile acid transporter (SLC10A2/ASBT). *Clinics and Research in Hepatology and Gastroenterology*, 41(5):509–515, October 2017. ISSN 2210-7401. doi: 10.1016/j.clinre.2017.02.001. URL http://www.sciencedirect.com/science/article/pii/S2210740117300487.

Huan Yan, Guocai Zhong, Guangwei Xu, Wenhui He, Zhiyi Jing, Zhenchao Gao, Yi Huang, Yonghe Qi, Bo Peng, Haimin Wang, Liran Fu, Mei Song,

Pan Chen, Wenqing Gao, Bijie Ren, Yinyan Sun, Tao Cai, Xiaofeng Feng, Jianhua Sui, and Wenhui Li. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife*, 1, November 2012. ISSN 2050-084X. doi: 10.7554/eLife.00049. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3485615/.

Z. R. Yang, R. Thomson, P. McNeil, and R. M. Esnouf. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 21(16):3369–3376, August 2005. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/bti534. URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformati

Philip Yeagle. *The Membranes of Cells, 3rd ed.* January 2016. ISBN 978-0-12-800047-2.

A. P. Yeh, A. McMillan, and M. H. B. Stowell. Rapid and simple protein-stability screens: application to membrane proteins. *Acta Crystallographica Section D: Biological Crystallography*, 62(4):451–457, April 2006. ISSN 0907-4449. doi: 10.1107/S0907444906005233. URL http://scripts.iucr.org/cgi-bin/paper?en5143. Number: 4 Publisher: International Union of Crystallography.

J. Zelano, S. Mikulovic, K. Patra, M. Kühnemund, M. Larhammar, L. Emilsson, R. N. Leão, R. Leao, and K. Kullander. The synaptic protein encoded by the gene Slc10A4 suppresses epileptiform activity and regulates sensitivity to cholinergic chemoconvulsants. *Experimental Neurology*, 239:73–81, January 2013. ISSN 1090-2430. doi: 10.1016/j.expneurol.2012.09.006.

T Zeuthen. Molecular mechanisms for passive and active transport of water. *International review of cytology*, 160:99–161, January 1995. ISSN 0074-7696. doi: 10.1016/s0074-7696(08)61554-5. URL https://doi.org/10.1016/s0074-7696(08)61554-5.

Yunying Zhao, Hongbo Yan, Ricardo Happeck, Tina Peiter-Volk, Huihui Xu, Yan Zhang, Edgar Peiter, Chloë van Oostende Triplet, Malcolm Whiteway, and Linghuo Jiang. The plasma membrane protein Rch1 is a negative regulator of cytosolic calcium homeostasis and positively regulated by the calcium/calcineurin signaling pathway in budding yeast. *European Journal of Cell Biology*, 95(3): 164–174, March 2016. ISSN 0171-9335. doi: 10.1016/j.ejcb.2016.01.001. URL http://www.sciencedirect.com/science/article/pii/S0171933516300024.

Xiaowan Zheng, Sean Ekins, Jean-Pierre Raufman, and James E. Polli. Computational Models for Drug Inhibition of the Human Apical Sodium-Dependent

Bile Acid Transporter. *Molecular Pharmaceutics*, 6(5):1591–1603, October 2009. ISSN 1543-8384, 1543-8392. doi: 10.1021/mp900163d. URL https://pubs.acs.org/doi/10.1021/mp900163d.

Mikhail Zhernenkov, Dima Bolmatov, Dmitry Soloviov, Kirill Zhernenkov, Boris P. Toperverg, Alessandro Cunsolo, Alexey Bosak, and Yong Q. Cai. Revealing the mechanism of passive transport in lipid bilayers via phonon-mediated nanometre-scale density fluctuations. *Nature Communications*, 7(1):11575, September 2016. ISSN 2041-1723. doi: 10.1038/ncomms11575. URL http://www.nature.com/articles/ncomms11575.

Xiaoming Zhou, Elena J. Levin, Yaping Pan, Jason G. McCoy, Ruchika Sharma, Brian Kloss, Renato Bruni, Matthias Quick, and Ming Zhou. Structural basis of the alternating-access mechanism in a bile acid transporter. *Nature*, 505(7484): 569–573, January 2014. ISSN 1476-4687. doi: 10.1038/nature12811.

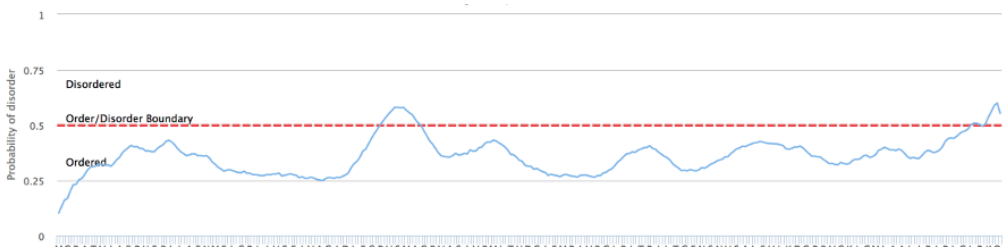
Xianqiong Zou, Dianzuo Wang, Guanzhou Qiu, Chaoneng Ji, Feng Jin, Maoqing Wu, Huarui Zheng, Xin Li, Liyun Sun, Yu Wang, Rong Tang, Robert Chunhua Zhao, and Yumin Mao. Molecular Cloning and Characterization of a Novel Human C4orf13 Gene, Tentatively a Member of the Sodium Bile Acid Cotransporter Family. *Biochemical Genetics*, 43(3):165–173, April 2005. ISSN 1573-4927. doi: 10.1007/s10528-005-1509-y. URL https://doi.org/10.1007/s10528-005-1509-y.

Appendix 1

SLC10A7 *Asticcacaulis taihuensis*

SLC10A7 *Asticcacaulis taihuensis* protein sequence and respective RONN disorder plot:

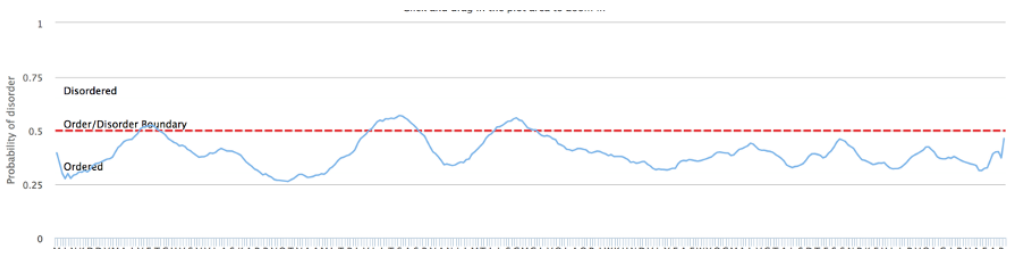
```
MKVGKVRIDAFLTGMVLAIILAALFPTPGVHGGFLHPELLTKLGIALIFFLNGAMLSFHALKDGVLRWRLHLIIQ
VATFVLFPVIGLVFLAVAGRWISPDLQLGFFYLCALPSTVSSSVAMTSAARGNVPVAVFNATISSLIGIVVTPLW
MSLLLKTSGHHLDVAGVFIDLFKWMVLPMIIGQVLRRWIGAWLHRRKKLAQTADRGTILLLVYTSFCDSFALNIW
SSTNLIVMLSVIAATLVLFFGVLLLLWKLCDLGGIAPAFRSAVVFCGTKKSLATGVPMASLIFAGHASGLGLILL
PIMIYHPLQLLICTPLASRWAKEAQA
```



SLC10A7 *Cupriavidus necator*

SLC10A7 *Cupriavidus necator* protein sequence and respective RONN disorder plot:

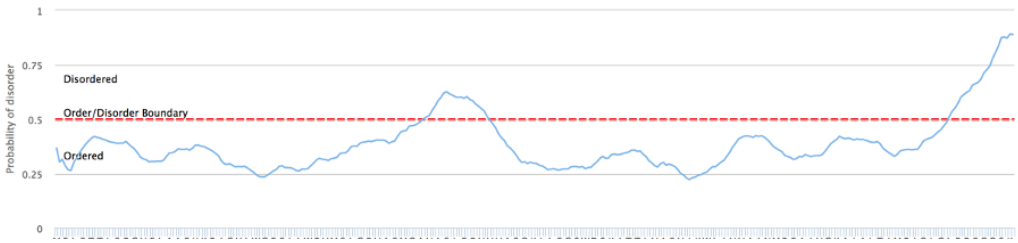
```
MSRIVRNLKKLYDLIDGFVLVMLTAIAIALVAPEIGTGKGPLHLGVVTSFGVALVFFLHGAALSRDKLVAGARHW
RLHVFVQVFTYVVFPAVGALLMLSLRHTLPPDLLLGVFFLCALPSTVSSSVAMTSMARGNVSGAIFNATISGLIG
MLVTPLLMGLVISASGASMPLGKALTGVALQLLLPFALGQLLRPLIGSWLAKKKHITNKIDRGVIVLIVYSSFCD
ATAEGLWHQYQWQTIGAVMLIAAVLLFVVLGFTTLTARRLGFSVEDEITAVFCGSKKSLANGIPMAKILFAGHPA
LGLLVLPLMVYHQLQLIVCSVIASRYANRDALLEDQAAARA
```



SLC10A7 *Escherichia coli*

SLC10A7 *Escherichia coli* protein sequence and respective RONN disorder plot:

```
MKLFRILDPFTLTLITVVLLASFFPARGDFVPFFENLTTAAIALLFFMHGAKLSREAIIAGGGHWRLHLWVMCSTFVLF
PILGVLFAWWKPVNVDPMLYSGFLYLCILPATVQSAIAFTSMAGGNVAAAVCSASASSLLGIFLSPLLVGLVMNVHGAG
GSLEQVGKIMLQLLLPFVLGHLSRPWIGDWVSRNKKWIAKTDQTSILLVVYTAFSEAVVNGIWHKVGWGSLLFIVVVSC
VLLAIVIVVNVFMARRLSFNKADEITIVFCGSKKSLANGIPMANILFPTSVIGMMVLPLMIFHQIQLMVCAVLARRYKR
QTEQLQAQQESSADKA
```



SLC10A7 *Escherichia coli* Q322 C-terminal Truncation

SLC10A7 *Escherichia coli* Q322 C-terminal Truncation protein sequence:

```
MKLFRILDPFTLTLITVVLLASFFPARGDFVPFFENLTTAAIALLFFMHGAKLSREAIIAGGGHWRLHLWVMCSTFVLF
PILGVLFAWWKPVNVDPMLYSGFLYLCILPATVQSAIAFTSMAGGNVAAAVCSASASSLLGIFLSPLLVGLVMNVHGAG
GSLEQVGKIMLQLLLPFVLGHLSRPWIGDWVSRNKKWIAKTDQTSILLVVYTAFSEAVVNGIWHKVGWGSLLFIVVVSC
VLLAIVIVVNVFMARRLSFNKADEITIVFCGSKKSLANGIPMANILFPTSVIGMMVLPLMIFHQIQLMVCAVLARRYKR
QTEQLQ
```

SLC10A7 *Escherichia coli* Q320 C-terminal Truncation

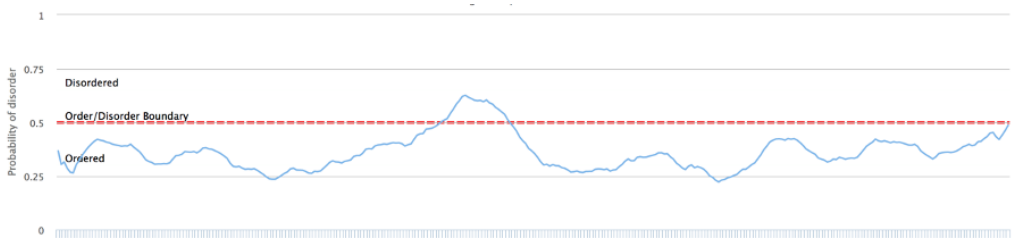
SLC10A7 *Escherichia coli* Q322 C-terminal Truncation protein sequence:

```
MKLFRILDPFTLTLITVVLLASFFPARGDFVPFFENLTTAAIALLFFMHGAKLSREAIIAGGGHWRLHLWVMCSTFVLF
PILGVLFAWWKPVNVDPMLYSGFLYLCILPATVQSAIAFTSMAGGNVAAAVCSASASSLLGIFLSPLLVGLVMNVHGAG
GSLEQVGKIMLQLLLPFVLGHLSRPWIGDWVSRNKKWIAKTDQTSILLVVYTAFSEAVVNGIWHKVGWGSLLFIVVVSC
VLLAIVIVVNVFMARRLSFNKADEITIVFCGSKKSLANGIPMANILFPTSVIGMMVLPLMIFHQIQLMVCAVLARRYKR
QTEQ
```

SLC10A7 *Escherichia coli* R316 C-terminal Truncation

SLC10A7 *Escherichia coli* protein sequence and respective RONN disorder plot:

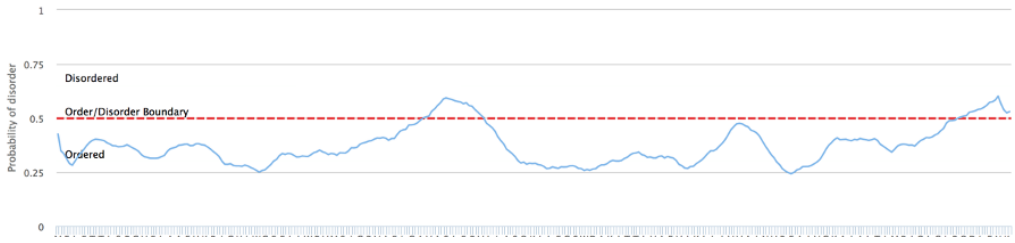
```
MKLFRILDPFTLTLITVVLLASFFPARGDFVPFFENLTTAAIALLFFMHGAKLSREAIIAGGGHWRLHLWVMCSTFVLF
PILGVLFAWWKPVNVDPMLYSGFLYLCILPATVQSAIAFTSMAGGNVAAAVCSASASSLLGIFLSPLLVGLVMNVHGAG
GSLEQVGKIMLQLLLPFVLGHLSRPWIGDWVSRNKKWIAKTDQTSILLVVYTAFSEAVVNGIWHKVGWGSLLFIVVVSC
VLLAIVIVVNVFMARRLSFNKADEITIVFCGSKKSLANGIPMANILFPTSVIGMMVLPLMIFHQIQLMVCAVLARRYKR
```



SLC10A7 *Klebsiella pneumoniae*

SLC10A7 *Klebsiella pneumoniae* protein sequence and respective RONN disorder plot:

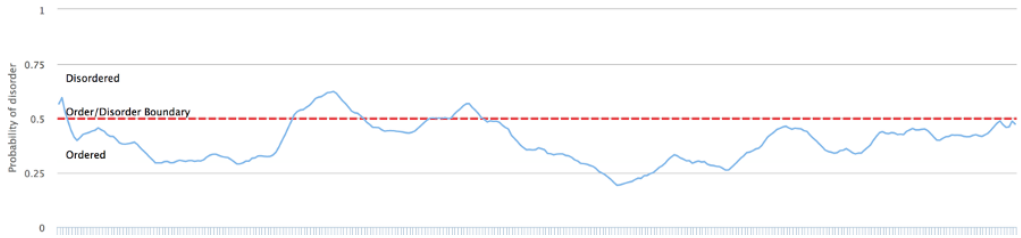
```
MKLFRILDPFTATLITVVLLASFFPARGAFVPFFEHLTTAAIALLFFMHGAKLSREAIIAGGSHWRLHLWVMCSTFILF
PLLGVLFAWWAPVNVDPMLYSGFIYLCILPATVQSAIAFTSLAGGNVAAAVCSASASSLLGIFVSPLLVGVLMNLHGAG
GSLEQVGKIMLQLLLPFVLGHLSRPWIGDWVAKHKKWIGKTDQTSILLVVYSAFSEAVVNGIWHKVGLGSLLFIVVVSL
VLLAIVIAVNVFVARRCGFNKADEITIVFCGSKKSLANGIPMANILFPTSILGIMVLPLMIFHQIQLMVCAVLARRYKQ
QTDARLAEEKANAAKA
```



SLC10A7 *Pseudomonas aeruginosa* R318 C-terminal Truncation

SLC10A7 *Pseudomonas aeruginosa* R318 C-terminal Truncation protein sequence and respective RONN disorder plot:

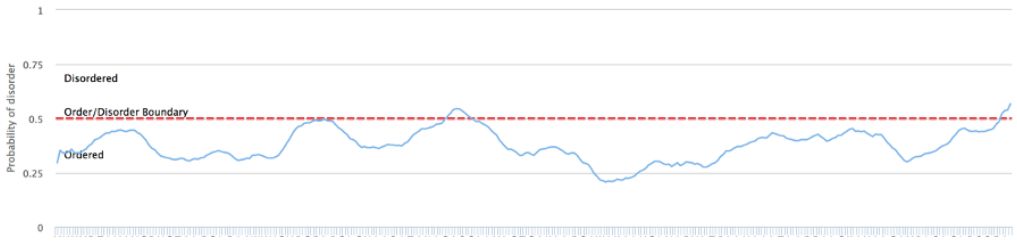
```
MSRPRFLPDNFTLALIATVLLATFLPCSGQTAVVFEWVTNIGIGLLFFLHGAKLSRQAIIAGMTHWRLHLLVFACTFVM
FPLLGLALKPALSPMVTPELYLGILFLCALPATVQSSIAFTSLARGNVPAAVCSASVSSLLGVFLTPLLVKLLLGAEGE
TGNALDAIGKITLQLLVPFIAGQVLRRWIGAWVERNKPVLRYVDQGSILLVVYTAFSAAVIQGLWHEVPWLALLGLTVA
CCVILALALVLTTVLARRLGFSKEDEITIVFCGSKKSLATGVPMAKVLFATSAVGPMVLPLMLFHQIQLMVCAVLAQRY
AR
```



SLC10A7 *Pseudomonas putida*

SLC10A7 *Pseudomonas putida* protein sequence and respective RONN disorder plot:

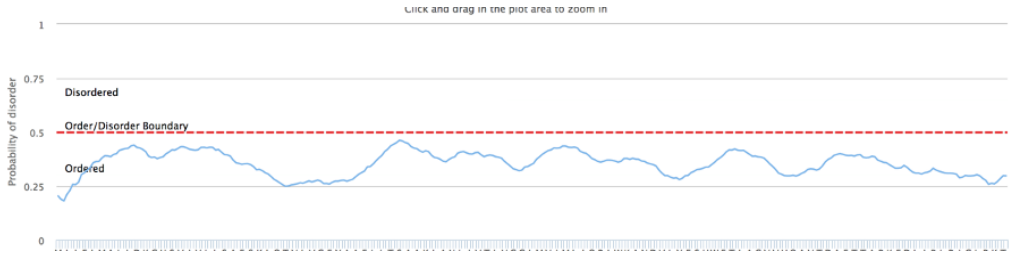
```
MDTVMKYLRMLFDNFTLALLGVVLIATVLPCSGDGAVYFGWLTNLAIGLLFFLHGAKLSREAIIAGAGHWRLHLLVFSC
TFVLFPLLGLAFKPLFVPLVGNELYLGILYLCALPATVQSAIAFTSLARGNVPAAICSAAASSLLGIFLTPLLVMLLLG
AGGDTGSGLDAVLKITLQLLVPFVAGQVARRWIGAWVKRNARWLKVVDQGSILLVVYTAFSEAVVTGLWHTVSPQHLAG
LFVVCAILLAVVLFGTRLLGKALGFDLEDRITILFAGSKKSLATGVPMAQVLFVGSGIGAMILPLMLFHQIQLMVCAVL
AQRYASREQAVEASAVS
```



SLC10A7 *Vibrio breoganii*

SLC10A7 *Vibrio breoganii* protein sequence and respective RONN disorder plot:

MQIIKKIKKEWFLVGMVMAIVLAVLTPDLGKTGGTLHLDSVTVLGIALVFFLHGLGLSPKAIVEGLSNWKLHLYV
QGATFVVYPLLWVVFGQGFLAYMPAALAFGFCYLFVLPSTISSSVAMTAIGKGNLPGAIFNASLSSILGVFITPL
LIQVFMGMEGAELNLMDSVISISKMLLLPMILGQIMRPWLLEWTQKHKAVVNKLDKYVILLIVYSAFCDSIDHGI
WSDFSPTLLLISALICLIVLMVMVHGIQWGARKVGFTHQDEVAAVFCGTKKTLAAGIPMAKVIFAHDPNLGMILL
PIMLYHPIQIFYCAVLANRYAKVATA



Appendix 2

ASBT *Leptospira biflexa*

ASBT *Leptospira biflexa* protein sequence and respective RONN disorder plot:

```
MLTRTEEILFAAMVFFLMVAMGSTLTIENFKKAVHSKKPLIVGVISQFGFMPLIAFGLAKSLDLSPLFSIGLILV
GCTPGGTTSNLLTYYAKGDVALSISMTITSTILATVMMPFLFWLYCSGFAENDIQIPYKSIVGSIFILIIPVLIG
IQIRSYNTRMALKIEKIGSYLGILMILFLLGVMVPKNLDILQITTWQMYLAAILITVLGYSFGYIFSRILNLSEK
QARTVSLETGIQNGPLTIAVILLSFSNSISNEILWMPLLYALFVPITSSIATYYFYLKSKQESKGQV
```