Towards the structural characterisation of equilibrative nucleoside transporters

Jessica Catherine Boakes

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Biomedical Sciences

June 2022

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 2 and **Chapter 4** contains methodologies, data, analysis, and written content adapted from the following co-first and first authored articles:

i. Harborne SPD*, Strauss J*, **Boakes JC***, Wright DL, Henderson JG, Boivineau J, Jaakola VP, Goldman A. IMPROvER: the Integral Membrane Protein Stability Selector. Sci Rep. (2020) Sep 16;10(1):15165.

Author Contributions: S.P.D.H. and A.G. conceived the study and planned the experiments. S.P.D.H. wrote the computational pipeline. J.S., J.C.B. and J.G.H. carried out literature searches for critical residues of *Cl*-PPase, hENT1 and hPTH1R, respectively. J.S. carried out mutagenesis and stability testing of *Cl*- PPase and developed a rapid screening strategy. S.P.D.H., D.L.W., J.C.B., J.B. and V.-P.J. carried out mutagenesis and stability testing of hENT1. J.G.H. carried out stability testing of hPTH1R. S.P.D.H., J.S. and A.G. wrote the manuscript. All authors assisted manuscript revision and editing.

The methods and the initial data for hENT1 (up until the expression and localisation of hENT1 in insect cells) from this publication are used in this thesis.

ii. **Boakes JC**, Harborne SPD, Ngo JTS, Pliotas C & Goldman, A. Novel variants provide differential stabilisation of human equilibrative nucleoside transporter 1 states. Front Mol Biosci. (2022) Nov (9) doi: 10.3389/fmolb.2022.970391

Author Contributions: S.P.D.H and A.G conceived the study. S.P.D.H, J.C.B and A.G planned the experiments. J.C.B and J.T.S.N carried out stability testing of hENT1. J.C.B carried out bacmid extraction and sequencing. J.C.B wrote the manuscript. All authors assisted in manuscript revision and editing.

The candidate devised and wrote the manuscript under the guidance of Prof. Adrian Goldman. The candidate made all the figures and tables. All written content and figures are used in this thesis.

Chapter 2 and Chapter 3 are partly based on previous unpublished work by former group member Dr. Steven Harborne.

> The candidate used previously established expression and purification protocols to produce wild type hENT1 and to inform further methodology developments. Please contact Prof. Adrian Goldman to obtain relevant data of previous work undertaken by Dr. Steven Harborne.

Acknowledgements

I will be forever grateful to have found myself a part of such a wonderful lab group. Thank you to all members of the Goldman Group, past and present, and all those on Astbury Level 6. Special thanks go to our former post-docs, to Dr. Maren Thomsen, Dr. Julie Heggelund and Dr. James Hillier for all your advice on all thing's protein purification and crystallography, to Maria, Jannik and Claudia for always trying to get me to be social and inviting me to events and activities, even though you know that I will always bail, to Andreas for coffee breaks, and to everyone else: Jack, Brendan, Craig, Katie, Alex, Ana, Anchal, Jamie and all those on Astbury 6 for making the lab such a welcoming place.

A big thank you goes to Dr. Steven Harborne for all your earlier work laying the foundation for my research, and for all your patience and support while I found my footing. Thank you to Danielle Wright for all your hard work, and thank you to Professor Peter Henderson and Dr Stephen Muench for being wonderfully understanding and supportive mentors.

Additional thanks go to Oli, Antony and all the co-workers at Salt Works for providing me with a wonderful environment to write this thesis in, thank you for keeping me caffeinated!

Special thanks go to all the children and families that I had the honour of working with in my past career. I don't think I would be on this path if it weren't for you.

To my incredible parkner Dean, thank you for moving fridges with me and for tolerating me barely being home recently. Despite your best efforts, I've still not quite mastered the art of pacing myself. Hopefully it's a skill I can develop in our next adventure. Thank you for always being the best.

Finally, and most importantly, I would like to thank my best friend Frankie, my sister Fiona, and my Mum. I don't think I'll ever be able to quite articulate, let alone summarise for an acknowledgements section, just how grateful I am to have you in my life. I know that I would certainly not be where I am or the person that I am today were it not for your love and support, especially over these last few years. Thank you for always being there for me and thank you for always believing in me.

Abstract

Human equilibrative nucleoside transporters (ENTs) represent a major pharmaceutical target for cardiac, cancer and viral therapies. Understanding the molecular basis for transport is crucial for the development of improved therapeutics through structure-based drug design. ENTs have been proposed to utilise an alternating access mechanism of action, similar to that of the major facilitator superfamily. However, ENTs lack functionally-essential features of that superfamily, suggesting they may use a different transport mechanism. Understanding the molecular basis of their transport requires insight into diverse conformational states. In this work I explored distinct but complementary approaches towards the stabilisation of ENTs, with the goal of structural characterisation.

First, I demonstrated that yields of wild-type human ENT isoform 1 (hENT1) obtained using previously optimised methodologies (expression in *Spodoptera frugiperda* cells and a two-step IMAC/protease purification) were not amenable to further investigations due to instability and poor yields. Second, I explored stabilising point mutations of hENT1. I identified four distinct variants of hENT1 with single-point mutations at the large intracellular loop (ICL6) and transmembrane helix 7 (TM7) that stabilise the *apo*-state (ΔT_m 0.7-1.5 °C). Furthermore, I found that two variants that specifically stabilise the inhibitor-bound state ($\Delta\Delta T_m$ 3.0-5.0 °C), supporting the role of ICL6 and TM7 in hENT1 gating. I also found that variant T336A (TM8) binds the inhibitor nitrobenzylthioinosine with a 7-fold lower affinity than wild-type. Therefore, this residue may help determine inhibitor and substrate sensitivity. Finally, I identified a homologue of hENT1 from the ascomycete *Byssochlamys spectabilis* (*Bs*ENT). *Bs*ENT is significantly more stable than hENT1 (ΔT_m of 17.3 ± 1.9 °C (p = <0.0001)) and can be purified from insect cells with high, homogenous and monodisperse yields, 3-5-fold higher than hENT1. Furthermore, initial crystallisation trials suggest that *Bs*ENT is amenable to structural characterisation. Thus, *Bs*ENT is a promising candidate for homology modelling of hENT1.

Table of Contents

List of figures

Appendix figures

List of tables

Appendix tables

List of abbreviations

ABC	ATP-binding cassette
AMP	Adenosine monophosphate
AP	Antarctic phosphatase
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BEVS	Baculovirus expression system
BIIC	Baculovirus infected insect cells
BRIL	Thermostabilised apocytochrome b_{562}
BSA	Bovine serum albumin
CHS	Cholesterol hemisuccinate
CMC	Critical micelle concentration
CNT	Concentrative nucleoside transporter
Constructs:	
CTGH	C-terminally TEV-GFP-His_6 tagged
NHGV	N-terminally tagged His-GFP-HRV
NHsfGT	N-terminally tagged His-super folder GFP-TEV
NHT	N-terminally tagged His-2TEV
CPM	N-[4-(7-diethylamino- 4-methyl-3-coumarinyl)phenyl] maleimide
Detergents:	
DDM	n-dodecyl-β-D-maltopyranoside
DM	n-decyl-β-D-maltopyranoside
DMNG	Decyl maltose neopentyl glycol
FC12	n-dodecyl-phosphocholine
LMNG	Lauryl maltose neopentyl glycol
OGNG	Octyl maltose neopentyl glycol
OM	n-octyl-β-D-maltopyranoside
DIBMA	di-isobutylene maleic acids
DMSO	Dimethyl sulfoxide
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
ECL	Extracellular loop
EDTA	Ethylenediaminetetra-acetic acid
ELISA	Enzyme-linked immunosorbent assay
ENTs	Equilibrative nucleoside transporters
*Bb*ENT	*Branchiostoma belcheri* equilibrative nucleoside transporter
*Bs*ENT	*Byssochlamys spectabilis* equilibrative nucleoside transporter
FUN26	Function unknown 26
*Gg*ENT	*Gallus gallus* equilibrative nucleoside transporter
hENT1	Human equilibrative nucleoside transporter isoform 1

hENT2	Human equilibrative nucleoside transporter isoform 2
hENT3	Human equilibrative nucleoside transporter isoform 3
hENT4	Human equilibrative nucleoside transporter isoform 4
*Rv*ENT	*Ramazzottius varieornatus* equilibrative nucleoside transporter
*Tt*ENT	*Tetrahymena thermophila* equilibrative nucleoside transporter
FACS	Fluorescence activated cell sorting
FPLC	Fast protein liquid chromatography
GFP	Green fluorescent protein
GLUT	Glucose transporter
GPCR	G protein coupled receptors
HA	Human influence virus hemagglutinin
HABA	4'-hydroxyazobenzene-2-carboxylic acid
HEK	Human embryo kidney
HPLC	High performance liquid chromatography system
HRP	Horseradish peroxidase
HRV 3C	Human rhinovirus 3C protease
ICH	Intracellular helix
ICL	Intracellular loop
IMAC	Immobilised metal affinity chromatography
IMP	Integral membrane protein
IMPROvER	Integral membrane protein stability selector
IPTG	Isopropyl β-d-1-thiogalactopyranoside
LB	Luria broth
LC-MS	Liquid chromatography-mass spectrometry
LCP	Lipidic cubic phase
MACS	Magnetic activated cell sorting
MFS	Major facilitator superfamily
MOI	Multiplicity of infection
MP	Membrane proteins
MW	Molecular weight
MWCO	Molecular weight cut off
NADs	Nucleoside analogue drugs
NBMPR	Nitrobenzyl mercaptopurine ribonucleoside/ Nitrobenzyl-thioinosine
NHS	Nucleoside proton symporters
NMR	Nuclear magnetic resonance
NT	Nucleoside transporter
OD	Optical density

Organisms

B. belcheri	*Branchiostoma belcheri*
B. spectabilis	*Byssochlamys spectabilis*
E. coli	*Escherichia coli*
G. gallus	*Gallus gallus*
P. pastoris	*Pichia Pastoris*

P. variotii	*Paecilomyces variotii*
R. varieornatus	*Ramazzottius varieornatus*
S. cerevisiae	*Saccharomyces cerevisiae*
S. fugiperda	*Spodoptera frugiperda*
T. thermophila	*Tetrahymena thermophila*
PA	Proliferation arrest
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PMF	Proton motive force
SAS	Small angle scattering
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEC-MALLS	Size exclusion chromatography with multi angle laser light scattering
SEM	Standard error of the mean
Sf9	*Spodoptera frugiperda*
SLC	Solute carrier
SMALPs	Styrene-maleic acid lipid particles
TEV	Tobacco etch virus
T_m	Apparent melting temperature
TM	Transmembrane alpha-helix
YPD	Yeast, peptone, D-glucose media for yeast cultures

Chapter 1 Introduction

1.1 Protein structure

Proteins are molecules that are made up of one or more amino acid polypeptide chain, and it is the sequence of these amino acids that determine the conformation of these polypeptide chains and thus, protein structure.

Amino acids contain an amino group, a side chain, and a carboxylic acid, which are linked to a single α-carbon atom (Cα). There are 20 different commonly occurring amino acids, differentiated by side chain, each with differing properties (nonpolar, uncharged polar, acidic, and basic). Polypeptide chains are synthesised through the joining of a carboxyl group and a free amino group of an incoming amino acid in a stepwise manner, to achieve linear chains with structural polarity (N- and C-termini) that feature covalent CO-NH bonds (known as peptide bonds)[1,2].

Individual peptide groups have rigid, planar structures, with no rotation around these peptide bonds. However, the amino and carboxyl bonds to the Cα within an amino acid allow for rotation, with the torsional angles described as ϕ and ψ, respectively. These ϕ and ψ rotations allow polypeptides to be highly flexible. However, because of steric collisions between atoms in each amino acid, most combinations of ϕ and ψ do not occur. Thus, steric constraints limit the conformational range[1,2].

Protein folding is further constrained by weak, non-covalent interactions: ionic and hydrogen bonds, and *van der Waals* attractions. While each of these non-covalent bonds are weak individually, many bonds can act in parallel to create a strong bonding arrangement. Thus, the stability of a folded shape is determined by the combined strength of large numbers of such non-covalent interactions. In addition to covalent and non-covalent interactions, hydrophobic effects are a major influence on protein folding. These effects result in proteins folding in a way to minimise the contact between hydrophobic non-polar side chains with water and amphipathic molecules. Therefore, the distribution of polar and non-polar amino acids is also an important influence on protein structure[1-3].

As a result of these interactions, folded proteins can adopt diverse and complex three-dimensional structures. However, there are two regular folding patterns that are commonly found: α-helices and β-sheets. Proteins structures typically contain a high proportion of these conformations (~60% on average). Both α-helices and β-sheets arise from hydrogen bonding

between the N-H and C=O groups in the polypeptide backbone to adopt regular, repeating conformations. This hydrogen bonding within the polypeptide backbone accounts for up to 70% of hydrogen bonding in proteins[1,2].

In a β-sheet hydrogen bonds are formed between peptide bonds in neighbouring chains. These chains may run in parallel or antiparallel direction, to produce a rigid sheet, with a right-hand twist, with the side chains alternately projecting above and below the plane of the sheet. This right-hand twist gives β-sheets an inherent curvature and subsequently, sheets of more than 6 strands often roll up to form β-barrels. In an α-helix strong hydrogen bonds are formed within groups in the same chain, with a hydrogen bond made between one amino acid and a neighbouring residue, four peptide bonds away. This results in a rigid, cylindrical conformation in which the core of the helix is tightly packed, and the side chains are oriented towards the outside of the helix. In both folds, the hydrophilic polypeptide backbone is shielded from the external environment by the projection of the side chains.

1.2 Membrane proteins and the membrane lipid bilayer

Cells and subcellular organelles are surrounded by biological membranes, which are comprised of a lipid bilayer that is heavily populated with proteins and carbohydrates. The membrane serves to contain cellular contents and maintain homeostasis through the separation of biological processes from their environment. The permeability of the membrane is determined by both lipid and protein components, with membrane proteins serving as the gatekeepers.

1.2.1 The lipid bilayer

The lipid bilayer is primarily comprised of phospholipids and sphingolipids, and as is seen in protein folding, hydrophobic effects influence lipid organisation within the bilayer. Here, two lipid monolayers arrange to orient the lipid polar headgroups to the surface and the lipid acyl chains forming a non-polar domain in between[4-6], and thus, the contact between the hydrophobic acyl chains with aqueous molecules is minimised[1-3]. The non-polar domain may be further stabilised by *van der Waals* attractions between the close-packed acyl chains[4].

The lipid composition of the two monolayers of a lipid bilayer may be distinctly different from one another and thus, provide an asymmetry to membranes. Furthermore, the lipid composition of the membrane bilayer can also vary by location and function. While phospholipids and sphingolipids are the most abundant membrane lipids, in eukaryotes the bilayer may also feature sterols, with cholesterol being the most abundant sterol in mammals and a major

component of plasma membranes[1-4,7]. The ratio of protein to lipid in the membranes can vary with membrane function[1-4]. For example, myelin membranes typically feature a higher proportion of lipids than protein (~80% and ~20%, respectively). Whereas mitochondrial inner membranes feature proportionally more protein than lipid (~75% and ~25%, respectively)[1-4].

1.2.2 Membrane proteins

Membrane proteins have diverse functions, such as the transport of molecules and transduction of signals across the membrane, determination of membrane morphology, and facilitating cellular adhesion. Owing to the diversity in membrane protein functions, specific proteins may only occur in specific membranes. Membrane proteins can be subdivided into two broad classifications, integral or peripheral, depending on their interactions with the membrane (Figure 1.1). Peripheral proteins, also called extrinsic membrane proteins, associate at the surface of the membrane, through interactions with the lipid polar head groups, or non-covalent interactions with integral membrane proteins[8,9].Their specific binding is mediated by highly conserved motifs and often by divalent cations. For example, ankyrin, which mediates the linkage of the cytoskeleton to the plasma membrane via interactions with integral membrane proteins, and is composed almost entirely of 'ANK' repeats[4]. Peripheral proteins can both regulate and be regulated by membrane lipids. Furthermore, many of these proteins undergo co-/post-translational modifications to covalently attach to lipids via fatty acids and prenyl groups[3,4]. A major class of lipid-anchored proteins are the glycosylphospholipid-anchored proteins (GPIs), which feature more than 150 human proteins with functions ranging from protease inhibition, cell signalling, adhesion, transcytotic transport, enzymatic function and complement regulation[10].

Integral membrane proteins are embedded into the bilayer and typically cannot be removed without disrupting the membrane[1,3,4]. Integral membrane protein folds typically expose surface residues that are suited for integration into or association with the hydrophobic non-polar region of the lipid bilayer, and thus they interact closely with nearby lipids as well as other proteins. In both α-helical and β-barrel folds hydrogen bonding within the hydrophobic backbone is fully satisfied, and the backbone is protected from unfavourable interactions with the external lipid environment. It is these characteristics of these folds mean that they are the predominant structures in all proteins that are integrated into the membrane lipid bilayer[3].

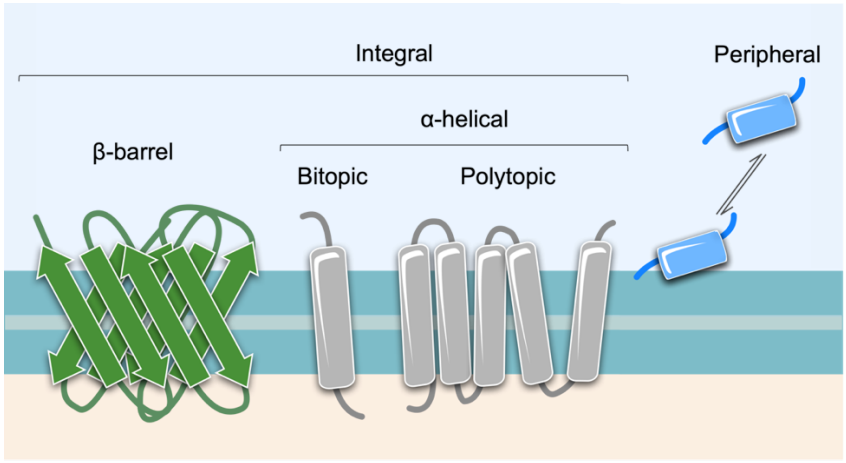


Figure 1.1 Membrane proteins.

Peripheral membrane proteins (blue) are proteins that do not interact with the hydrophobic core of the lipid-bilayer. Instead, they bind at the membrane surface through interactions with lipid polar head groups, or non-covalent interactions with integral membrane proteins[8,9]. An important family of peripheral membrane proteins is phospholipases. Phospholipases cleave phospholipids in the bilayer and some, for example phospholipase C and A_2, produce secondary messengers and therefore, are key regulatory enzymes[3,4].Conversely, integral membrane proteins (IMPs) (grey and green) are proteins that are permanently embedded in the lipid bilayer. Furthermore, transmembrane proteins may bitopic ("single-pass") or polytopic, where they span the bilayer multiple times[11]. Transmembrane proteins are largely α-helical (grey), and bundles of α-helices are found almost exclusively in subcellular compartment and cytoplasmic membranes[3,12]. β-barrel transmembrane proteins (green) are found almost exclusively in the outer membranes of gram-negative bacteria, mitochondria and chloroplasts[3,12,13].

1.2.3 The role of the lipid bilayer and integral membrane proteins

1.2.3.1 Hydrophobic matching

As discussed in 1.2.2, integral membrane proteins feature hydrophobic surfaces that allow for (or require) incorporation into the hydrophobic core of the lipid bilayer. Through hydrophobic effects, these hydrophobic surfaces are coupled to the lipids within the bilayer (Figure 1.2). This is known as hydrophobic matching, and it is this coupling that influences protein sorting and the integration into, secretion through, and folding of proteins within the bilayer[14]. Hydrophobic matching can also influence the lipid structure in the membrane and can contribute to membrane distortion[5,7]. For example, if mismatching occurs, the lipid bilayer can distort to avoid unfavourable exposure of the hydrophobic regions of the membrane protein (Figure 1.2). In addition, hydrophobic mismatch may cause the protein to undergo adaptations, such as conformational rearrangements or tilting, to avoid exposure[5,7,14].

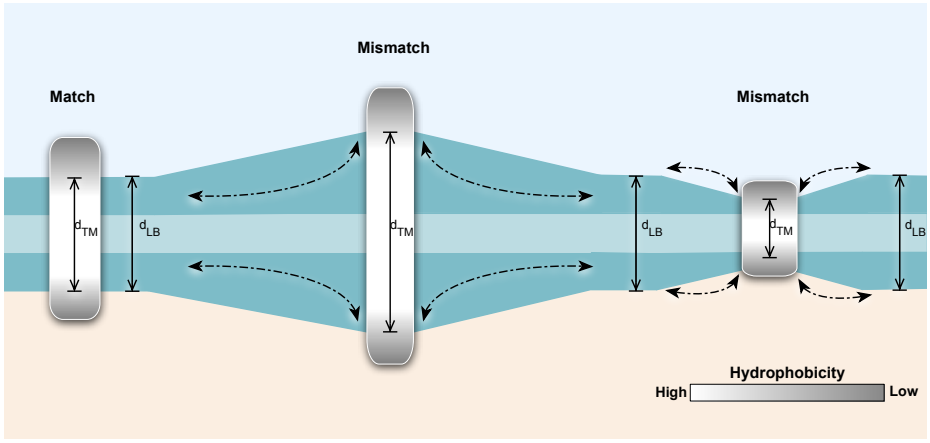


Figure 1.2 Hydrophobic matching in the lipid bilayer.

Hydrophobic matching is where the hydrophobic surface of the transmembrane domain (dTM) matches the length of the lipid bilayer (dLB). When mismatch occurs, this may cause lipid stretching or compression (dashed arrows) to avoid exposure of protein hydrophobic domains to aqueous environments, thus resulting in membrane distortions and curvatures.

1.2.3.2 Lipid-protein interactions

Membranes are dynamic structures and there are constant movements within the bilayer, both laterally, in the plane of the two-dimensional matrix, and transversely, across the bilayer[4,8,14,15]. Hydrophobic matching (1.2.3.1) can be accommodated for because of this inherent dynamic nature. Furthermore, these dynamics mean that the lipid bilayer can feature diverse, complex and time dependent lipid compositions and spatial arrangements[16]. Subsequently, lipid composition influences the local bilayer environment and characteristics, such as lateral pressure, membrane fluidity and packing, surface charge distribution, or the segregation of membrane microdomains. For example, the fused ring system within sterols provide a rigidity that is greater than encountered in other membrane lipids. Therefore, lateral microdomains (such as the purported lipid rafts), that are enriched with cholesterol and sphingolipids, have less fluidity and increased thickness than the bulk lipids in the bilayer. These microdomains are suggested to function as platforms for protein and lipid associations and are also typically enriched with GPI-anchored proteins (1.2.2). Similarly, phosphatidylinositol phosphates (PIPs) and other anionic lipids can create an overall negative electrostatic field, which is likely to facilitate the association of peripheral proteins to the inner membrane surface[16]. Thus, these domains, with differing lipid enrichments/types have an important role in cellular functions. In addition, movements within the membrane allow for temporal protein-protein and protein-lipid associations that are important to membrane function, and studies have shown that membrane protein functionality is influenced by specific lipid-protein interactions, suggesting possible lipid roles as allosteric modulators[17,18] (Figure 1.3).

While, the lipid bilayer contributes significantly to membrane protein structure, function, and regulation[6,16,19-23] (Figure 1.3), the topic is incredibly diverse and complex and is beyond the scope of this study. Therefore, this will not be discussed in any further detail here.

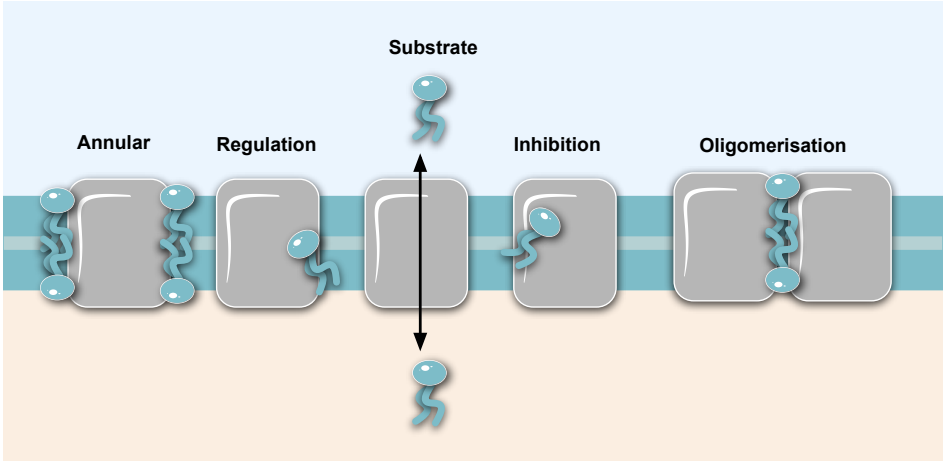


Figure 1.3 The relationships of integral membrane proteins and the lipid bilayer.

Annular lipids modulate local environmental factors, such as lateral pressure, membrane fluidity and packing, and segregation of microdomains, such as lipid rafts. Specific lipid-protein interactions can contribute to activity regulation or may function as substrates or inhibitors and can promote oligomerisation. For example, interactions between the oligosaccharide:H^+ symporter LacY and the polar headgroup of the phospholipid phosphatidylethanolamine (PE) have been shown to be crucial for proton coupling[18]. In addition, lipid binding at the dimer interface has been shown to stabilise the functional oligomerisation of several transporters and receptors, such as interactions with the polar headgroup of PE and the purine transporter UapA[22]. Finally, phosphatidylinositol bis-phosphate (PIP_2) has been shown to be a direct modulator of several membrane proteins, such as the receptor tyrosine kinase, epidermal growth factor receptor[24]; and the K^+ channels KcsA[17], Kir2.2[25] and KvAP[26].

1.2.4 Pharmacological importance of integral membrane proteins

Integral membrane proteins make up approximately ~30% of the human proteome[27], and thus represent a significant target for drug discovery. ~60% of all drug targets are located at the cell surface[28,29]. However, most therapeutics that bind membrane proteins do so within solvated regions outside the lipid bilayer[30]. The human G-protein coupled receptors (GPCRs) represent the largest and most frequently targeted family of membrane proteins[31], and over 50% of membrane protein drug targets come from only four key families; GPCRs, nuclear receptors, and ligand-gated and voltage-gated ion channels[28].

1.3 Membrane transport

The lipid bilayer is impermeable to most molecules and compounds, except for a few small, electrically neutral and lipid soluble molecules, such as oxygen and carbon dioxide. These small molecules can freely diffuse across membranes down their concentration gradient *via* simple

diffusion. However, to allow the uptake of nutrients and therapeutic agents and the removal of waste and harmful products, the transport of charged and/or large molecules with low lipid solubility is essential. The transport of these molecules is facilitated by membrane transport proteins[32] (Figure 1.4). Despite their biological importance, transport proteins are significantly underrepresented as targets of current available therapeutics[29,33-36]. Transport proteins account for less than 10% of all human drug targets, with solute carriers and active transporters accounting for only 2.8% and 1.5%, respectively[31,37,38]. Pores and channel proteins facilitate transport of small molecules at a rapid rate. They may function with low specificity as seen in bacterial porins, or very high specificity as seen in the Na^+ and K^+ ion channels. As discussed, the highly specific ion channels represent a large percentage of membrane protein drug targets, such as gamma-aminobutyric acid channels as a target for neuromodulating therapies, and transient receptor potential channels in the treatment of neuropathies[28]. Thus, ion channels are therapeutically important[28]. However, their function is distinct from that of membrane transporter proteins and will not be discussed in any further detail here.

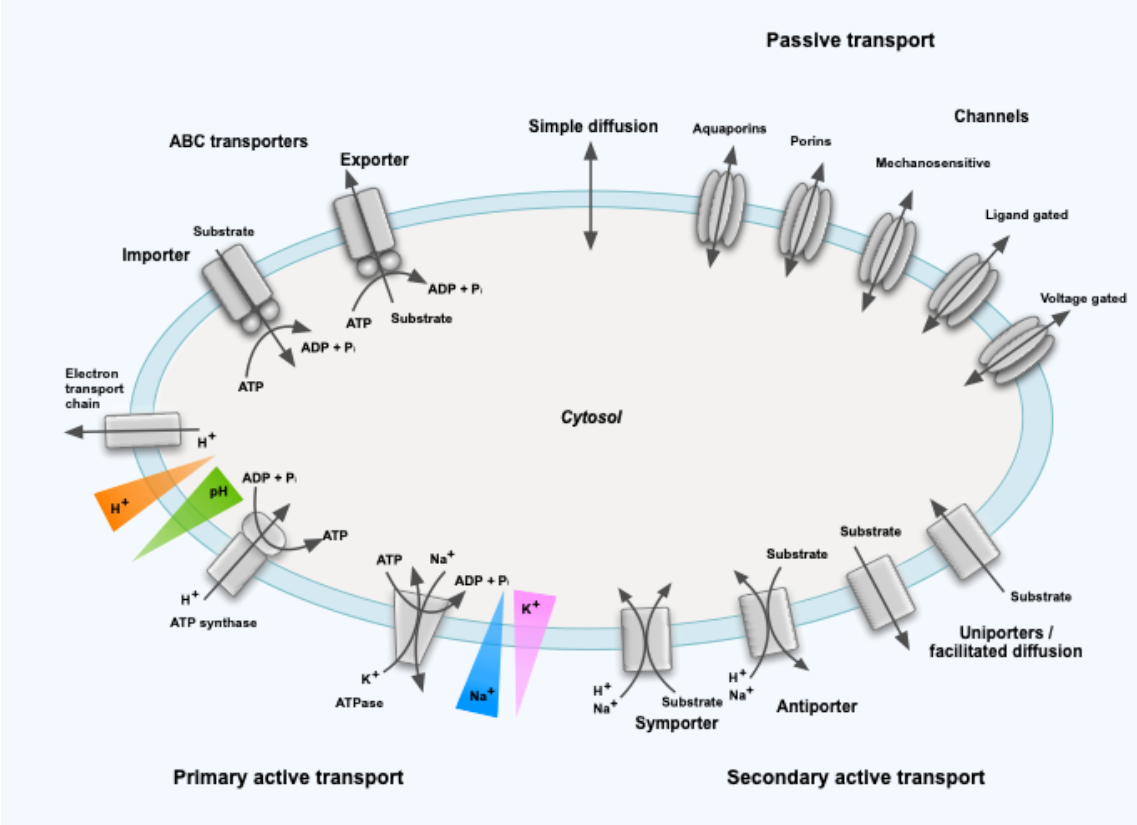


Figure 1.4 Mechanisms of membrane transport.

Primary active transporters directly couple transport of a substrate to an energy producing process. ABC transporters utilise the energy generated from ATP hydrolysis to drive transport. ATP synthase utilises the electrical potential and pH gradient (green), established during respiration, for the transport of protons down their electrochemical gradient (orange). Proton transport is coupled to the synthesis of ATP. In eukaryotes, the P-type Na^+/K^+ H^+ ATPase establishes an alternate system, the Na^+ gradient, during ATP hydrolysis. Here, the hydrolysis of one ATP molecule is coupled to the movement of three Na+ ions to the outside of the cell and two K^+ ions to the inside of the cell and establishes the high intracellular K^+ gradient (pink) and high extracellular Na^+ gradient (blue). These gradients are used to drive the symport and antiport in secondary active transport systems. Uniporters are an additional subclass that transport by facilitated diffusion. Passive transporters, such as channels, may be highly specific and gated by ions or ligands, or may be modulated by stressors such as mechanical force. Porins in bacterial outer membranes allow for the rapid transport of small polar solutes, such as nutrients and waste products. Aquaporins allow for the passage of water molecules at extremely high rates and are selectively expressed in cells that perform rapid transport of water. Small, uncharged, lipophilic molecules can be transported by simply diffusion.

1.3.1 Facilitated diffusion

For every molecule requiring transport there is, generally, a membrane protein with high specificity that assists the diffusion of a single molecule down its concentration gradient, *via* facilitated diffusion[39]. No additional energy is required to transport the solute and the method provides solute transport at a rate that is orders of magnitudes faster than simple passive diffusion. However, the rate of transport achieved in facilitated diffusion is insufficient for competitive metabolism. Therefore, the need for an increased rate of transport gave rise to the

evolution of active transport mechanisms[39] (Figure 1.4). Active transport mechanisms couple metabolic energy to a substrate to catalyse the movement against its electrochemical gradient and, depending on the energy used, are subdivided into primary or secondary active transport[40-42]. While insufficient for competitive metabolism, in environments where high concentrations of solutes are already established, i.e., intracellular compartments in eukaryotes or the blood-cell interface, transport often occurs by facilitated diffusion[39].

1.3.2 Primary active transport

Primary active transport directly couples the transport of a substrate to an energy producing process, such as respiration or ATP hydrolysis. For example, the ATP binding cassette (ABC) transporters utilise the free energy that is released during ATP hydrolysis to drive the transport of a substrate[43,44] (Figure 1.4). ABC transporters represent a major class of primary active transporters. They form one of the largest of all protein families and can be found in all kingdoms of life[43,44]. Canonical ABC transporters feature two hydrophobic transmembrane domains, and two hydrophilic nucleotide binding domains. They have broad and diverse substrate specificities, with regards to both size and chemical nature, allowing the translocation of small ions to large macromolecules, such as lipids, oligopeptides, and oligosaccharides. Furthermore, they can function as importers or exporters. However, importers are suggested to be exclusively found in prokaryotic organisms[43,44]. Using an alternating access mechanism of transport that is similar to the mechanism seen in the major facilitator superfamily of secondary transporters (which will be discussed in more detail in 1.4.2.2), ABC transporters undergo conformational rearrangements of the transmembrane domains, transitioning between inward and outward facing, to drive substrate translocation across the membrane[41,43-48]. However, the distinct difference between primary active transport in ABC transporters, and secondary active transport in MFS, is that ABC transporters feature a catalytic domain. These domains, the nucleotide binding domains, perform ATP hydrolysis and use the free energy that is released from this reaction to drive the large conformational rearrangements that are required for transport[41,43,47].

Another example is that of the F-type H^+ ATP synthase, which is localised to the inner mitochondrial membrane in eukaryotes. During respiration, energy is released in the form of electrons following the breakdown of hydrogen from energy rich sources, such as glucose. The transport of these free electrons across the closed membrane to an electron acceptor, *via* the electron transport chain, establishes a pH gradient and electrical potential across the membrane. The established gradient, known as the proton motive force (PMF), is a store of

energy that allows the transport of protons back across the membrane down their electrochemical gradient *via* the ATP synthase (Figure 1.4), which couples the transport of protons to the synthesis of ATP from ADP and organic phosphate[39].

1.3.3 Secondary active transport

Secondary active transport utilises the potential energy of an ion electrochemical gradient generated by primary active transport, such as the Na^+ gradient or the H^+ gradient associated to the PMF, to couple the transport of a substrate to the transport of an ion [40,42,49,50]. Here, the ion is the driving substance, and the binding of the ion facilitates the binding of the substrate[40,51]. Therefore, if an ion and substrate gradient are both present, the ion would drive the direction of transport regardless of the direction of the substrate gradient, thereby enabling uphill transport. This coupling of one-to-one in secondary active transport is consequently, a more energy-efficient system than that of primary active transport.

The direction of substrate transport relative to the direction of the driving substance allows for further subdivision into symporters and antiporters (Figure 1.4). While symporters and antiporters both transport both the substrate and ion(s), in symport, the direction of the transport of the substrate and the ion is coupled. Meaning that they move in the same direction and are transported together. However, in antiport, transport of the substrate and the ion are in opposing directions and the binding of one is dependent on the prior release of the other. Uniporters are considered an additional subclass, in which they are thought of as secondary active transporters that have lost the energy coupling[51]. If there is an absence of driving substance, only the single substrate molecule is transported down its concentration gradient, *via* facilitated diffusion.

Many secondary active transporters are monomeric polypeptides comprising of 12 transmembrane α-helices (TMs), although some transporters may have 10-14 TMs and may also be comprised of multimeric subunits[40,42,49,50]. Furthermore, transporter multimers may form further complexes with other membrane proteins to establish powerful and polyspecific transport systems, i.e. the tripartite resistance-nodulation-cell division efflux pumps from *Escherichia coli* which are comprised of a trimeric inner membrane protein (AcrB), a membrane fusion protein (AcrA) and an outer membrane protein (TolC)[52]. Transporter proteins with shared structural folds have been shown to participate in different mechanisms of secondary active transport[40], and a number of PMF-driven transporters have been shown to behave like uniporters when under a strong concentration gradient of the substrate[53]. Therefore, it is

proposed that there are no unique structural features that distinguish uniporters, symporters, or antiporters.

1.4 Transporter proteins

1.4.1 Solute carriers

The solute carrier (SLC) superfamily is the second largest family of membrane proteins in the human genome[34]. The human genome encodes for over 550 SLC transporters[18], and more than 80 human SLC transporters have been implicated in rare, monogenic disorders[36]. SLC transport mechanisms are predominantly facilitative or secondary active, and the inhibition of transport is a commonly used strategy for the currently available SLC targeted therapeutics[36]. SLCs can be subdivided into 65 families[54,55], with family assignment determined by sequence identity. However, many SLCs have an undefined relationship and the range of specificities within a single family may be broad. Furthermore, analysis of primary sequence structure by hidden Markov models identified that approximately one-third of all SLCs also share classification with the major facilitator superfamily (MFS)[56].

1.4.2 Major facilitator superfamily transporters

The MFS transporters are one of the largest and most ubiquitous transporter superfamilies[57]. In prokaryotes, ~25% of transporters belong to the MFS family, and in the human genome genes that encode for at least 110 MFS proteins have been identified[58]. MFS proteins transport a wide range of substrates, such as ions, carbohydrates, lipids, amino acids, nucleosides, and a number of other small hydrophilic molecules[59]. The number of MFS proteins relative to the vast and diverse number of substrates that they transport is quite small, so it is proposed that poly-specificity is likely to be a common characteristic of MFS transporters[39,53,60]. Over the past 10 years, there have been several high-resolution structures published of MFS transporters in a series of diverse conformations[61-67]. These structures have provided insight into the molecular basis of transport and enabled the modelling of the mechanism of action in MFS[50-53,57,58,68,69].

1.4.2.1 MFS Transporter architecture

The canonical architecture/topology of MFS comprises 12 transmembrane α-helices which are organised into two domains, an N-terminal, and a C-terminal domain, with a pseudo-symmetrical 6+6 topology. In each of these domains the six helices are arranged in a pair of 3+3 inverted repeats, and the two domains joined by a conserved core fold. Experimental evidence

indicates that the three-helix bundles represent the basic structural and functional unit in MFS transporters[68] (Figure 1.5). The first helix in each repeat (i.e., TM1, 4, 7 and 10) contribute to the formation of the central cavity, which serves as both the binding site and the transport path for substrates. These helices are also often observed as discontinuous helices which undergo rearrangement (bending and unfolding) for the transition to an intermediate occluded state upon substrate binding. The second set of helices (i.e., TM2, 5, 8 and 11) form the side walls of the central cavity and are usually long and curved; these helices are directly involved in inter-domain conformational changes. The third set of helices (i.e., TM3, 6, 9 and 12) do not contribute to the formation of the cavity but instead are located at either end of the transmembrane core. Unlike the other two groups, this third helix group is less likely to move during major conformational changes. These helices are in contact with the lipid bilayer and are involved in interactions with the surrounding membrane[53,69,70]. MFS transporters also have a highly conserved motif, the A-motif, at the intracellular loop between TM2 and TM3 and/or TM8 and TM9 (Figure 1.5 and Figure 1.6). The A-motif (GXXX(D/E)(R/K)XG(R/K)(R/K)) is the most conserved motif in MFS. Mutational studies have shown that mutations at the acidic fifth residue and/or substitutions of the first glycine for bulky residues dramatically inhibits activity. However, transport was minimally affected following mutation of the basic residues or the glycine at the eighth position.[71]. This motif is also seen to interact with TMs to provide conformational stabilisation[53,68,71] (Figure 1.6). Therefore, the A-motif is essential for the transport activity in many MFS transporters.

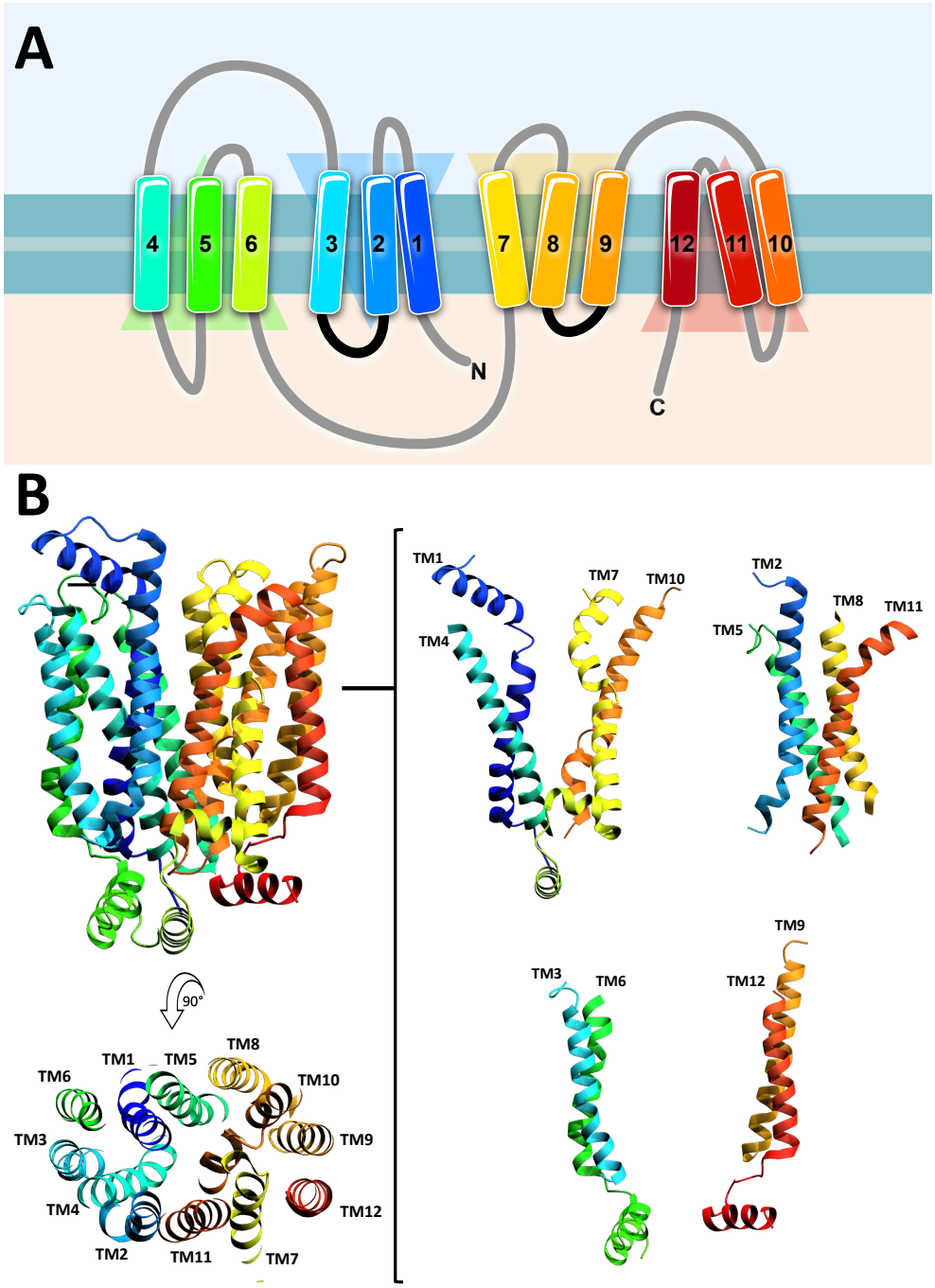


Figure 1.5 MFS Transporter architecture.

(A) The 12 α-helices of MFS transporters are shown with a rainbow representation, with blue-green depicting the N-terminus and yellow-red the C-terminal. The organisation into three helical bundle repeats (TM1-3, 4-6, 7-9 and 10-12) are represented by triangles. Sites of the A-motif, at the intracellular loop between TM2 and TM3 and/or TM8 and TM9, are highlighted in black. (B) The arrangement of the helices in the structure of a representative MFS uniporter, the mammalian glucose transporter GLUT3 (PDB: 4ZWB). The arrangements of the first, second and third helix in each three-helix bundle repeat are shown, demonstrating the formation of the central cavity by the discontinuous first helices, the side walls of the central cavity by the long and curved second helices, and the third helices located to the outside of the transmembrane core.

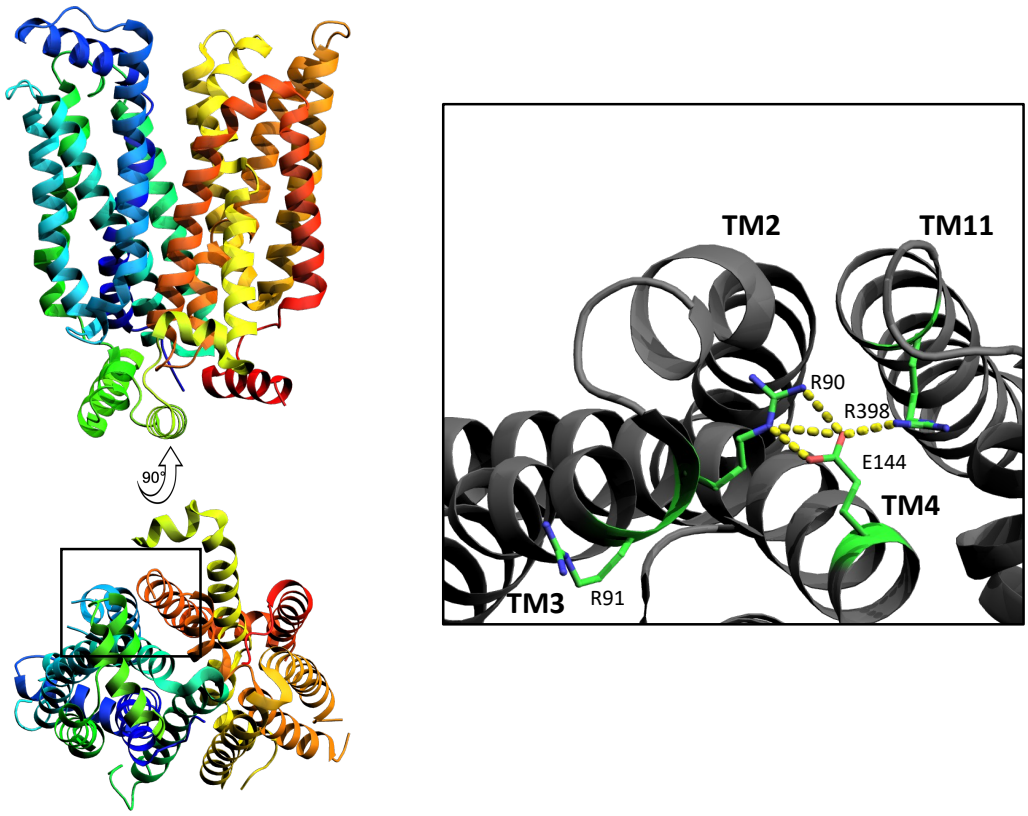


Figure 1.6 The A-motif of MFS.

In the structure of a representative MFS uniporter, GLUT3, in an outward-facing conformation (PDB: 4ZBW), R90 and R91 of the A-motif at the loop between TM2 and TM3 is seen to interact with charged residues on the TM4 and TM11.

1.4.2.2 The alternating access mechanism of action

The general model for substrate translocation by all MFS involves an alternating access mechanism, in which the transporter undergoes large conformational rearrangements to adopt one of two major alternating conformations, outward-facing (C_o) and inward-facing (C_i)[51,53,58,69]. Based on the understanding of the molecular basis of transport, informed by experimentally solved structures of MFS, the alternating access model can be further classified into three differing mechanisms: rocker-switch, rocking bundle, and elevator[18,39,72] (Figure 1.7). While these defined states represent major conformations of the transport cycle, transitions throughout the cycle requires the transporter to adopt a series of discrete intermediate states (Figure 1.8). During the large conformational rearrangements gating interactions, such as salt bridge formations, are established when residues are moved sufficiently close to one another and break as they move apart. These interactions contribute to these differences in intermediate states and have also been shown to represent sites of function and regulation[51,58,73-75].

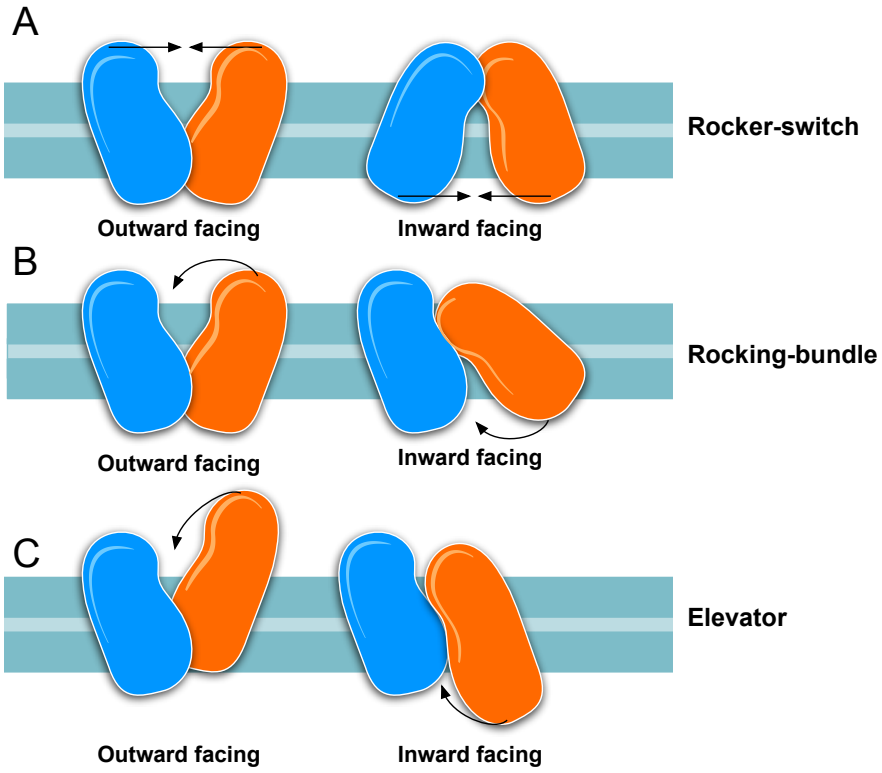


Figure 1.7. Alternating access models of MFS transporters.

A single binding site within a polar cavity is accessible to only one side of the membrane at a time. (A) In the rocker-switch mechanism, structurally similar N-terminal (blue) and C-terminal (orange) domains maintain their overall structure as approximate rigid bodies, and both domains undergo a symmetrical change of relative position. (B) In the rocking-bundle mechanism, structurally dissimilar N- and C-terminal domains undergo asymmetrical rearrangement, in which one 'transport' domain (orange) rearranges against a less labile 'scaffold' domain (blue)[18]. In both rocking-switch and rocking-bundle model rearrangements occur over a rotation axis that crosses the central substrate-binding site at the domain interface[51]. (C) The elevator mechanism also utilises a 'scaffold' N-terminal domain and a 'transport' C-terminal domain. However, here the two domains are highly divergent, and the N-terminal domain contributes to a state of oligomerisation, thus remains static and fixed. In all models, these large conformational transitions can also result in hydrophobic mismatching with the local lipid bilayer. Subsequently, this results in the stretching or compression and bending/curving of the local lipid bilayer[5,7,14,20]. In addition to driving substrate transport, these large conformational changes are also proposed to permit signal transduction[76].

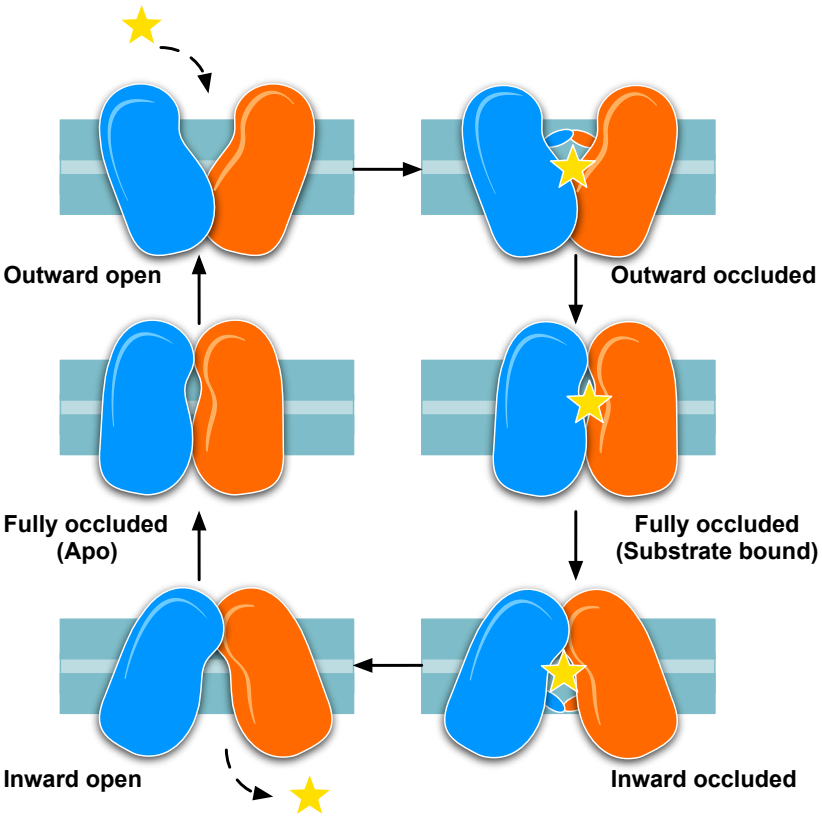


Figure 1.8 An example of the rocker switch transport cycle of an MFS uniporter.

A single binding site is accessible to only one side of the membrane, representing an open state. Binding of the substrate (yellow star) induces a series of interactions between the helices of the central cavity which results in a narrowing of the pore and the formation of a thin gate, representing an intermediate occluded state. In some cases, this binding causes the N- and C-terminal domains to undergo subtle rotations, resulting in a more compact conformation than in the substrate-free (*Apo*) state. Even if compacted, these occluded states still occupy a distinct inward-facing or outward-facing conformation[51,77]. This occluded state and gate formation results in a change in binding affinity between the two domains which destabilises the interface. Destabilisation of the interface results in tilting of the domains, which in turn initiates the closing of the open pore and the occupation of a fully occluded state. This is followed by the formation and opening of a pore on the opposing side, and finally, the release of the substrate[18,51,53,58].

1.4.2.3 The energetics of MFS transport

MFS members are predominantly secondary active transporters, although the mammalian glucose transporter (GLUT) family represent a seminal MFS uniporter family. It is suggested that the MFS substrate binding and transport process is analogous to that in induced-fit enzyme catalysis[41]. In which, initially the substrate binds poorly to the transporter in the open, ground-state. However, substrate binding triggers rearrangements that result in a transition to an occluded state, which in turn increases binding affinities and free energy. This energy is used to further drive the large conformational rearrangements required to transition from the ground state of one major conformation, to the excited state of the opposing major conformation. Subsequently, the reversion from an excited state to a ground state is driven by the conformational energy stored in the excited state and is more or less automatic[40,53]. However, the inherent nature of uniport requires a transition between the two major conformational states to occur in the absence of substrate. Therefore, it is proposed that the major conformational states of uniporters are energetically similar, and that the energy landscape for the transition between states is relatively flat. It is predicted that it is only the interactions of the substrate on binding or release that increases the probability of transition, and thus promotes efficient transport. In symporters and antiporters, as the molecules being transported are intrinsically higher in energy, PMF etc, the energy barrier between the ground and excited state is expected to be higher[41,53,78].

1.5 Nucleoside Transporters

Nucleosides are essential signalling molecules in purinergic signalling pathways. These pathways modulate many cellular processes. Adenosine signalling is the most common and significant of the purinergic signalling pathways and participates in many processes in the cardiovascular and central nervous systems, such as vasodilation, inflammation, and neuromodulation[79-81]. Adenosine signalling is mediated by GPCRs (A_1, A_{2A}, A_{2B}, A_3), and aberrant signalling has been implicated in numerous pathologies[81-85]. Nucleosides are also precursors for nucleotides synthesised *via* salvage pathways. These salvage pathways are crucial in cells that lack *de novo* biosynthetic pathways, such as erythrocytes, leukocytes, bone marrow cells and some cells in the brain. Synthesis *via* salvage pathways accounts for 90% of the daily nucleotide synthesis in humans[86]. Nucleotides have a variety of roles such as nucleic acid precursors, energy currency, secondary messengers in signal transduction, activated metabolic intermediates and components of cofactors (Figure 1.9).

Therefore, the uptake of nucleosides is vital for many important biological processes. Owing to the relatively hydrophilic nature of nucleosides, their uptake is dependent on membrane transport proteins. Therefore, nucleoside transporters mediate nucleoside and nucleotide homeostasis, and thus modulate biological processes by regulating the concentration and the availability of signalling molecules to cell-surface receptors. Therefore, nucleoside transporters represent an important drug target[86-89]. In eukaryotes, there are two evolutionarily unrelated classes of membrane transport proteins that mediate the transport of nucleosides: the SLC28 family of concentrative nucleoside transporters (CNTs), and SLC29 family of equilibrative nucleoside transporters (ENTs)[86,90,91] (Figure 1.9). CNTs and ENTs are not found in Gram-negative bacteria. Instead, nucleoside transport is mediated by proteins such as the nucleoside specific outer membrane transporter Tsx[92], and nucleoside proton symporters (NHS), such as NupG and NupC in *E. coli*[86,93]. Like CNTs and ENTs, NupC and NupG are evolutionarily distinct. NupC is a member of the CNT family[94], but NupG is unrelated to ENTs. However, NupG has been shown to an MFS member. In 2021 the first structures of NupG were determined in an inward-open conformation, with a topology consistent with MFS transporters[93]. However, owing to the limited structural insight into further conformational states required in MFS transport, the molecular basis for the mechanism of action remains undetermined.

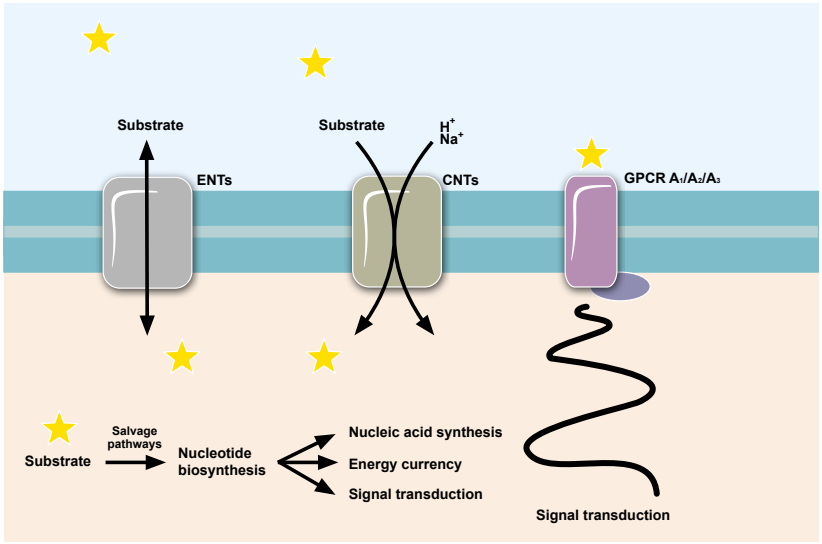


Figure 1.9 Overview of nucleoside transport.

(A) Equilibrative nucleoside transporters (ENTs) and concentrative nucleoside transporters (CNTs) mediate nucleoside and nucleoside homeostastis, and thus regulate the concentration and availability of signalling molecules, such as adenosine (yellow), to cell surface receptors, such as GPCRs.

Like nucleosides, purine nucleobases are vital metabolites in salvage pathways. In mammals, ENTs also play a significant role in the uptake of nucleobases[90,95,96] and as of June 2022, ENTs are the only identified and characterised nucleobase transporters in humans[95]. The transport of

nucleobases in other species and kingdoms is typically mediated by members of the nucleobase:cation symporter-1 (NCS-1) and the nucleobase/ascorbate transporter (NAT) families. Both NCS-1 and NATs are members of the amino acid-polyamine-organocation (APC) superfamily, which is the second largest superfamily of secondary transporters after MFS[97]. They are evolutionarily distinct from MFS, and thus will not be discussed in further detail in this work.

1.5.1 SLC28, the concentrative nucleoside transporter family

SLC28 transporter proteins are a family of sodium or proton dependent CNTs which are found in eubacteria and eukaryotes. There are three members of the CNT family, CNT1, CNT2 and CNT3 (SLC28A1, SLC28A2 and SLC28A3, respectively), with each of the members conferring different substrate specificities[91,98]. In mammals CNT1 is pyrimidine selective, but with high adenosine affinity. It is found primarily in epithelial tissues and is localised to the apical membrane. It is suggested that the apical CNT1 works in conjunction with basolateral membrane localised ENTs for the mediation of transepithelial flux of nucleosides[98-100]. While the function of this epithelial co-operativity is not well characterised, the presence of CNT1 at the apical membrane allows for the uptake of nucleosides from the lumen against a concentration gradient, Thus, this co-operativity likely allows for the intestinal absorption of nucleosides and nucleoside analogues, and the renal recovery of essential nucleosides from urine. Furthermore, this co-operativity may serve as mechanism to maintain crucial nucleoside pools for cells that depend on salvage pathways for nucleotide biosynthesis, such as those in the brain.

CNT2 is purine selective and is widely distributed, with localisation to the plasma membrane. CNT3 is the most extensively studied CNT. It has the broadest tissue distribution and most abundant expression of the CNTs. It is also localised to the plasma membrane, although intracellular localisation has also been observed. Furthermore, CNT3 has also been shown to function with either proton or sodium dependence, with sodium/nucleoside cotransport exhibiting a unique stoichiometry of 2:1[101-105]. Structures have been solved for the human CNT3[105] (Figure 1.10) and a bacterial homologue[103] using cryo-EM. These structures demonstrated CNTs form a trimeric homodimer and utilise a multi-step elevator-like mechanism for an alternating-access mechanism of transport, with independent substrate transport achieved in each protomer[102,105].

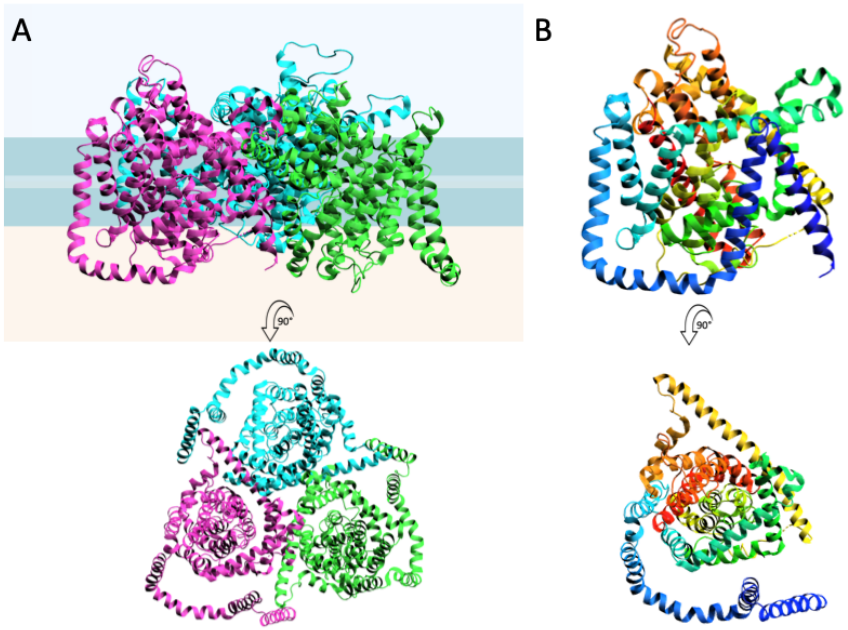


Figure 1.10 The structure of concentrative nucleoside transporters.

Panel (**A**) shows the trimeric homodimer of the human concentrative nucleoside transporter (PDB: 6KSW) solved by cryo-EM. Each of the monomers highlighted in a distinct colour (pink, blue and green). Panel (**B**) shows a rainbow representation of each of the monomers, with blue-green colouring depicting the N-terminus, and yellow-red the C-terminus. Both structures are presented at a view perpendicular to the membrane in the upper panel, and top-down in the lower panel.

1.5.2 SLC29, the equilibrative nucleoside transporter family

SLC29 transporter proteins are a family of sodium-independent ENTs that are found exclusively in eukaryotes. In mammals there are four members of ENTs, ENT1, ENT2, ENT3 and ENT4 (SLC29A1, SLC29A2, SLCA3 and SLCA4, respectively)[90,99,106]. Representative ENTs have been identified in plants, fungi, insects, nematodes, and protozoa[99,106,107]. Efforts to understand the structure and function of ENTs, and nucleoside transporters as a whole, was initiated by Professor Stephen Baldwin and Professor James Young in 1981[108]. Since then, the functions of ENTs have been well characterised. Following their discovery in 1997, ENT1 and ENT2 have been the most extensively studied of the human ENTs (hENTs)[109,110]. hENT3 and hENT4 were discovered later in 2005[111] and 2006[112], respectively, and are less well understood. Structurally, all ENTs are proposed to share the same membrane architecture, comprising of 11 TMs, an intracellular N-terminus, and an extracellular C-terminus[113] (Figure 1.11). Furthermore, all ENTs are predicted to have a conserved architecture that is shared with MFS transporters.

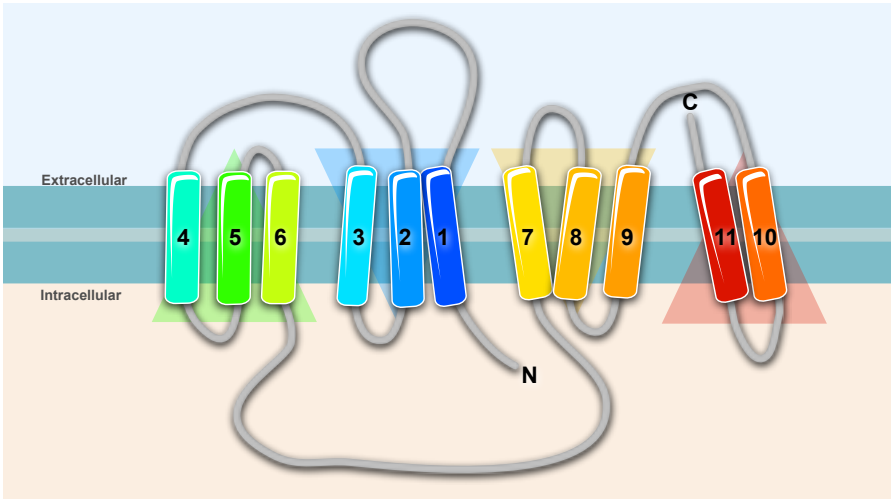


Figure 1.11 The topology of equilibrative nucleoside transporters.

The 11 α-helices of ENTs are shown with a rainbow representation, with blue-green depicting the N-terminus and yellow-red the C-terminal. The organisation into three and two helical bundle repeats (TM1-3, 4-6, 7-9 and 10-11) are represented by triangles. Owing the absence of TM12, ENTs have an extracellular C-terminus. In addition to a different orientation of the C-terminus in comparison to MFS transporters, ENTs are expected to have distinct differences in helix and fold arrangements.

1.5.2.1 The human ENT3 and ENT4

hENT3 is widely expressed across many tissue types but is particularly abundant in heart and liver tissues. It is primarily localised to intracellular membranes. Owing to this intracellular localisation, nucleoside transport in hENT3 is pH dependent and optimal transport is achieved at pH 5.5. It is broadly selective for purines and pyrimidines, with some selectivity for nucleobases[111,114,115]. The clinical relevance of hENT3 has been explored extensively, and mutations and dysfunction have been shown to contribute to lysosomal storage disorders[114,116-118].

hENT4 is an evolutionarily divergent ENT. It is widely distributed across tissue types, with an abundance in heart and skeletal muscles. It is also primarily localised to the plasma membrane, although under normal physiological conditions, it only transports organic cations. Thus, hENT4 is more commonly known as a plasma membrane monoamine transporter (PMAT)[90,99,108]. The transport of nucleosides by hENT4 is optimal at pH 5.5. Therefore, hENT4 is suggested to be an insignificant member of the ENT family under normal physiological conditions. It is suggested that hENT4 instead contributes to nucleoside transport under acidotic conditions where the activity of hENT1 is repressed, such as hypoxia and ischemia[112,119]. While ENTs are broadly considered to transport *via* facilitated diffusion, owing to their pH dependence both hENT3 and hENT4 are suggested to be proton-dependent transporters. However, the function of hENT3 and hENT4 remains poorly understood.

1.5.2.2 The human ENT1 and ENT2

hENT1 and hENT2 are both widely expressed and localised to the plasma membrane, although both are also found in nuclear membranes and the endoplasmic reticulum. hENT2 is also abundant in skeletal muscles. In addition, both hENT1 and hENT2 are also localised to the basolateral membranes in epithelial cells. As previously discussed, this epithelial basolateral localisation is believed to work in conjunction with apically localised CNT1 for the mediation of transepithelial nucleoside flux for the co-operative absorption/recovery and clearance of essential purinergic signalling molecules[98,99]. Both hENT1 and hENT2 are broadly selective in the transport of pyrimidine and purine nucleosides, although hENT2 has a greater ability to transport nucleobases, and hENT1 has a higher affinity for adenosine, with a K_m ~40 µM (vs hENT2 K_m ~ 140 µM)[87]. In addition to substrate specificity, hENT1 and hENT2 are functionally distinguished by their sensitivity to inhibition by the nucleoside analogue nitrobenzylmercaptopurine ribonucleoside (NBMPR). hENT1 is sensitive to inhibition by NBMPR at nM concentrations, whereas hENT2, along with hENT3 and hENT4, are only weakly inhibited and are considered NBMPR insensitive. NBMPR is not a therapeutically relevant compound due to significant off target effects. However, the high affinity and high specificity for hENT1 means that NBMPR is widely used in functional analysis of hENTs. Furthermore, owing to a 1:1 binding ratio, NBMPR can be used to quantify hENT1.

Table 1.1 Overview of human equilibrative nucleoside transporters

Human gene	Family name	Length (bp)	SLC29A1 seq. ID	Transport type	pH^{opt}	Permeants	Expression	Function	Localisation
SLC29A1	hENT1	456	-	Facilitated diffusion	7.0	Nucleosides (Adenosine K_m ~40 µM)	Broad tissue distribution	Nucleoside transport	Plasma membrane
SLC29A2	hENT2	456	46%	Facilitated diffusion	7.0	Nucleosides (Adenosine K_m ~140 µM) Nucleobases	Broad tissue distribution, abundant in skeletal tissue.	Cytoprotection of muscoskeletal system	Plasma membrane
SLC29A3	hENT3	475	29%	Possibly proton dependent	5.5	Nucleosides Adenine	Broad tissue distribution, abundant in heart and liver	Intracellular transport	Intracellular endosomal and mitochondrial membranes
SLC29A4	hENT4 (PMAT)	530	18%	Possibly proton dependent	5.5	Adenosine Organic cations	Broad tissue distribution, abundant in heart and skeletal	Intracellular transport during hypoxia and ischemia	Plasma membrane

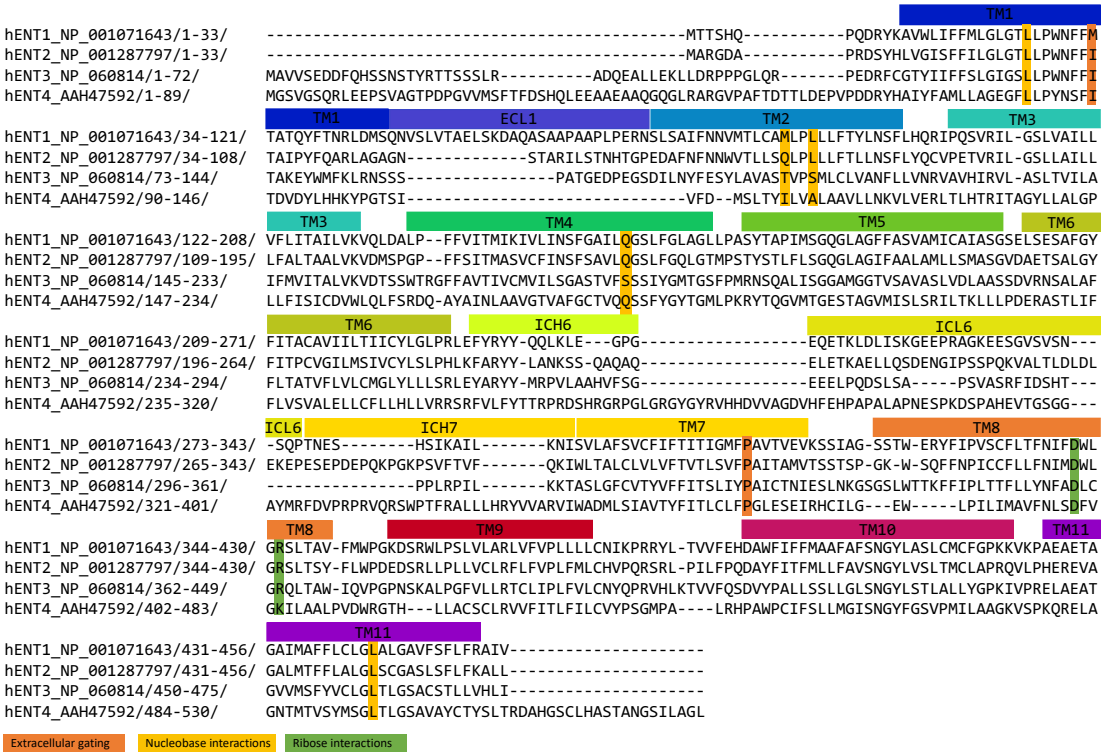


Figure 1.12 Sequence alignment of human ENT isoforms 1-4.

Equivalent residues to those that have been proposed as being involved in extracellular gating interactions in hENT1 are highlighted in orange, and those involved in interactions with the purine and ribose moiety of NBMPR are highlighted in yellow and green, respectively.

1.5.2.3 The biological and pharmacological importance of ENTs

As previously discussed, ENTs contribute to the modulation of several important biological processes by regulating the concentration and the availability of signalling molecules to cell-surface receptors. Adenosine is recognised as a potent cardioprotective agent and neuromodulator[81,83,120]. During periods of metabolic and physiological stress endogenous levels of adenosine increase, which triggers retaliatory actions, such as inhibition of platelet aggregation, anti-inflammation, and vasodilation to protect against damage[79,121-123]. Owing to the significance of adenosine signalling and the preference for adenosine transport, hENT1 is an important target for important adenosine reuptake inhibitors such as dipyridamole, dilazep and draflazine, as used in cardiac therapies. Furthermore, due to elevated adenosine levels triggering neuromodulation responses, ENT targeted adenosine reuptake inhibition has also been proposed as a strategy for the treatment of several neurological disorders and the potentiation of the analgesic effects of opioid adjuncts[86-88].

In addition to inhibition of transport, the uptake of nucleoside analogue drugs (NADs) is another distinct strategy used in the pharmacological targeting of ENTs. NADs therapeutic effect is

exerted through the disruption of nucleic acid repair and synthesis, either *via* nucleic acid integration and subsequently interfering with enzyme function, or direct interactions and inhibition of enzymes. These strategies are utilised in several anticancer and antiviral therapies, with the uptake dependant on the transport *via* ENTs. For example, the hENT2 specific antiviral AZT targets reverse transcriptase, thereby inhibiting viral DNA synthesis. The anticancer cladribine is incorporated into DNA in its active form, thereby halting chain elongation by DNA polymerase, and the anticancer agent gemcitabine inhibits ribonucleotide reductase, thereby inhibiting *de novo* DNA synthesis[88,124]. Both cladribine and gemcitabine are hENT1 specific. Furthermore, it has been demonstrated that there is a correlation between the high levels of hENT1 expressed in cholangiocarcinoma's and increased survival rates following treatment with anticancer therapeutics[125-128]. Therefore, hENT1 expression is often suggested as a predictive biomarker for chemosensitivity to predict the likelihood of therapeutic effect. However, attempts to elucidate biomarker profiles are inherently complex and multifactorial. Furthermore, downregulation of ENT expression or selection of ENT-deficient cells contributes to mechanisms of the significant and ongoing challenge of chemotherapeutic resistance.

ENTs have also been identified as a potential target for antiparasitic drugs. Members of the *Plasmodium* species, which are human pathogenic protists, are typically purine auxotrophic[129]. Therefore, they depend on purine salvage pathways for survivability and the pathogenesis of malaria. 90% of clinical malaria cases are due to transmission *via Plasmodium falciparum or plasmodium vivax*, with *P. falciparum* responsible for approximately 80% of malaria related deaths[130]. The *Plasmodium* species has four ENT family members with low similarity to hENT1 (<17%). ENT1 has been identified as the primary purine import pathway for purine salvage in *Plasmodium* species, but the potent hENT1 inhibitors such as NBMPR and dipyridamole showed little to no efficacy against *Pf*ENT1 or *Pv*ENT1. Previous studies in the development of *Plasmodium* ENT1 inhibitors have identified compounds with up to 1000-fold higher affinity for *pf*ENT1 than is measured in hENT1[129,131]. Thus, the development of *Plasmodium* ENT1 inhibitors is suggested to be a suitable target for antimalarial drug development, with minimal effects on hENT1.

The development of a broad range of new therapeutics is an essential and promising avenue for overcoming chemotherapeutic resistance, improving therapeutic efficiency and efficacy, and treating further pathologies[86,87,132]. However, although ENT targeting is an important therapeutic approach for many pathologies, previous inhibitor studies have demonstrated that

ENTs have low tolerance for diversity in interactions required for high affinity binding. This emphasises the importance of structure guided drug design[133].

1.6 The structure of hENT1

In 2019 the first, and as of June 2022 the only, structures of an ENT were solved by X-ray crystallography[134] (Figure 1.13). Two structures of hENT1 were published in an outward-facing conformation in complex with two chemically distinct inhibitors; PDB: 6OB6, with the nucleoside analogue NBMPR, and PDB:6OB7, with the adenosine reuptake inhibitor dilazep.

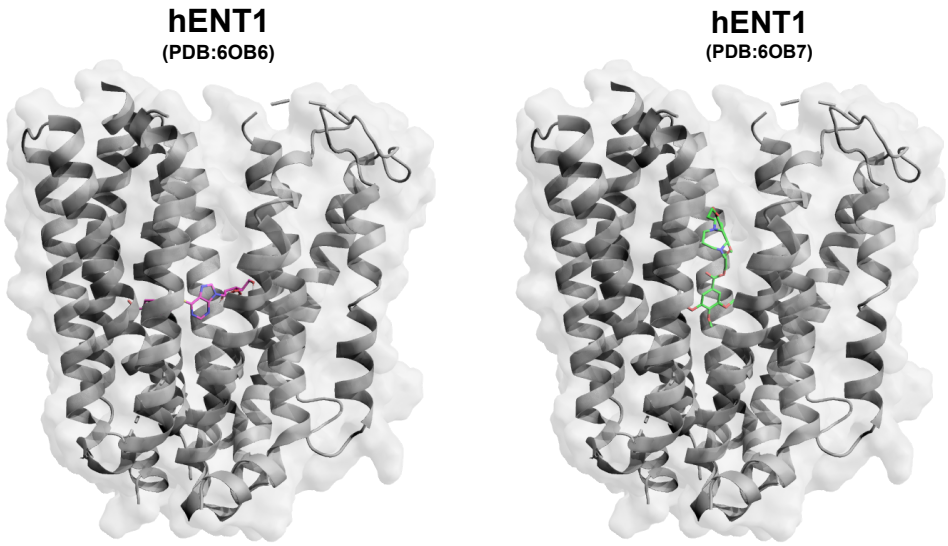


Figure 1.13 Structures of hENT1 solved by X-ray crystallography.

Structures of hENT1 in outward facing conformations, with the chemically distinct inhibitors NBMR (pink) and dilazep (green) bound, respectively.

Comparison of the hENT1 structure to a representative outward-facing MFS X-ray structure of the human GLUT3[77] demonstrated that, as predicted, the fold of hENT1 matches TM1-11 out of 12 in MFS transporters[134]. However, owing to the absence of TM12 in ENTs, features of hENT1 distinct from those canonical in MFS transporters were identified. In hENT1 TM9 is arranged to occupy the space that is shared by TM9 and TM12 in MFS transporters (Figure 1.14).

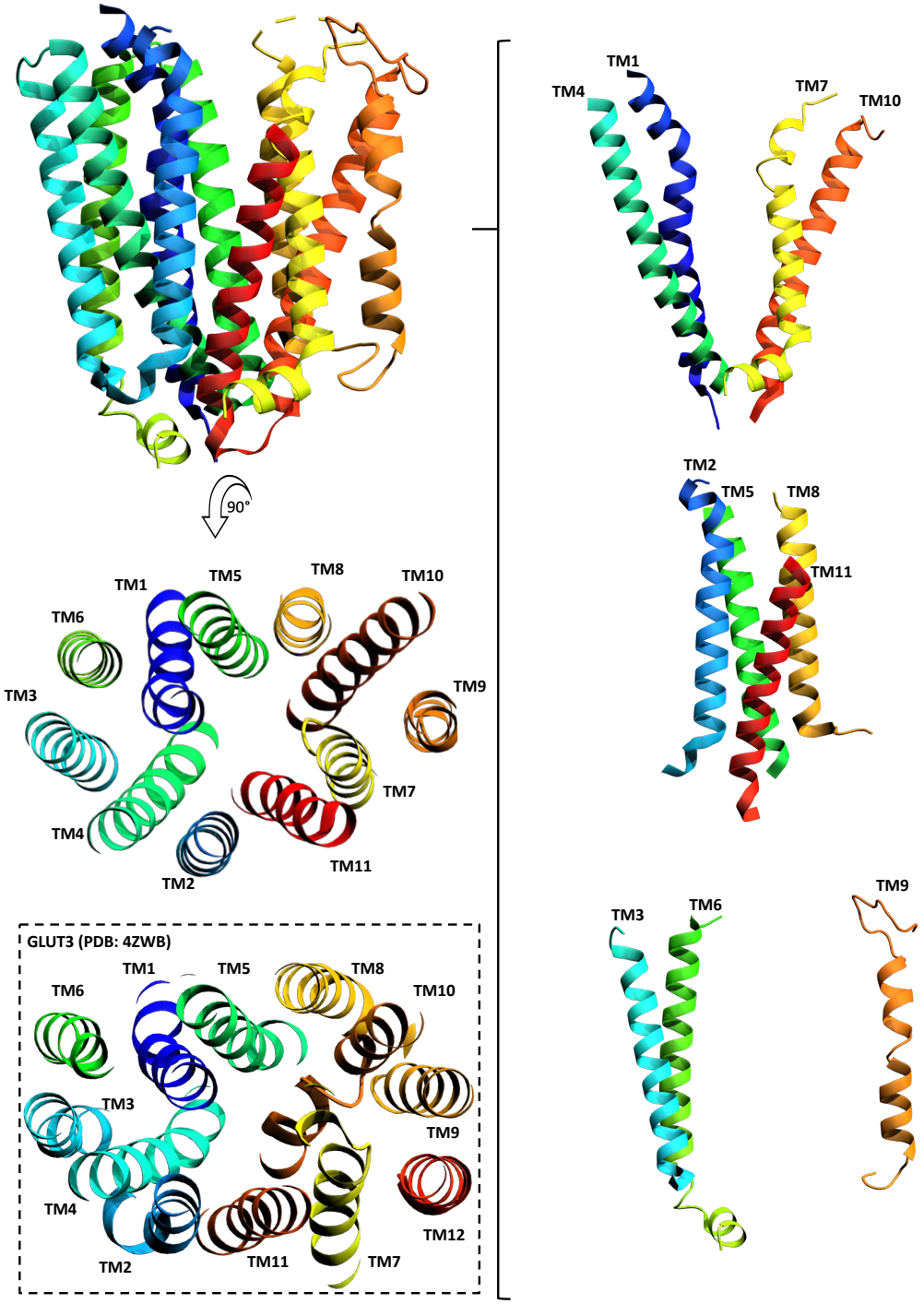


Figure 1.14 Arrangement of the helices of hENT1 and GLUT3.

As predicted, the fold of hENT1 is similar to that of MFS. The 11 TMs are arranged into an N-terminal domain (TM1-6) and a C-terminal domain (TM7-11), with a pseudo-symmetrical 6+5 topology. Within these domains there are a pair of 3+3 and 3+2 inverted repeats. The first and second helix in each repeat (TM1, 4, 7 and 10, and TM2, 5, 8 and 11, respectively) form the central cavity, and the third helices are located to the outside of the transmembrane core. In the absence of TM12, TM9 is arranged to occupy the space that is shared by TM9 and TM12 in GLUT3 (dashed box).

Along with the difference in the observed symmetry, ENTs lack the A-motif, which is highly conserved in MFS (Figure 1.6) and is essential for transport activity[53,68,71]. Instead, ENTs have an

extensive network of hydrophobic contacts in the intracellular regions of TM4, TM5, TM10 and TM11 (Figure 1.15). These hydrophobic contacts occlude access to the central cavity from the intracellular side, and are further stabilised by additional polar and charged interactions[134] (Figure 1.15, bottom panel).

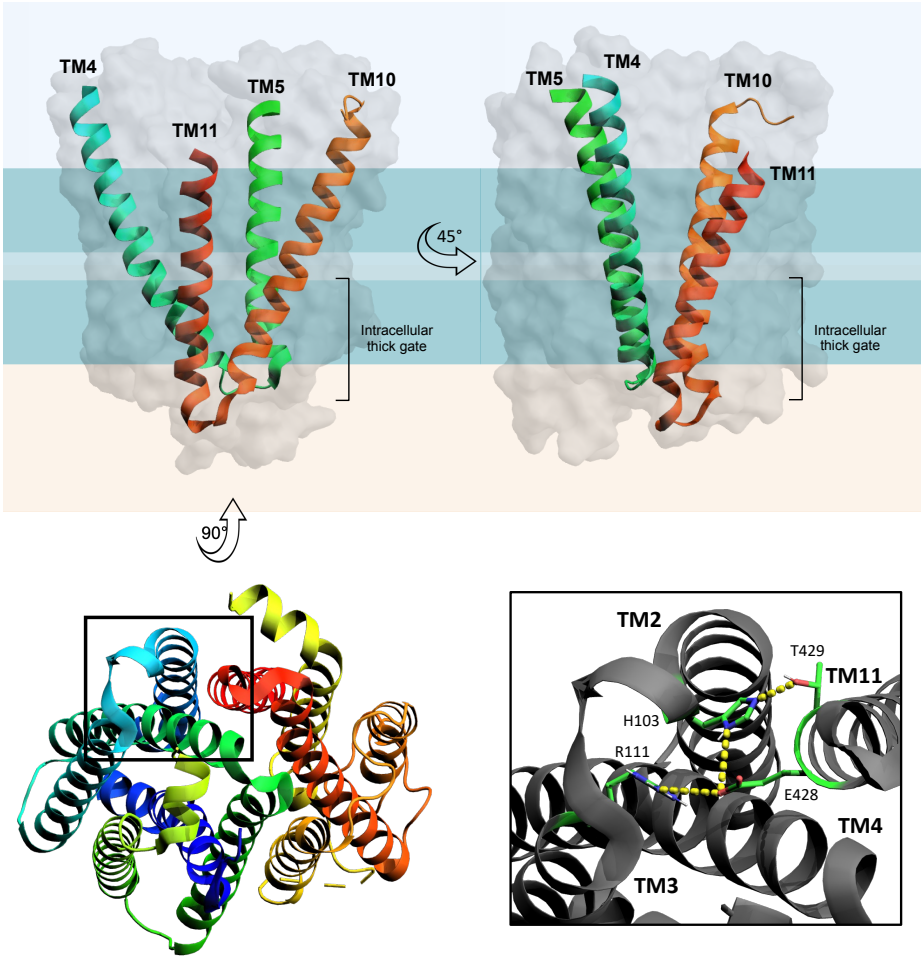


Figure 1.15 The intracellular thick gate of hENT1.

Hydrophobic interactions between TM4, TM5, TM10 and TM11 occlude access to the central cavity from the intracellular side and contribute to the formation of an intracellular thick gate. Polar and charged interaction between H103, R111, E428 and T429 further stabilise the intracellular gating[134].

1.6.1 Gating interactions at the extracellular domain

Upon substrate binding, hENT1 is suggested to transition to the occluded state *via* gating interactions at the extracellular region. As is seen in several MFS models[50,51,58], it is suggested that subtle rearrangements at the extracellular region of TM7 form a narrow constriction point which prevents the substrate from being freed from the extracellular side. In hENT1 these interactions are mediated by M33 on TM1 and P308 on TM7 (Figure 1.16A). M33 has previously been shown to be a crucial residue for permeant selectivity and sensitivity to the inhibitors

dilazep and dipyridamole[135,136]. This residue is highly conserved in mammalian ENT1s. However, ENT2, which is significantly less sensitive to inhibition by dipyridamole and dilazep, has isoleucine at the equivalent position. P308 is conserved across all mammalian ENTs. The role of M33 in dilazep sensitivity is supported by the dilazep-bound structure, with hydrophobic interactions observed between M33 and the central diazepane ring (Figure 1.16B). This interaction in turn sterically hinders the complete formation of the extracellular thin gate, thus inhibiting transitions throughout the transport cycle. This blocking of the formation of the extracellular gate is a commonly utilised approach in competitive inhibitors of MFS and sodium-coupled neurotransmitter transporters[137].

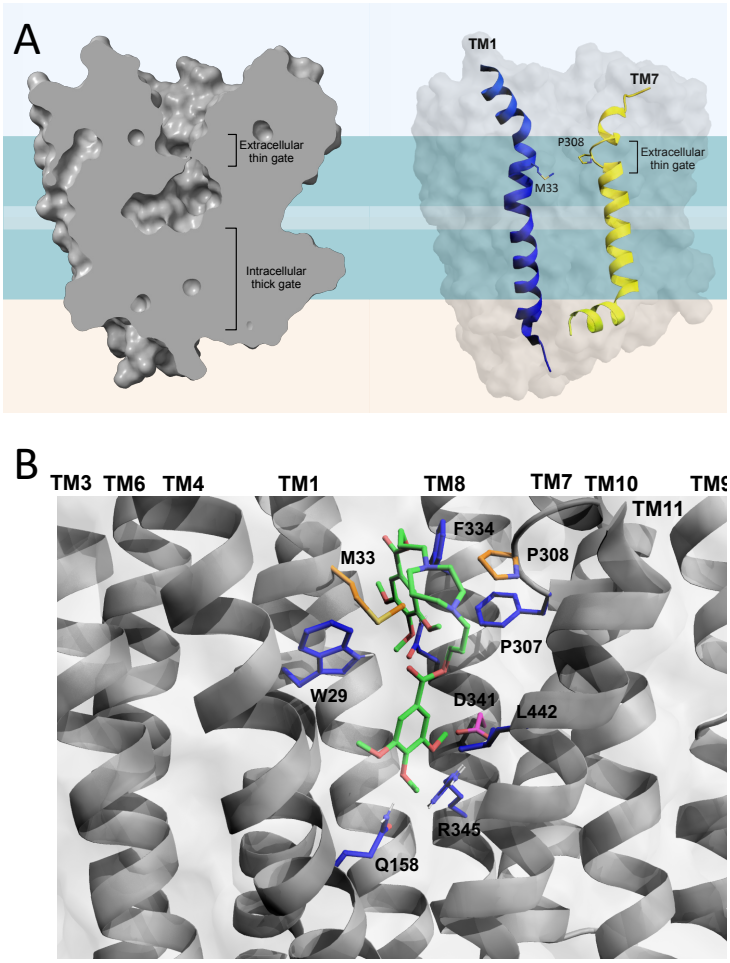


Figure 1.16 The extracellular thin gate of hENT1 and dilazep inhibition.

(A) (Left) A clipped view of hENT1 perpendicular to the membrane shows the density of both the extracellular thin gate and intracellular thick gate. (Right) TM1 and TM7 are represented as ribbons, with the side chains of M33 and P308 shown. (B) A close up view in to the central cavity of hENT1 (PDB 6OB7), TM2 is removed for clarity. Dilazep (green) inhibits the formation of the extracellular thin gate through hindering interactions between M33 and P308 (orange). Instead, M33 cradles the central diazapene ring of dilazep. Additional dilazep interactions are shown in blue, such as F334 and P307 forming π-π stacking interactions with a trimethoxyphenyl ring, and W29 and Q158 interacting with the second thrimethoxynbenzoic acid ring.

1.6.2 The central cavity: substrate recognition, inhibitor binding and the transport pathway

Within the central cavity of the NBMPR-bound structure, interactions between hENT1 and the thioinosine moiety of NBMPR are suggested to represent the mechanism for adenosine recognition. A series of hydrophobic interactions from TM1, TM2 and TM11, and a single polar interaction from TM4 surround the purine ring, and charged residues on TM8 interact with the ribose moiety[134] (Figure 1.17). Studies have previously identified M89 and L92 on TM2, L442 on TM11[138-141] and the highly conserved charged residues R345 and D341 on TM8[142] as conferring inhibitor sensitivity and nucleoside transport efficiency. Additionally, residues F334 and N338 on TM8 have previously been identified as conferring substrate transport efficiency and inhibitor sensitivity[143]. In the dilazep-bound structure F334, along with P307 from TM7, forms π-π stacking interactions with a trimethoxyphenyl ring of dilazep (Figure 1.16B). No interactions with F334 are observed in the NBMPR-bound structure, or N338 in either the NBMPR or dilazep-bound structure. Therefore, these residues are likely to line the substrate transport pathway and may mediate interactions during translocation to the central cavity[134].

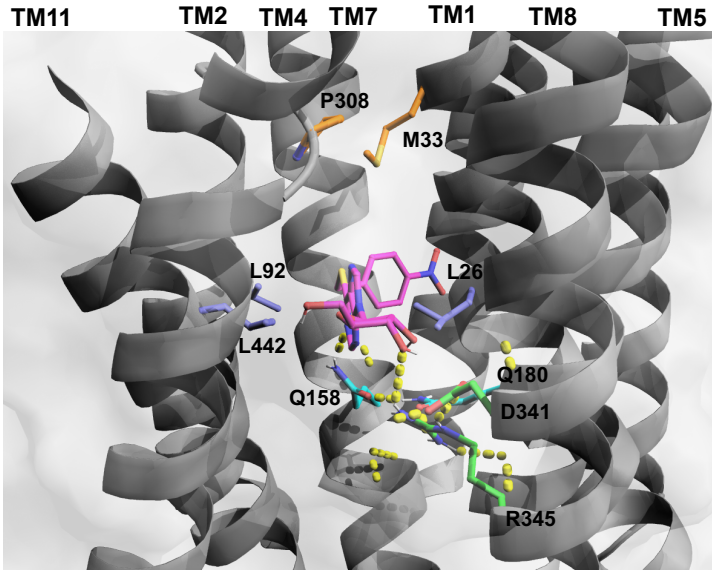


Figure 1.17 The central cavity of hENT1.

A close up view in to the central cavity of hENT1 (PDB 6OB6), TM10 is removed for clarity. In the NBMPR bound structure the occluded state is established by the formation of the extracellular thin gate by M33 and P308 (orange). Hydropobic (lilac) and polar (cyan) interactions surround the purine ring of NBMPR (pink). Charged residues (green) on TM8 interact with the ribose moeity.

1.6.3 Unique mechanisms of inhibition

Previous studies have identified G154 from TM4 as a key determinant of NBMPR isoform selectivity[144], and the determination of the molecular basis of NBMPR binding identified a

mechanism of inhibition which is suggested to be unprecedented[134]. In the NBMPR-bound structure the *p*-nitrobenzyl ring of the NBMPR occupies a deep hydrophobic pocket which extends from the central cavity and protrudes into the N-terminal domain of hENT1. This hydrophobic pocket occupies a space between TM1, TM3 and TM4, and the *p*-nitrobenzyl ring of NBMPR is directly abutted by G154. In the NBMPR insensitive isoforms hENT2-4 the residues at the equivalent position are cysteine or serine. The bulkier side chains of these residues cause the constriction of this pocket and hinder the binding of the *p*-nitrobenzyl ring. The occupation of this pocket likely prevents the conformational reorientation of the N-terminal domain required for the transition from an outward-facing state to an inward-facing state during transport. Therefore, it is suggested that NBMPR functions as a competitive inhibitor for allosteric control of the conformational transition[134].

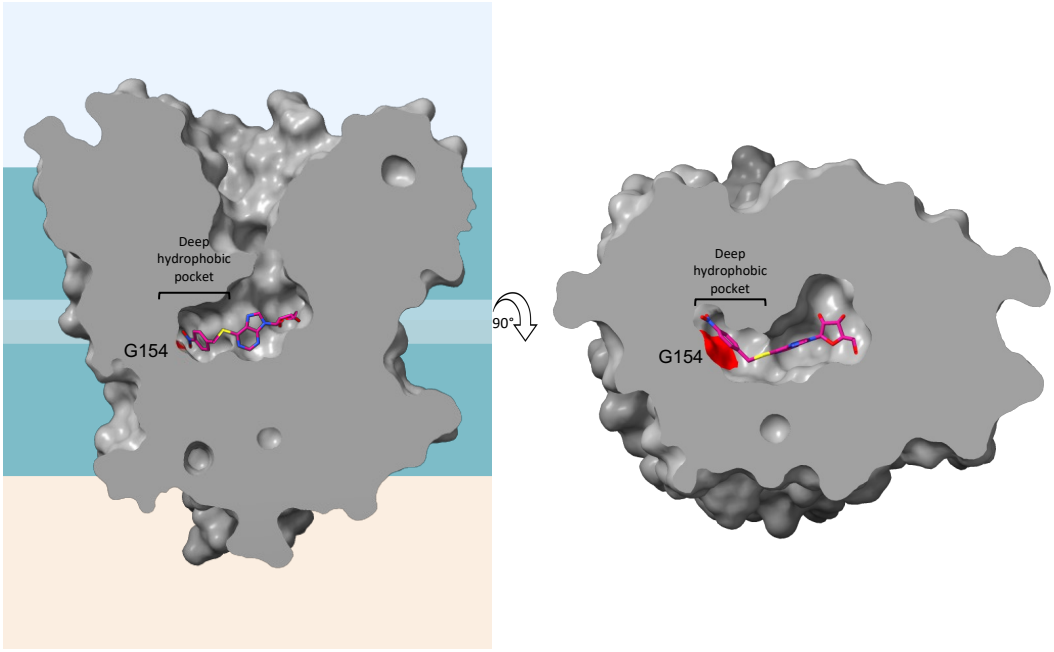


Figure 1.18 The deep hydrophobic pocket of hENT1.

A clipped view of the deep hydrophobic pocket In the NBMPR-bound structure(PDB 6OB6), with a view perpendicular to the membrane and a top-down view. The *p*-nitrobenzyl ring of the NBMPR (pink) extends away from the central cavity and protrudes into the N-terminal bundle, where it is directly abutted by G154 (red). In NBMPR-insensitive isoforms the residue at the equivalent psotion is a cysteine or serine. These 'bulkier' residues thus constrain this space and hinder the formation of this pocket[134].

1.6.4 The loops of hENT1

As a polytopic membrane protein, hENT1 has a series of loops which connect the sequential TMs. Topological predictions propose that the extracellular loop connecting TM1 and TM2 (ECL1), and the intracellular loop connecting TM6 and TM7 (ICL6) are the largest of the loops[113]. Furthermore, mutational studies have shown that these loops mediate ENT function. N-linked

glycosylation of N48 in the ECL1 has been determined to be critical for NBMPR sensitivity and substrate transport efficacy[145]. The ECL1 is absent from both hENT1 structures (Δ49-73), presumably due to not being resolved because of disorder.

Studies on loop-deletion mutants of hENT1 demonstrated that deletion of the ICL6 has minor effects on the kinetic properties of hENT1-mediated uptake of nucleosides and NBMPR binding[146]. However, a significant reduction in the efficiency of protein folding and trafficking was seen. Therefore, it was suggested that while the loops contribute to protein trafficking, they are not essential for transport activity and instead contribute to the fine-tuning of transport regulation, predicting that any modulating effects are subtle. This study of loop-deletion mutants proposed that ICL6 is 'dispensible'[146]. The ICL6 is absent from both hENT1 structures due to its deletion (Δ243-274) contributing to a crystallisable hENT1 construct[134].

1.6.4.1 The role of the ICL6

The ICL6 of hENT1 is predicted to be unstructured and has several charged residues that impose polarity[147]. The flexible conformation of the ICL6 likely allows for interactions between the loop and other proteins that may contribute to activity and function of hENT1. It has been shown that the ICL6 can be phosphorylated at several serine residue sites by several kinases[148-150]. Phosphorylation of membrane proteins has been shown to contribute to the facilitation of conformational rearrangements, thereby enabling the transition between different states for the translocation of substrates, as well as inhibitor binding[149]. For example, similar domains in loops of the multidrug resistance protein MRP1[151] and the type 2B Ca2+-ATPase ACA8[152] have been shown to serve as regulatory domains that are sensitive to conformational changes. Alanine mutations of the predicted sites of casein kinase II phosphorylation in hENT1 (T248 and S254) showed minimal effects on protein localisation, thereby suggesting that these residues contribute to other functions[149,150]. Furthermore, phosphorylation of S281 in hENT1 by protein kinase A or protein kinase C has been shown to mediate ENT1 transport[148]. Additionally, kinase phosphorylation of hENT1 has been shown to regulate ethanol sensitivity, with acute ethanol exposure inhibiting adenosine reuptake[153,154]. Therefore, the ICL6 is believed to be important for the fine tuning of transport regulation. However, the mechanism by which this achieved remains unknown.

1.6.5 Implications for our understanding of ENTs

In both of the structures available, hENT1 adopts a distinctly outward-facing conformation[134]. Therefore, hENT1 is proposed to utilise the canonical MFS alternating access mechanism of

action in which transporters occupy one of two major conformations, outward-facing or inward-facing. However, the understanding of the molecular basis of this mechanism requires the understanding of extended conformational states, or at the very least, the opposing major conformation[51,53,58]. Defined mechanisms of alternating access in MFS transporters suggests that the transition from apo to substrate bound state is mediated by subtle rearrangements. Loop regions of ENTs have been demonstrated to be crucial for function and may also modulate their actions by subtle rearrangements[148-150,153]. Recently there have been significant developments in the prediction of protein structure *via* computational methods[155-157]. However, these methods cannot confidently predict structures of ENTs. Attempts to model the ECL1 and ICL6 results in a model with very low model confidence in these regions (Figure 1.19).

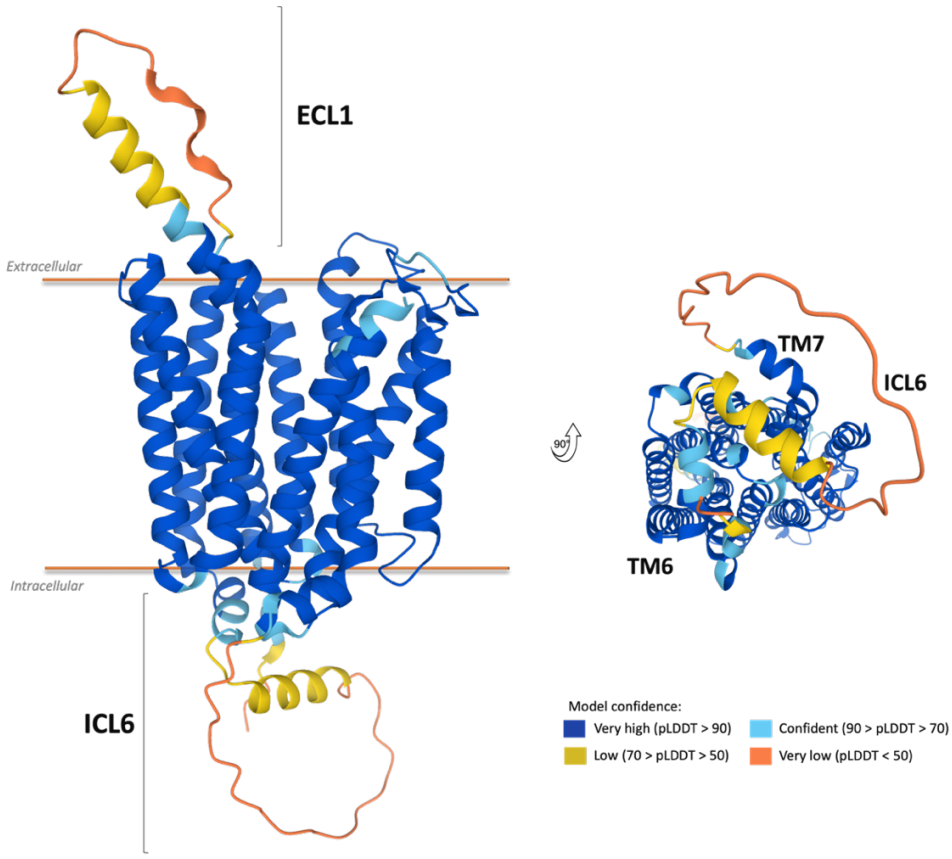


Figure 1.19 Structural prediction of the full length structure of hENT1.

The ECL1 and ICL6 are absent from the structures of hENT1 in complex with inhibitors NBMPR (PDB: 6OB6) and dilazep (PDB: 6OB7). A structural prediction of the full-length wild type hENT1 was obtained from the Alphafold Protein Structure Database[158]. Model confidence is ranked, on a per-residue confidence score (pLDDT), as very high (dark blue, pLDDT > 90), confident (light blue, pLDDT >70), low (yellow, pLDDT > 50) and very low (orange, pLDDT < 50).

Structural characterisation of loop-inclusive and/or *apo*-state structures would provide insights into the molecular basis of modulation and regulation. Moreover, as canonical MFS features

such as the A-motif are absent, hENT1 has been proposed to utilise a mechanism of transport that is distinct from MFS transporters. This further highlights the incompatibility of MFS transporters as homology models for ENTs and reinforces the need for loop inclusive ENT structures in diverse conformational states.

1.7 Challenges in the structural studies of membrane proteins

Despite their biological and pharmacological importance, there is relatively sparse structural information about IMPs. As of June 2022 fewer than 1% of all unique protein structures in the protein data bank (PDB) represent unique eukaryotic membrane protein structures[13]. Efforts towards the high-resolution structural characterisation of membrane proteins is often hindered due to bottlenecks at every step[159]. Eukaryotic IMP structural characterisation continually proves challenging due to difficulties faced in protein expression, purification, stabilisation, and structural analysis. Furthermore, achieving the yield and purity required for structural studies requires the optimisation of each of these steps on a protein-by-protein basis[160,161].

1.7.1 Membrane protein expression

Typically, the intrinsic concentrations of membrane proteins that are of pharmacological interest are typically very low. Therefore, the utilisation of a recombinant overexpression system for their study is essential. Furthermore, as the lipid-bilayer has been shown to affect protein function[5,7,14], this also influences the expression systems used. In addition, the native biosynthesis machinery and post-translational modifications, such as the N-linked glycosylation in hENT1, contribute to the production of functional protein[145]. Put simply, expression hosts most like that of the native host produce higher yields of functional protein. Therefore, eukaryotic membrane proteins typically require eukaryotic expression systems. However, as the complexity of the expression system increases, the yield typically decreases[162,163]. Furthermore, as the complexity of the protein of interest and the expression system increases, so does the associated cost and time required.

1.7.1.1 Yeast expression systems

Of the available eukaryotic expression systems, yeast expression may offer a good middle-ground for eukaryotic proteins[164]. Yeast expression systems are significantly cheaper than insect or mammalian expression and are amenable to large scale growths, although they may lack the appropriate post translational modifications required for functional protein. *Saccharomyces cerevisiae* has been well established as a successful system for the production of mammalian

transporter proteins for structural studies[165], and a number of the structures of human GLUTs were solved following expression in *S. cerevisiae*[62,166]. *Pichia pastoris* is also a highly successful expression system for eukaryotic membrane proteins due to it being more tolerant of modifications often utilised in stabilisation efforts. *P. pastoris* is also capable of more complex glycosylation than is achieved by *S. cerevisiae*, as well as the ability to achieve a significantly higher biomass[164]. Comparisons of membrane protein expression, both in terms of yield and functionality, in *P. pastoris* to more complex insect and mammalian cell systems has been shown to have comparable efficiency[160,167]. Thereby, supporting yeast as a viable alternative to more complex and more expensive systems.

1.7.1.2 Insect and mammalian expression systems

Of the mammalian MFS structures available as of June 2022, the majority were solved following expression in insect and mammalian cell lines. At present, GLUTs remain the only human MFS structures solved following expression in any yeast systems. Therefore, while yeast expression systems offer viable alternatives, there is still a significant requirement for more complex cell lines. *Spodoptera frugiperda* (*Sf*9) and *Trichoplusiani* (Hi5) are the most commonly used insect cell lines, with *Sf*9 being the most prevalent in the contribution to structural characterisation of eukaryotic membrane proteins[167]. Human embryo kidney (HEK) and Chinese hamster ovary are the most used mammalian cell lines[160,162,167]. However, owing to their time-consuming nature, expense, and typically low yields, they contribute significantly less to solved structures of eukaryotic membrane proteins.

1.7.2 Engineering stabilised membrane proteins

Structural characterisation of membrane proteins requires not only the efficient expression of targets, but also the extraction and isolation of proteins in a homogenous, monodisperse, functional and stable state[168]. Protein instability is a significant problem in the efforts towards structural characterisation and can result in significant losses during the workflow. There are several different approaches that can be utilised to engineer a protein construct that is more amenable to expression and downstream processing. One of the most easily applied approaches for improved expression is the codon optimisation of the target protein sequence for the target expression system. For example, *P. pastoris* has been found to express more stable protein from codon optimised genes[164].

1.7.2.1 Truncations and fusions

As structural disorder gives rise to structural instability, another simple approach to protein stabilisation is the truncation or removal of flexible and disordered domains, such as the ICL6 in hENT1[134,169]. Similarly, as membrane proteins are inherently hydrophobic, it has been shown that protein modifications to include ordered soluble protein fragments, known as fusion partners, can improve stability. However, the location of any fusion must be considered to allow the preservation of the fold of the partner protein, but without disruption to the protein of interest[170]. Fusion partners, such as T4 Lysozyme (T4L) or thermostabilised apocytochrome b_{562} (BRIL), have been used extensively in the structural characterisation of GPCRs. As of June 2022, over half of the available structures for Class A GPCRs include a fusion partner[13,170].

In addition to increasing the stability of the protein of interest, fusion partners can serve as protein biomarkers and purification tools. Green fluorescent protein (GFP) is a widely used fluorescent fusion partner. It is a soluble protein comprising of a beta-barrel which surrounds a central chromophore that can form without requiring any auxiliary cofactors, enzymes or substrates other than molecular oxygen[171]. GFP remains conformationally stable in the typical conditions encountered in protein expression and purification, and fluorescence emission is only quenched on denaturation. Therefore, GFP can serve as a sensitive indicator of protein folding, as well as allowing for the visualisation of protein localisation *in vivo*[172].

1.7.2.2 Stabilising point mutations

In addition to fusion partners, over two thirds of GPCR structures have at least one stabilising point mutation[13,170]. Furthermore, the construct that Wright & Lee developed for the crystallisation of hENT1 had three point mutations (L168F, P175A and N288K) that, in addition to the truncation of ICL6, reportedly contributed towards significant stabilising effects on hENT1[134].The well-established approach of scanning-mutagenesis facilitates the identification of mutants with improved stability, and thus improved expression, purification, and structural characterisation[173,174]. It has been shown that a single point mutation can significantly improve the expression of membrane proteins[175-178]. While highly successful, the scanning-mutagenesis approach is labour and cost intensive, and thus is another point at which progress towards structural characterisation hits a bottleneck. This has given rise to the development of computational pipelines that predict the likelihood of stabilising effects of specific amino acid changes[173,179-182].

1.7.3 Homology modelling

A complementary approach to protein engineering is the study of homologous proteins. The identification of homologues may allow for the use of simpler and cheaper expression systems and may contribute to structural characterisation through 'mass production' of a target. Furthermore, the identification of a stable homologue may consequently produce a protein that is more amenable to processes towards structural characterisation[173]. Homology models have successfully provided the foundation for the understanding of the molecular basis of transport of several mammalian transporters, including CNT3 (1.5.1)[77,104,183].

1.7.4 Binding partners: Antibodies, antibody derivatives and nanobodies

The binding of partner proteins, such as antibodies and antibody derivatives, has also contributed to the structural characterisation of a number of challenging eukaryotic membrane protein targets, such as the SLC and MFS members PepT2[184] and Glut5[66]. Conventional antibodies consist of two heavy and two light chains, and each chain has a variable domain for antigen binding. Antibodies and their derivatives are widely used as affinity reagents for the detection, selection, and stabilisation of proteins. However, the large-scale production of conventional antibodies and antibody fragments is costly and lengthy. Furthermore, the portfolio of validated antibodies is limited and there is an ongoing need for binding partners with improved specificity, affinity, and stability[185-189].

1.7.4.1 Nanobodies

Nanobodies are the single domain fragments of antibodies that contain the full antigen-binding capacity of heavy chains only that occur naturally in camelids. As a single domain fragment nanobodies are small in size, at ~110 amino acids (~15kDa). Due to their small size, nanobodies can target areas that are not accessible to conventional antibodies and antibody fragments. Nanobodies are highly specific and bind to targets with nM affinity. In addition, they are easily produced in bacterial and yeast expression systems[185-187,190,191].

Nanobody binding facilitates the stabilisation of flexible regions and specific, preferred conformations. Therefore, nanobodies can be used as 'conformational selectors' and can contribute to the determination of the conformational ensemble required for understanding the mechanism of action[191-193]. Furthermore, the binding of nanobodies also contributes to a protein complex that is more amenable to structural analysis. In X-ray crystallography nanobodies can act as crystallographic chaperones by increasing protein polar surface, which in turn improves

the likelihood of crystal contact formation[176,187]. In addition, recent advancements in the unique technology for reformatting nanobodies into megabodies has been suggested to offer a solution to overcome challenges of size limitation and preferential orientation faced in cryo-EM[194]. Therefore, nanobodies are valuable tools in the stabilisation and characterisation of challenging targets.

1.7.4.1.1 Synthetic nanobodies

The generation of matured nanobodies depends on animal systems, with large amounts of protein (typically 1-2 mg) required for repeat immunisations for the affinity maturation. Animal immunisation typically requires stable protein that can remain properly folded for extended periods of time and at elevated temperatures[188,189]. However, following immunisation, detergent solubilised membrane proteins may be denatured due to detergent dissociation[187]. In addition, ethical considerations restrain the use of certain compounds used in the stabilisation of challenging targets, such as ligands or the detergent digitonin, due to their toxicity in animal models[188]. Therefore, while nanobodies are incredibly valuable tools, the generation of matured nanobodies against membrane protein targets presents several challenges. These limitations have given rise to the development of synthetic nanobody libraries, such as those developed by the Kruse lab[186] and the Seeger lab[188,189], for *in vitro* selection. These *in vitro* synthetic libraries forgo the need for immunisation, and thus several of the difficulties encountered for nanobody selection Therefore, sybodies are proposed to be a more accessible avenue for nanobody discovery and are especially suited for challenging membrane targets. Furthermore. while not a specifically matured immune nanobody, efficient screening reportedly allows for the selection gives rise to binders with nM affinities[186,188,189].

1.7.5 Extraction from the lipid bilayer

As already discussed, membrane transporter proteins are inherently hydrophobic and dynamic. Functional and structural studies of IMPs typically take place in aqueous solution. Therefore, studies of IMPs initially require their extraction from the bilayer. However, as membrane protein function is influenced by the general local bilayer environment and by specific lipid-protein interactions, extraction from the lipid bilayer is likely to be detrimental for the function and stability of IMPs.

1.7.5.1 Detergents

Historically, detergent solubilisation has been the most utilised and successful method of IMP extraction. Detergents are small amphipathic molecules that disrupt lipid interactions in the bilayer, causing lipids to dissociate and be replaced by detergent molecules at the hydrophobic protein surface[195]. At appropriate concentrations detergent molecules self-assemble into micelles, thereby extracting membrane proteins into soluble proteomicelles[196]. The most popular class of detergent in the structural characterisation of IMPs are the nonionic alkyl maltopyranosides, with n-dodecyl-β-D-maltopyranoside (DDM) and n-decyl-β-D-maltopyranoside (DM) being the most successful detergents in this class. However, detergent systems are a poor mimic of the native lipid environment and are, consequently, frequently destabilising and inactivate the captured protein[197-200].

Furthermore, short chain detergents are frequently more destabilising than their long chain counterparts. However, long chain detergents present their own challenges. Long chains give rise to larger micelles which may obstruct solvent accessible regions of the protein that contribute to binding sites of partners, such as nanobodies[201]. Additionally, the large micelle size may hinder structural analysis through limiting crystal contact formation required in protein crystallisation or obscuring the view of solvent accessible domains in cryogenic electron microscopy (cryo-EM). Conversely, while short chain detergents may be more destabilising, their small micelle size is more amenable to high resolution structural determination. Therefore, finding the optimal detergent for both solubilisation and structural characterisation is essential. Additional steps during handling to allow for detergent exchange can allow for the utilisation of different detergents during different experimental points.

1.7.5.2 Lipid mimetics and maleic acid lipid particles

Owing to the role of lipids on IMP stability and function, improving the native lipid-like environment is a key focus in the advancement of structural characterisation of IMPs. There are now several alternatives available which better mimic the native lipid bilayer.

1.7.5.2.1 Exogenous lipid-containing systems

Bicelles[202] and liposomes[203] are exogenous lipid-containing systems in which IMPs can be reconstituted into. Bicelles are planar lipid bilayers that are surrounded by a stabilising and solubilising detergent ring, and they are often used in the NMR study of IMPs in a lipid environment since they can orient in a magnetic field[204]. Liposomes are detergent free spherical

lipid-bilayer vesicles[203]. The spherical nature of liposomes makes them an important component in the functional characterisation of transporter proteins, as this allows for the bilayer to partition substrates, and thus has uses in transport assays. However, bicelles and liposomes typically yield heterogenous populations. Therefore, they are not typically amenable to structural studies[205]. Nanodiscs are an alternative exogenous lipid containing system that offer a solution to the issue of heterogeneity[206]. Nanodiscs are discoidal lipid bilayers which are surrounded by two stabilising and solubilising amphipathic helical protein belts, termed membrane scaffold proteins[207]. They offer a significant advantage over bicelles because of a degree of control offered in the size of the particles, which is determined by the length of the membrane scaffold proteins[208]. Therefore, nanodiscs can be made homogenous and monodisperse, and thus are more amenable to structural characterisation[206]. Recently, nanodiscs have been utilised in the structural characterisation of the challenging mammalian MFS transporters ferroportin[209,210] and major facilitator superfamily domain containing 2A[211].

1.7.5.2.2 Lipid-free systems

Peptidiscs and amphipols are two further lipid mimetic systems. However, these systems do not contain lipids. Instead, peptidiscs have amphipathic bi-helical peptides[212] and amphipols have amphipathic polymers[213]. In both systems the amphipathic molecules associate to the hydrophobic surface protein in a way that is like that of detergents. While these systems do not offer the lipid interactions available in other systems with the inclusion of exogenous lipids, they offer several advantages over detergents. For example, both peptidiscs and amphipols offer improved stability, in terms of both protein stability and stability of system association. Additionally, amphipols form a compact layer on the target protein, and thus overcome issues faced in the 'bulkiness' of detergent micelles. Furthermore, the hydrophobic associations of amphipols are stable in the absence of competing surfactants but are freely miscible with other surfactants. This means that they are also amenable to reconstitution or displacement into other systems[205,213]. Moreover, amphipols cannot efficiently compete with protein-lipid interactions. Therefore, dissociated lipids can rebind to membrane proteins, thereby facilitating a more native environment.

A8-35 is the most extensively used amphipol[213,214]. A8-35 has short (~35 residues) ungrafted carboxylic acid chains. Some of the carboxylic groups are derivatized with octyl chains, giving A8-35 its amphipathic profile, and some with isopropyl groups, reducing the charge density along the polymer[215]. The remaining groups are charged in aqueous solution, thus contributing to the high water solubility of the polymer[215]. However, factors that affect the charge, such as

low pH or the presence of multivalent cations, increases hydrophobicity, and causes aggregation.

1.7.5.2.3 Native lipid systems

While each of the discussed systems offer unique advantages, all require initial detergent solubilisation of the protein from the membrane before exchange into the target system. Therefore, the challenges of detergent solubilisation remain. Furthermore, as interactions between specific lipids have been shown to contribute to protein function[25,216,217], the study of membrane proteins in the most native-like environment is essential for our understanding of the molecular basis of the mechanism of action. This need may not be adequately met through the utilisation of exogenous lipids. The development of maleic acid lipid particles has allowed for significant advancements in the study of membrane proteins in a native environment. Styrene-maleic acid lipid particles (SMALPs)[218] and derivatives, such as di-isobutylene maleic acids (DIBMA), can extract IMPs directly from the lipid-bilayer, and thus retain specific (and non-specific) local lipid interactions. This also eradicates the need for detergents altogether. Like nanodiscs, the maleic acid particles function as a scaffold that surrounds and stabilise the lipid disc, thus they can be thought of as a type of native nanodisc. However, their usage is limited due to incompatibilities with downstream processes because of their sensitivity to low pH and divalent cations, in addition to their absorption at 260 nm[218-222].

1.7.6 Methods of structural characterisation

There are three main methods for the structural characterisation of membrane proteins (X-ray crystallography, cryo-EM, and nuclear magnetic resonance (NMR)), each with their own advantages and disadvantages. Historically, X-ray crystallography has been the most robust and successful method for protein structural characterisation and accounts for 88% of the structures currently available in the PDB[223]. NMR and cryo-EM collectively account for most of the remaining structures. Depositions to the PDB over the last decade have shown that the number of structures being solved using cryo-EM has grown at an explosive rate, while the trend in the number of structures solved using X-ray crystallography and NMR has remained relatively flat[223]. This explosion in popularity is due to technological advancements and the 'resolution revolution' meaning that significantly lower quantities of protein are required compared to other methods, and the method can generate structures solved with near-atomic resolution[224-226].

While contributing significantly to advancements in the structural characterisation of challenging membrane protein targets, current cryo-EM methods still face significant hurdles,

such as the previously mentioned size limitation and preferential orientation[227]. Therefore, X-ray crystallography remains a popular method for structural characterisation[199,228]. However, X-ray crystallography is not without its' own challenges. The methodology requires protein to be in a crystalline form, and the generation of high-resolution structures and diffraction patterns requires a well-ordered crystal structure[200]. For the structural characterisation of membrane protein these requirements present significant challenges[229,230]. As discussed, structural characterisation of membrane proteins requires their initial extraction from the lipid bilayer. However, the methods used for extraction, such as detergent solubilisation, directly affects crystal formation. The presence of detergent micelles negatively influences characteristics such as solvent content, crystal packing and crystal order, which subsequently leads to poor crystals and diffraction[198].

Recently lipidic cubic phase (LCP) crystallisation, which allows for the exchange of detergent molecules for lipids, has grown in popularity. Owing to the absence of detergent micelles and the supplementation with lipids, LCP crystallisation promotes the formation of crystals with characteristics that are more likely to produce good quality diffraction and provide a better mimic of the native lipid bilayer[198,231,232]. LCP crystallisation has contributed to the solving of several challenging membrane protein targets, including the structures of hENT1[134].

1.7.7 Summary

The challenges faced in the study of membrane proteins are multifaceted and consequently, require a combination of approaches to overcome them. This is seen in the large number of the available GPCR structures where, as discussed, the combination of N- and C-terminal truncations, fusion partners, stabilising point mutations, antibody derivatives or nanobodies, and structural characterisation in LCP have made significant contributions to recent structures[170,174,176,197]. Moreover, the combined approaches of stabilising point mutations, truncations and LCP crystallisation directly contributed towards the generation of the first structures of an ENT[134]. However, there is no 'one size fits all' approach, no way of predicting how amenable a target (be that wild-type or engineered) may be to expression, purification, or structural characterisation. Therefore, efforts towards the characterisation of challenging membrane targets often requires the optimisation of multiple factors, on a protein-by-protein basis.

1.8 Aims and objective of this project

As described previously (1.6.5), hENT1 is proposed to utilise an alternating access mechanism of action. However, distinct differences between ENTs and canonical MFS structures supports the suggestion that ENTs utilise a mechanism of action that is distinct from that MFS transporters. Furthermore, due to the unique structural features of ENTs there is a lack of suitable homology models, and current methods for the computational prediction of ENT structures generates models with overall poor fit and low confidence. Therefore, while the available structures of hENT1 contribute to significant advancements in our understanding of ENTs, there is still a large gap in our knowledge. Without structures in the major opposing conformation, in addition to structures of full-length ENTs, the molecular basis by which transport is achieved remains unknown and ENT mechanisms cannot be properly addressed. Therefore, the overarching aim of my project was the stabilisation of equilibrative nucleoside transporters, for the purposes of producing ENTs that were more amenable to structural characterisation by X-ray crystallography. To achieve this, I explored three distinct but complementary avenues. First, the identification of nanobodies to allow for stabilisation, conformational selection and to serve as crystallisation chaperones. Second, stabilisation of hENT1 by point mutations. Third, the identification and characterisation of a thermostable homologue of hENT1.

Chapter 2 Methods and materials

2.1 Materials

2.1.1 Chemicals

All chemicals used were purchased from Sigma Aldrich, Melford Laboratories, VWR International, Thermo Fisher Scientific, or Anatrace unless stated otherwise.

2.1.2 Buffers and solutions

All buffers were prepared using Milli-Q® water unless stated otherwise. All buffers were filtered using 0.22 µm or 0.45 µm membrane filters (GE Healthcare Life Sciences) or syringe filters (Acrodisc®, Pall Corporation or Minisart, Sartorius™). Buffers prepared for use in Fast Protein Liquid Chromatography (FPLC) were also degassed prior to use. A list of commonly used buffers and solutions is detailed in Table 2.1. All other buffers used are described in the text.

Table 2.1 Table of commonly used buffers

Buffer	Composition
Phosphate buffered saline (PBS)	10 mM Na_2HPO_4, 2.7 mM KCl, 1.8 mM KH_2PO_4, 137 mM NaCl, pH 7.4
5x sodium dodecyl sulphate (SDS) loading dye	500 mM Tris HCl pH 6.8, 10% (w/v) SDS, 500 mM dithiothreitol (DTT), 50% (w/v) glycerol,0.05% (w/v) bromophenol blue
SDS running buffer	25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% (w/v) SDS
TAE	40 mM Tris-acetate pH 8.5, 1 mM ethylenediaminetetra-acetic acid (EDTA)
Tris buffered saline (TBS)	20 mM Tris-HCl pH 7.6, 150 mM NaCl
TBT with Tween (TBS-T)	20 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% (v/v) Tween 20

2.1.3 Media

All media used for the culturing of *Escherichia coli* (*E. coli*), *S. cerevisiae* and *P. pastoris* were prepared in Milli-Q® water and sterilised by autoclaving. Where necessary, components such as antibiotics, carbon sources, etc. were sterilised by 0.22 µm filtration and added after autoclaving. Agar plates were prepared by adding 1.5% (w/v) agar. Table 2.2 details commonly used media that are referred to in this thesis. All other media used are described in the text.

Table 2.2 Table of commonly used media

Media	Composition
Luria broth (LB)	0.5% (w/v) yeast extract, 1% (w/v) tryptone, 1% (w/v) NaCl
Super Optimal broth with Catabolite repression (SOC)	0.5% (w/v) yeast extract, 2% (w/v) tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose
YPD media	1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) D-glucose

2.1.4 Primers

All primers used were standard, desalted primers that were synthesised by Eurofins. A list of primers used in this project are given in Table 2.3 and Table 2.4.

Table 2.3 Table of all cloning primers

Primer ID	Direction	Sequence 5' -> 3'	T_a (°C)	Application
#1.1	Forward	GGATTATTCATACCGTCCCA	62	Amplification of hENT1 insert
#1.2	Reverse	ATTGGGACAACGCCAGTG		
#1.3	Forward	TTATTTTCAGGGCGGCGCCGGCGCTAGCCCGGTGCTAGACCTCCTC	66	Amplification of *Bb*ENT insert for cloning into pFastBac
#1.4	Reverse	TCGACAAGCTTGCATTAGCCTGCAGTTATTATTAGATGGAGGCGGTCAGAG		
#1.5	Forward	TTATTTTCAGGGCGGCGCCGGCGCTAGCGAGGCTCGTCAGCCTGCT	68	Amplification of *Bs*ENT insert for cloning into pFastBac
#1.6	Reverse	TCGACAAGCTTGCATTAGCCTGCAGTTATTATTAGGCGGAAGCAGCCAAG		
#1.7	Forward	TTATTTTCAGGGCGGCGCCGGCGCTAGCACCACTCGTGACGGTCCTC	66	Amplification of *Gg*ENT insert for cloning into pFastBac
#1.8	Reverse	TCGACAAGCTTGCATTAGCCTGCAGTTATTATTAGATCAGGATTTGGAACAGGAAGG		
#1.9	Forward	TTATTTTCAGGGCGGCGCCGGCGCTAGCTTCCACAAGTCTGCTCCAG	62	Amplification of *Rv*ENT insert for cloning into pFastBac
#1.10	Reverse	TCGACAAGCTTGCATTAGCCTGCAGTTATTATTACACGGTCAGGGG		
#1.11	Forward	TTATTTTCAGGGCGGCGCCGGCGCTAGCGAGCTGGCTAAGCCTTTG	60	Amplification of *Tt*ENT insert for cloning into pFastBac
#1.12	Reverse	TCGACAAGCTTGCATTAGCCTGCAGTTATTATTACTCTTTAGCCAGAGG		
#1.13	Forward	CTGCAGGAATTCGATATCAAG	56	Linearisation of pDDGFP plasmid
#1.14	Reverse	ACCAGCAGAAGTACCTTGAAA		
#1.15	Forward	AACTTGTACTTTCAAGGTACTTCTGCTGGTGAGGCTCGTCAGCCTGCT	68	Amplification of BsENT insert for cloning into pDDGFP
#1.16	Reverse	ATCGATAAGCTTGATATCGAATTCCTGCAGTTATTATTAGGCGGAAGCAGCCAAG		
#1.17	Forward	GACCGGTCTTGCTAGATTCTAATCAA	67	Linearisation of pPICZB plasmid for cloning
#1.18	Reverse	ATGATGCATGAATTCCTCGTTTCG		
#1.19	Forward	GAATTCATGCATCATCACCATCACCATCACCACAG	67	Amplification of BsENT insert for cloning into pPICZB
#1.20	Reverse	CTAGCAAGACCGGTCTTATTATTAGGCGGAAGCAGCCAAG		
#1.21	Forward	GTTTAAACGCTGTCTTGGAACC	65	Linearisation of pPICZB-NHGV-*Bs*ENT plasmid for cloning
#1.22	Reverse	TGTCAGTTTTGGGCCATTTG		

Table 2.4 Table of all sequencing primers

Primer ID	Sequence 5' -> 3'
#2.1	ATGACAACCAGTCAC
#2.2	CACAATTGCCCGGAAC
#2.3	CCAAAGATCCAAACG
#2.4	CAAATGTGGTATGGCTGATT
#2.5	CATGGCTGGACAGGGTGTCG
#2.6	GTGACATAACTAATTACA
#2.7	GCGACTGGTTCCAATTGACAAG
#2.8	CGTGACCACATGGTCCTTCTTG
#2.9	CTTCTTCCTGATGGTCATGGTG
#2.10	TCTGGCCTTCCTGATCTGGAAC
#2.11	GAAGCTGCTGGTGGTTTCATG

2.1.5 Genes

Double-stranded DNA GeneArt® Strings™ were synthesised by Thermo Fisher Scientific and codon optimised for expression in Sf9 cells. The sequences of all synthetic genes are given in FASTA format in Appendix Table 1, with the NCBI accession number and protein parameters as determined by ProtParam, ExPASy[233].

2.2 General methods of molecular biology

2.2.1 DNA methods

2.2.1.1 E. coli transformations

All transformations were done following the standard heat shock protocol with OmniMAX™ competent cells (generally prepared in-house). ~100 ng of plasmid DNA was added to 50 µL competent cells. Cells were incubated on ice for 20 minutes, heat shocked for 45 seconds at 42 °C, and returned to ice for a further 2-minute incubation. 1 mL of sterile LB or SOC media was added to the cells and then incubated for 1 hour at 37 °C at 200 rpm. Cells were then harvested by centrifugation (3,000xg, 5 minutes, 20 °C), resuspended in sterile media, and then applied to selection plates. Colonies typically formed after being incubated for 12 hours at 37 °C.

2.2.1.2 Agarose gel electrophoresis

To visualise DNA, 1% (w/v) agarose gels with 0.5 x SYBR™ Safe DNA Gel Stain (Invitrogen) were prepared in 1 x TAE buffer. Samples were prepared in 6 x Purple Gel Loading Dye (New England BioLabs) and then loaded alongside either a 1 kb DNA ladder (New England BioLabs) or GeneRuler DNA ladder mix (Thermo Fisher Scientific). Gels were run in 1 x TAE buffer at a constant 100 V then visualised using a G:BOX (Syngene).

2.2.1.3 Polymerase chain reaction

All polymerase chain reactions (PCRs) were performed using Q5® High-Fidelity 2X Master Mix (New England Biolabs) and a T100™ Thermal Cycler (Bio-Rad). In general, samples were prepared in Milli-Q® to a final concentration of 0.5 µM forward primer, 0.5 µM reverse primer, 1X Q5® High-Fidelity Master Mix, and 0.5 ng/µL template DNA. PCR was performed as per the manufacturer's recommendation. Annealing temperatures were determined using the NEB T_m calculator (https://tmcalculator.neb.com). All primer sequences are listed in Table 2.3.

For colony PCR, following cloning, colonies were picked with a sterile pipette tip and streaked on a selection plate 5-10 times in a 5 µL aliquot of 0.5 µM forward primer, 0.5 µM reverse primer, 1X Q5® High-Fidelity Master Mix, and Milli-Q®. PCR thermocycling was performed as above, but the initial denaturation step at 95 °C was increased to 7 minutes. The pipette was then transferred to 5 mL LB selection media and PCR positive clones were incubated overnight at 37 °C, 200 rpm.

2.2.1.4 Linearisation of DNA

Generally, vectors and inserts were linearised either by restriction digest or PCR. For restriction digest, using the appropriate restriction enzymes, reactions were prepared in CutSmart® Buffer (New England Biolabs). Digestions were incubated, and heat inactivated as per the manufacturer's recommendations. For the linearisation of vectors, inverse PCR was performed and was achieved using the relevant primers, as detailed in Table 2.3. Linearised products were separated by agarose gel electrophoresis (2.2.1.2). The bands at the expected molecular weight were visualised using a Safe Imager™ 2.0 Blue Light Transilluminator (Invitrogen), cut from the gel, and the DNA isolated using a Nucleospin® Gel and PCR Clean Up Kit (Macherey-Nagel), as per the manufacturers protocol. DNA yields were determined using the 'dsDNA' application on a DS-11 Spectrophotometer (DeNovix®).

2.2.1.5 Ligation-independent cloning

For ligation-independent cloning, inserts were generated by PCR (2.2.1.3) using primers designed using the Clontech In-Fusion Cloning Primer Design Tool (https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools). 20-50 ng of linearised vector and insert were prepared at a 2:1 molar ratio of insert:vector and mixed with 1 x In-Fusion® cloning mix (Clontech Laboratories). Reactions were incubated for 20 minutes at 50 °C, and then 1 µL was transformed into OmniMAX™ *E. coli* cells (2.2.1.1). Positive clones were identified using colony PCR (2.2.1.3) and plasmid DNA was isolated from the corresponding overnight culture using a Nucleospin® Plasmid Miniprep kit (Macherey-Nagel). Successful assembly was verified by Sanger sequencing (Eurofins Genomics) of the entire open reading frame.

2.2.2 Protein methods

2.2.2.1 Polyacrylamide gel electrophoresis

Samples for denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) were prepared in 5 x reducing SDS sample buffer and incubated at room temperature for 10 minutes. Samples were then loaded into a Mini-Protean® TGX (4-20%) precast gel (Bio-Rad) alongside a Color Prestained Protein Standard (Broad Range, New England BioLabs). Proteins were separated by running at a constant voltage of either 120 or 150 V in SDS-PAGE running buffer, and then visualised using either in-gel GFP fluorescence, by Western blotting (2.2.2.2), or Coomassie staining (Quick Coomassie Stain, Generon). Western blots and Coomassie stained gels were imaged using a G:BOX (Syngene), using the appropriate settings. GFP fluorescence was imaged using either a G:BOX Chemi XX6 with Blue LEDs (Syngene), an iBright FL 1500 (Thermo Fisher Scientific) or an ImageQuant™ 800 (Cytiva), using the appropriate settings.

2.2.2.2 Western blotting

Following separation by SDS-PAGE (2.2.2.1), proteins were transferred onto PVDF or nitrocellulose membranes using Trans-Blot® Turbo™ Transfer Packs and Transfer System (Bio-Rad). Membranes were then blocked in TBS-T with 3% (w/v) bovine serum albumin (BSA), either for 1 hour at room temperature, or overnight at 4 °C, with rolling. Blocked membranes were incubated with either a horseradish peroxidase (HRP)-conjugated anti-His antibody (1:2,000) (R&D Systems), or a primary rabbit anti-hENT1 antibody (1:10,000) (HPA012384, Atlas Antibodies) for 1 hour at room temperature, with rolling. Membranes were then washed three times with TBS-T for 15 minutes at room temperature, with rolling. Where appropriate, membranes were further incubated with a secondary HRP-conjugated goat anti-rabbit antibody (1:5,000) (Jackson ImmunoResearch Europe) for 1 hour at room temperature, with rolling, followed by three washes with TBS-T. Membranes were developed using Clarity™ Western ECL Substrate (Bio-Rad) and imaged with a G:BOX (Syngene).

2.3 Insect cell culture and the baculovirus expression system

2.3.1 Cell maintenance

Sf9 insect cells were maintained at mid-log phase in Insect-XPRESS™ Protein-free media with L-glutamine (Lonza) or Sf9-900 II Serum Free Media (Gibco™) in Erlenmeyer cell culture flasks at 27 °C, 120 rpm. Cell viability was assessed by staining with trypan blue stain (Invitrogen), and

cells maintained at >95% viability. Cell densities were maintained at 0.5 – 2.0 x 10^6 cells/mL, and cell densities were counted either using a Countess™ II Automated Cell Counter (Invitrogen) or manually, using a haemocytometer.

2.3.2 Bacmid DNA preparation

DH10Bac™ *E. coli* (Invitrogen) cells were transformed with pFastBac™ (Thermo Fisher Scientific) plasmids using a modified heat shock transformation protocol (2.2.1.1). Following the heat-shock and incubation on ice, transformed cells were recovered in LB medium overnight at 37 °C, 200 rpm. Cells were then applied to LB agar plates with 50 µg/mL kanamycin, 10 µg/mL tetracycline, 7 µg/mL gentamycin, 100 µg/mL X-Gal and 40 µg/mL isopropylthio-β-galactoside (IPTG), for the selection of recombinant clones by blue-white screening. Plates were incubated for a minimum of 24 hours at 37 °C. Suspected white colonies were selected for positive clones, indicating disruption of the β-galactosidase gene (*lacz*) by successful homologous recombination into the bacmid DNA. White colonies were re-streaked onto a fresh selective plate following inoculation of 2 mL LB media with 50 µg/mL kanamycin, 10 µg/mL tetracycline and 7 µg/mL gentamycin. Cultures were incubated overnight at 37 °C, 200 rpm. After confirmation of a blue-white phenotype, cells were harvested by centrifugation (3,000xg, 10 minutes, 20 °C). Bacmid DNA was isolated using a protocol adapted from the Bac-to-Bac® Expression System Manual (Invitrogen). Cells were lysed using a Nucleospin® Plasmid Miniprep Kit (Macherey-Nagel), and the lysate clarified twice by centrifugation at 11,000xg for 10 minutes. Bacmid DNA was precipitated in 40% (v/v) isopropanol and pelleted by centrifugation (11,000xg, 10 minutes, 20 °C). The pellet was carefully washed with 200 µL 70% (v/v) ethanol, added dropwise to the side of the tube opposite to the pellet. Following a final centrifugation (11,000xg, 10 minutes, 20 °C) the bacmid DNA was stored under 50 µL fresh 70% (v/v) ethanol and used immediately for transfection into Sf9 cells.

2.3.3 Sf9 transfection with bacmid DNA and small-scale expression testing

Following isolation, bacmid DNA was transfected into Sf9 cells using an adherent cell culture protocol scaled for 12-well plate format. In summary, mid-log phase Sf9 cells with >95% viability were diluted to a density of ~0.55 x 10^6 cells/mL in Insect XPRESS™ Protein-free Insect Cell Medium (Lonza). Cells were distributed in 2 mL aliquots into each well of a poly D-lysine coated 12-well flat bottom cell culture plate (Sarstedt) and left to adhere to the surface for 30 minutes at 27 °C. Two wells were prepared per construct. Bacmids were prepared for transfection by aspiration of the ethanol (2.3.2), being allowed to air dry, then resuspended in 30 µL sterile Milli-

Q®. 10 µL bacmid DNA was added to 150 µL Insect-XPRESS Protein-free media (Lonza) with 5 µL XtremeGENE High Performance Transfection Reagent (Roche) and incubated for 30 minutes at 20 °C. The transfection mixture was then added dropwise to the adherent cell cultures. Un-transfected cell control and cell free media control wells were also prepared. Plates were incubated for 72 hours at 27 °C and then the media containing recombinant baculovirus particles was collected as virus V0. Viruses were stored in LightSafe tubes at 4 °C.

The cells were supplemented with fresh Insect XPRESS™ Protein-free Insect Cell Medium (Lonza) and incubated for a further 72 hours at 27 °C to assess the efficiency of transfection and expression. Cells were then imaged using an EVOS FL Auto 2 Imagining System (Thermo Fisher Scientific) using both bright-field and GFP fluorescence modes ($\lambda_{excitation}$ 470-522$_{nm}$, $\lambda_{emission}$ 525-550$_{nm}$) for confirmation of GFP expression and plasma membrane localisation. Media was then removed and discarded, and the cells were harvested by scraping the wells and resuspending cells in 1 x PBS). Cells were lysed by sonication and then analysed by SDS-PAGE with fluorescence imaging (2.2.2.1) and anti-His Western blot (2.2.2.2).

2.3.4 Virus amplification

Virus amplification cultures were set up to both increase the volume of virus available for further infections and to increase the viral titre. Fresh, mid-log phase Sf9 cells with >95% viability were diluted to 1.0 x 10^6 cells/mL in Insect XPRESS™ Protein-free Insect Cell Medium (Lonza) in Erlenmeyer flasks, and 25 mL of cells were infected with 3 mL of V0. Cells were incubated at 27 °C, 120 rpm and cell density and viability was monitored daily. Infected cell densities were maintained at <1.5 x 10^6 cell/mL until proliferation arrest (PA) was achieved, + 48 hours. At PA +48 the media was harvested following centrifugation (300xg, 5 minutes, 20 °C) and stored in LightSafe tubes at 4 °C as virus V1.

2.3.5 Viral titre test and preparation of baculovirus-infected insect cells

To determine the optimal infection and the day of best expression, fresh mid-log phase Sf9 cells with >95% viability were diluted to 1.0 x 10^6 cells/mL in Insect XPRESS™ Protein-free Insect Cell Medium (Lonza). In multiple Erlenmeyer flasks, 100 mL of cells were infected with V1 at 1:4,000, 1:2,000 or 1:1000. Cells were incubated at 27 °C, 120 rpm and cell density and viability was monitored daily to determine the titre of virus that allowed for one cell doubling event before PA, which corresponds to a multiplicity of infection (MOI) of <1.

For the cultures with a MOI <1, baculovirus-infected insect cell (BIIC) stocks were prepared to allow for long term storage of viruses, and for easy and efficient large-scale expression in Sf9 cells. At PA + 24, half of the cells were harvested by centrifugation (200xg, 10 minutes, 20 °C) and resuspended in sterile filtered cryopreservation medium (90% (v/v) Insect XPRESS™ Protein-free Insect Cell Medium (Lonza), 10 g/L bovine serum albumin (BSA), 10% (v/v) dimethyl sulfoxide (DMSO) to a density of 1.0 x 10^7 cells/mL. 1 mL BIIC aliquots were prepared in cryopreservation vials, and the vials stored in a Mr. Frosty™ Freezing Container (Thermo Fisher Scientific) for 48 hours at -80 °C before being transferred to liquid nitrogen for long term storage.

Following the removal of cells for BIIC preparations, the remaining cells were maintained until PA + 148. Cell density and viability was monitored daily, and equivalent samples were also taken for analysis in SDS-PAGE with fluorescence imaging (2.2.2.1) and anti-His Western blot (2.2.2.2) to assess the day of optimum protein expression.

2.3.6 Large-scale expression in Sf9 cells

All large-scale expressions in Sf9 cells were achieved using BIICs as a virus source. One BIIC was used per 1 L of Sf9 cells at 1.0 x 10^6 cells/mL in Erlenmeyer flasks. To infect cells, BIICs were thawed quickly through warming by hand and diluted into 50 mL pre-warmed insect cell medium before being add to large-scale cultures. 100 x Antibiotic-Antimycotic (Gibco™) was added to a final concentration of 1 x to prevent bacterial and fungal contamination. Infected cultures were incubated until day of best expression. Using this method, infected cells typically doubled once before achieving proliferation arrest. All constructs used in this work had a day of best expression of PA +48, so for all large-scale expression cells were harvested 72 hours after infection. Cells were harvested by centrifugation (700xg, 30 minutes, 4 °C), washed in ice-cold 1 x PBS pH 7.4 with 1 x cOmplete EDTA-free protease inhibitors (Roche) and flash frozen in aliquots equivalent to 250 mL – 1 L of Sf9 cell cultures. Cell pellets were stored at -80 °C.

2.4 Standardised protein purification and characterisation methods

The following give the overview of protocols used throughout this work; experimental specifics are given in detail later in the relevant sections.

2.4.1 Protein purifications

All steps were performed at 4 °C unless otherwise stated. Generally, Sf9 cell pellets were thawed on ice and then resuspended in 0.01 x equivalent volume of hypertonic buffer (40 mM Tris-HCl

pH 7.4, 5% (w/v) glycerol, 3 mM DTT and 1 x Pierce Protease Inhibitors (EDTA-free, Thermo Fisher Scientific) or cOmplete™ Protease Inhibitor Cocktail (EDTA-free, Roche)). Resuspended cells were lysed with 30-40 strokes in a Dounce glass homogeniser with a tight plunger (Wheaton). Protein was solubilised by the addition of 1% (w/v) detergent. Solubilisation was performed for 1 hour at 4 °C with rolling, and solubilised protein was isolated by centrifugation at 200,000xg for 1 hour at 4 °C in a Ti 45, Ti 50.2, or Ti 70.1 rotor (Beckman Coulter). The supernatant was incubated with Ni^{2+}-NTA resin (Thermo Fisher Scientific), at 1 mL resin / 1L of Sf9 cell culture equivalent, pre-equilibrated in purification buffer 1 (50mM HEPES pH 7.4, 400 mM NaCl, 20 mM imidazole pH 8.0, 5% (w/v) glycerol, 3 mM DTT and 1 x Pierce Protease Inhibitors (EDTA-free, Thermo Fisher Scientific) or cOmplete™ Protease Inhibitor Cocktail (EDTA-free, Roche) and detergent at 5-10 x critical micelle concentration (CMC)), for 1 hour at 4 °C with rolling. The supernatant and resin were then applied to an Econo-Pac® (Bio-Rad) gravity flow column and the flow through collected. The resin was washed with 20 x resin bed volumes of purification buffer 1, followed by 10 x resin bed volumes of purification buffer 2 (50mM HEPES pH 7.4, 50 mM NaCl, and detergent at 5-10 x CMC). Resin was resuspended in a volume of purification buffer 2 twice that of the resin volume, and human rhinovirus (HRV) 3C or tobacco etch virus (TEV) protease added to a final concentration of 0.05 mg/mL. Protease digestion was incubated for 1-2 hours at 4 °C with end over end turning. The digested sample was then applied to an Econo-Pac® (Bio-Rad) gravity flow column, the flow through collected and the protein eluted with 10 x resin bed volumes of purification buffer 2. Any proteins that remained on the Ni^{2+}-NTA resin were eluted with purification buffer 3 (50mM HEPES pH 7.4, 50 mM NaCl, 250 mM imidazole pH 8.0, and detergent at 5-10 x CMC). Purification samples were applied to SDS-PAGE gel and fractions containing protein of interest were identified by Coomassie staining or Western blot, performed following the previously described protocols (2.2.2). Protein fractions were combined and concentrated to <1 mL using either a Vivaspin™ polyethersulfone membrane concentrator (Cytiva) or an Amicon® Ultracel regenerated cellulose membrane concentrator (Millipore), with a 50 kDa molecular weight cut off (MWCO), pre-equilibrated in protein elution buffer. Protein concentration was determined by UV-Vis spectrophotometry ($A280_{nm}$) using the 'protein' application on a DS-11 spectrophotometer (DeNovix®).

2.4.2 Protein modifications

2.4.2.1 Exchange into amphipol A8-35

To determine the optimal ratio for exchange from detergent into amphipol A8-35, a series of samples at differing mass ratio of purified protein:A8-35 were prepared; 1:0, 1:1, 1:2, 1:3, 1:4,

1:5 and 1:10. Samples were incubated for 20 minutes at 4 °C with inversions. The 1:0 sample was prepared in duplicate to serve as two differing controls. Following exchange, free detergent was removed by the addition of Bio-Beads™ SM-2 (Bio-Rad) at 0.2g/mL and overnight incubation at 4 °C with inversions. Bio-Beads™ were omitted from one of the 1:0 control samples. Samples were then aspirated, and any precipitated protein removed by centrifugation (20,817xg, 20 minutes, 4 °C). Efficiency of exchange was then analysed by $A280_{nm}$. Once the optimal ratio for exchange was determined the protocol was repeated using that mass ratio only for the full protein sample, small volumes were reserved for the inclusion of 1:0 controls

2.4.2.2 Biotinylation

Biotinylation was achieved using EZ-Link™ Sulfo-NHS-SS-Biotin (Thermo Scientific™), following the manufacturers protocol. In brief, a 20-fold molar excess of EZ-Link™ Sulfo-NHS-SS-Biotin was added to purified protein and incubated at 4 °C for 4 hours. Unreacted biotin was removed by dialysing in protein elution buffer, using a Slide-A-Lyzer™ 3 mL cassette with a 20k MWCO (Thermo Scientific™).

To determine biotin labelling efficiency, a 4'-hydroxyazobenzene-2-carboxylic acid (HABA)/biotin displacement assay was performed following the procedure detailed by G-Biosciences[234] (https://info.gbiosciences.com/blog/how-much-biotin-is-coupled-to-my-protein-the-haba/avidin-assay). In summary, the absorbance at 500_{nm} (A_{500}) of a HABA/avidin solution (0.3 M HABA, 0.3 mM, 0.5 mg/mL avidin) solution in a cuvette with a 1 cm path length was recorded. Biotinylated protein was added in a 1:10 dilution and the change in $A500_{nm}$ was monitored, A_{500} values were recorded once they remained steady for 15 seconds. The A_{500} of the biotin free sample was corrected using a correction factor of 0.9, and then the amount of biotin in the sample was determined using Beer Lambert Law ($A_\lambda = \varepsilon_\lambda bC$), where $A = \Delta A_{500}$ and $\varepsilon = 34{,}000$. The concentration of biotin in the sample was then related to the concentration of protein to inform labelling efficiency.

2.4.3 Radioligand binding

A working stock of [Benzyl-3H]-nitrobenzylthioinosine (PerkinElmer) ([3H]-NBMPR) was prepared by diluting stock [3H]-NBMPR to a final concentration of 2 µCi/mL (64 nM). A working stock of dipyridamole (1 mM) was prepared in DMSO. Generally, 500 µL of purified protein at 0.2 mg/mL or 500 µL whole Sf9 cells resuspended in 1 x ice-cold PBS with 1 x cOmplete protease inhibitors, EDTA-free (Roche) or 1 x Pierce Protease Inhibitors, EDTA-free (Thermo Fisher Scientific) were incubated on ice. Protein samples were divided into ten 50 µL aliquots for three

technical repeats of three assay conditions: (1) no inhibitor (-/-), (2) 32 nM [3H]-NBMPR (+/-) and (3) 32 nM [3H]-NBMPR and 20 µM dipyridamole (+/+). All conditions were incubated for 1 hour, on ice. Samples were applied to a GF/B filter (Whatman) pre-equilibrated in ice cold 1 x PBS on a vacuum manifold (Promega). The liquid was pulled through the filters under vacuum and washed three times with 2 mL of ice cold 1 x PBS. Filters with 0, 1, 2.5, 5, 7.5, 10, 12.5 and 15 µL of [3H]-NBMPR working stock (64 nM) were also prepared to allow for the determination of scintillation counting efficiency and to fit a standard curve. Filters were incubated overnight at room temperature in 10 mL of Ultima Gold scintillant (PerkinElmer). Radioactive disintegrations from bound [3H]-NBMPR in samples were quantified in counts per minute using a TriCarb scintillation counter (PerkinElmer) using 10-minute reads, which were performed twice. Background (−/−) and non-specific binding (+/+) were subtracted from the [3H]-NBMPR (+/-) values to determine protein specific binding. The specific radioactive signal was normalised against either the concentration of protein or the in-gel protein-linked signal intensity.

2.4.4 Stability assay: Ten-temperature challenges

(15 mL equivalent) cell pellets were thawed on ice and resuspended in the corresponding resuspension buffer to a final volume of 500 µL. Detergent was added to 1% (w/v). Solubilisation was performed for 1 hour at 4 °C with inversions, and solubilised protein was isolated by centrifugation at 20,817xg for 1 hour at 4 °C in a F45 rotor (Eppendorf). 10 x 50 µL aliquots of supernatant were prepared, one aliquot was retained on ice (4 °C) while the remaining nine aliquots were incubated in a T100 thermal cycler (Bio-Rad) at a single temperature (30, 35, 40, 45, 50, 55, 60, 65 or 70 °C) for 10 minutes, followed by a 10-minute incubation at 4 °C. Following heating, precipitated protein was removed by centrifugation (20,817xg, 1 hour, 4 °C). 40 µL of each supernatant was then transferred to a fresh well of a 96-well PCR plate and 10 µL of 5x SDS loading dye was added. SDS-PAGE was performed, and the protein visualised using fluorescence imaging (2.2.2.1). Signal intensity of ENT-linked GFP bands were quantified using ImageJ. Temperature challenges were performed in biological triplicates.

2.4.4.1 Data-fitting and statistical analysis

The T_m of the ENT in each condition was obtained by plotting the fluorescence signal of each temperature point after normalisation to the on-ice sample. Data were fit with a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting in GraphPad Prism 9.0. The inflection point of the fitted curve represents the temperature at which half of the protein is denatured and is assigned as T_m. ΔT_m was calculated by subtracting the relevant negative

control T_m value. The standard error of the mean (SEM) was calculated for all values, with error propagation factored in for ΔT_m using the following equation:

$$SEM_{\Delta T_m} = \sqrt{\left(SEM_{\Delta T_{m-ve\ control}}\right)^2 + \left(SEM_{\Delta T_m}\right)^2}$$

Statistical analysis was performed using ordinary one-way ANOVA with a Dunnet follow-up test for multiple comparisons in GraphPad Prism 9.0.

2.5 hENT1

2.5.1 Purification of hENT1 from Sf9 cells

The methods for the expression and purification of an N-terminally tagged 8xHis, 2x TEV protease recognition sequence (ENLYFQ|G) hENT1 (NHT-hENT1) were established by a previous member of the Goldman group, Dr. Steven Harborne. Typically, purifications of cell pellets equivalent to 1.2-6 L of Sf9 cell culture were performed as described in 2.4.1, with solubilisation by 1% (w/v) LMNG with 1 mg/mL cholesterol hemi succinate (CHS) and 20 µM NBMPR. 0.01% (w/v) LMNG, 10 µg/mL CHS and 20 µM NBMPR were included in all subsequent purification buffers (Table 2.5). Protease digestion was performed by incubation with TEV protease at a final concentration of 0.05 mg/mL for 2 hours at 4 °C, with inversions. Eluted protein was combined and concentrated to <1 mL using a 50 kDa MWCO Vivaspin™ concentrator (Cytiva), pre-equilibrated with hENT1 purification buffer 2 (Table 2.5). Final concentrations typically achieved 2-4 mg/mL, as estimated by absorbance at $A280_{nm}$. Samples from throughout the purification were analysed by SDS-PAGE (2.2.2.1) with Coomassie staining and anti-hENT1 Western blot (2.2.2.2). Concentrated samples were typically highly heterogeneous, and thus additional size exclusion chromatography (SEC) steps were performed.

Table 2.5 Buffers used in the purification of hENT1

		Purification			
	Hypertonic	**1**	**2**	**3**	**SEC**
Tris HCl pH 7.4 (mM)	40	0	0	0	0
HEPES pH 7.4 (mM)	0	50	50	50	50
NaCl (mM)	0	400	50	50	150
Imidazole pH 8.0 (mM)	0	20	0	250	0
Glycerol (% (w/v))	5	5	0	0	0
Protease Inhibitor[a] (x)	1	1	0	0	0
DTT (mM)	3	0.1	0	0	0
CHS (µg/mL)	0	10	10	10	5
LMNG (% (w/v))	0	0.01	0.01	0.01	0.005
NBMPR (µM)	20	20	20	20	10

[a]Pierce Protease Inhibitors (EDTA-free, Thermo Fisher Scientific) or cOmplete™ Protease Inhibitor Cocktail (EDTA-free, Roche)

2.5.2 Size exclusion chromatography

SEC was performed at 4 °C on an ÄKTA FPLC protein purification system (GE Healthcare). A Superdex 200 10/300 column (GE Healthcare) was pre-equilibrated with 2 x column volumes of filtered Milli-Q® H_2O, followed by 2 x column volumes of hENT1 SEC buffer (50 mM HEPES pH 7.4, 150 mL NaCl, 10 µM NBMPR, 5 µg/mL CHS and 0.005% (w/v) LMNG). Concentrated protein sample (<1 mL at 2-4 mg/mL) was centrifuged (10,000xg, 10 minutes, 4 °C) to remove aggregated protein, and injected onto the column. The column was run at 0.35 mL/min. Fractions containing protein of interest were applied to SDS-PAGE gel (2.2.2.1) and confirmed by Coomassie staining or Western blot (2.2.2.2). Protein fractions were combined and concentrated using a 50 kDa MWCO Vivaspin™ concentrator (Cytiva), pre-equilibrated in SEC buffer (Table 2.5). Final protein sample concentration was determined by absorbance at $A280_{nm,}$ and typical final yields obtained were 30-50 µL at 2-6 mg/mL

2.5.2.1 Size exclusion chromatography with multiple angle laser light scattering

Size exclusion chromatography with multiple angle laser light scattering (SEC-MALLS) was performed at 20 °C on a high-performance liquid chromatography (HPLC) protein purification system equipped with UV/VIS and fluorescent detectors (Shimadzu) and a DAWN HELEOS MALS detector (Wyatt Technology). Analysis was performed on a SEC purified sample of NHT-hENT1 in LMNG (30 µL at 2.13 mg/mL) using a Superdex200 10/300 column pre-equilibrated with 2 x column volumes of filtered Milli-Q® H_2O, followed by 2 x column volumes of SEC buffer (Table 2.5). Molecular weights were determined with the Astra package (Wyatt Technology) using conjugate analysis for the determination of the fraction of protein in complex with detergent

micelles. Conjugate analysis refractive index (dn/dc) LMNG modifier was therefore assigned at 0.132 mL/g. The method was calibrated using alcohol dehydrogenase standards. Protein of interest UV extinction coefficient was assigned at 1.135 mL/(mg.cm).

2.5.3 Purification optimisations

2.5.4 Generating an immobilised adenosine resin

Immobilised AMP resin was obtained with two different immobiliser attachment points on the nucleobase (Jena Bioscience), 8-amino-hexyl-AMP and N6-(6-amino)hexyl-AMP. To remove phosphate from the ribose moiety, resin samples (5-7 mM) were incubated with 500 units Antarctic phosphatase (AP) (EC 3.1.3.1, New England Biolabs) for 90 minutes at 37 °C, followed by heat inactivation for 2 minutes at 80 °C. Resin samples were taken at 45 minutes and 90 minutes and a Baginksi[235] assay was performed to determine the rate of phosphate release. 1 mL 7% (w/v) ammonium (hepta)molybdate was added to 10 mL acid solution (3% (w/v) ascorbic acid, 0.5 M HCl) and the solution incubated for 1 hour on ice. In a 96 well microplate, 60 µL of solution was added to 50 µL of the resin sample, and reactions incubated for 20 minutes, on ice. 90 µL sodium arsenite solution (20 mg/mL sodium arsenite, 20 mg/mL trisodium citrate and 0.02% (v/v) glacial acetic acid) was then added, and the reactions incubated for 2 hours at room temperature. The plate was read at absorbance 860_{nm}. Inorganic phosphate standards were included to allow for the fit of a standard curve and determination of free phosphate from the resin.

2.5.4.1 Adenosine affinity purification

Two Sf9 cell pellets equivalent to 300 mL culture were thawed on ice and then resuspended in 0.01 x equivalent volume of hypertonic buffer (Table 2.5). Resuspended cells were lysed with 30-40 strokes in a Dounce glass homogeniser with a tight plunger (Wheaton). Protein was solubilised by the addition of 1% (w/v) LMNG with 1 mg/mL CHS. Solubilisation was performed for 1 hour at 4 °C with rolling, and solubilised protein was isolated by centrifugation at 200,000xg for 1 hour at 4 °C in a Ti 45 rotor (Beckman Coulter). One supernatant was incubated with AP hydrolysed 8-amino-hexyl-AMP resin, and the second with the AP hydrolysed N6-(6-amino)-hexyl-AMP resin, both resins were pre-equilibrated in purification buffer 1, without NBMPR (Table 2.5). Samples were incubated for 1 hour at 4 °C with rolling. Supernatants and resins were applied to an Econo-Pac® gravity flow column (Bio-Rad). The flow through was collected and the resins washed with 20 x resin bed volumes of purification buffer 1, without NBMPR (Table 2.5), followed by 20 x resin bed volumes of purification buffer 2 (Table 2.5) and finally, 20 x resin bed

volumes of adenosine and adenosine monophosphate (AMP) elution buffer (50 mM HEPES pH 7.4, 50 mM NaCl, 0.1 M adenosine, 0.1 M adenosine monophosphate, 0.01% (w/v) LMNG and 10 ng/mL CHS). Samples from throughout the purification were analysed by SDS-PAGE (2.2.2.1) with Coomassie staining and anti-His Western blot (2.2.2.2).

2.5.4.2 Buffer optimisation

Generally, purifications of cell pellets equivalent to 1-2 L of Sf9 cell culture were performed as described in 2.4.1, with solubilisation by 1% (w/v) LMNG with 1 mg/mL CHS and 20 µM NBMPR. 5% (w/v) glycerol was included in purification buffer 2 (Table 2.5). Protease digestion was performed by incubation with TEV protease at a final concentration of 0.05 mg/mL for 2 hours at 4 °C, with inversions. Eluted protein fractions were combined and concentrated to <500 µL using a Vivaspin™ 50 kDa MWCO concentrator (Cytiva), pre-equilibrated in purification buffer 2 (Table 2.5). Final yields typically achieved 2-3 mg/mL, as estimated by Pierce™ BCA Protein Assay. Samples from throughout the purification were analysed by SDS-PAGE (2.2.2.1) with Coomassie staining and anti-His Western blot (2.2.2.2). Concentrated samples were typically highly homogenous, and thus no further purification steps were required.

2.5.5 Radioligand binding assay

Purified hENT1, obtained following 2.5.4.2, was diluted to 0.25 mg/mL and a single concentration radioligand binding assay was performed as detailed in 2.4.3. An unrelated membrane protein was included as a negative control. The specific radioactive signal was normalised to the concentration of protein, as estimated by $A280_{nm}$.

2.5.6 Generating nanobodies against hENT1

2.5.6.1 Camelid affinity matured nanobodies

The purification of cell pellets equivalent to 10 L of Sf9 cell culture was performed as described in 2.5.1. SEC was performed using the two-step purification product of 1 mL at 4 mg/mL. SEC elutions were combined and concentrated to 200 µL at 2.0 mg/mL, as estimated by $A280_{nm}$. Immunisation and nanobody selection with this protein was performed by the Steyaert Lab at VIB-VUB, and they generously provided me with the plasmids and sequences for in-house expression and validation (Appendix Table 2).

2.5.6.1.1 Nanobody expression in *E. coli*

Plasmids were transformed and amplified in OmniMAX™ cells as per 2.2.1.1. Plasmids were isolated and then heat shock transformed (2.2.1.1) into *E. coli* WK6 cells (genotype: *F' lacI^q Δ(lacZ)M15 proA$^+$B$^+$ traD36I Δ(lac-proB) galE rpsL*). Transformed cells were applied to ampicillin selective LB agar plates with 2% (w/v) D-glucose. Positive colonies were used to inoculate 10 mL of ampicillin selective LB media with 2% (w/v) D-glucose and 1 mM $MgCl_2$ and then incubated overnight at 37 °C, 200 rpm. 5 mL of the overnight cultures were then used to inoculate 1 L of ampicillin selective TB media with 0.1% (w/v) glucose and 2 mM $MgCl_2$. Cells were then incubated at 37 °C, 200 rpm until an OD_{600} of 0.8-1.2 was achieved. Cells were then induced with 1 mM IPTG and then incubated overnight at 28 °C, 200 rpm. The OD_{600} of induced cells were measured before being harvested by centrifugation (700xg, 15 minutes, 20 °C), flash frozen and stored at -80 °C. Samples of cell suspensions pre and post induction were analysed by SDS-PAGE (2.2.2.1) and anti-His Western blot (2.2.2.2).

2.5.6.1.2 Purification of nanobodies

Cell pellets were resuspended in TES buffer (0.2 M Tris pH 8.0, 0.5 mM EDTA, 0.5 M sucrose). 15 mL TES was used per 1 L culture at OD_{600} of 25. Resuspended pellets were incubated for 1.5 hours at 4 °C with rolling. Additional TES, diluted 4-fold in H_2O, was added at twice the volume as the original resuspension, and the samples incubated for a further 1.5 hours at 4 °C with rolling. The periplasmic fractions were isolated by harvesting the supernatant by centrifugation (1000xg, 30 minutes, 4 °C). The supernatants were then incubated with 1 mL Ni^{2+}-NTA, pre-equilibrated with phosphate buffer 1 (50 mM Na phosphate, 1 M NaCl, pH 7.0), and then applied to an Econo-Pac® gravity flow column (Bio-Rad). The flow through was collected and the resin washed with 20 x resin bed volumes of phosphate buffer 1. Nanobodies were then eluted in 10 x resin bed volumes of competition elution buffer (50 mM Na phosphate, 150 mM NaCl, 300 mM imidazole, pH 7.5), followed by 10 x resin bed volumes of low pH elution buffer (50 mM NaAc, 1 M NaCl, pH 4.5). Low pH elution fractions were instantly neutralised by the addition of 1M Tris-HCl pH 7.5 to a final dilution of 170 mM. The fractions containing protein were identified by A_{280}, then combined and concentrated using a 5 kDa MWCO Vivaspin™ concentrator (Cytiva), pre-equilibrated in 1 x PBS, to 500 µL at 3.5-12.7 mg/mL as estimated by A_{280}. Nanobody yields were analysed by SDS-PAGE with Coomassie staining (2.2.2.1) and anti-His Western blot (2.2.2.2). For full extraction of nanobodies, the above purification was repeated a second time.

To improve the homogeneity of the purification yields, the above expression and purification was repeated but with the inclusion of 30 mM imidazole in phosphate buffer 1.

2.5.6.1.3 Nanobody validation

To determine whether nanobodies bound to hENT1, complex formation was assessed using Ni^{2+}-NTA capture of the His-tagged nanobodies following incubation with a tag-free hENT1. hENT1 was purified from cell pellets equivalent to 2 L of Sf9 cell culture as described in 2.5.1, except DDM was used in place of LMNG. Solubilisation was performed using 1% (w/v) DDM, and 0.05% (w/v) DDM was included in all buffers. hENT1 purification elutions were combined and concentrated to 300 µL at 1.0 mg/mL, as estimated by $A280_{nm}$. hENT1 was then incubated with the nanobodies purified in 2.5.6.1.2 in a 1:1.2 molar ratio[187]. The samples were incubated for 2 hours, on ice and then incubated with Ni^{2+}-NTA, pre-equilibrated with purification buffer 2 (Table 2.5), for 2 hours at 4 °C, with rolling. Samples were applied to Mini-Bio-Spin® columns (Bio-Rad) and washed with 20 x resin bed volumes purification buffer 2 (Table 2.5), and then 10 x resin bed volumes of nanobody elution buffer (2.5.6.1.2). Flow through, wash and elution samples were analysed by SDS-PAGE (2.2.2.1) with anti-His and anti-hENT1 Western blot (2.2.2.2).

2.5.6.1.4 Modifications of hENT1 for improved nanobody selection

2 x 2 L equivalent Sf9 cell pellets were purified as described in 2.5.4.2, but one purification was performed with solubilisation in 1% (w/v) DDM in place of LMNG, and all purification buffers for that condition contained 0.05% (w/v) DDM in place of LMNG. The final yield from the DDM purification was 70 µL at 4 mg/mL, and the final yield from the LMNG purification was 150 µL at 1.7 mg/mL. Following the protocol as detailed in 2.4.2.1, the LMNG purified sample was exchanged into A8-35 using a molar mass ratio of 1:2. The A8-35 and DDM samples were then split into two equivalent volumes and one volume from each condition was labelled with EZ-Link™ Sulfo-NHS-SS-Biotin (Thermo Scientific™) following the methods as described in 2.4.2.2. Dialysis of unreacted biotin was performed in purification buffer 2 (Table 2.5), detergent was omitted from the dialysis of the A8-35 to prevent detergent re-associating to the protein.

2.5.6.2 Yeast surface display

2.5.6.2.1 Fluorophore labelling

Ahead of screening a nanobody library using a yeast surface display system, with nanobody expression and nanobody-antigen binding analysed by magnetic and fluorescent activated cell sorting (MACS and FACS, respectively), the protein of interest and a nanobody specific anti-human influence virus hemagglutinin (HA) were labelled with fluorescent dye. Fluorophore labelling was achieved using DyLight™ NHS Labelling Dyes 488 and 650 (Thermo Scientific™), following the manufacturers protocol. In brief, a 10-15-fold molar excess of DyLight™ Dye was added to purified hENT1 or anti-HA IgG and incubated for 1 hour at 4 °C. Unreacted dye was removed by dialysing in purification buffer 2 (Table 2.5), using a Slide-A-Lyzer™ 3 mL cassette with a 20k MWCO (Thermo Scientific™).

2.5.6.2.2 Nanobody library expansion

2 mL aliquots at 10^{10} cells/mL aliquots of yeast cells encoding for the yeast display nanobody library were generously provided by the Kruse Lab (Harvard Medical School). As per the protocol outlined in McMahon *et al*.[186] cells were thawed on ice and then resuspended in 1 L of Yglc -Trp selective media pH 4.5 (0.67% (w/v) yeast nitrogen base, 100 units/mL Pen-Strep, 20% (w/v) D-glucose, 35 mM citric acid monohydrate, 40 mM sodium citrate, L-alanine 0.15 mg/mL, L-arginine 0.15 mg/mL, L-asparagine 0.15 mg/mL, L-aspartic acid 0.15 mg/mL, L-cysteine 0.15 mg/mL, L-glutamic acid 0.15 mg/mL, L-glutamine 0.15 mg/mL, L-glycine 0.15 mg/mL, L-histidine 0.15 mg/mL, L-isoleucine 0.15 mg/mL, L-leucine 0.75 mg/mL, L-lysine 0.15 mg/mL, L-methionine 0.15 mg/mL, L-phenylalanine 0.15 mg/mL, L-proline 0.15 mg/mL, L-serine 0.15 mg/mL, L-threonine 0.15 mg/mL, L-tyrosine 0.15 mg/mL, L-valine 0.15 mg/mL, adenine 0.04 mg/mL, inositol 0.15 mg/mL, p-aminobenzoic acid 0.01 mg/mL and uracil 0.15 mg/mL). Cells were incubated overnight at 30 °C, 230 rpm. Overnight cultures were then diluted in an additional 2 L of Yglc -Trp selective media pH 4.5 and incubated for 48 hours at 30 °C, 230 rpm to achieve stationary phase. Cells were then harvested by centrifugation (3,500xg, 5 minutes, 20 °C), then resuspended with Yglc -Trp selective media pH 4.5, supplemented with 10% (v/v) DMSO, to a final density of 10^{10} cells/mL. 2 mL aliquots were prepared in cryopreservation vials and stored at -80 °C.

2.5.6.2.3 Nanobody expression

A single aliquot of cells was thawed on ice and used to inoculate 1 L Yglc -Trp selective media pH 6.0. The culture was incubated at 30 °C, 230 rpm until an OD_{600} of 1.0 was reached. Cells were then harvested by centrifugation (3,500xg, 5 minutes, 20 °C) and then resuspended in 1 L Yglc -Trp selective media pH 6.0, but with 2% (w/v) galactose instead of 2% (w/v) glucose. Cells were incubated for up to 72 hours at 30 °C, 250 rpm.

2.5.6.2.4 Nanobody expression testing and analysis of non-specific binding

After 72 hours incubation, three aliquots of 1 x 10^6 cells were washed with 500 µL purification buffer 2 (Table 2.5) supplemented with 0.1% BSA and harvested by centrifugation (3,500xg, 1 minute, 4 °C). Cells were then resuspended in 100 µL purification buffer 2 with 0.01% (w/v) LMNG. 0.5 µg anti-HA labelled with DyLight™ 650 was added to one sample. To a second sample hENT1 labelled with DyLight™ 488 was added to a final concentration of 1 µM, with 0.5 µg anti-HA labelled with DyLight™ 650. The third sample was retained as unstained cells, free of antigen to act as a negative control. Samples were incubated for 15 minutes at 4 °C, then washed twice with 500 µL purification buffer 2 with 0.01% (w/v) LMNG. Final pellets were resuspended in 100 µL purification buffer 2 and analysed by flow cytometry. Flow cytometry was performed using a CytoFLEX (Beckman Coulter) using the 488_{nm} laser with a 525/40 band pass filter for the analysis of cells stained with DyLight™ 488, and the 640_{nm} laser with a 660/20 band pass filter for the analysis of cells stained with DyLight™ 650. Unstained, antigen free cell samples were used to establish population gates.

To try to overcome non-specific binding, the above methods were repeated with an additional blocking step before incubation with hENT1. Cells were incubated with 1 x PBS with either 1% (w/v) BSA or 0.005% (w/v) LMNG for 15 minutes at 4 °C. Cells were then washed with 500 µL 1 x PBS, then harvested by centrifugation (3,500xg, 1 minute, 4 °C) for incubation with antigen, as above.

2.6 IMPROvER

2.6.1 Expression cultures for stability testing

All virus stocks available in storage at 4 °C from work on IMPROvER variants of hENT1 by previous members of the Goldman lab were collected and catalogued. For each virus available, three biological repeat expression cultures were set up using fresh mid-log phase Sf9 cells, diluted to

a density of approximately 1.0 x 10^6 cells/mL in 15 mL pre-warmed Insect-XPRESS™ Protein-free Insect Cell Medium (Lonza). Cultures were set up in 50 mL Erlenmeyer flasks. To standardise the workload, viruses were used at the following ratios for infection: V0 at 1:10, V1 at 1:20 and V2 at 1:40. Cultures were infected with virus, and incubated for 72 hours at 27 °C, 270 rpm. Cells were then harvested by centrifugation at (250x*g*, 10 minutes, 20 °C) and stored at -20 °C until use.

2.6.2 Ten-temperature stability assays

Ten-temperature stability assays were performed using all cell pellets generated, following the method described in 2.4.4. Cell pellets were resuspended in resuspension buffer (1 x PBS pH 7.4 with 1 x Pierce Protease Inhibitors (EDTA-free, Thermo Fisher Scientific)). Cell resuspensions were solubilised by the addition of 1% (w/v) DDM. Once viruses/variants of interest were identified, 30 mL expression cultures were repeated using the corresponding viruses, in four biological repeats. Cells were harvested as 2 x 15 ml equivalent pellets. To determine the stabilising effects of NBMPR, repeat ten-temperature stability assays were performed. One pellet was suspended in resuspension buffer supplemented with 20 µM NBMPR, and the second with resuspension buffer free of NBMPR. Assays were performed using three biological repeats.

2.6.3 Bacmid DNA validation from transfected Sf9 cells

2.6.3.1 Bacmid DNA extraction

To allow for the assignment of variant identity to the data generated in the ten-temperature assay, bacmid DNA was extracted from both whole and insoluble fractions of Sf9 cells using a NucleoSpin® Plasmid isolation mini kit (Macherey-Nagel) following the manufacturer's protocol for plasmid DNA isolation of low-copy plasmids from *E. coli*. However, where manufacturer's guidelines suggest use of 5-10 mL of a saturated *E. coli* Luria broth (LB) culture, I substituted '1.0 x 10^6 cells/mL of Sf9 culture' for 1 mL saturated *E. coli* culture. In this work, 15 mL Sf9 cultures at 1.0 x 10^6 cells/mL were processed as equivalent to 15 mL saturated *E. coli* cultures. 5-15 mL volumes of Sf9 cells were used with success. The pellet achieved following lysis is characteristically more glutinous than is seen with *E. coli* and therefore requires additional care during aspiration. I sent this initially-extracted bacmid DNA for sequencing as per standard protocols. However, this was unsuccessful and resulted in low signal to noise ratios and overlapping peaks in chromatograms. This is possibly due to mixed DNA present in the samples or packing of the DNA in a way that was incompatible with the techniques used[236]. I therefore decided to PCR-amplify the extracted bacmid DNA before sequencing.

2.6.3.2 DNA amplification and sequencing

The construct used for the expression of hENT1 variants has a C-terminal TEV-GFP-His[8] (hENT1-CTGH) and was originally cloned into a pFastBac™ Expression Vector (Invitrogen). PCR amplification of the hENT1 gene portion of the construct was performed using primers #1.1 and 1.2 (Table 2.3) and using the method as described in 2.2.1.3. PCR clean-up was performed using a Nucleospin® Gel and PCR Clean Up Kit (Macherey-Nagel), as per the manufacturers protocol. The final DNA products were sent for Sanger sequencing (Eurofins) using primers #2.1 and 2.02 (Table 2.4). Sequencing was analysed using GeniousPrime 2021.2.2.

2.6.4 Radioligand binding assay

Single cell pellets for the variant T336A and wild-type hENT1 were resuspended in resuspension buffer (phosphate buffered saline pH 7.4 and EDTA-free protease inhibitor cocktail) to a final volume of 500 µL. The optical density of each cell resuspension was measured at 600_{nm} to inform normalisation and a sample of the cell resuspension was analysed by SDS-PAGE with fluorescence imaging (2.2.2.1). The intensity of the signal of the hENT1-linked GFP band was quantified using FIJI/ImageJ. A single concentration radioligand binding assay was performed as detailed in 2.4.3. A non-transfected Sf9 cell pellet was included as a negative control. The specific radioactive signal was normalised for each sample against the OD_{600} and the intensity of the in-gel GFP-linked band for each corresponding hENT1 variant. These normalised values were then scaled relative to wild type values.

2.7 Homologues of hENT1

2.7.1 Bioinformatics

The sequences of putative homologues of hENT1 were selected for by first determining species of interest through literature searches to identify thermotolerant or extremophile species. BLAST® (https://blast.ncbi.nlm.nih.gov/Blast.cgi, https://fungi.ensembl.org/index.html, http://ciliate.org/index.php/home/welcome) was then performed against hENT1 to identify proteins of interest (Appendix Table 1). Constructs for expression were designed in GeneiousPrime.2018.1. Synthetic double stranded DNA GeneArt® Strings™ were purchased from Invitrogen, and all genes were codon optimised for expression in Sf9 cells (2.1.5). A multiple sequence alignment was constructed of all selected ENT versions, aligned using the multiple sequence comparison by log- expectation (MUSCLE) tool and alignments visualised using Jalview. The secondary structure of all were predicted using Phobius[237] transmembrane

topology predictor. Parameters for the theoretical purified proteins were predicted using ProtParam (ExPASy)[233].

2.7.2 Expression in Sf9 cells

Five hENT1 homologue synthetic GeneArt® Strings™ were assembled into a pFastBac™ (Thermo Fisher Scientific) vector encoding a human rhinovirus (HRV) 3C-cleavable N-terminal His_8-GFP tag (NHGV-*xx*ENT). The vector was linearised by restriction digest using $NheI^+$ and $PstI^+$ (New England Biolabs) as per the manufacturer's recommendation (2.2.1.4), and gene inserts were amplified using the primers as detailed in Table 2.3. Ligation independent cloning was performed as described in 2.2.1.5, colony PCR was performed to select for positive clones using the primers as used for the amplification of the insert. Plasmids were isolated from the positive clones and submitted for sequencing to Eurofins Genomics using primers #2.3 and #2.4.

Positive plasmids were transformed into *E. coli* DH10Bac™ cells. Bacmids preparation, Sf9 cell transfection and virus amplifications were performed as described in 2.3.2 to 2.3.4. Titre tests were performed as described in 2.3.5, and cells that were not used for the preparation of BIICs were used for solubilisation screens.

2.7.3 Solubilisation screening

Solubilisation trials were performed using whole cells resuspended in hypertonic buffer (Table 2.5) to an OD_{600} of 1.0. Cells were lysed with a Dounce glass homogeniser (Wheaton), and 1 mL aliquots were prepared per condition. The detergents screened are detailed in

Table 2.6, and all detergents were screened in the presence and absence of the cholesterol mimic CHS. Detergents were added to a final concentration of 1% (w/v), and CHS at 1 mg/mL. A control sample containing no detergent was prepared, and a sample containing the detergent n-dodecyl-phosphocholine (Fos-choline-12 (FC12)) was prepared to provide an approximation of the upper solubilisation limit. Samples were incubated for 1 hour at 4 °C with end over end turning, and then centrifuged at 280,000xg for 1 hour at 4 °C using a TLA-100 rotor (Beckman Coulter). The soluble and insoluble fractions were separated and analysed by SDS-PAGE (2.2.2.1) with anti-His Western blot (2.2.2.2). In-gel band intensities were quantified by densitometry using Fiji/ImageJ, and the solubilisation efficiency was determined by relating the amount of signal in the soluble fraction to the total signal of soluble and insoluble combined. Reported solubilisation efficiencies in this preliminary screen correspond to a single measurement.

Table 2.6 Detergents included in the solubilisation screening of hENT1 homologues.

Detergent		Formula	MW	CMC % (w/v) (in H_2O)
n-dodecyl-phosphocholine	(FC12)	$C_{17}H_{38}NO_4P$	351.5	0.047
lauryl maltose neopentyl glycol	(LMNG)	$C_{47}H_{88}O_{22}$	1005.19	0.001
decyl maltose neopentyl glycol	(DMNG)	$C_{43}H_{80}O_{22}$	949.08	0.0034
octyl glucose neopentyl glycol	(OGNG)	$C_{27}H_{52}O_{12}$	568.69	0.058
n-dodecyl-β-D-maltopyranoside	(DDM)	$C_{24}H_{46}O_{11}$	510.6	0.0087
n-decyl-β-D-maltopyranoside	(DM)	$C_{22}H_{42}O_{11}$	482.6	0.087
n-octyl-β-D-maltopyranoside	(OM)	$C_{20}H_{38}O_{11}$	454.4	0.89

2.8 *Bs*ENT

2.8.1 Solubilisation screens

Further solubilisation screens were performed for the homologue of interest, '*Bs*ENT' from *Byssochlamys spectabilis*. Additional small scale expression cultures were set up using fresh mid-log phase Sf9 cells, diluted to a density of approximately 1.0 x 10^6 cells/mL in 30 mL pre-warmed Insect-XPRESS™ Protein-free Insect Cell Medium (Lonza). Cultures were set up in 50 mL Erlenmeyer flasks. Cells were infected with V1 at 1:4000, incubated for 72 hours at 27 °C, 270 rpm. Cells were then harvested by centrifugation at (250x*g*, 10 minutes, 20 °C) and stored at -20 °C until use.

A ten-temperature stability assay was performed for *Bs*ENT as described in 2.4.4, in experimental triplicates. Cells were thawed on ice and resuspended in in hypertonic buffer (40 mM Tris-HCl pH 7.4, 1 x Pierce Protease Inhibitors (EDTA-free, Thermo Fisher Scientific) and 3 mM DTT) and solubilised in 1% (w/v) DDM. The determined apparent T_m was then used to perform a single temperature challenge, in which 32 different solubilisation and buffer conditions were screened: solubilisation in 1% (w/v) LMNG, DMNG, DDM and DM; with and without 1 mg/mL CHS; in hypertonic buffer A, B, C or D (Table 2.7).

Table 2.7 Buffer composition used in the extended solubilisation screen

	Buffer composition
A	Hypertonic (40 mM Tris-HCl pH 7.4, 1 x Protease Inhibitor[a], 3 mM DTT)
B	Hypertonic + 20 µM NBMPR
C	Hypertonic + 5% (w/v) glycerol
D	Hypertonic + 20 µM NBMPR + 5% (w/v) glycerol

[a]Pierce Protease Inhibitors (EDTA-free, Thermo Fisher Scientific) or cOmplete™ Protease Inhibitor Cocktail (EDTA-free, Roche)

One cell pellet per hypertonic buffer condition was thawed on ice and resuspended in 2 mL hypertonic buffer (A, B, C or D). Resuspensions were then split into eight 250 µL aliquots and

solubilised with 1% (w/v) detergent (LMNG-, DMNG-, DDM- and DM-), with 1 mg/mL CHS where appropriate (LMNG+, DMNG+, DDM+ and DM+). Solubilisations were incubated for 1 hour at 4 °C with inversions and then solubilised protein was harvested by centrifugation (20,817x*g,* 1 hour, 4 °C). Supernatant was split into two x 100 µL aliquots. One aliquot was retained on ice while the second aliquot was incubated at 60 °C for 10 minutes in a T100 thermocycler (Bio-Rad), then at 4 °C for 10 minutes. Following the temperature challenge all aggregated protein was removed by centrifugation (20,817x*g,* 1 hour, 4 °C). The supernatant was then measured for GFP fluorescence using a QFX Fluorometer (DeNovix®) ($\lambda_{excitation}$ 470_{nm}, $\lambda_{emission}$$514\text{-}567_{nm}$). Non-transfected Sf9 cell samples were included to determine background cell fluorescence. The specific fluorescence of the 58.4 °C sample was then related to the on-ice control and all 32 conditions were analysed collectively. Data was generated in three biological repeats. All samples for the most stabilising condition, DMNG -, and the least stabilising, DDM +, were then validated by SDS-PAGE with fluorescence imaging for the visualisation of *Bs*ENT-linked GFP signal. The most stabilising conditions, 'DMNG –' in buffers A, B, C and D, then had a full ten-temperature challenge performed, as performed for DDM. Data was generated in three biological repeats, and data analysis performed as described in 2.4.4

2.8.2 Large scale expression and purification of *Bs*ENT

Large scale expression of *Bs*ENT was achieved using the BIIC infection method as described in 2.3.6. Generally, purifications of cell pellets equivalent to 2-4 L of Sf9 cell culture were performed as described in 2.4.1, with solubilisation by 1% (w/v) DMNG, and 0.02% (w/v) DMNG and 10% (w/v) glycerol included in all purification buffers (Table 2.8). Protease digestion was performed by incubation with HRV protease at a final concentration of 0.05 mg/mL for 2 hours at 4 °C, with inversions. Eluted protein fractions were combined and concentrated to <500 µL using an Amicon® regenerated cellulose 50 kDa MWCO concentrator (Millipore), pre-equilibrated in purification buffer 2 (Table 2.8). Final yields typically achieved 4-20 mg/mL, as estimated by $A280_{nm}$ and Pierce™ BCA Protein Assay. Samples from throughout the purification were analysed by SDS-PAGE (2.2.2.1) with Coomassie staining and anti-His Western blot (2.2.2.2).

SEC was performed at 4 °C on an ÄKTA FPLC protein purification system (GE Healthcare). A Superdex 200 10/300 column (GE Healthcare) was pre-equilibrated with 2 x column volumes of filtered Milli-Q® H_2O, followed by 2 x column volumes of *Bs*ENT SEC buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 0.01% (w/v) DMNG). Concentrated protein sample (<1 mL at 2-4 mg/mL) was centrifuged (10,000xg, 10 minutes, 4 °C) to remove aggregated protein, and injected onto the

column. The column was run at 0.35 mL/min. SEC determined that concentrated two-step purification samples were typically highly homogenous and eluted with a monodisperse peak. Therefore, additional SEC purification was not performed as standard.

Table 2.8 Buffers used in the purification of *Bs*ENT

		Purification			
	Hypertonic	**1**	**2**	**3**	**SEC**
Tris HCl pH 7.4 (mM)	40	0	0	0	0
HEPES pH 7.4 (mM)	0	50	50	50	50
NaCl (mM)	0	400	50	50	150
Imidazole pH 8.0 (mM)	0	20	0	250	0
Glycerol (%(w/v))	10	10	10	10	0
Protease Inhibitor[a]	1	1	0	0	0
DTT (mM)	3	0.1	0	0	0
DMNG (% (w/v))	0	0.02	0.02	0.02	0.01

[a]Pierce Protease Inhibitors (EDTA-free, Thermo Fisher Scientific) or cOmplete™ Protease Inhibitor Cocktail (EDTA-free, Roche)

2.8.3 Crystallisation trials

*Bs*ENT was purified from 4 L equivalent Sf9 pellets as described in 2.8.2 and the yields were separated into two. One sample was further concentrated to 21 µL at 16.4 mg/mL ($A280_{nm}$) and was used to prepare sitting drop vapour diffusion crystallisation trials in the commercial sparse-matrix screen MemGold2 (Molecular Dimensions). 96 well MRC 2 drop plates were set up with a reservoir volume of 70 nL and a 100 nL sitting drop of a 1:1 ratio of protein to mother liquor using a Mosquito liquid handler (TTP Labtech).

The second protein aliquot was used to prepare LCP crystallisation trials, with monoolein as the host lipid. Protein was enriched in the mesophase following the cubicon method[238]. In brief, monoolein was molten at 40 °C, and 2 x 100 µL Hamilton syringes were pre-warmed to 40 °C. A third 100 µL Hamilton syringe was kept on ice. The cooled syringe was loaded with 20 µL protein solution, a pre-warmed syringe was loaded with 10 µL molten monoolein, and then the syringes connected by a coupler. Protein was then pushed into the lipid by steady mixing at ~1-2 strokes per seconds, for 3-5 minutes. The mesophase and aqueous phase were separated and isolated into one syringe each. The syringes were then un-coupled, and the protein-depleted aqueous phase dispensed into an empty PCR tube. The same volume of protein as was removed in the aqueous phase syringe was then loaded into the empty syringe, the syringes coupled, and the method repeated. Enrichments were repeated until all protein was used, ensuring that the ratio of lipid:protein was maintained at 1:2 throughout. Once all protein was enriched in the

mesophase, LCP was prepared by adding additional molten monoolein to the mesophase so that the final ratio of lipid:protein was 3:2 and mixing until the solution turned clear. The final LCP was 30 µL at 24.5 mg/mL (the 'equivalent' stock solution/protein concentration excluding lipid was 61.2 mg/mL). The commercial sparse-matrix screen MemGold, MemMeso, MemTrans, MemSys and MemStart (Molecular Dimensions) were diluted to 70% (v/v) for better compatibility with LCP. 50 nL LCP aliquots were dispensed by a NT8® crystallisation robot (Formulatrix) onto glass sandwich plates (Jena Biosciences) and overlaid by 800 nL of mother liquor. All plates were sealed with glass covers. The aqueous phase samples were analysed by SDS-PAGE to observe for any protein loss (2.2.2.1).

Vapour diffusion and LCP plates were incubated at 20 °C in a temperature controlled automated imaging system (RockImager® 1000, Formulatrix). Plates were imaged using second order nonlinear imaging of chiral crystals (SONICC) with second harmonic generation (SHG) and UV two-photon excited fluorescence (UV-TPEF) imaging. Crystals grown by vapour diffusion were harvested by Dr. Maren Thomsen, cryo-cooled in liquid nitrogen, and shot at the i24 beamline at the Diamond Light Source.

2.8.4 Expression in *S. Cerevisiae*

2.8.4.1 Cloning of constructs

The *Bs*ENT construct for expression in *S. cerevisiae* was achieved by the ligation independent cloning of a *Bs*ENT insert from the pFastBac™-NHGV plasmid into a pDDGFP vector that encodes for a N-terminal 8xHis-tag, followed by superfolder GFP and a TEV protease cleavage sequence (NHsfGT), to produce an pDDGFP-NHsfGT-*Bs*ENT expression vector. The plasmid was generously provided by my colleague Dr. Jannik Strauss and has a region that encodes for a membrane bound pyrophosphatase (PPase). The plasmid was linearised by inverse PCR (primers #1.13 and 1.14) to allow for the excision of the (PPase) and to facilitate the ligation independent cloning. Initial vector linearisation was unsuccessful at the calculated primer T_m of 61 °C, so small scale (5 µL) PCR was repeated using a temperature gradient of 55-70 °C at the annealing step. This determined that the optimal annealing temperature was 55.4 °C, so a repeat PCR was performed with the optimised annealing temperature. The *Bs*ENT insert was linearised by PCR amplification (primers #1.15 and 1.16). Ligation independent cloning was performed as described in 2.2.1.5, colony PCR was performed to select for positive clones using the same primers as *Bs*ENT insert amplification, and isolated plasmids were submitted for sequencing to Eurofins Genomics using primers #2.3, 2.5 and 2.6.

2.8.4.2 Transformation and expression testing

Transformation was performed using two strains of *S. cerevisiae*; BJ1991 (genotype: *MATα, pep4-3, prbl-1122, ura3-52, leu2, trpl*) and FGY217 (genotype: *MATα, ura3-52, pepΔ4, lys2Δ201)*. And all *S. cerevisiae* transformations were done following a standard heat transformation and LiAc, single-stranded carrier DNA, polyethylene glycol (PEG) method from Gietz & Schiestl[239].

In summary, strains of *S. cerevisiae* were incubated on YPD plates (1% (w/v) yeast extract, 2% (w/v) peptone, 2 % (w/v) D-glucose, 1.5% (w/v) agar) for 72 hours at 30 °C. 5 mL YPD liquid media (1% (w/v) yeast extract, 2% (w/v) peptone, 2 % (w/v) D-glucose) was inoculated with a single colony and incubated overnight at 30 °C, 200 rpm. 1 mL from the overnight starter culture was harvested per transformation and cells were harvested by centrifugation (3,000xg, 5 minutes, 20 °C). Cells were washed in three times in Milli-Q® water before being resuspended in 360 µL yeast transformation buffer (33.3% (w/v) PEG3350, 10 mM LiAc, 100 µg denatured salmon sperm, 125-250 ng DNA) and then heat shocked for 40 minutes at 42 °C. Cells were then harvested by centrifugation (3,000xg, 5 minutes, 20 °C) and resuspended in sterile MilliQ®. Cells were then applied to synthetic complete dropout media minus uracil (SCD-Ura) selection plates (0.67% (w/v) yeast nitrogen base, 2% (w/v) D-glucose, 2% (w/v) agar, 0.1 mg/mL carbenicillin, 0.41 mg/mL L-threonine, 0.10 mg/mL L-phenylalanine, 0.06 mg/mL L-lysine, 0.04 mg/mL of L-arginine, 0.04 mg/mL L-histidine, 0.04 mg/mL L-leucine, 0.04 mg/mL L-methionine, 0.04 mg/mL L-tryptophan, 0.04 mg/mL L-tyrosine, 0.04 mg/mL adenine) and incubated at 30 °C until colonies formed (typically ~72 hours).

2.8.4.3 Expression

Single yeast colonies were used to inoculate 5 mL starter cultures in SCD-Ura media and incubated overnight at 30 °C, 250 rpm. Overnight cultures were then diluted to OD_{600} of 0.1 in SCD-Ura with reduced glucose (0.1% (w/v) instead of 2% (w/v)) and incubated at 30 °C until an OD_{600} of 0.2-1.0 was reached. Protein expression was induced by the addition of 2% (w/v) galactose and incubated for up to 24 hours at 30 °C, 250 rpm. Cells were monitored hourly for OD_{600} and GFP fluorescence, measured using a QFX Fluorometer (DeNovix®) ($\lambda_{excitation}$ 470_{nm}, $\lambda_{emission}514\text{-}567_{nm}$). Cells were also imaged using an EVOS FL Auto 2 Imagining System (Thermo Fisher Scientific), using both bright-field and GFP fluorescence modes ($\lambda_{excitation}$ $470\text{-}522_{nm}$, $\lambda_{emission}$ $525\text{-}550_{nm}$) for visualisation of GFP expression and plasma membrane localisation.

2.8.5 Expression in *P. Pastoris*

2.8.5.1 Cloning of constructs

The *Bs*ENT construct for expression in *P. pastoris* was achieved by the ligation independent cloning of the full expression cassette for NHGV-*Bs*ENT the pFastBac™ plasmid into a pPICZB vector, to generate a pPICZB-NHGV-*Bs*ENT expression vector. The pPICZB plasmid was generously provided by my colleague Yue Ma and has a region that encodes for an ion channel. The plasmid was linearised by inverse PCR (primers #1.17 and 1.18) to allow for the excision of the ion channel and to facilitate the ligation independent cloning. The *Bs*ENT insert was linearised by PCR amplification (primers #1.19 and 1.20). Ligation independent cloning was performed as described in 2.2.1.5, colony PCR was performed to select for positive clones using the same primers as *Bs*ENT insert amplification, and isolated plasmids were submitted for sequencing to Eurofins Genomics using primers #2.7-2.11.

2.8.5.2 Transformation

Transformation was performed using two strains of *P. pastoris*; SMD1163 and SuperMan5 (genotype: *his4 pep4Δ, prb1Δ*). And all *P. pastoris* transformations were done following a standard electroporation transformation method, with LiAc and DTT pre-treatment, by Wu & Letchworth[240].

In summary, the recombinant pPICZB-NHGV-*Bs*ENT was linearised by inverse PCR (primers #1.21 and 1.22) as per 2.2.1.4. Glycerol stocks of *P. pastoris* were used to inoculate 10 mL YPD media and starter cultures were incubated overnight at 30 °C, 250 rpm. The overnight cultures were diluted to OD_{600} 0.5 in 100 mL of fresh YPD media and incubated at 30 °C, 250 rpm until OD_{600} reached 1.0-2.0. The number of cells were calculated using OD_{600} of 1.0 = 5 x 10^7 cells/mL. Cells were then harvested by centrifugation (1,500xg, 5 minutes, 4 °C) and then resuspended in sterile filtered LiAc/DTT solution (100 mM LiAc, 10 mM DTT, 600 mM sorbitol and 10 mM Tris HCl pH 7.5) to 1 x 10^8. Cells were incubated for 30 minutes at 20 °C, 100 rpm and then harvested by centrifugation (1,500xg, 5 minutes, 4 °C). To ensure that the cells are suspended in a salt-free and osmotically stabilising condition, cells were then washed three times by resuspension in ice cold 1 M sorbitol (1.5 mL per 5 x 10^8 cells), harvesting by centrifugation (1,500xg, 5 minutes, 4 °C). Cells were then resuspended in ice cold 1 M sorbitol to 1 x 10^{10} cells/mL and stored on-ice until transformation.

2 mm electroporation cuvettes were pre-chilled on ice. 100 ng linearised plasmid DNA and 80 µL of prepared cells were added to the cuvette and incubated for a further 5 minutes on ice. Cells were then pulsed using a Gene Pulser (Bio-Rad) set to 1.5 kV, 25 µF capacitance and 186 Ω resistance. Following electroporation, 1 mL of ice cold 1 M sorbitol was added, and the cuvettes incubated for 3 hours at 30 °C. Settled yeast was then aspirated from the bottom of the cuvette chamber and applied to a Zeocin™ (Gibco™) and sorbitol selective YPD plate (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) D-glucose, 1.5% (w/v) agar, 1 M sorbitol and Zeocin™ 0.2 mg/mL). Plates were incubated for 72 hours at 30 °C.

2.8.5.3 Expression testing

Colonies were used to inoculate 500 µL YPD media and cultures were incubated overnight at 30 °C, 250 rpm. Cells were harvested by centrifugation (1,000xg, 5 minutes, 4 °C) and then resuspended in 500 µL induction media (2 x yeast nitrogen base, 100 mM potassium phosphate buffer pH 6.0 and 0.5% (v/v) methanol). Cultures were incubated for 7 hours at 30 °C, 250 rpm and then imaged using an EVOS FL Auto 2 Imagining System (Thermo Fisher Scientific), using both bright-field and GFP fluorescence modes ($\lambda_{excitation}$ 470-522$_{nm}$, $\lambda_{emission}$ 525-550$_{nm}$) for visualisation of GFP expression. Cells were then imaged using an LSM880+ Airyscan Inverted Confocal Microscope (Zeiss) using both brightfield and GFP imaging to determine GFP localisation. Confocal microscopy was performed by Dr. Ruth Hughes at the University of Leeds.

2.8.6 NBMPR binding

Using *Bs*ENT purified as described in 2.8.2, and hENT1 purified as described in 2.5.4.2 and in the absence of NBMPR, a single concentration [3H]-NBMPR binding assay was performed, as described in 2.5.5. Both protein samples were diluted to ~0.25 mg/mL, and an unrelated membrane protein was included as a negative control. The specific radioactive signal was normalised to the concentration of protein, as estimated by A280$_{nm}$.

Following on from this, a saturation binding assay for *Bs*ENT was performed. *Bs*ENT was further diluted to ~0.1 mg/mL. A working stock of [3H]-NBMPR solution was prepared by diluting stock [3H]-NBMPR to a final concentration of 32 µCi/mL (1 µM). Serial dilutions of 24, 16, 8, 3, 1.5 and 0.3 µCi/mL (0.75, 0.5, 0.25, 0.1, 0.05 and 0.01 µM) [3H]-NBMPR were then prepared in-ice cold 1 x PBS. A working stock of dipyridamole (1 mM) was prepared in DMSO. Protein was divided into 48 x 25 µL aliquots for three technical repeats per concentration of [3H]-NBMPR tested, and a 0 µm [3H]-NBMPR negative control, in two assay conditions: (1) [3H]-NBMPR (+/-) and (2) 32 nM [3H]-NBMPR and 20 µM dipyridamole (+/+). All conditions were incubated for 1 hour, on

ice. Samples were applied to GF/B filters (Whatman) pre-equilibrated in ice cold 1 x PBS on a vacuum manifold (Promega). The liquid was pulled through the filters under vacuum and washed three times with 2 mL of ice cold 1 x PBS. Filters with 0, 1, 2.5, 5, 7.5, 10, 12.5 and 15 µL of the 3 µCi/mL (0.1 µM) [3H]-NBMPR working stock were also prepared to allow for the determination of scintillation counting efficiency. Filters were incubated overnight at room temperature in 10 mL of Ultima Gold scintillant (PerkinElmer). Radioactive disintegrations from bound [3H]-NBMPR in samples were quantified in counts per minute using a TriCarb scintillation counter (PerkinElmer) using 10-minute reads. The total (+/-) and non-specific (+/+) binding were analysed by a One site – total and non-specific saturation binding curve in GraphPad Prism 9.0, and baseline correction performed by deduction of the non-specific binding.

2.8.7 Substrate screening

To try to identify endogenous substrates of *Bs*ENT, I opted to explore the thermostabilising effects of putative substrates using ten-temperature assays. Thermostabilising effects are reported as the difference between the determined T_m in resuspension buffer, and resuspension buffer + substrate (ΔT_m). The nucleosides and nucleobases selected for screening, and the concentrations in which they were used, were determined by compound solubility. Due to differing solvent requirements, different resuspension buffer conditions were required. The compounds screened and their corresponding resuspension buffers are detailed in Table 2.9. All conditions were solubilised by the addition of 1% (w/v) DDM. The remaining steps were performed as described in 2.4.4.

Table 2.9 Substrates and buffer conditions for thermostability screening of putative endogenous substrates of *Bs*ENT.

Resuspension buffer	Substrate/additive
pH 7.4 (1 x PBS, 1 x Protease Inhibitors)	Adenosine (7.5 µM or 5 mM)
	Cytidine (5 mM or 100 mM)
	Uridine (5 mM or 100 mM)
pH 7.4 (1 x PBS, 1 x Protease Inhibitors, 0.0005% (v/v) formic acid)	Adenine (7.5 µM)
	Guanosine (7.5 µM)
pH 8.2 (1 x PBS, 1 x Protease Inhibitors, 5 mM NaOH)	Uracil (5 mM)

Chapter 3 Expression and purification of hENT1

3.1 Introduction

While the first available structures of hENT1 have provided advancements in our understanding of ENTs, without additional structures the mechanism of action cannot be properly addressed. Therefore, there is still a significant amount to be gained from the pursuit of structures of ENTs. However, as discussed in 1.7, efforts towards the high-resolution structural characterisation of eukaryotic IMPs continually proves challenging due to bottlenecks at every step[159]. Achieving the yield and purity required for structural studies requires optimisation of protein expression, purification, stabilisation, and structural analysis on a protein-by protein basis[160,161].

3.1.1 Summary of previous work on hENT1

My project began as a continuation of the work of a previous member of the research group, Dr Steven Harborne, and my industrial collaborator Dr Veli-Pekka Jaakola. Dr Jaakola and a member of his research group, Dr Shahid Rehan, had previously established protocols for the expression and purification of hENT1 from Sf9 cells[241]. Dr Harborne continued with further optimisation and construct designs, during which N- vs C-terminal tags (FLAG, Strep, His_6) and fusion partners (BRIL and GFP) were investigated for expression in Sf9 cells. This concluded with the final construct design of an N-terminal His_6 tag, followed by two TEV protease recognition sequences (NHT-hENT1). This construct allows for a two-step purification protocol (immobilised metal affinity chromatography (IMAC) and protease digestion) for the elution of a tag free hENT1. Solubilisation screens identified LMNG supplemented with CHS as being the most stabilising detergent for the purification of NHT-hENT1. Further purification optimisation, such as investigating different IMAC resins, resin wash and elution conditions, and TEV protease digestion conditions, contributed to the establishment of the final protocol as detailed in section 2.5.1. However, yields obtained with the established protocols were not sufficient for structural analysis.

3.1.2 Aims and strategy

The work presented in this chapter focused on the investigation into the performance and further optimisation of the previously established protocols for the purification and structural analysis of hENT1. Furthermore, I set out to explore the use of nanobodies for the stabilisation and conformational selection of hENT1, in addition to their potential as crystallisation chaperones. To this end, I explored two distinct options for nanobody selection; the selection of

immune generated, affinity matured camelid nanobodies, and the use synthetic nanobody libraries for *in vitro* selection. Each of these methods have distinct advantages and disadvantages. Moreover, where challenges remain within each of the approaches, I set out to identify and address the limitations of the experiments with the current NHT-hENT1 construct.

3.2 Results

3.2.1 Purification of hENT1

3.2.1.1 Efficiency of a two-step purification

The expression and purification of hENT1 followed the previously established protocol, as described in section 2.5.1. Owing to difficulties faced in qualitative analysis of hENT1 as discussed in the upcoming sections of this chapter, accurate quantitative analysis of hENT1 is also a challenge. However, based on final yields and relative losses, hENT1 expression by Sf9 cells is estimated as 1 mg/L (Table 3.1). Overall, the efficiency of the existing purification protocol is poor. Following the solubilisation protocol that was established by Dr. Harborne, only ~30% solubilisation is achieved. In addition, IMAC efficiency is low, with ~35% of the solubilised protein lost in the flow through. Furthermore, ~60% of the bound protein precipitates on the resin following TEV protease digestion (Table 3.1). The average yields achieved following concentration are 0.35 mg /L culture, as estimated by $A280_{nm}$. However, the eluates are highly heterogenous (Figure 3.1A: El.) despite successful washing of the resin for the removal of proteins with non-specific and low-affinity binding (Figure 3.1A: Washes 1 and 2).

Table 3.1 Efficiency of the two-step purification of hENT1, presented as mg of hENT1 per L of Sf9 cell culture[a].

	Purification repeats					
	1	**2**	**3**	**4**	**5**	**6**
Total (mg/L)	0.91	0.93	1.06	0.94	1.03	0.90
Solubilised (mg/L)	0.25 (28%)	0.20 (22%)	0.38 (36%)	0.25 (27%)	0.34 (33%)	0.26 (29%)
Bound to resin (mg/L)	0.14 (55%)	0.14 (68%)	0.22 (58%)	0.16 (65%)	0.23 (67%)	0.15 (58%)
Eluted (mg/L)	0.06 (43%)	0.05 (38%)	0.10 (47%)	0.06 (35%)	0.10 (44%)	0.07 (45%)
Heterogeneity	High	High	High	High	High	High

[a]Quantities are estimates based on final purification yields and relative losses incurred throughout each purification step, as determined by in-gel hENT1 signal intensity in anti-hENT1 western blot

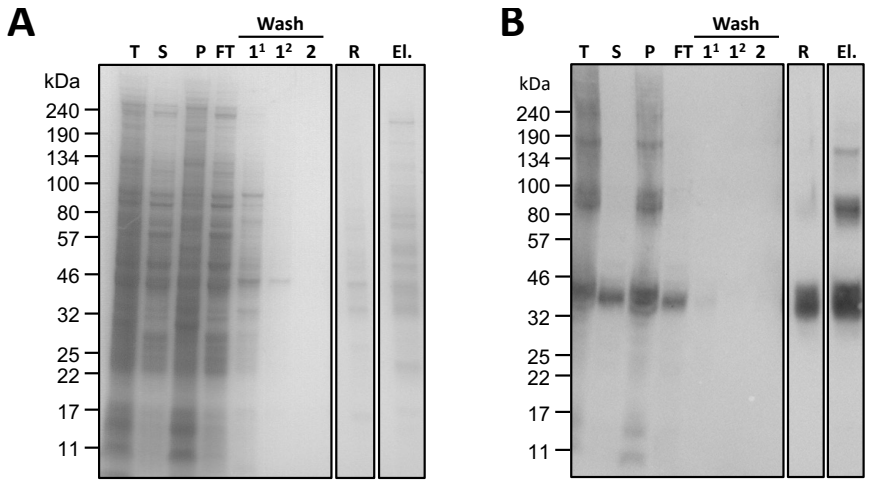


Figure 3.1 Two step purification of hENT1.

A representative SDS-PAGE gel of the two-step purification of hENT1 with (A) Coomassie staining and (B) anti-hENT1 Western blot. T = total, S = supernatant, P = pellet, FT = flow through, 1^1 = 10 resin bed volume high salt, low imidazole wash, 1^2 = 10 resin bed volume high salt, low imidazole wash, 2 = low salt wash, R = resin post digestion elution, El. = post digestion elution.

hENT1 yields with improved homogeneity that are obtained from subsequent SEC purification steps can help to estimate the quantity of hENT1 in heterogeneous fractions. This can be achieved by normalising samples with a known concentration (determined by A_{280}) to corresponding Western blot densities, and then relating these to hENT1 specific band densities from earlier purification steps. This method of quantification was used to determine losses and yields for fractions with high heterogeneity. This suggests that the two-step yield of hENT1 specifically is closer to 0.05-0.1 mg/L of culture. Therefore, non-hENT1 proteins contribute significantly to the heterogeneity of the elution. This is supported by the SDS-PAGE images (Figure 3.1). Therefore, additional proteins with some degree of nickel-affinity are possibly binding to the resin. IMAC purification had previously been investigated by Dr. Harborne, and elution heterogeneity was not improved when performed using alternative resins (data not shown). Consequently, yields from a two-step purification protocol require further, or alternate, purification steps to achieve sample qualities amenable to downstream processes.

3.2.1.2 SEC purification

The SEC trace (Figure 3.2A) of concentrated eluate following a two-step purification (Figure 3.1: El) shows a heterogeneous sample, but with monodisperse peaks. In the SEC trace shown in Figure 3.2A, fractions from the peak with a retention volume of ~11 mL were isolated and analysed by SDS-PAGE gel (Figure 3.2B and C). Anti-hENT1 Western blot confirms that these fractions contain hENT1, and Coomassie staining suggests that these fractions have improved homogeneity (Figure 3.2B). However, following concentration, the final average yields are

extremely low at 0.03 mg/L culture, as estimated by $A280_{nm}$ (Table 3.2). Furthermore, in the total purifications performed following this methodology, only 1.14 mg of protein was obtained from over 25 L of Sf9 cell culture. Therefore, while SEC does improve homogeneity, the current protocols are incredibly inefficient, and thus are not sustainable.

Table 3.2 Yields and heterogeneity of purification elutions, pre and post SEC

	Purification repeats					
	1	**2**	**3**	**4**	**5**	**6**
Sf9 culture eq. (L)	2.4	5.4	5.4	4.8	6.0	1.2
Final volume (µL)	30	30	50	100	120	50
Final conc. (mg/mL)	2.13	5.3	7.7	1.5	2.7	0.9
Final total yield (µg)	65	160	390	150	320	50
Post SEC yield (mg/L)	0.03	0.03	0.07	0.03	0.05	0.04

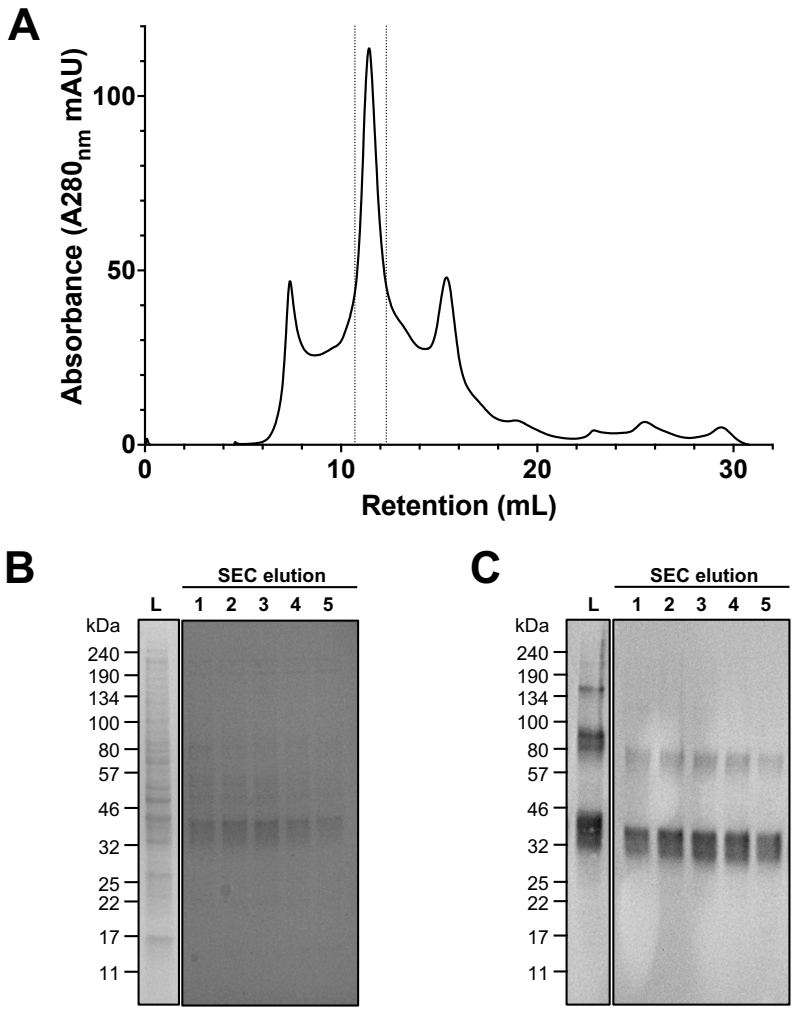


Figure 3.2 SEC purification of hENT1.

(A) SEC trace of post digestion elution sample. The peak containing the fractions analysed further in gel is highlighted by dotted lines. SDS-PAGE gel of highlighted SEC fractions with (B) Coomassie staining and (C) anti-hENT1 Western blot. L = concentrated protein sample applied to the SEC column, and 1-5 = the fractions of interest from the peak with the 11 mL retention volume.

3.2.1.3 The presentation of hENT1 in SDS-PAGE

3.2.1.3.1 Differing molecular weight presentations

Western blot using anti-hENT1 (Figure 3.1B and Figure 3.2C) produces a broad hENT1 specific band at 32-46 kDa. The physiologically relevant form of hENT1 is predicted to be a monomer (53 kDa). However, integral membrane proteins often elute on SDS-PAGE at lower than their expected MW[242,243]. Therefore, the broad band at 32-46 kDa probably indicates a monomeric species. Additional bands are observed in the total, pellet, and elution samples at ~80-100 kDa and ~160 kDa (Figure 3.1B and Figure 3.2C), which could indicate oligomeric species. In a previous study by Rehan & Jaakola[241,244] they also observed these higher MW species of hENT1. However, their additional analysis of NBMPR binding following fractionation on gel filtration identified that these species were non-functional and thus had been identified as SDS-resistant aggregates. Therefore, the higher MW species present in Western blot are likely similar SDS-resistant aggregates, rather than further quaternary structures.

As higher MW species of hENT1 observed in Western blot are likely aggregates, the presence or absence of these bands could be interpreted as an indicator of stability. Thus, the absence of these higher MW bands in the supernatant and flow through (Figure 3.1B) may suggest that hENT1 is more stable in these conditions.

3.2.1.3.2 Band broadening

In addition to the higher MW bands, all conditions except supernatant and flow through have distinctly broad/double bands at each of the different MW (Figure 3.1B and Figure 3.2C). hENT1 is known to have a site of N-linked glycosylation at N48[145]. Therefore, these broad bands could represent sub-species, such as with and without post translational modifications. Dr. Harborne had previously investigated whether de-glycosylation of purified hENT1 using Endo H and PNGase F resolves this broad band profile. However, this did not yield any improvements (data not shown).

The broadening of the bands appears to coincide with the increased presence of higher MW species. Therefore, it is possible that the broadening of the bands represents aggregation or degradation, rather than different expression products. Additionally, while higher MW aggregates are absent from the resin sample, the presence of a broad monomeric band could suggest that there are also concerns with protein stability in this condition and thus contributes to the significant losses during with protein precipitating on the resin.

3.2.1.3.3 Summary of the in-gel presentation of hENT1

The observed differences between the supernatant and flow through samples and the remaining samples could simply be attributed to differences in sample loading on the gel, with the broad and higher MW bands simply being more visible in more concentrated samples. However, there are distinct differences between the buffer conditions used to generate the supernatant and flow through, and those used to elute the protein. Therefore, it is possible that the observed differences are a true representation of stability/instability, rather than simply a handling error. Furthermore, the return of higher MW SDS-resistant aggregates in the eluates following IMAC suggests that there is something in these later conditions specifically that contributes to a change in the in-gel presentation. These identifiable and distinct differences between conditions allowed for quick and easy investigating of buffer optimisation for these later purification steps (3.2.4).

3.2.2 Establishing the oligomeric state of hENT1 in LMNG

While it is likely that the higher MW species of hENT1 observed in the Western blots can be attributed to SDS-resistant aggregation, size exclusion chromatography with multiple angle laser light scattering (SEC-MALLS) analysis was performed on a SEC purified sample of hENT1 in LMNG to confirm monomeric status, and to investigate the degree of aggregation.

3.2.2.1 Size exclusion chromatography - multiple angle laser light scattering (SEC-MALLS)

In the SEC-MALLS trace peak 1 (Figure 3.3B) represents detergent micelles containing protein and has a calculated MW of 39 kDa ±11%. Peak 2 represents empty LMNG micelles. The calculated mass of the protein in peak 1 is lower than that of hENT1 (53 kDa). However, using the method as detailed in 2.5.2.1, it is difficult to achieve independent resolution of empty detergent micelles and protein-detergent micelles due to the size similarity of the components. This results in overlapping peaks, as shown in Figure 3.3, and consequently contributes to inaccuracy of the calculated values. Therefore, here the SEC-MALLS analysis can only provide insight into oligomeric status and cannot be used to comment on protein size and subsequently, degradation. In these results, the sample appears to be homogeneous, with a monodisperse peak present in the UV trace representative of monomeric hENT1. Additionally, there is no evidence of hENT1 oligomerisation in SEC purified samples. Therefore, the higher MW species of hENT1 observed in Western blot are likely SDS-resistance aggregation and are unlikely to be physiologically relevant. Very little aggregation is present in the SEC-MALLS trace, with only a small contribution to the UV trace at the void volume (5-minute retention time, Figure 3.3A).

However, additional aggregates may have been removed during the sample preparation. Therefore, a reduced amount of aggregation may have been loaded on to the column for analysis.

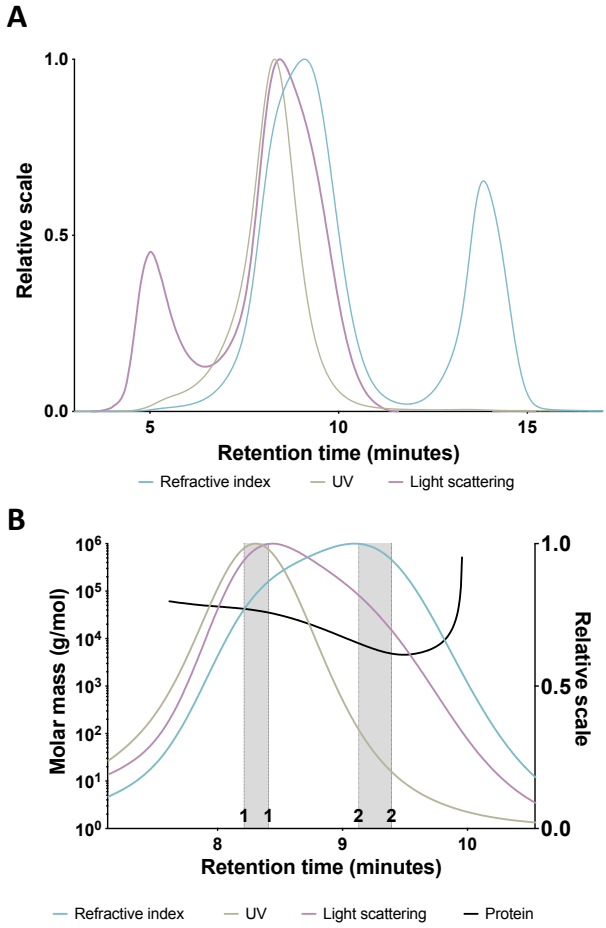


Figure 3.3 SEC-MALLS trace of hENT1 in LMNG.

(A) The SEC-MALLS trace shows some protein aggregation which elutes at the column void, observed as a minor shoulder in the UV trace at ~5 minutes retention time. (B) A view of the peak assignments with the molar mass calculations applied. The light scattering and refractive index traces both suggest the presence of two overlapping peaks. In the refractive index the shoulder of the trace is assigned as peak 1, and the apex of the trace is assigned peak 2. However, in the light scattering the apex of the trace is within peak 1, and the shoulder of the trace in peak 2. The UV trace further differentiates the two peaks, with a monodisperse peak within the peak 1 assignment. Any UV contribution in peak 2 is due to poor resolution and is unlikely to be a true UV contribution. Therefore, peak 1 can be assigned as the protein-detergent micelle peak and peak 2 the empty detergent micelle peak.

3.2.3 Optimisation of hENT1 purification

As SEC-MALLS suggests that hENT1 is present as a monomer, my focus moved to the optimisation of hENT1 purification. Here I used two distinct approaches. First, I investigated the use of substrate immobilised on resin. A similar methodology was used in the structural characterisation of the β2-adrenergic receptor[245] for the selection of functional protein. I was

interested in this approach for both potential improvements to specificity and efficiency of binding in relation to that achieved in the His-affinity IMAC. In addition, specific selection of functional protein would hopefully produce a final hENT1 sample with improved stability. Secondly, following on from opportunities for buffer optimisation as discussed in 3.2.1.3.3, I explored the buffer conditions used for the TEV-protease digestion and protein elution steps. I was interested in whether modifying the buffers at these steps so that they better resemble those of earlier steps will improve stability and consequently, the presentation of hENT1 bands in an anti-hENT1 Western blot.

3.2.3.1 Substrate affinity chromatography

3.2.3.1.1 Generating an immobilised adenosine resin

Two resins with adenosine mono phosphate (AMP) immobilised at differing attachment points were dephosphorylated to produce immobilised adenosine resins. A phosphate release assay (Figure 3.4A) suggests that only 20% dephosphorylation was achieved. Therefore, the final resins are predicted to contain both immobilised adenosine (Figure 3.4B and C) and immobilised AMP, in a 1:4 ratio.

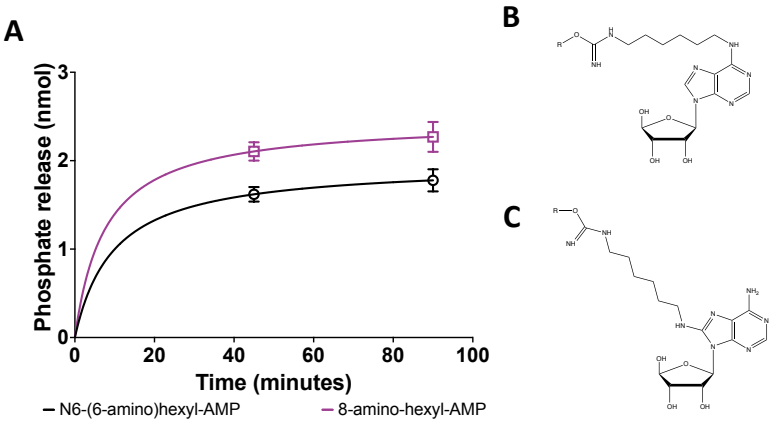


Figure 3.4 Generation of immobilised adenosine resins.

(A) Antarctic phosphatase was able to hydrolyse AMP as indicated by the ~2 nmol of free phosphate per 10 nmol AMP resin. While hydrolysis was successful, immobilised adenosine will be among immobilised AMP at a ratio of 1:4. The structures of the two different theoretical adenosine resins, with (B) N6-(6-amino)hexyl and (C) 8-amino-hexyl attachment points.

3.2.3.1.2 Adenosine resin affinity purification of hENT1

Small scale purifications were performed using the immobilised adenosine resins, with two theoretical elution conditions tested. First, I competitively eluted hENT1 using a buffer containing the high affinity and hENT1 specific inhibitor NBMPR (Figure 3.5: Wash 2). This was then followed by elution with a buffer containing increased concentrations of AMP and adenosine for the affinity elution of any other non-hENT1 specific binders (Figure 3.5: AA).

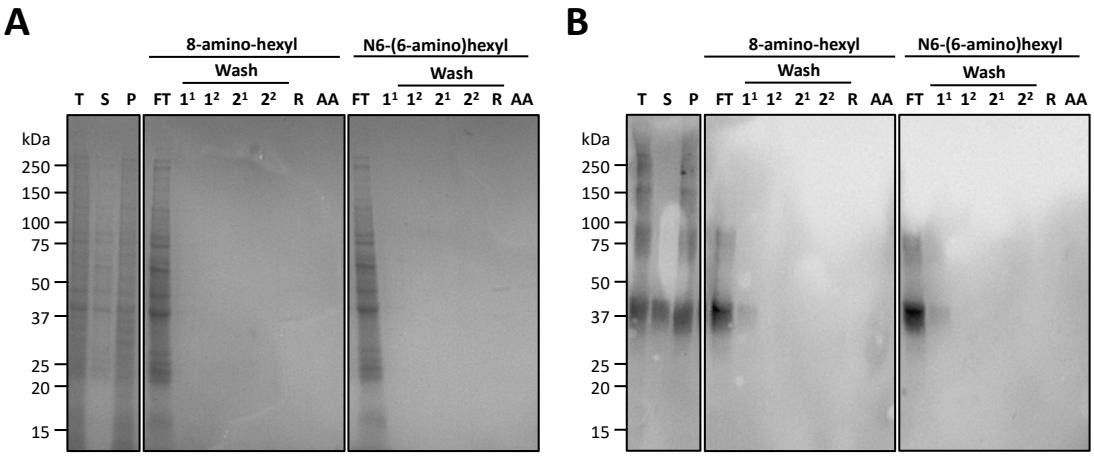


Figure 3.5 Substrate affinity purification of hENT1.

SDS-PAGE of the adenosine affinity purification of hENT1 with (A) Coomassie staining and (B) anti-hENT1 Western blot. T = total, S = supernatant, P = pellet, FT = flow through, 1^1 = 10 resin bed volume high salt wash, 1^2 = 10 resin bed volume high salt wash, 2^1 = 10 resin bed volume NBMPR wash, 2^2 = 10 resin bed volume NBMPR wash, R = resin post NBMPR elution, AA = 10 resin bed volume adenosine and AMP wash.

Coomassie stained SDS-PAGE suggests that all hENT1 eluted in the flow through and the first resin wash (Figure 3.5B: FT and Wash 1[1]), and that no protein was eluted in the NBMPR wash (Figure 3.5A and B: Wash 2). Therefore, binding of hENT1 to the resin was unsuccessful. As discussed in 0, previous studies have shown that ENTs have poor tolerance for chemical diversity in the type of interactions that lead to the high affinity binding of ligands[133]. Therefore, the modifications of the adenosine and the bulky nature of the immobilisation linker may impair or abolish binding by hENT1. While the affinity of hENT1 for adenosine (K_m 40 µM) may simply be too low to prevent dissociation from the resin during washes, the elution of nearly all protein at the flow through suggests that minimal binding of any protein to the resin occurred. Very minimal amounts of protein of any kind bound to the resin overall (Figure 3.5A: R). Anti-hENT1 Western blot suggests that any protein that did bind to the resin is not hENT1 (Figure 3.5B: R). Therefore, it is possible that the resin is either too specific and there are no other proteins present in the solubilised fraction with affinity for AMP or adenosine, or the dephosphorylation has inactivated the resin and abolished any binding capacity, for hENT1 or otherwise. Owing to these results this approach was not taken forward for further optimisation.

3.2.4 Buffer optimisation

One of the distinct differences between the buffer condition in the supernatant and flow through, and the buffer in the TEV protease digestion and protein elution is the glycerol content, with it being absent from the buffers in the latter conditions (2.5.1 and Table 3.3). Glycerol is known to contribute towards membrane protein stabilisation[246,247]. Therefore, purification with buffers from the existing protocol, supplemented with 5% (w/v) glycerol, was explored.

Table 3.3 A summary of the purification buffers and in-gel presentation of hENT1

		Purification buffer		
	Hypertonic	**1**	**2**	**SEC**
Tris HCl pH 7.4 (mM)	40	0	0	0
HEPES pH 7.4 (mM)	0	50	50	50
NaCl (mM)	0	400	50	150
Imidazole pH 8.0 (mM)	0	20	0	0
Glycerol (% (w/v))	5	5	0	0
Protease Inhibitor (X)	1	1	0	0
DTT (mM)	3	0.1	0	0
CHS (µg/mL)	0	10	10	5
LMNG (% (w/v))	0	0.01	0.01	0.005
NBMPR (µM)	20	20	20	10
Purification step	Solubilisation, IMAC	IMAC wash	Elution	Elution
SDS-resistant aggregates?	Insoluble only	N/A	Yes	Yes
Broad bands?	Insoluble only	N/A	Yes	Yes

As observed previously, the anti-His Western blot (Figure 3.6A) suggests that the SDS-resistant aggregates are absent from the supernatant and flow through samples. The Coomassie image (Figure 3.6B) supports the volume of washes being adequate for the removal of non-specific binding to the Ni^{2+}-NTA resin, as was seen in the original purification protocol (Figure 3.2A and B). However, following the inclusion of 5% (w/v) glycerol, protein that elutes following TEV protease digestion is significantly more homogeneous than obtained in the absence of glycerol. The homogeneity of the sample is sustained following concentration, and final yields of 0.14 mg/L culture are obtained. Furthermore, higher MW species and thus, SDS-resistant aggregates are absent (Figure 3.6B: hENT1). Therefore, glycerol appears to improve sample homogeneity and stability. However, a broad and diffuse band remains in the eluted sample, and thus suggest that issues with protein heterogeneity, aggregation, and/or degradation remain (Table 3.4).

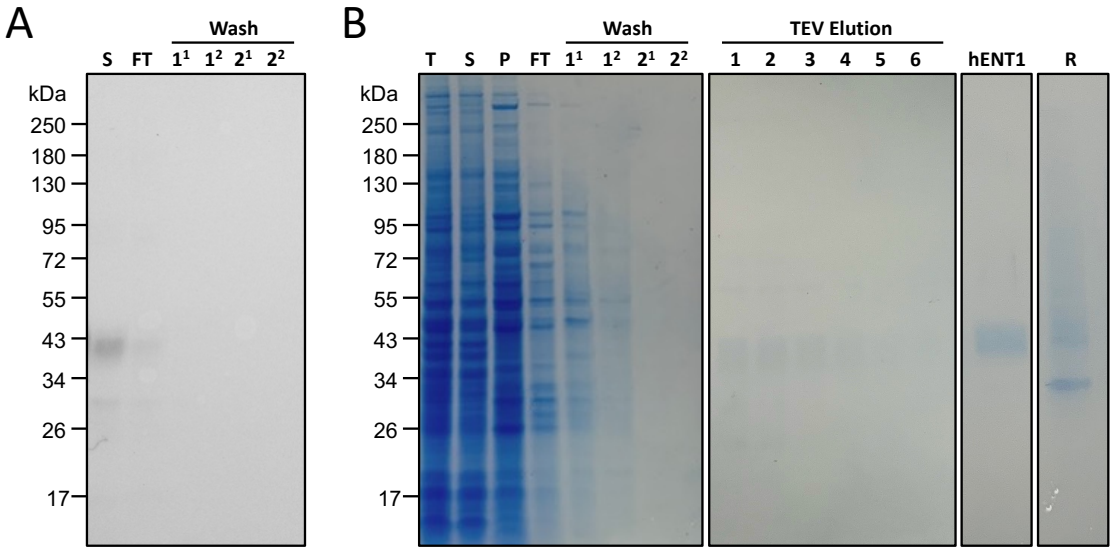


Figure 3.6 Buffer optimised purification of hENT1.

SDS-PAGE images of the buffer optimised purification of hENT1 with (A) anti-His Western blot and (B) Coomassie staining. Only Coomassie stained images are available for eluted hENT1 due to the expiration of anti-hENT1 antibodies used for Western blot. Anti-His is suitable for all purification steps prior to TEV protease cleavage of the N-terminal His6 tag. T = total, P = pellet, S = supernatant, FT = flow through, 1^1 = 10 resin bed volume high salt and low imidazole wash, inclusive of glycerol, 1^2 = 10 resin bed volume high salt and low imidazole wash, inclusive of glycerol, 2^1 = 10 resin bed volume low salt wash, inclusive of glycerol, 2^2 = 10 resin bed volume low salt wash, inclusive of glycerol, 1-6 = post TEV digestion elution fractions, hENT1 = combined and concentrated elution fractions, R = resin post digestion elution.

Table 3.4 Summary of the performance of the original, and a glycerol supplemented, purification of hENT1.

		Original	Glycerol supplemented
Pre SEC	Total (mg/L)	0.90	0.90
	Solubilised (mg/L)	0.27 (30%)	0.32 (35%)
	Bound to resin (mg/L)	0.17 (65%)	0.22 (70%)
	Eluted (mg/L)	0.07 (45%)	0.14 (65%)
	Heterogeneity	High	Low
Post SEC	Yield (mg/L)	0.04	
	Heterogeneity	Low	N/A
SDS-resistant aggregates?		Yes	No
Broad bands?		Yes	Yes

3.2.5 Investigating the quality of purified hENT1

3.2.5.1 Radioligand binding

To investigate the ability of purified protein to bind NBMPR, a single concentration radioligand binding assay was performed. hENT1 was purified following the glycerol supplemented protocol (2.4.3), but with the exclusion of NBMPR in all buffers. In this assay it was determined that for each pmol of purified hENT1 (as determined by $A280_{nm}$), 0.85 ± 0.10 pmol of [3H]-NBMPR is bound (Figure 3.7). hENT1 binds NBMPR in a 1:1 ratio. Therefore, while the broad band

presentation in-gel causes concern regarding degradation and/or instability, this assay suggests that almost all the purified hENT1, 85%, can bind NBMPR.

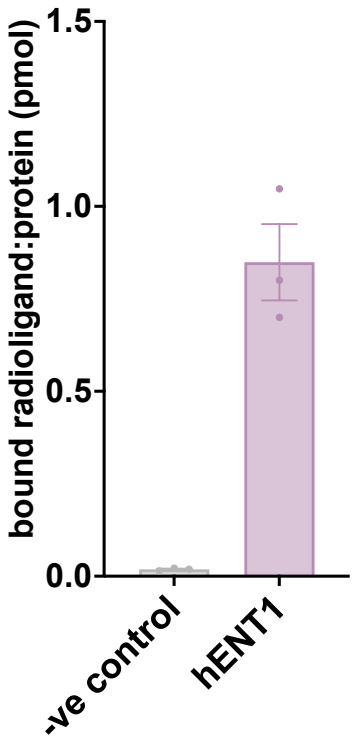


Figure 3.7 Specific [3H]-NBMPR binding of hENT1.

Relative amount of [3H]-NBMPR specifically bound to a purified negative control membrane protein in DDM, and to purified hENT1 in LMNG, as a pmol:pmol ratio. Error bars represent SEM, n = 3. Negative control membrane protein binds [3H]-NBMPR at 0.018 ± 0.002 per pmol of protein, and hENT1 binds [3H]-NBMPR at 0.85 ± 0.10 per pmol of protein.

3.2.5.2 Mass spectrometry of purified hENT1

To investigate the heterogeneity of the monomeric hENT1, analysis using liquid chromatography-mass spectrometry (LC-MS) was attempted. The rationale for LC-MS was to assess whether there was one MW species present, and the persistent broad band presentation is simply an in-gel artefact, or if there is a mix of hENT1 sub-species within the 32-46 kDa MW range. hENT1 was purified following the glycerol supplemented protocol (2.5.4.2). However, analysis could not be performed on purified protein due to recurrent issues with protein precipitation. Similarly, MS analysis was attempted for the detection and localization of sites of glycosylation. However, the same issues regarding protein precipitation in solution, paired with the lack of defined bands in-gel for excision and analysis, meant that analysis could not be performed.

3.2.6 Nanobody stabilisation of hENT1

Owing to the poor results of hENT1 purification, despite the efforts towards optimisation, I decided to investigate the stabilisation of hENT1 through binding partners, such as antibody derivatives. As discussed in 1.7.4.1, nanobodies are valuable tools in the stabilisation and structural characterisation of challenging membrane targets. Therefore, the use of nanobodies was a key area of interest and two distinct approaches to nanobody selection were explored.

3.2.6.1 Camelid affinity matured nanobodies

I was successful in obtaining a place on an Instruct-ERIC funded workshop, Instruct4Nanobodies, in partnership with the Steyaert Lab, Vlaams Instituut voor Biotechnologie, Vrije Universiteit Brussel (VIB-VUB), Belgium for the camelid generation of nanobodies against hENT1. I provided the Steyaert Lab with purified hENT1 for camelid immunisation, with the intention of performing the screening and selection of nanobodies for hENT1 during the workshop. hENT1 was solubilised in LMNG and purified in the presence of NBMPR, following the protocols as described in 2.5.6.1. As per the protocols described by Pardon *et al*[187] camelid immunisation was performed and a diverse, affinity matured nanobody library was prepared from camelid blood samples. Unfortunately, owing to limit time availability during the workshop the screening and selection was unable to be completed by myself. However, this was generously performed by the Steyaert Lab at a later date. The immune library was cloned for phage display, and selection and enrichment of target specific nanobodies was performed by phage panning. 46 individual clones from the enriched sub-library were isolated and screened using enzyme-linked immunosorbent assay (ELISA). 9 hENT1 specific clones were identified. Sequencing determined that, owing to high similarity in the CDR3 sequence, these 9 clones represented 5 families: CA14400, CA14401, CA14402, CA14404 and CA14405. The Steyaert Lab provided me with the plasmids for the five nanobodies for expression and purification in *E. coli* WK6 cells.

3.2.6.1.1 Expression of nanobodies in WK6 cells

All nanobodies were expressed in WK6 cells using a 4-hour isopropyl β-d-1-thiogalactopyranoside (IPTG) induction at 37 °C; high cell densities were achieved (OD_{600} = 8.5-11). Anti-His Western blot confirms that all nanobodies are expressed well following induction (Figure 3.8), with no pre-induction leaky expression observed.

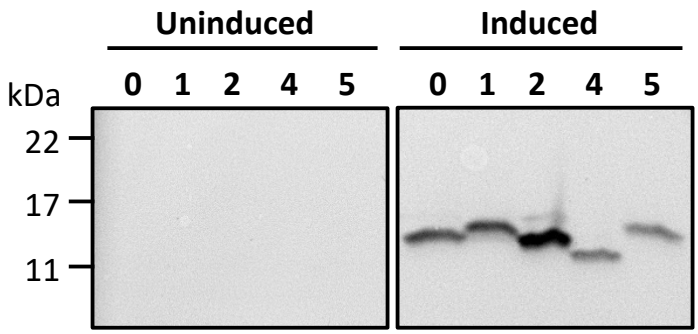


Figure 3.8 Expression of nanobodies in WK6 cells.

Anti-His Western blot of CA14400 (0), CA14401 (1), CA14402 (2), CA14404 (4) and CA14405 (5) before and after induction with IPTG. Typically, nanobodies have a MW of ~15 kDa and the expected MW of CA14400-05 ranges from 13.8 kDa and 15 kDa.

3.2.6.1.2 Purification of nanobodies

The nanobodies are expressed with a C-terminal His-tag, to allow for IMAC purification, and are cloned behind a pre-signal (PelB signal sequence) that directs them to the periplasmic space. As per advice from the Steyaert lab, as detailed in 2.5.6.1.2, two extractions and purifications of the periplasmic fraction were performed to achieve maximal yields. Additionally, each purification used two differing elution approaches to identify the optimal elution conditions: an initial competitive elution using a buffer containing imidazole, followed by a low pH elution.

$A280_{nm}$ analysis of all fractions suggested that the initial competitive elution following the first extraction was adequate for nanobody elution. The Coomassie stained gel also demonstrates that the yield and heterogeneity of the samples worsened quite significantly following the second extraction (Figure 3.9). However, in-gel SDS-PAGE with Coomassie staining and anti-His Western blot showed that all fractions had a persistent non-His contaminant at ~27 kDa (Figure 3.9). The first extraction elution of CA14400 is markedly more heterogeneous than those of the remaining nanobodies. Additionally, CA14404 has a broad, smeared band that runs from 10-20 kDa. While the gel image shown in Figure 3.9 is too overloaded for high resolution of the nanobody bands, the presentation of the nanobody band in the CA14404 is unlikely to be attributed to an artefact of overloading. The profile of the band in CA14404 is not present in any of the other samples which are equally, if not more, overloaded. While heavily ‘smeared’, there is some definition at 10 and 20 kDa. This may represent nanobody dimers for CA14404.

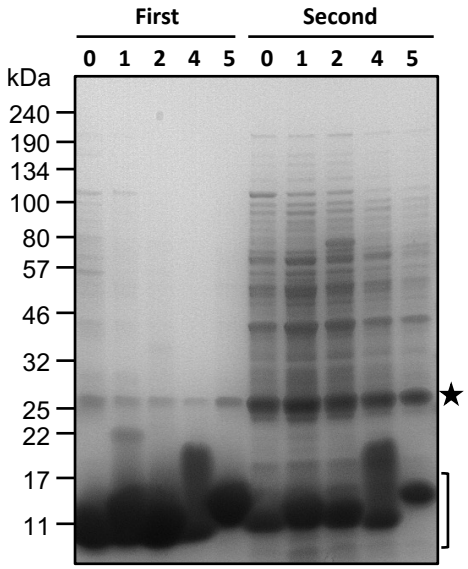


Figure 3.9 Competitive elution of nanobodies following the IMAC of the first and second extraction of the periplasmic fraction.

Coomassie stained SDS-PAGE gel of the first, competitive elution from the first and second purifications for CA14400 (0), CA14401 (1), CA14402 (2), CA14404 (4) and CA14405 (5). While all samples are overloaded, there is a distinct difference in the heterogeneity of the elutions obtained following the first and second extractions. The discussed ~27 kDa contaminant is highlighted by the star, and the range of nanobody MWs is indicated by a bracket.

3.2.6.1.3 Optimisation of nanobody purification

The purification protocol, as detailed in 2.5.6.1.2, uses a high NaCl buffer in pre-equilibration and in the pre-elution washes, which should inhibit and remove any non-specific binding to the Ni^{2+}-NTA resin. However, this would not overcome any low affinity specific binding. Therefore, to improve the homogeneity of the nanobody elutions, purifications were repeated with the inclusion of 30 mM of imidazole in all pre-equilibration and pre-elution buffers. Coomassie stained SDS-PAGE (Figure 3.10) shows that the inclusion of imidazole in these preliminary buffers provides significant improvements to elution homogeneity, increasing purity from 70-85% to >95%.

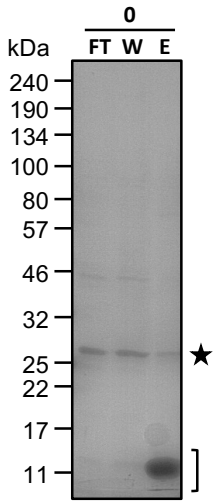


Figure 3.10 Optimisation of the competitive elution of nanobodies.

A representative Coomassie stained SDS-PAGE gel of the flow through (FT), wash (W) and elution (E) following purification, with the inclusion of 30 mM imidazole in the preliminary elution buffers. The discussed ~27 kDa contaminant is highlighted by the star, and the range of nanobody MWs is indicated by a bracket.

While the inclusion of imidazole does improve elution homogeneity, the ~27 kDa contaminant remains in the CA14400 (Figure 3.10) and CA14401 elution, although the amount is significantly reduced. Following this protocol, CA14402 and CA14405 eluted as a homogeneous and monomeric nanobody. Furthermore, following the optimised purification I obtained yields of 1.3, 7.5, 8.5, 7.0 and 4.7 mg/L for CA14400, CA14401, CA14402, CA14404 and CA14405, respectively, which were adequate for follow up experiments. Therefore, no further optimisations were performed.

3.2.6.1.4 Validation of nanobodies for hENT1 specificity

Nanobodies are expected to have nM affinities for their specific antigens[186,189,190]. Therefore, if there is specificity, co-incubation of nanobody and antigen should result in the formation of a high affinity complex. As the nanobodies are expressed and purified with a His-tag, and purified hENT1 is free of tags, I investigated using IMAC for the capture of nanobody-hENT1 complexes by the nanobody C-terminal His-tag. In the conditions as detailed in 2.5.6.1.3 I would expect that, were high affinity complexes formed, I would observe both complex components in high concentrations in the same sample on a denaturing SDS-PAGE gel. This would be true whether formation of a complex hinders the association of the nanobody His-tag to the resin or not. Were a complex to form, but Ni^{2+}-NTA affinity is inhibited then the majority of both nanobody and hENT1 would be present in the flow through and washes. Similarly, if the complex forms and

Ni^{2+}-NTA affinity is retained, then I would expect to see the majority of both nanobody and hENT1 in the competitive elution.

Figure 3.11 suggests that there is a small amount of nanobody being eluted with hENT1 in the flow through and washes of CA14401 and CA14404, and in CA14400 and CA14402 but at a lesser degree. Similarly, there is also a small amount of hENT1 present in the competitive elution for some of the nanobodies. However, owing to the small amounts of each of these I believe this is more likely non-specific and is an artefact of inefficient IMAC capture of the nanobodies and incomplete resin washes. This could be validated by further investigations, such as first capturing the nanobodies by Ni^{2+}-NTA before incubation with hENT1, and more thorough washes before the elution of the nanobodies. However, should these 'co-elution' bands represent true, specific binding, the relative values and affinities are too low to be of use for downstream applications. Therefore, these nanobodies were not explored any further.

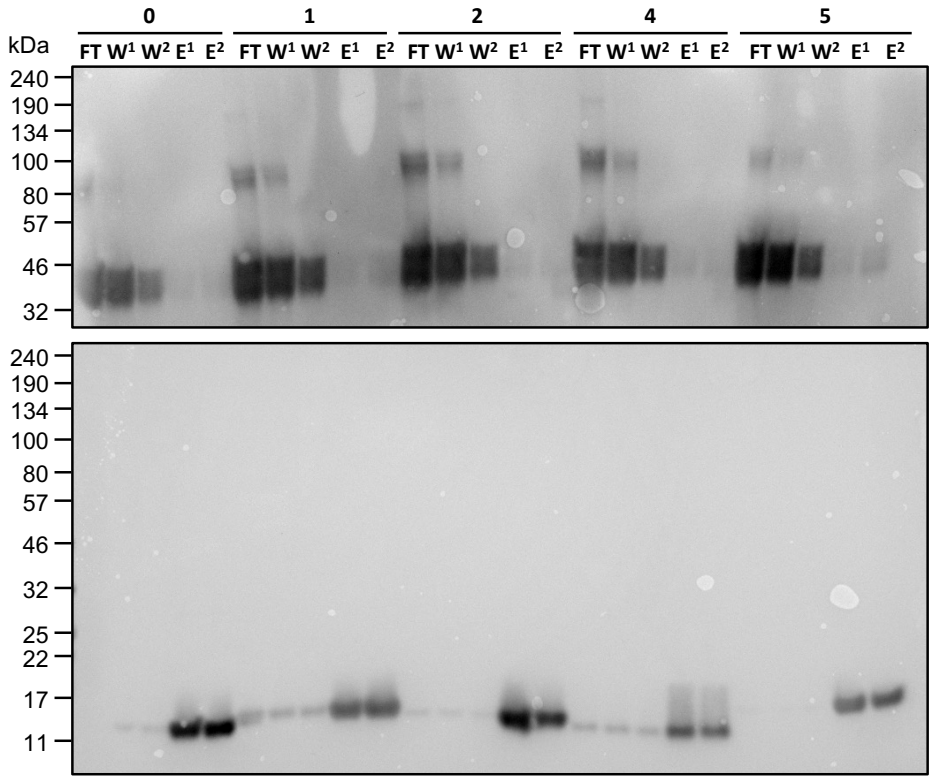


Figure 3.11 Western blots of attempts at hENT1 and nanobody complex formations.

Anti-hENT1 (top panel) and anti-His (bottom panel) Western blot of IMAC of hENT1 and nanobody co-incubations. FT = flow through, W^1 = 10 bed volume low salt wash, W^2 = 10 bed volume low salt wash, E^1 = first elution, and E^2 = second elution. In all conditions, the majority of hENT1 elutes at the FT, W^1 and W^2, and the majority of nanobodies elute at E^1 and E^2.

3.2.6.2 Optimisation of hENT1 for improved nanobody selection

The initial selection that was performed used protein that was available at the time, a purification optimised, tag and modification free hENT1 in LMNG. However, attempts to validate hENT-nanobody binding suggests that there is little-no affinity observed. As nanobody binding requires correctly folded protein, the absence of binding could be due to hENT1 being unfolded and the nanobodies not recognising linearised epitopes. However, the demonstration of NBMPR binding by purified hENT1 (3.2.5.1) supports hENT1 being folded. Therefore, it is more likely that the initial selection of hENT1 specific nanobodies was inefficient. To improve specificity of selection, further optimisations were required.

Furthermore, for optimal selection of nanobodies with desired properties, it is recommended that parallel selections in differing conditions and using different methods are performed[187]. Repeated rounds of selection in the same condition potentially limits the diversity of the final yields, with each round possibly enriching a specific population[186]. Therefore, differing immobilisation techniques and lipid mimetic systems were investigated to allow for repeated and diverse screenings to be performed in parallel.

In addition to lipid mimetics, hENT1 purified in DDM was also intended to be included in repeat screens in an effort to improve epitope accessibility. LMNG has a reported micelle size of ~90 kDa[248]. However, LMNG has also been shown to form unconventional micelles at high concentrations[249], with reported micelle sizes ranging from 235-622 kDa[250]. Therefore, the use of DDM, which forms smaller and more consistent micelles (~70 kDa[248]) should reduce micelle obstruction. Other detergents, such as n-Decyl-β-D-maltopyranoside (DM) or n-Octyl-β-D-glucopyranoside (OG), could provide micelles that were smaller still (~40 and ~25 kDa, respectively[248]). However, as DDM was found to stabilise hENT1 (data not shown) and purification in DDM had also been previously optimised, no other detergents were pursued.

3.2.6.2.1 Exchange into a lipid mimetic system

Of the lipid mimetic systems discussed in section 1.7.5.2, owing to the stability of both the protein and the system association, the compact size, freely miscible nature, and the ease of the exchange protocol[213,214,251,252], amphipol A8-35 was identified as the system of interest. As per the protocol detailed by Zoonens *et al*[252], a series of antigen: A8-35 mass ratios were investigated to determinate the optimal ratio for exchange, and bio-beads were investigated for the removal of free detergent as to prevent micelle reformation. The efficiency of exchange and

detergent removal was assessed by absorbance spectra and relative protein concentrations, as determined by $A280_{nm}$.

The absorbance spectrum demonstrates that without exchange into A8-35, the incubation of detergent solubilised hENT1 with bio-beads causes the protein to precipitate. Protein precipitation is demonstrated by the loss of a peak at $A280_{nm}$, as is seen in the positive control (Figure 3.12). A8-35 absorbs at ~220_{nm} and this A8-35 peak overlaps with the peptide bond peak that is seen in protein absorbance spectrums. Therefore, if protein exchange into A8-35 is successful, this results in an increase in the $A220_{nm}$ peak, and the reduction in the $A280_{nm}$ is likely due to protein precipitation due to exchange inefficiency. All ratios investigated were successful in the exchange into A8-35. The percentage of protein that remains soluble was calculated for each condition (Table 3.5), and a 1:2 mass ratio was determined as the minimum and optimal ratio for exchange, with 93% hENT1 recovered following exchange and detergent removal (Table 3.5).

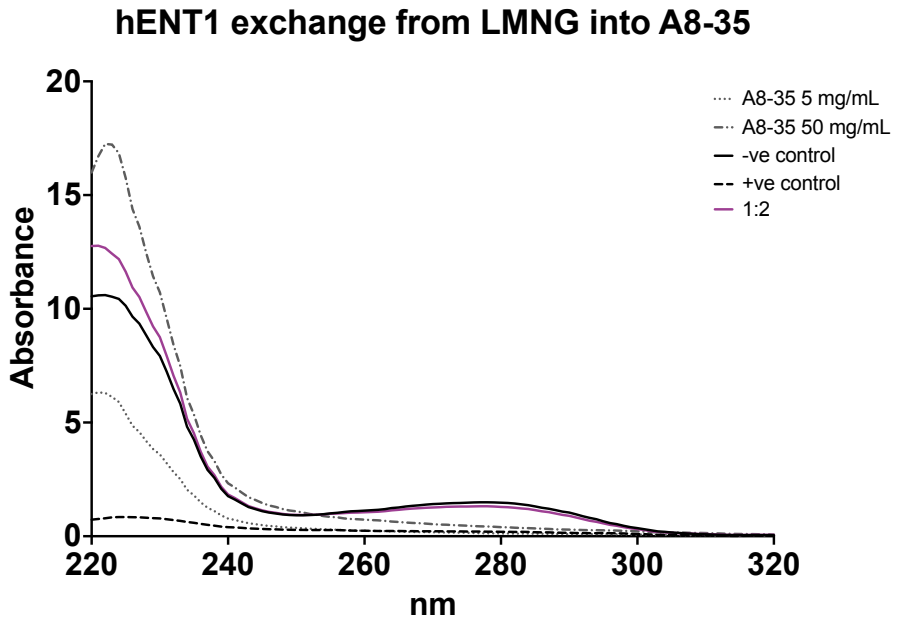


Figure 3.12 Absorbance trace of hENT1 exchanged from LMNG into A8-35.

A8-35 control samples at 5 and 50 mg/mL show contributions to absorbance at ~225-230_{nm} only, with no contributions at 280_{nm}. The negative protein control was incubated without A8-35 or biobeads. Exchange into A8-35 results in an increase in the absorbance at ~225-230_{nm} relative to the negative control. Similarly, precipitation of protein, and thereby efficiency of exchange, results in a reduction in the absorbance at 280_{nm}. The positive protein control was incubated with biobeads only. This results in the precipitation of all protein, as shown by the loss of absorbance at 220_{nm} and 280_{nm}.

Table 3.5 Efficiency of hENT1 exchange from LMNG into A8-35.

Mass ratio for exchange (hENT1:A8-35)	% of hENT1 recovered[a]
1:0 (no exchange, with BioBeads™)	9%
1:0 (no exchange, without BioBeads™)	100%
1:1	94%
1:2	93%
1:3	87%
1:4	81%
1:5	83%

[a] A_{280} of hENT1 after A8-35 exchange, BioBead™ incubation and high-speed centrifugation was related to the A_{280} of the same sample before exchange

3.2.6.2.2 Immobilisation

To prevent the possible denaturation of hENT1 during selection due to aspecific adsorption onto solid surface, alternative immobilisation methods were explored. As per the recommendation in Pardon *et al.*[187] biotinylation was explored, to allow for the specific capture of biotinylated hENT1 on a solid phase coated NeutrAvidin plate, thereby preventing denaturation and increasing specificity. hENT1 was labelled with the thiol-cleavable amine-reactive biotinylating agent, EZ™-Link Sulfo-NHS-SS-Biotin (Thermo Fisher Scientific). A HABA/Avidin assay[234] determined that hENT1 was successfully labelled with the EZ™-Link Sulfo-NHS-SS-Biotin in a 1:1 ratio. The use of a cleavable biotin agent will also allow for the specific elution of an hENT1-nanobody complex using reducing agents, thus further contributing to improved specificity.

3.2.6.3 Synthetic nanobody libraries

As discussed in section 1.7.4.1.1, because of the need for large amounts of stable protein and the limitations in the use of compounds that may be toxic to animals[188,189], the generation and selection of matured nanobodies against membrane protein targets presents several challenges. This gave rise to the development of synthetic nanobody libraries, such as those developed by the Kruse lab[186] and the Seeger lab[188,189], for *in vitro* selection. These synthetic libraries are proposed to be a more accessible avenue for nanobody discovery and are especially suited for challenging membrane targets. Therefore, the selection of synthetic nanobodies was explored in parallel to the generation and selection of affinity matured camelid nanobodies.

3.2.6.3.1 Currently available synthetic nanobody libraries

The synthetic library developed by the Kruse lab uses surface display in *S. cerevisiae*. Here, the nanobodies are presented extracellularly and are anchored to the cell wall by a protein featuring a human influenza virus hemagglutinin (HA) epitope. Thus, using antigen and fluorophore

labelling, this system allows for the selection and enrichment of specific nanobodies by magnetic or fluorescent cell sorting (MACS and FACS, respectively). Therefore, it is the most suitable for in-house *in vitro* working. I set out to explore screening and selection of hENT1 binders using the yeast display system. The yeast display synthetic nanobody library was generously provided by the Kruse lab.

3.2.6.3.2 Yeast display of a synthetic nanobody library

As with immune nanobody libraries, rounds of selection are performed in which nanobodies are incubated with an antigen. Following incubation, specific binders are recovered, amplified, and then selection is performed again. As the successive rounds of selection progress, nanobodies without the desired activity are cleared and the specificity of nanobodies for the antigen should increase. In the Kruse yeast display system expression is under a galactose inducible promoter. Therefore, to obtain cells displaying nanobodies for selection, cells are induced in a galactose media and are harvested 48-72 hours post induction. Using the HA epitope of the cell wall anchor protein, a nanobody expression test is performed using a fluorescently labelled anti-HA antibody and flow cytometry analysis.

3.2.6.3.2.1 Experimental design, considerations, and preparations

As with the immune nanobodies being selected against hENT1 by the Steyaert lab, the same considerations regarding epitope accessibility and selection specificity apply here. However, in this process the nanobody-antigen complexes are captured by MACS and/or FACS, rather than on a solid phase plate. Therefore, along with biotin labelling, fluorophore labelling of hENT1 was explored. Owing to the influences of detergent micelles as discussed in 1.7.5.1, two purifications of hENT1 were performed. One purification was performed in DDM, as to investigate the influence of a detergent micelle smaller than that predicted for LMNG. The second purification was performed in LMNG, but final protein yields were exchanged into A8-35. hENT1 (both DDM and A8-35) and anti-HA were labelled with the spectrally distinct fluorophores, DyLight™ 488 NHS Ester ($\lambda_{excitation}$ 493_{nm}, $\lambda_{emission}$ 518_{nm}) and DyLight™ 650 NHS Ester ($\lambda_{excitation}$ 652_{nm}, $\lambda_{emission}$ 672_{nm}) (Thermo Fisher Scientific), respectively. All samples were determined to be labelled in a 1:1 ratio.

3.2.6.3.2.2 Assigning background and population gating

Cells were grown to a high density (OD_{600} =25) in galactose media for the induction of nanobody expression and harvested after 72 hours. Various conditions were analysed for the

determination of expression, populations within flow cytometry, and to investigate non-specific binding. Cell population analysis by flow cytometry of a sample of unlabelled cells 72 hours post induction served as a negative (and quality) control and allowed for the assignment of population gates for subsequent analysis (Figure 3.13). Analysis of forward *vs* side scatter (Figure 3.13: P1) demonstrated a homogeneous population, and the monodisperse cluster observed in the height vs area analysis suggests the presence of a homogeneous, single cell population (Figure 3.13: P2) free of doublet cells. Background fluorescence contributions from unlabelled cells, within the spectra for DyLight™ 488 and DyLight™ 650, were determined to assign fluorophore specific gates (Figure 3.13: P3 and P4, respectively), and to subsequently establish boundaries for specific fluorophore signal analysis (Figure 3.13: Q1). Q1-LL contains unlabelled cells with no hENT1, or anti-HA bound, a population within Q1-LR represents cells with only anti-HA bound. A population within Q1-UR represents cells with both hENT1 and anti-HA bound, and a population within Q1-UL represents cells with only hENT1 bound. As the literature suggests that <25% of cells are expected to express nanobodies, expression and confirmation that only specific binding of hENT1 to nanobodies is occurring would be demonstrated by an empty Q1-UL quadrant, Q1-LR <25%, Q1-UR < Q1-LR, and the remainder of the population in Q1-LL.

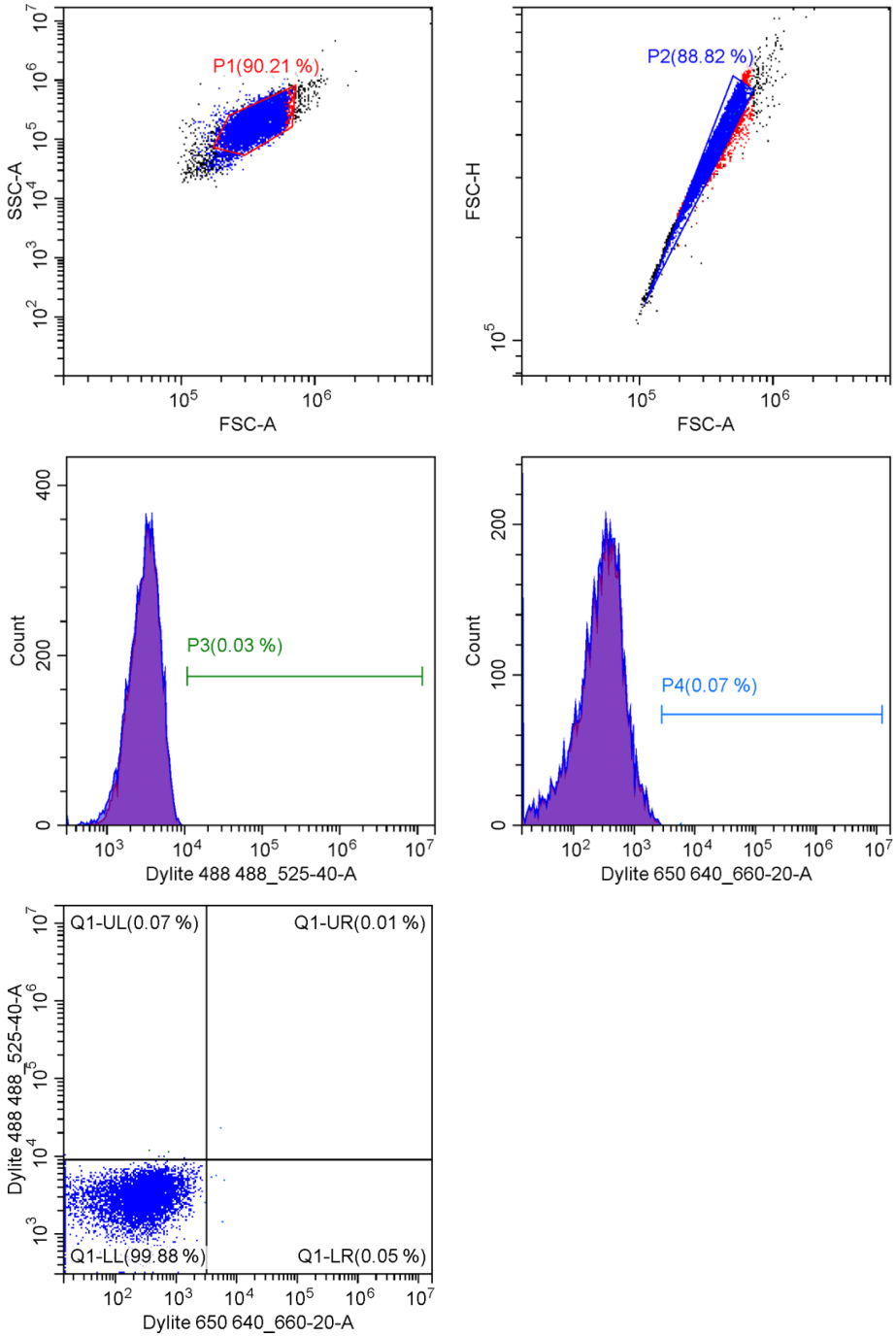


Figure 3.13 Flow cytometry analysis of unlabelled yeast cells, 72 hours post induction.

Unlabelled yeast cells are used to establish the gates for population analysis of subsequent samples. Forward (FSC-A) vs side scatter (SSC-A) analyses the diversity of cell types in a sample. Forward scatter area (FSC-A) vs height (FSC-H) analyses the singularity of cells within a sample, thus allowing for the exclusion of doublet cells. Population 1 (P1) and 2 (P2) are the assignment of the homogenous population of interest. Incidences and intensity of fluorophore signal were used to establish background contributions within the P2 population, and to assign specific fluorophore signal gates for DyLight™ 488 (P3) and DyLight™ 650 (P4), respectively. Analysis of both fluorophore signal intensities allows for the establishment of quadrants, to assign background (Q1-LL), DyLight™ 488 (Q1-UL), DyLight™ 650 (Q1-LR) or dual DyLight™ 488 and DyLight™ 650 (Q1-UR) events.

3.2.6.3.2.3 Expression test and assessment of non-specific binding

Following incubation with labelled anti-HA, as discussed, cells expressing nanobodies are represented by a defined population shift into Q1-LR. Nanobodies are expressed by 8.5% of the total cells (Figure 3.14B). The literature suggests that nanobody expression occurs in <25% of the cell population, but 8-12% expression is considered low[186]. However, low expression can be compensated for during the initial selection round by using a larger number of cells to ensure that maximum library diversity is available for selection. The confirmation of nanobody expression by analysis of fluorophore signal confirms that DyLight™ labelling was successful and is appropriate for analysis. In addition, although expression is low, 72 hours growth post induction is adequate for nanobody expression.

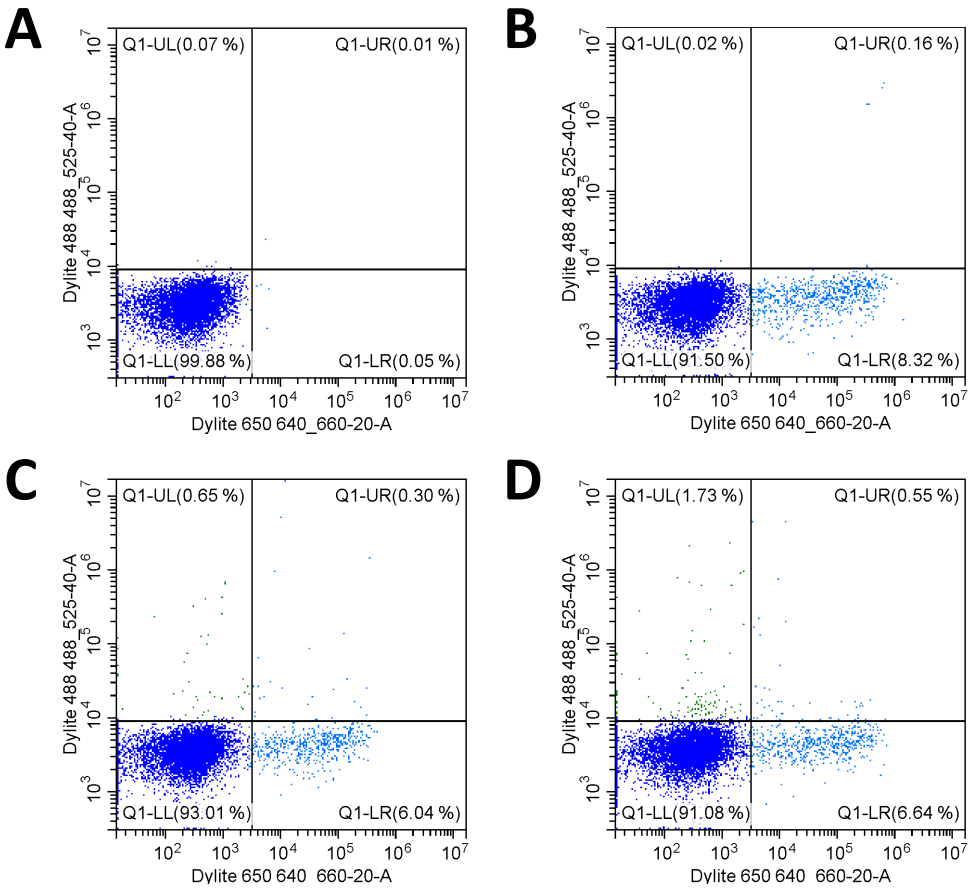


Figure 3.14 Assessing non-specific binding by hENT1.

(A) An induced but unlabelled cell sample is a negative control. This allows for the establishment of the quadrants for analysis of specific intensities. (B) Induced cells following incubation with anti-HA labelled with DyLight 650 indicates nanobody expression by an increase in the population within Q1-LR. Induced cells following incubation with anti-HA labelled with DyLight 650 and hENT1 labelled with DyLight 488, in LMNG (C) or A8-35 (D), indicate non-specific binding by an increase in the population within Q1-UL.

Following the protocol detailed by McMahon *et* al.[186] cells expressing nanobodies were incubated with both labelled anti-HA and labelled hENT1. Following incubation, in both hENT1 in DDM and hENT1 in A8-35, there is a small shift in population into Q1-UR (Figure 3.14C and D, respectively) that could indicate specific hENT1-nanobody binding. However, there is also significant increase in the population of Q1-UL, which is representative of hENT1 binding independent of nanobodies, and thus represents non-specific binding. Therefore, while the population in Q1-UR suggests dual hENT1-anti-HA binding, it most likely represents specific nanobody-anti-HA binding occurring with non-specific hENT1 binding. Consequently, the population in Q1-UR likely represents coincidental, rather than specific, hENT1-nanobody binding. Overall nanobody expression agrees with that determined in anti-HA only analysis (~6.3-7.2% vs 8.5%, respectively), although here the expression percentage is lower still. However, as discussed, this could be mitigated in later steps through the inclusion of higher cell numbers.

3.2.6.3.2.4 Troubleshooting non-specific binding

In the original protocol 0.1% (w/v) bovine serum albumin (BSA) is included in the selection buffer as a blocking agent to inhibit non-specific binding. Therefore, I explored whether increasing the concentration of BSA would further reduce non-specific binding. However, as detergents can bind to BSA, the presence of BSA could also be inducing non-specific binding in the detergent hENT1 sample (Figure 3.14C). In the protocol for selection using the Seeger lab synthetic library, as detailed in Zimmerman *et al*.[189], this is accounted for by the inclusion of detergent (0.1% (w/v) or 3 x critical micelle concentration (CMC), 'whichever is higher') in the selection buffer. Amphipols can adsorb at the surface of cells[253,254]. Therefore, the non-specific binding of A8-35 hENT1 (Figure 3.14D) may be a result of direct interactions between the amphipols and the yeast cell surface. Consequently, different blocking conditions were explored in parallel for the purposes of identifying conditions that affected non-specific binding.

3.2.6.3.2.4.1 Blocking with bovine serum albumin (BSA)

An additional blocking step of the cells with 1% (w/v) BSA prior to incubation of hENT1 incubation was performed. This results in an increase in non-specific binding for both hENT1 in detergent and in A8-35 (Figure 3.15B and C). As detergents are known to bind to BSA, the increase in BSA giving rise to an increase in non-specific binding is unsurprising. hENT1 in A8-35 shows significantly more non-specific binding following the blocking with BSA (Figure 3.15C). However, there is an overall shift upwards of the entire cluster. This shift is also observed in the

hENT1 in detergent condition (Figure 3.15B), albeit to a lesser degree. Therefore, there has been a change in the total population, not just the non-specific population. Thus, the increase to the non-specific binding by hENT1 in A8-35 is unlikely to be as high as the suggested 24% (Figure 3.15C). This overall shift in total population is reflected in the population gating for the hENT1 specific fluorophore DyLight 488 (P3) (Figure 3.16). The histograms across all conditions retain monodisperse populations, but the intensity in absorbance for the entire population is shifted.

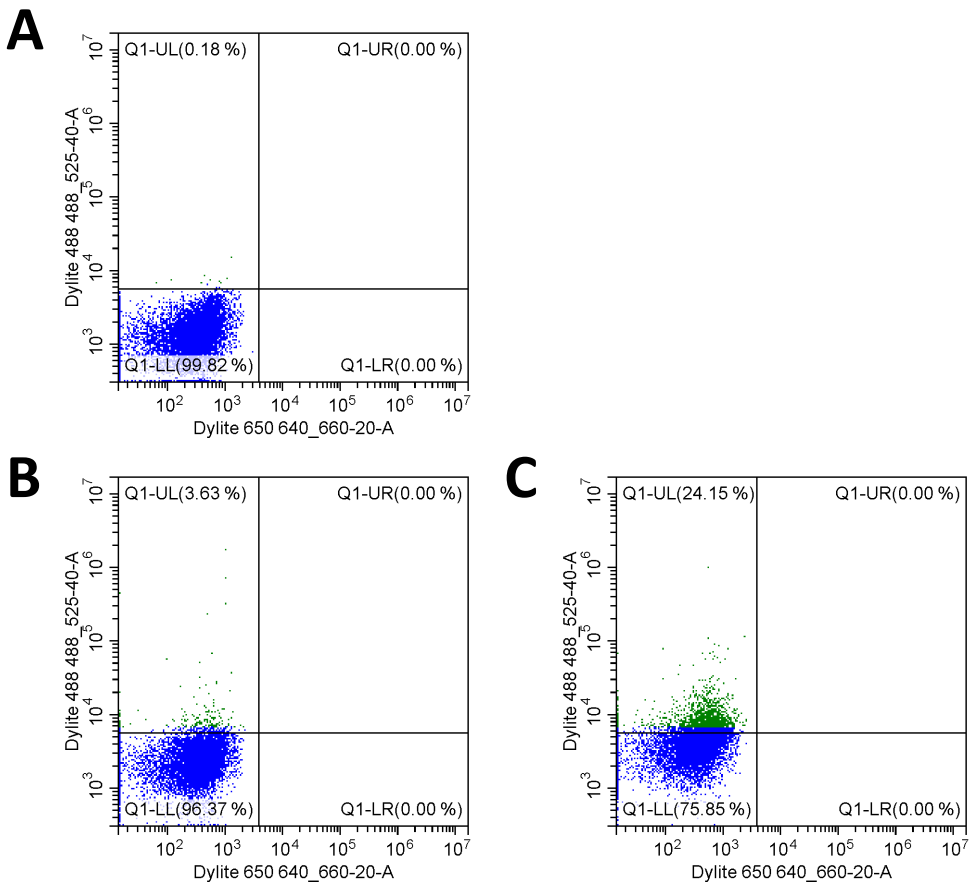


Figure 3.15 Assessing non-specific binding by hENT1 following blocking with BSA.

Anti-HA is absent from this data due to limited reagents being available. However, as this analysis is looking at the patterns and shifts in populations into Q1-UL, along with any overall changes in the clustering of events/cells, the absence of anti-HA, while not ideal, does not affect interpretation. (A) An induced but unlabelled cell sample is a negative control, and thus allows for the establishment of the quadrants for analysis of specific intensities. Induced cells following incubation with hENT1 in LMNG (B), or A8-35 (C) indicate non-specific binding by an increase in the population within Q1-UL.

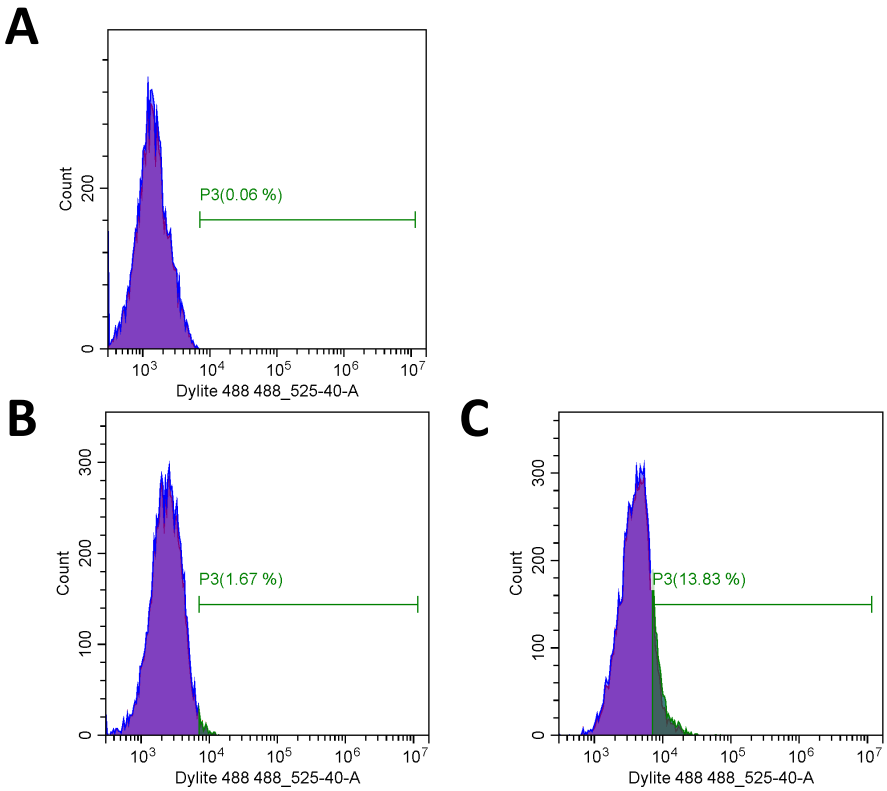


Figure 3.16 Assessing the total population shift following blocking with BSA.

(A) An induced but unlabelled cell sample is a negative control, and thus allows for the establishment of the P3 gating for counts of incidences with specific DyLight 488 intensities. Induced cells following incubation with hENT1 in LMNG (B), or A8-35 (C) are counted for incidences of specific DyLight 488 signal, within the P3 gating.

3.2.6.3.2.4.2 Blocking with detergent

As blocking with BSA yielded no improvement to non-specific binding for hENT1 in either detergent or A8-35, blocking with detergent (minimum 0.1% (w/v)) as per the protocol detailed by Zimmerman *et al.*[189] was investigated. As amphipols can be displaced when in the presence of an excess of detergent, hENT1 in A8-35 was excluded from this investigation as to not have further mixed populations. The results following detergent blocking are like those observed with the original protocol with no blocking steps (Figure 3.17 and Figure 3.14, respectively), with a similar degree of non-specific binding observed between the two conditions. Therefore, blocking with detergent does not improve specificity. However, nor does it affect the overall population, as blocking with BSA appears to do.

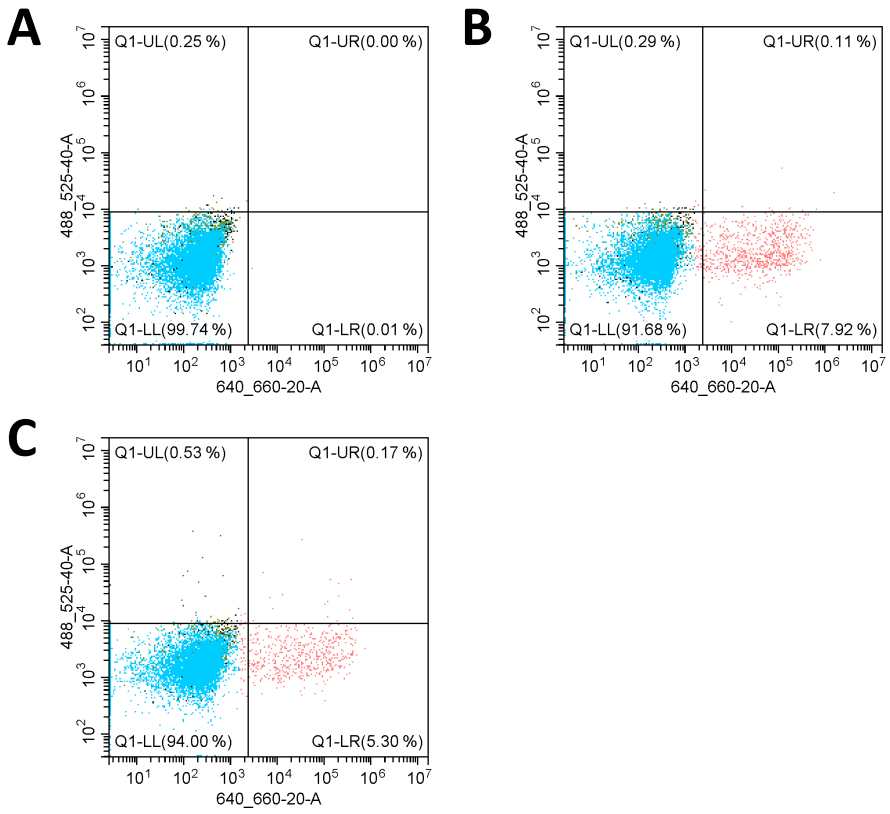


Figure 3.17 hENT1 binding following blocking with detergent.

(A) An induced but unlabelled cell sample is a negative control, and thus allows for the establishment of the quadrants for analysis of specific intensities. (B) Induced cells following incubation with anti-HA labelled with DyLight 650 indicate nanobody expression by an increase in the population within Q1-LR. (C) Induced cells following incubation with anti-HA labelled with DyLight 650 and hENT1 labelled with DyLight 488 (in LMNG) indicate non-specific binding by an increase in the population within Q1-UL.

3.3 Discussion

3.3.1 Purification of hENT1

The data presented in this chapter suggest that using the existing construct, expression methods and extensively optimised purification protocols, purified hENT1 is unstable and aggregation prone. SEC-MALLS analysis suggests that hENT1 is purified as a monomer. However, SDS-PAGE analysis suggests that SDS-resistant aggregation occurs at almost all stages of purification. In an effort to improve the stability of hENT1, the inclusion of glycerol in all purification buffers was explored.

The optimised buffer conditions were found to reduce incidences of presumed SDS-resistant aggregation, which present as higher MW species in anti-hENT1 Western blot[241]. Furthermore, yields following two step purification were significantly more homogeneous, and thus no longer required an additional SEC step for the removal of non-hENT1 proteins. As the improved

homogeneity reduces the need for additional purification steps, and the overall improved stability is presumed to reduce the losses incurred during two-step purification, this provides an overall increase to the final yields obtained (0.14 mg/L). However, while yields obtained are almost five times that of the original protocol (0.03 mg/L), the final yields are still very low. Furthermore, the issue of in-gel band broadening and protein heterogeneity remains present in the final eluted sample. Radioligand binding determined that despite the in-gel presentation, purified hENT1 can still bind NBMPR in a near equimolar ratio (0.85 ± 0.10 pmol [3H]-NBMPR bound per pmol hENT1). Therefore, purified hENT1 is likely to be folded correctly. Further investigations into protein post translational modifications, aggregation, and degradation by MS and LC-MS were attempted. However, this was unsuccessful, due to significant aggregation rendering the samples unsuitable for analysis.

Therefore, while the addition of glycerol in the elution buffers improves the homogeneity and yield of the final sample, there remains a significant concern regarding protein quality and heterogeneity. Consequently, despite considerable efforts to optimise the purification of hENT1, current yields are not amenable to structural analysis. Moreover, due to poor purification efficiency, continuing to work with wild-type hENT1 following these protocols is simply not cost-effective, nor viable. Owing to this overall poor performance, nanobodies were explored for the stabilisation of hENT1.

Additional techniques could be explored to help determine the oligomeric state of hENT1. The small-angle scattering (SAS) techniques SAXS and SANS use X-ray and neutron scattering profiles to provide information on the the size, shape, structure, oligomerization state, conformational heterogeneity, flexibility, and intrinsic disorder of biological macromolecules, and their conformational responses to changes in external conditions[255-258]. SAS experiments can also be performed using small, dilute sample volumes and can support the coupling to other techniques, such as SEC, ion exchange chromatography and MALLS, to allow for optimal (monodisperse) sample analysis[257,258]. Therefore, analysis by SEC coupled to SAS may a promising route to probe the oligomeric state of hENT1. Furthermore, SAS techniques can accommodate challenging systems, such as those not amenable to structural characterisation by other high-resolution techniques (i.e., X-ray crystallography, cryo-EM and NMR) due to intrinsic flexibility and disorder. Moreover, using SANS in combination with SAXS, deuterated components, and/or contrast variation, structural information from membrane proteins in lipid or detergent complexes can be separated from the detergent and lipid parts[259-264]. Therefore, SAS may also

be able to offer further insights into hENT1 structures and conformations, in addition to the oligomeric state.

3.3.2 Selection of camelid affinity matured nanobodies

Camelid affinity matured nanobodies were successfully raised against hENT1, and an initial round of selection identified five nanobody families suggested to bind to hENT1. However, attempts to validate these nanobodies demonstrated that they showed no specificity for hENT1. Initial nanobody selection was performed using tag and modification free hENT1 in LMNG. This detergent solubilisation may contribute to the lack of success in selecting hENT1 specific binders.

As detergents associate to protein hydrophobic surfaces at the transmembrane domains and form spherical and ellipsoid protein-detergent micelles, these detergent micelles likely obstruct solvent accessible regions of the protein and hinder epitope accessibility[201]. Furthermore, hENT1 is predominantly transmembrane α-helical and has limited extracellular domains. Therefore, there are likely limited opportunities for binding sites that protrude beyond the micelle. Moreover, as discussed in 1.7.5.1, while long chain detergents are often more stabilising than their shorter chain counterparts, longer chains give rise to larger and possibly more obstructive micelles. In addition, detergent aggregates have been shown to capture antibody derivatives[265] and the Steyaert Lab advised that they often encounter issues with LMNG non-specifically capturing nanobodies. Therefore, longer chain and 'bulkier' detergents such as LMNG, while being the most stabilising for hENT1, may be the least optimal for specific selection of nanobodies. Finally, the high affinity binding between nanobody and antigen is dependent on folded proteins, with poor binding observed in unfolded protein and linear epitopes[187]. Nanobody selection was performed using non-specific adsorption onto the solid surface of an ELISA plate. However, this approach can result in partial denaturation of proteins[187,266]. Therefore, specific, high affinity binding for folded hENT1 may not have been achieved during selection.

To overcome micelle steric hinderance of epitopes, opportunities for modification and selection optimisation were identified, including purification of hENT1 in a detergent with a smaller and more consistent micelle, and the exchange of hENT1 from LMNG into A8-35. In addition, to improve immobilisation conditions and to increase the specificity of selection both conditions were modified to have a cleavable biotin linker. Repeat parallel selections were to be performed with the hopes of improving specificity of the selected nanobodies for hENT1. I was due to perform repeat selections during a visit to the Steyaert Lab at the VIB-VUB, Belgium. However,

due to restrictions that arose because of Covid-19, this visit was unable to take place. The Steyaert Lab generously agreed to perform the selections on my behalf, and I provided them with the modified hENT1 samples. Unfortunately, due to time constraints, this selection has not yet been performed.

3.3.3 Selection of hENT1 binders from a yeast display, synthetic nanobody library

While affinity matured camelid nanobodies would likely produce more desirable hENT1 binders, synthetic libraries were investigated for the purposes of increasing the chance of finding hENT1 binders, to increase the diversity of selected nanobodies, and to also provide an in-house platform for *in vitro* nanobody selection for other projects in the lab. However, significant difficulties with non-specific binding were encountered and could not be overcome by blocking of non-specific binding with 0.1% (w/v) free detergent or 1% (w/v) BSA.

While the data presented could be attributed to overall poor specificity for hENT1, unrelated membrane proteins (an MFS member MaTap, and the pyrophosphatase TmPPase) were included throughout the analysis and the same patterns of non-specific binding were observed (Appendix Figure 1 and Appendix Figure 2). Furthermore, the degree of non-specific binding in these unrelated proteins was far more significant than that observed in hENT1. Moreover, a soluble protein was also included as a 'membrane protein' negative control and did not display the same degree of non-specificity as the membrane protein samples (Appendix Figure 3). Thus, the issues encountered with non-specificity here are likely a uniquely membrane protein challenge.

Owing to these difficulties I reached out to the Kruse lab to seek advice. Following discussions with Dr. McMahon and Prof. Kruse they suggested to explore differing free detergent concentrations or to use solubilised membranes as a blocking agent. It was reported that previous selections for GPCRs were significantly improved by the inclusion of extracted Sf9 membrane fractions, likely due to the presence of large amounts of 'negative control' proteins. Additionally, the inclusion of a non-induced control was recommended for screening for non-specific binding. However, owing to time constraints, because of Covid-19 and the prioritisation of other more promising avenues of work, this has not been explored any further.

3.3.4 The publication of the first structures of an ENT and N- glycosylation of hENT1

In 2019 the first structures of an ENT were solved by Wright & Lee, using X-ray crystallography[134]. The hENT1 construct used for crystallisation included three stabilising point mutations, the

truncation of the ICL6[134], a C-terminal GFP-FLAG-His[10] and was expressed in HEK293 GnTI- cells. As discussed in 1.7.1, expression hosts most like the native host are more likely to produce functional protein. Therefore, HEK293 is likely better equipped to produce hENT1 with essential for function modifications, such as the N-linked glycosylation of N48. While Sf9 cells are not capable of producing complex N-glycans, instead producing high mannose glycans[267], the hENT1 expressed by Sf9 presented in the work in this chapter and in previous work by Rehan & Jaakola[241] has been shown to be functional. An advantage of the specific use of HEK293 GnTI- may be the homogeneity of the N-linked glycosylation, in addition to/rather than the complexity of the glycosylation. HEK293 GnTI- cells lack N-acetyl-glucosaminyltranferase I activity, which allows for the expression of proteins that are homogeneously N-glycosylated with Man5GlcNAc2 N-glycan. This homogeneous N-linked glycosylation has been shown to contribute to the overexpression of wide variety of mammalian membrane proteins and has made them more amenable to structural characterisation[268]. Therefore, it is possible that expression of homogenously N-glycosylated hENT1 produced protein that was more homogenous and stable, and thus, more amenable to purification. However, Wright & Lee deglycosylated the final hENT1 product before crystallisation[134]. There is no information available in the publication as to the influence of de-/glycosylation on protein stability or amenability to crystallisation. Previous studies have discussed the 'glycosylation problem', where large chemical post translational modifications are expected to increase surface entropy and inhibit crystalisation[269,270]. Therefore, while it is possible that expression of homogenously N-glycosylated hENT1 produced protein that was more amenable to purification, it is also possible that the deglycosylation contributed to an hENT1 more amenable to crystallisation.

Dr. Harborne previously suspected N-linked glycosylation as contributing to the presentation of broad hENT1 bands in-gel, as encountered in the historic purifications of hENT1 within our lab. However, as investigations into enzymatic de-glycosylation did not yield any improvements and current yields are not amenable to analysis by MS, we are unable to determine whether this broad band presentation is due to glycosylation or other protein quality issues. An alternative approach to determining overall glycosylation influence on hENT1 expressed by Sf9 could be using alternate cell lines, such as Mimic™Sf9 cells (ThermoFisher Scientific). These cells are modified to stably express various mammalian glycosyltransferases. Therefore, these cells could be used to explore the influence of glycan complexity on the stability of hENT1 expressed by Sf9. However, as it is possible that the expression of hENT1 with homogeneous glycosylation, rather than complex glycosylation, is a more significant contributor to protein stabilisation the heterogeneity of glycosylation in Mimic™Sf9 cells may still present issues. Therefore, simply

transferring into the more native like expression system and attempting to replicate the expression in HEK293 GnTI- as detailed by Wright & Lee[134] may be the most efficient approach.

3.3.5 Summary

As discussed in 3.1.1, the research into hENT1 initiated in lab by Dr. Harbone built upon established work by Dr. Rehan and Dr. Jaakola[241]. The works presented in this chapter pre-date the first solved structures of hENT1 in 2019. Therefore, the approaches undertaken for optimisation and the stabilisation of hENT1 were done so without influence of methodologies which have since been shown to contribute to the solved structures. However, as touched upon in 3.3.4, there are distinct differences in the methodologies used by Wright & Lee[134] that could be explored for further optimisation of hENT1 expression and purification.

Furthermore, while the structures published by Wright & Lee have contributed to significant advancements in our understanding of hENTs, there is still a large gap in our understanding of their mechanism of action. For example, N48, the site of N-linked glycosylation in hENT1 is located at the ECL1. However, the ECL1 is absent from the available structures due to poor resolution. Consequently, insights into the molecular basis for the role of ECL1, and more specifically, N-linked glycosylation of N48, are lacking. Therefore, as it is possible that the N-linked glycosylation of N48 contributes significantly to stabilisation of hENT1, structural characterisation of hENT1, inclusive of essential for function loop regions, is key for our understanding of hENT1 mechanism of action.

Therefore, the optimisation of hENT1 purification remains an area of interest, and the use of nanobodies as stabilising binding partners, conformation selectors and crystallisation chaperones remain a promising area, albeit one with a significant amount of work still needing to be done. While I had planned to perform some of this follow up work myself by performing further rounds of selection against the affinity matured camelid nanobodies and investigating further avenues for overcoming the non-specific binding issues encountered in the yeast display system, unfortunately I have not had an opportunity to pursue this. Therefore, while the purification of an hENT1 amenable to structural characterisation was not achieved in this work, this chapter lays the foundation for future investigation and optimisation.

Chapter 4 Identifying thermostabilising point mutations of hENT1

4.1 Introduction

As discussed in 1.7.2.2, it has been shown that a single point mutation can significantly improve the expression of membrane proteins, as is seen in the human serotonin transporter (PDB: 5I6Z)[271], the cystic fibrosis transmembrane conductance regulator (PDB: 6D3R)[272], the Mg2+ transporter CorC (PDB: 7CFF)[273], amongst many others[175-178]. The labour and cost intensive nature of the scanning-mutagenesis approach has given rise to the development of computational pipelines that predict the likelihood of stabilising effects of specific amino acid changes[173,179-182]. The computational pipeline IMPROvER, the integral membrane protein stability selector which combines bioinformatic techniques with data from previous GPCR stabilisation studies, was developed by Dr. Harborne as a general-purpose tool for selecting stabilising variants of α-helical membrane proteins[134]. Following the user input of the protein target sequence, IMPROvER combines bioinformatic techniques with data from previous GPCR stabilisation studies to produce a ranked list of predicted stabilising variants[180]. In addition, known critical residues can be pre-emptively excluded from the selection to ensure function is not negatively affected, and comparative homology models can be supplied to optimise and tailor IMPROvER's performance[180].

4.1.1 IMPROvER

IMPROvER uses three distinct modules for analysis: deep-sequence, model-based and data-driven[180]. In the deep-sequence module, large sequence alignments across evolutionarily distinct homologues determine the natural frequency of an amino acid at a specific position. Comparison of the protein target sequence to the consensus from alignment allows for the identification of infrequently used and potentially detrimental amino acid residues. The module then suggests a substitution for an amino acid that is more frequently used[180,274,275]. In the model-based module, *in silico* thermal melting scans are performed on homology models to determine the free energy change (ΔG). Then *in silico* alanine scans and repeat thermal melting scans are used to estimate the ΔΔG between wild type and an alanine variant for all positions in the sequence. Positions are ranked by average stabilising effect, and the averaged top 20 positions are then changed to every other amino acid. The ΔΔG of these models are then estimated, and then the results for each position are ranked and the best amino acid substitution at each site is selected. Finally, in the data-driven module, a scoring matrix from GPCR based data sets is applied to the target sequence and predictions for stabilising mutations

are made. The scoring matrix is based on general rules established from previous GPCR stabilisation studies, with ~2000 amino acid data points, and is a summary of the trends in stability with consideration of amino acid identity, topology, conservation, lipid contact prediction and helix contact prediction. Following these three modules, IMPROvER provides a list of ranked variants and associated primer sequences for site-directed mutagenesis.

4.1.2 Summary of previous work

IMPROvER selected 205 variants of hENT1, with 116 variants selected from the data driven module, 68 from the deep sequence module, and 21 from the model-based module. This analysis of hENT1 by IMPROvER pre-dates the published structures of hENT1[134] or the publication of Alphafold[155], and thus only limited literature reported critical residues and homology models were available for this analysis. Of the 205 suggested variants, eight were eliminated as being at residues critical for function, and five variants appeared in more than one module. Therefore, in the final output, IMPROvER selected 192 variants of hENT1 for predicted stabilising effects[180].

As discussed in 3.1.1, Dr. Harborne had previously optimised constructs for the expression of hENT1 in Sf9 cells. An N-terminally tagged His_8-TEV hENT1 is used for large scale expression. However, Dr. Harborne set out to use a GFP-based assay for the investigations into thermostability, using a methodology adapted from Ashok *et al*.[276]. Therefore, an alternative construct with a GFP tag was required. During previous construct optimisations it was found that a C-terminal His_6-GFP-TEV tagged hENT1 (hENT1-CTGH) had improved protein stability. However, the GFP could not be removed following incubation with TEV protease (data not shown). While this C-terminally GFP-tagged hENT1 is not amenable to elution of a tag-free construct, and thus structural characterisation, this construct was appropriate for the purposes of assessing thermostability of IMPROvER suggested variants.

Using the IMPROvER generated primers, Dr Harborne performed site-directed mutagenesis using the hENT1-CTGH pFastBac™ plasmid, all variant sequences were confirmed by Sanger sequencing. The plasmids were transformed into *E. coli* DH10Bac cells for the propagation of bacmid DNA, which was then isolated and used to transfect Sf9 cells. Transfection and expression of hENT1 was confirmed by GFP signal localised to Sf9 cell plasma membranes (data not shown). As recombinant baculovirus particles are secreted by infected cells, media was harvested to allow for further infections of Sf9 cells. Repeat small scale infections were performed to allow for the increase of viral titre of the harvested media and to increase

expression. The success rate at each of these steps was low, and thus some of these steps had multiple rounds performed to increase the number of positive results. In total, Dr. Harborne was able to produce 21 variants of hENT1 that were expressed in Sf9 cells. The project was then taken over by a MA student, Danielle Wright, who then performed additional rounds of site-directed mutagenesis, bacmid generation and Sf9 cell transfections in the hope of increasing the success rate and number of variants available for stability testing. Danielle Wright produced an additional 20 variants of hENT1 in Sf9 cells. In total, 95 single point mutations were successfully introduced into hENT1 by site directed mutagenesis. However, only 75 bacmids were generated and only 41 variants were expressed and localised to the plasma membrane of Sf9 cells[180].

A ten-temperature challenge was performed using wild type hENT1 to establish a baseline of thermostability and to inform the next screening steps. As outlined in 2.6.2, hENT1 was solubilised in DDM and independent aliquots of solubilised hENT1 were heated at 4, 30, 35, 40, 45, 50, 55, 60, 65 or 70 °C. Following heating, protein that had unfolded was removed by high-speed centrifugation, and protein that remained in solution was analysed by SDS-PAGE with fluorescence imaging for the visualisation of hENT1-linked GFP signal. Protein was analysed using densitometry analysis, and band intensity at each temperature was related to the 4 °C. This data was then used to determine an apparent melting temperature (T_m), at which 50% of the protein is unfolded. This T_m was then used to perform a single temperature challenge. In the single temperature challenge, all remaining variants were incubated at the T_m of wild type, instead of the full ten-temperature range, and the surviving variant-linked GFP signal was related to a 4 °C control, and then related to wild type. Relative single temperature data was analysed collectively and 10 variants with fluorescence survival ≥ the 95% confidence interval were identified. These 10 variants were then taken forward for further analysis. Cell cultures were prepared for all ten variants of interest, cells were harvested by centrifugation and then stored at -20 °C until analysis was performed.

4.1.3 Aims and strategy

Unfortunately, owing to time constraints, neither Danielle Wright nor Dr. Harborne were able to perform the final round of stability assays for all 10 variants of interest. Therefore, I set out to investigate these variants for thermostability using the ten-temperature assay, as was performed for wild type. In addition, I set out to investigate the binding of NBMPR by all variants by exploring both radioligand binding assays and additional ten-temperature stability assays in the presence of NBMPR.

4.2 Results

4.2.1 Initial round of thermostability assays.

I performed ten-temperature assays for all 10 variants that had been identified during the single temperature challenge. The apparent T_m values suggested that all variants had a T_m above that of wild type, with ΔT_m ranging from 0.4 to 4.2 °C (data not shown). Furthermore, using a radioligand binding assay as detailed in 2.4.3 and similar to that used in 3.2.5.1, when variant binding was normalised to wild type, for every 1 pmol of [3H]-NBMPR bound by wild type hENT1, variants of hENT1 bound 0.04 to 0.74 pmol (data not shown). Therefore, I observed a range of effects on thermostability and NBMPR binding. There was one mutant that was of particular interest which seemed to be a highly thermostabilising variant but was significantly worse at binding [3H]-NBMPR. Therefore, we proposed that this variant could be a good candidate for efforts towards determining an *apo* structure of hENT1. I set out to explore this variant as a candidate for structural characterisation. However, as discussed in 4.1.2, the expression cassette in which all variants were originally prepared in was not amenable to the purification of a tag-free construct, and thus the generation of a construct for purification would require new site-directed mutagenesis or subcloning of the gene to be performed. I opted for the latter approach.

Subcloning of variant 'X' from the CTGH-pFastBac™ to the NHGT-pFastBac™ was successful and the recombinant plasmid was confirmed by Sanger sequencing. However, while sequencing confirmed that target integration was successful, it also highlighted that the identity of the variant was incorrect (variant 'X' was variant 'Y'). There are many steps between the generation and validation of a variant plasmid, and the expression of the variant in Sf9 cell, and thus there are many opportunities for handling errors to occur. Therefore, I set out to establish a method to validate the bacmid DNA and investigate whether this was a case of an odd one out, or whether more inconsistencies existed.

4.2.2 Validation of variant identity through the sequencing of bacmid DNA and the generation of thermostability data

Previous work by Dr. Harborne and Danielle Wright identified 41 variants that were expressed and localised to the plasma membrane of Sf9 cells. For these 41 variants, there were 67 virus stocks available for the infection of Sf9 cultures. Small scale expression tests using all available viruses found that only 48 viruses resulted in expression of protein with correct localisation to the plasma membrane (Figure 4.1). The remaining viruses were either contaminated or were too high a titre and resulted in significant cell death. Bacmid DNA was successfully isolated from

whole cells but could not be sequenced. However, the bacmid DNA isolated from whole cells was amenable to PCR amplification. PCR amplification and sequence determination by Sanger sequencing was successful for all 48 variants, although some variants required several rounds of sequencing before the mutation was identified because the quality of the initial DNA from the bacmid preparation was variable, possibly due to the nature of the pellet (see 2.3.2).

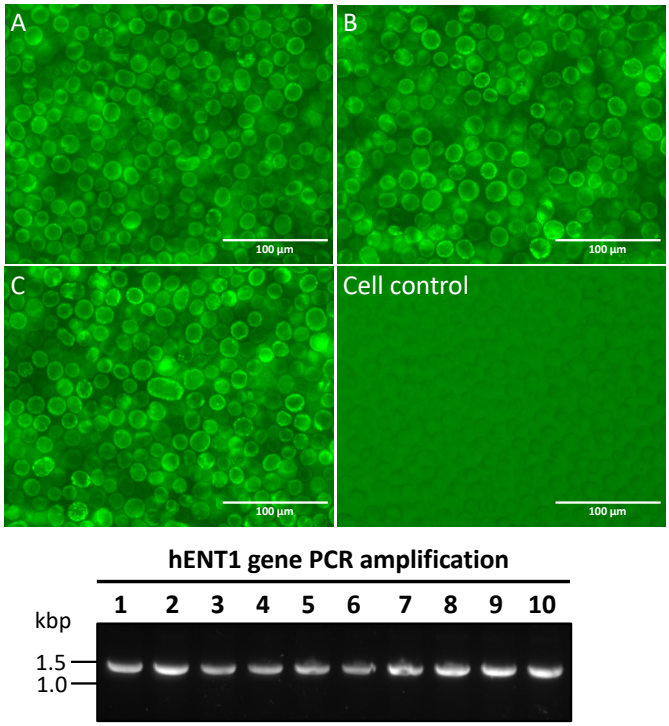


Figure 4.1 Fluorescence imaging of Sf9 cells transfected with baculovirus and PCR amplification of bacmid DNA.

(A. B, C) An example of Sf9 cells showing expression of GFP tagged hENT1 variants, with localisation to the plasma membrane, along-side an un-transfected cell control. Bacmid DNA was isolated from whole cells and PCR was performed for the amplification of the hENT1 gene. An example agarose gel (bottom panel) shows successful amplification of the hENT1 gene by the presence of a band at ~1.4 kbp. These amplified inserts were then sent for Sanger sequencing.

Of the 48 bacmids sequenced, I found 20 unique variants (inclusive of wild type). However, there were several inconsistencies in the results. 50% the sequencing results agreed with the original variant identity (virus X = variant X) and the remaining 50% disagreed, but to differing extents. For example, some virus sources were found to agree with the original variant identity, but other virus sources with the same original identity were found to be a different variant (virus X^1 = variant X, but virus X^2 = variant Y). However, some were in total disagreement (virus X^1 = variant Y, and virus X^2 = variant Z). Unfortunately, wild-type sequencing results were included in those with disagreements, and so while some wild type viruses were found to be wild-type, but others were found to be variants.

The cell pellets that I performed the initial stability assays on were not prepared by me, and I did not know which virus source was used to generate the cell pellet in the first place. Therefore, I was unable to simply re-assign variant identity from the validated sequencing results to the results from the initial round of thermostability assays (4.2.1) and then reanalyse the data. Therefore, to allow this avenue of investigation to continue, ten-temperature thermostability assays had to be repeated for all 48 viruses and new data generated (Table 4.1 and Figure 4.2).

To ensure that all data could be analysed appropriately, data were generated with a virus ID and only following the assay was variant ID assigned, based on the sequencing results of bacmid DNA isolated from the insoluble pellet. As with bacmid isolated from whole cells, the DNA isolated from the insoluble pellet could not be sequenced but was amenable to PCR amplification. PCR amplification and sequence determination by Sanger sequencing was successful for all 48 variants, and in agreement with the results of the small-scale expression sequencing, 20 unique variants (inclusive of wild type) were identified (Table 4.1). Where repeats of variants were identified the data were combined, thus accounting for the variance of the n value, with 3-15 repeats.

Table 4.1 Results of ten-temperature melting analysis of *apo*-hENT1 variants with identities validated by DNA sequencing.

Validated ID	T_m (°C)[a]	T_m SEM (± °C)[b]	Δ T_m (°C)[c]	Δ T_m SEM (±°C)[d]	Repeats (n =)	Status	Significance (P value)[e]
G305A	43.5	0.4	1.5	0.5	8	Stabilising	ns (p = 0.0794)
M306T	43.2	0.5	1.1	0.6	5	Stabilising	ns (p = 0.6440)
K263A	43.0	0.6	1.0	0.7	5	Stabilising	ns (p = 0.7620)
E264A	42.7	0.6	0.7	0.7	7	Stabilising	ns (p = 0.9502)
WT	42.0	0.3	0.0	0.5	15	N/A	N/A
N30F	42.0	0.5	0.0	0.6	10	Neutral	ns (p = 0.9997)
K283R	41.1	0.5	-0.9	0.6	12	Destabilising	ns (p = 0.3553)
A401L	41.0	0.5	-1.0	0.6	9	Destabilising	ns (p = 0.3059)
T336A	40.9	0.5	-1.1	0.6	12	Destabilising	ns (p = 0.2046)
I282V	40.8	0.3	-1.2	0.5	12	Destabilising	ns (p = 0.0744)
R233F	40.3	0.5	-1.8	0.6	3	Destabilising	ns (p = 0.1341)
Q246A	40.0	0.4	-2.0	0.6	6	Destabilising	** (p = 0.0047)
R233L	38.6	0.8	-3.4	0.9	6	Destabilising	**** (p = <0.0001)
T126G	38.4	0.8	-3.7	0.8	3	Destabilising	*** (p = 0.0001)
G225V	37.2	0.3	-4.8	0.4	3	Destabilising	**** (p = <0.0001)
S152L	37.1	0.4	-4.9	0.6	3	Destabilising	**** (p = <0.0001
A88L	36.9	0.4	-5.1	0.5	3	Destabilising	**** (p = <0.0001)
G207A	36.7	0.3	-5.3	0.5	3	Destabilising	**** (p = <0.0001)
F153A	36.5	0.8	-5.5	0.9	3	Destabilising	**** (p = <0.0001)
L27E	36.1	0.6	-5.9	0.7	6	Destabilising	**** (p = <0.0001)

[a]Average T_m was calculated from individual T_m estimated for each individual repeat by fitting with a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting in GraphPad Prism 9.0
[b]Standard error of the mean (SEM) shown.
[c]Δ T_m represents variant relative to wild type.
[d]Error calculated and propagated as detailed in Methods and materials: 2.4.4.
[e] Statistical analysis performed on Tm data using ordinary one-way ANOVA with a Dunnet follow-up test for multiple comparisons. Significance indicators: (ns) p = 0.1234, (*) p = 0.0332, (**) p = 0.0021, (***) p = 0.0002, (****) p = <0.0001

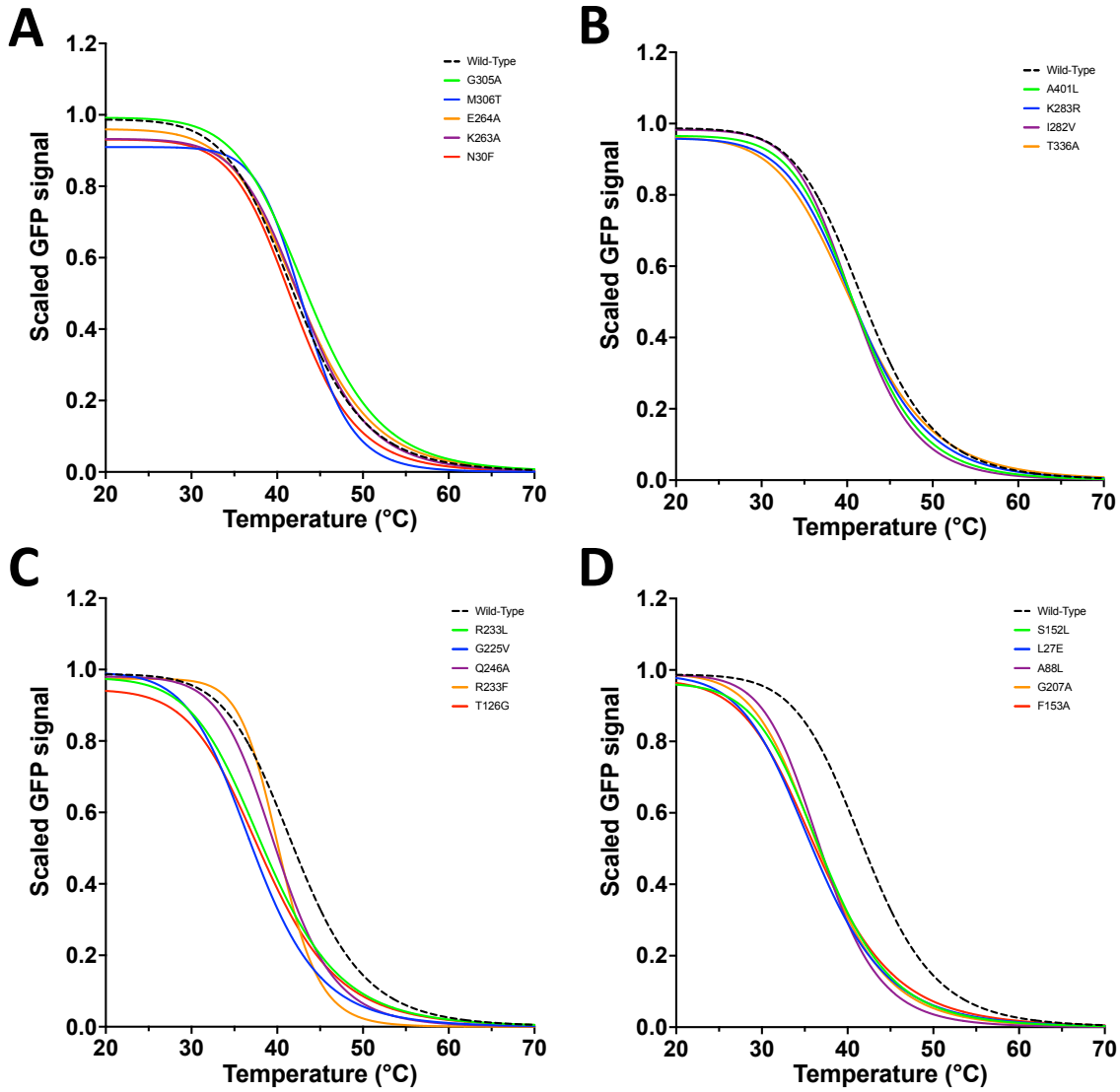


Figure 4.2 Ten-temperature stability curves for all *apo*-hENT1 variants.

Protein that remained in solution after challenge in a ten-temperature melting curve was quantified using in-gel GFP fluorescence. The curves are subdivided into the following groups for clarity of presentation: (A) Tm = 42.0 – 43.5 °C, (B) Tm = 40.8 – 41.1 °C, (C) Tm = 37.2 – 40.3 °C, and (D) Tm = 36.1 – 37.1 °C. hENT1 variant curves were initially collected in biological triplicates, as detailed in 2.6.2. However, following variant validation through bacmid DNA extraction, PCR amplification and Sanger sequencing, the number of repeats increased for several variants. Data shown is the average of at least n = 15 experiments for Wild type, and an average of n = 3-12 experiments for the variants. Data in each curve are normalised to the intensity of the sample incubated on ice. Error bars are representative of SEM. Data were fit with a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting in GraphPad Prism 9.0.

4.2.3 Comparing the stability of hENT1 variants between *apo* and inhibitor-bound states

4.2.3.1 Wild type hENT1 and identifying variants of interest

The T_m determined for *apo* wild type hENT1 was 42.0 ± 0.3 °C (Table 4.1, Figure 4.2 and Figure 4.3). Using a ± 0.6 °C cut-off (average SEM ΔT_m (Table 4.1)), four hENT1 variants were identified

as stabilising, one additional variant was identified as neutral, and the remaining 14 variants were destabilising (Table 4.1, Figure 4.2 and Figure 4.4C). Variants that were not significantly destabilising were taken forward for further investigation (Table 4.1). A401L and R233F were intended to be included in the next step, but unfortunately limited virus volumes were available, and expression of protein and preparation of additional cell pellets was unsuccessful. Therefore, further investigations were only able to be performed on N30F, K263A, E264A, I282V, K283R, G305A, M306T and T336A, in addition to wild type.

Stability assays in the presence of 20 µM NBMPR were performed on these eight variants and wild type to investigate differences in the difference in the T_m between the *apo* and NBMPR-bound states (ΔT_m) (Figure 4.3A and Table 4.2). The ΔT_m for NBMPR-bound relative to *apo* wild type hENT1 was 5.0 ± 0.8 °C (Figure 4.3B, Table 4.2 and Figure 4.4A and C). The $\Delta\Delta T_m$ value was determined for each variant to demonstrate the difference in stability for the NBMPR-bound relative to *apo*, relative to that seen for wild-type (5.0 ± 0.8 °C) (Figure 4.4D). Using a ± 1.3 °C cut-off (average SEM Δ T_m (Table 4.2)) four hENT1 variants (K263A, N30F, I282V, M306T) were identified as having an increased stabilising effect on the NBMPR-bound state, three variants were neutral (K283R, G305A, E264A), and one variant was destabilising (T336A) (Figure 4.4D).

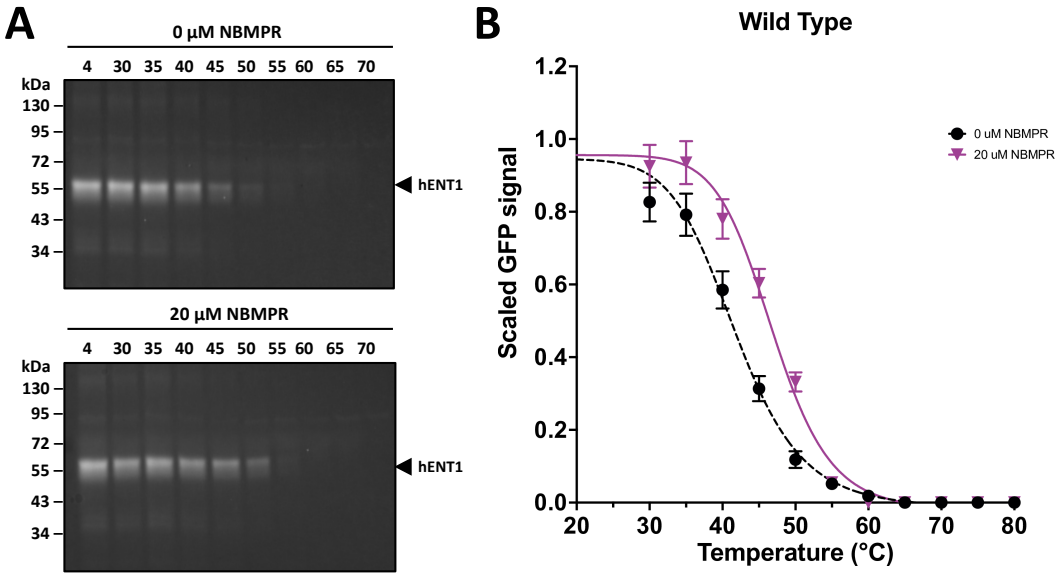


Figure 4.3 Stability curves for wild type hENT1 in both *apo* and NBMPR-bound states.

(A) Representative SDS-PAGE of hENT1 following a ten-temperature challenge. Following incubation at ten-temperature points, (4, 30, 35, 40, 45, 50, 55, 60, 65 and 70 °C), with 0 µM and 20 µM NBMPR, surviving hENT1-linked GFP signal is visualised using fluorescence imaging. hENT1 is present at ~57 kDa, and hENT1 specific signal reduces as temperature increases. The intensity of the protein that remained in solution after the temperature challenge was quantified and data normalised to the 4 °C control. (B) Wild type apo-state and NBMPR. Data shown is the average of at least n = 3 experiments. Data were fit with a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting in GraphPad Prism 9.0

Table 4.2 Δ T_m results of melting analysis of NBMPR-bound hENT1 variants relative to *apo*.

Validated ID	Δ T_m (°C)[a]	Δ T_m SEM (±°C)[b]	Repeats (n =)
K263A	10.0	1.5	3
N30F	9.9	2.0	3
I282V	8.0	1.6	3
M306T	6.7	1.5	3
K283R	6.2	1.7	3
Wild Type	**5.0**	**0.8**	**3**
G305A	4.0	0.7	3
E264A	4.7	1.3	3
T336A	0.3	0.8	6

[a]ΔT_m represents NBMPR, bound T_m relative to apo T_m
[b]Error calculated and propagated as detailed in 2.4.4.

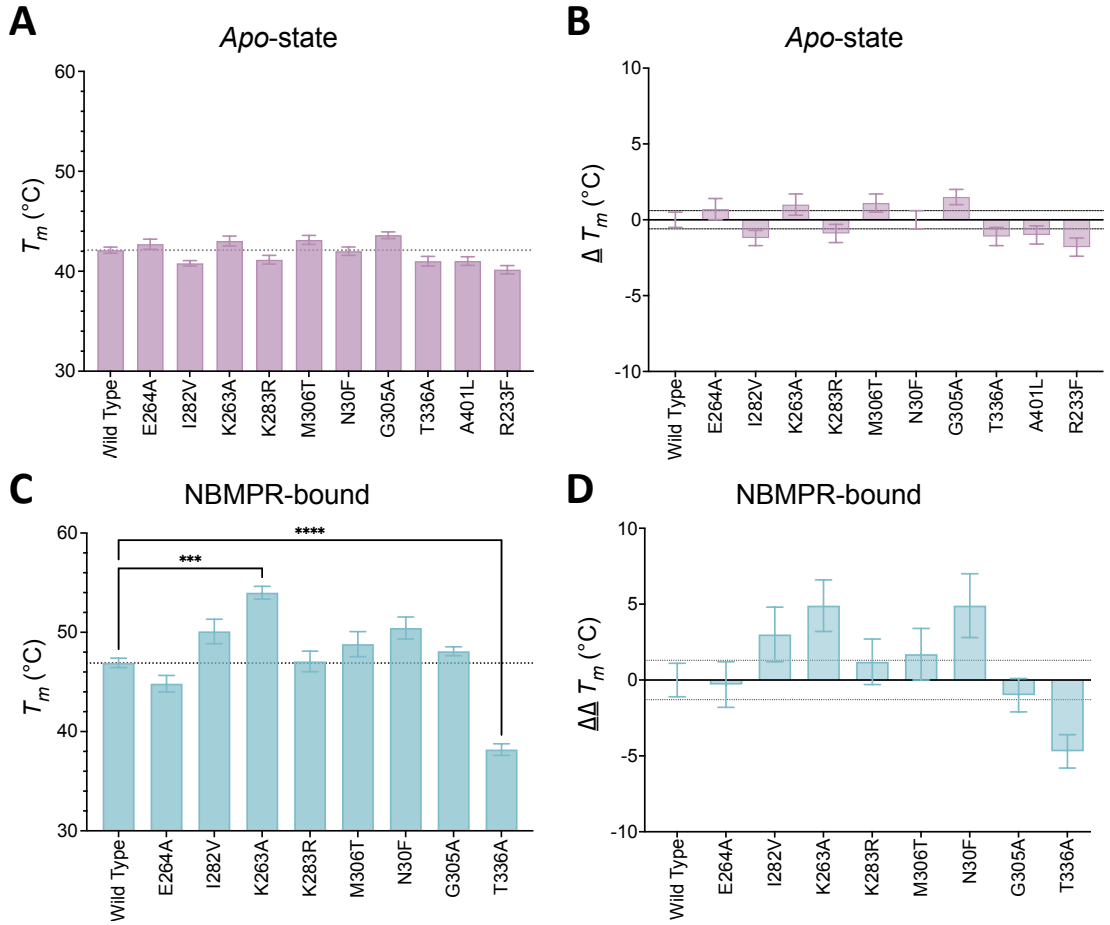


Figure 4.4 Analysis of the T_m values for hENT1 variants in the *apo* and NBMPR bound state.

(A) T_m of all hENT1 variants in an *apo*-state. Data shown is the average of at least n = 3 experiments. Error bars are representative of SEM. (B) The average SEM of the *apo* ΔT_m (variant relative to wild), ± 0.6 °, is used to assign a cut-off for assigning variants as stabilising or destabilising, the upper and lower bounds of the indicated by the dotted line. Error bars are representative of error propagation as detailed in 2.4.4.1. No statistical significance was observed for any variants that were identified as stabilising. (C) T_m of all non-significantly destabilising hENT1 variants (excluding A401L and R233F) in an NBMPR-bound state. Data shown is the average of at least n = 3 experiments. Error bars are representative of SEM. K263A and T336A are statistically significantly different to wild type, with (***) p = 0.0001 and (****) p = <0.0001, respectively. (D) The average SEM of the ΔT_m (NBMPR relative to *apo*), ± 1.3 °C, is used to assign a cut-off for assigning variants as stabilising or destabilising, the upper and lower bounds of the indicated by the dotted line. Error bars are representative of error propagation as detailed in 2.4.4.1.

4.2.4 Rationalising the effect of the variants on the stability of hENT1, in NBMPR-bound and *apo* states

As the first structures of hENT1 were published in 2019 and included a structure with NBMPR bound (PDB: 6OB6)[134], this allowed me to map all mutations (Figure 4.5) and to perform *in silico* mutagenesis for the rationalisation of the effects observed on thermostability. In mapping all variants, this highlights that none of the mutations to residues in the N-terminal domain in the *apo* state are stabilising, and that most of the mutations screened in this domain in this work are significantly destabilising. This could suggest that the N-terminal domain is less tolerant of mutations and may support differing roles between the N-terminal domain and the C-terminal domain.

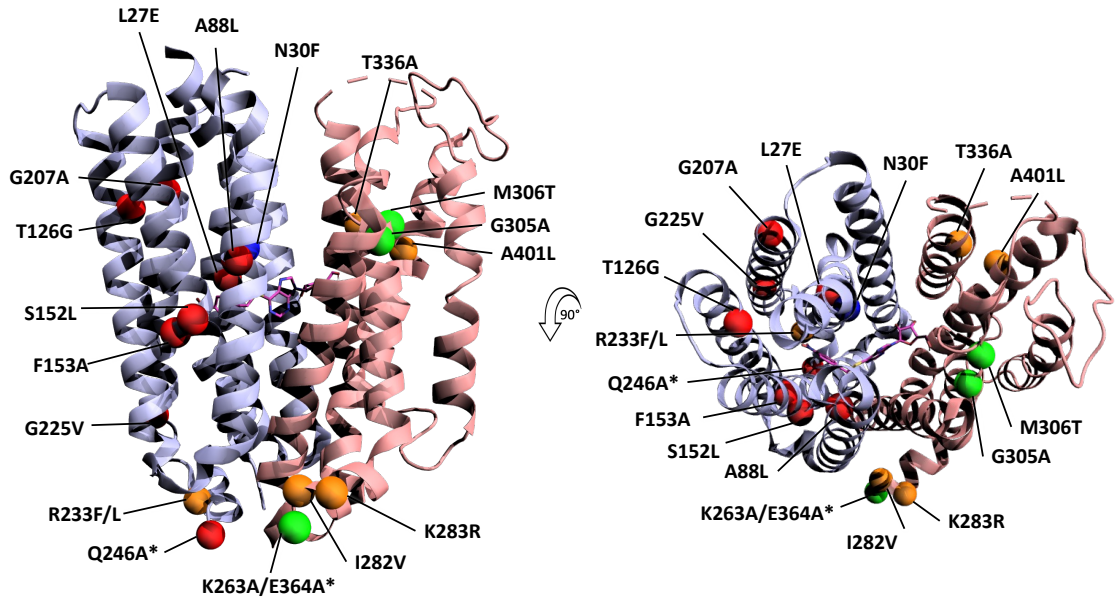


Figure 4.5 The location of the mutations for all variants of hENT1.

hENT1 (PDB: 6OB6) is shown with a side on and top-down view. The N-terminal domain is shown in blue, and the C-terminal domain is shown in pink. The locations of the mutations for each of the variants of hENT1 identified following bacmid DNA validation are mapped to the structure and are shown as spheres. Stabilising variants are shown in green, the neutral variant N30F is shown in blue, non-significantly destabilising variants are shown in orange, and significantly destabilising variants in red. NBMPR is shown in pink. *As, Q246A, K263A and E264A are in the ICL6, which is missing from the structures, the nearest residue is shown instead.

4.2.4.1 Mutation of TM1 and the central cavity

Variant N30F, which is located on TM1, has a neutral effect on the *apo*-state, Δ T_m of 0.0 ± 0.6 °C (Table 4.1 and Figure 4.4A and B). However, this mutation provides considerable stabilisation to the NBMPR-bound state, ΔΔT_m 4.9 ± 2.1 °C (Table 4.2, Figure 4.4C and D, and Figure 4.6A). Therefore, N30F has an *apo*-neutral but NBMPR-bound stabilising effect. In the NBMPR-bound

structure (PDB: 6OB6), interactions at TM1 are seen to contribute to both inhibitor binding and the gating mechanism which represents the occluded state in the outward-facing conformation[134] (1.6.1). The hydrophobic residues L26 on TM1, along with M89 and L92 on TM2, and L442 on TM11 are shown to surround the purine moiety of NBMPR. M33 on TM1 and P308 on TM7 form a narrow constriction point which prevents the NBMPR from releasing freely into the extracellular side (Figure 4.6B), thereby forming extracellular gating interactions[134].

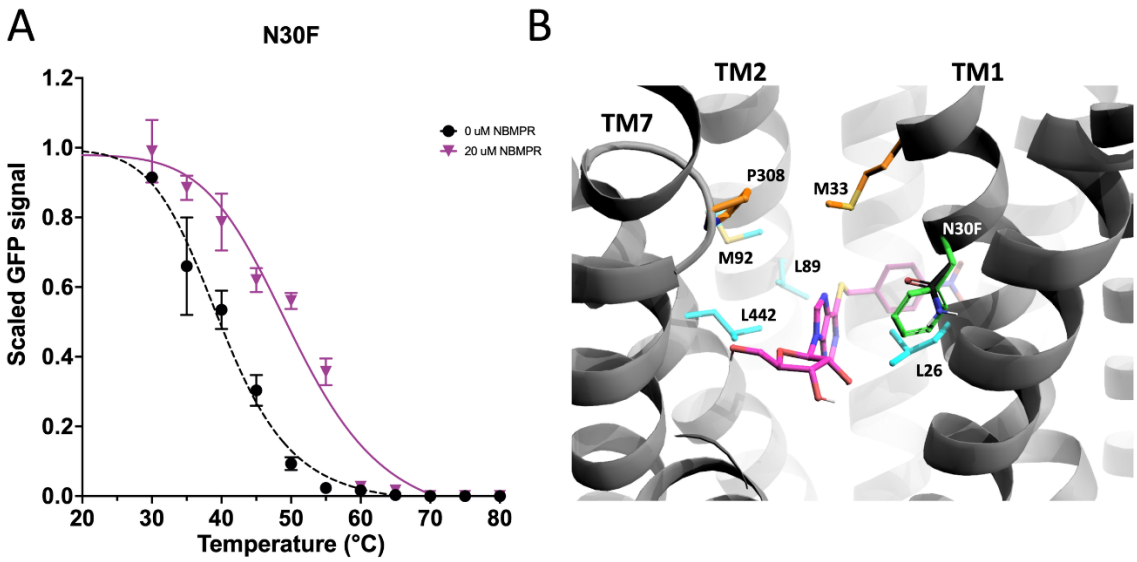


Figure 4.6. Investigation and rationalisation of variant N30F stabilisation of the NBMPR-bound state.

(A) Data generation, fitting and error analysis performed as detailed in wild type. Data shown for N30F *apo*-state is the average of n = 10 experiments, and N30F NBMPR-bound is the average of n = 3 experiments. (B) A close-up view into the central cavity of hENT1 with TM8 removed for clarity. Side chains of residues involved in the surrounding of the purine moiety of NBMPR (pink) are shown in cyan. Residues involved in the formation of the extracellular thin gate are shown in orange. Native N30 is shown in black, and *in silico* mutagenesis of N30F is shown in green.

N30 faces into the central cavity and sits one helix turn above L26, and 0.75 helix turn below M33 (Figure 4.5 and Figure 4.6B). Therefore, N30F may contribute to specific stabilisation of the NBMPR-bound state in several ways. Substitution of a polar side chain to a bulky, hydrophobic side chain increases the overall hydrophobicity of the central cavity and may support the hydrophobic environment around the NBMPR purine moiety. Additionally, the aromatic ring of the phenylalanine may contribute to π-π stacking interactions with the purine moiety. Furthermore, the bulky substitution of phenylalanine may protrude into the central cavity and therefore, may support the occlusion at the extracellular side established by the gating interactions between M33 (TM1) and P308 (TM8).

4.2.4.2 Mutation of TM7

G305 and M306 are located towards the end of TM7 that is nearest to the extracellular membrane surface and face towards TM11 and TM9, respectively (Figure 4.5 and Figure 4.7C and D). Individual mutations at these residues, G305A and M306T, stabilise the *apo*-state by 1.5 ± 0.5 °C and 1.1 ± 0.6 °C (Table 4.1, Figure 4.2 and Figure 4.4A and B). Additionally, both variants result in stabilisation in the NBMPR-bound state that is similar to that seen in wild-type, $\Delta\Delta T_m$ - 1.0 ± 1.1 °C and 1.7 ± 1.7 °C, respectively (Table 4.2, Figure 4.4C and D, and Figure 4.7A and B). It is possible that both G305A and M306T stabilise the *apo*-state by improving the helix-packing interactions between TM11 and TM9.

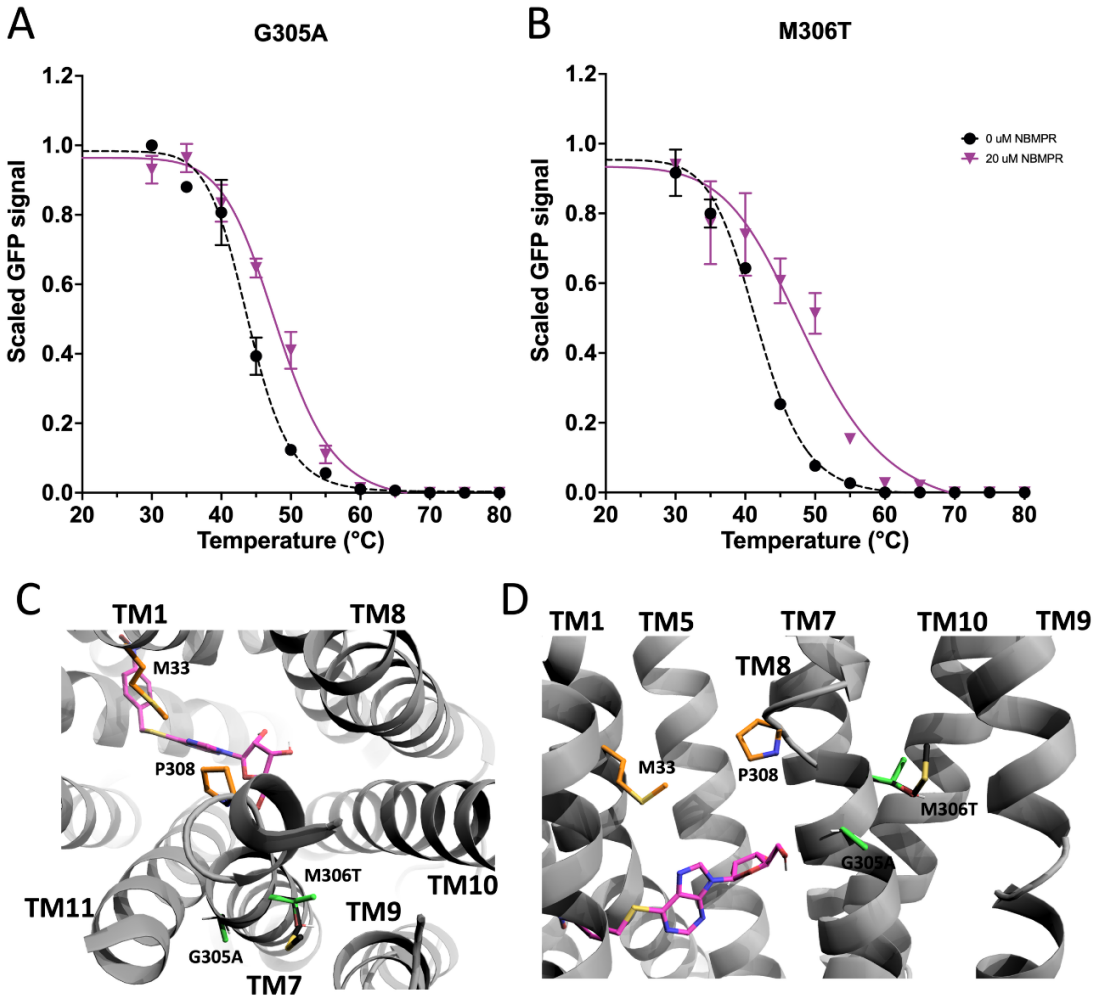


Figure 4.7 Investigation and rationalisation of variants G305A and M306T stabilisation of hENT1.

Data generation, fitting and error analysis performed as detailed in wild type. (A) Data shown for G305A *apo*-state is the average of n = 8 experiments, and G305A NBMPR-bound is the average of n = 3 experiments. (B) Data shown for M306T *apo*-state is the average of n = 5 experiments, and M306T NBMPR-bound is the average of n = 3 experiments. (C) A top-down and (D) a close-up perpendicular view into the central cavity of hENT1 (TM11 was removed in D for clarity). NBMPR is shown in pink. Residues involved in formation of the extracellular thin gate are shown in orange. Native residues are shown in black, and *in silico* mutagenesis is shown in green.

Variants I282V and K283R are also located on TM7. There is a kink in the intracellular region of both TM6 and TM7 that results in a short transverse helix that bridges the connection between the ICL and the TMs (Figure 4.5 and Figure 4.8C and D). I282V and K283R are located at this helix/bridge region of TM7 (ICH7). K283 faces towards the cytosol and I282 faces towards a hydrophobic region of TM2 and TM11 (Figure 4.8D). In the *apo*-state both I282V and K283R are destabilising, ΔT_m -1.2 ± 0.5 °C and -0.9 ± 0.6 °C, respectively (Table 4.1, Figure 4.2 and Figure 4.4A and B). However, I282V stabilises the NBMPR-bound state more than wild type, $\Delta\Delta T_m$ 3.0 ± 1.8 °C (Table 4.2, Figure 4.4C and D, and Figure 4.8 A and B), thus has an *apo*-destabilising but NBMPR-bound stabilising effect. This I282V mutation may contribute to NBMPR-bound state stabilisation through tightening of hydrophobic interactions within this region, possibly with both TMH2 and TMH11, and the lipid bilayer. For the NBMPR-bound state, K283R results in stabilisation similar to that seen for wild type, $\Delta\Delta T_m$ 1.2 ± 1.5 °C (Table 4.2, Figure 4.4C and D, and Figure 4.8 A and B). The retention of stabilisation by NBMPR suggests that, despite this mutation being destabilising for the *apo*-state, the protein is still able to interact with NBMPR in a way that provides wild type-like stabilisation.

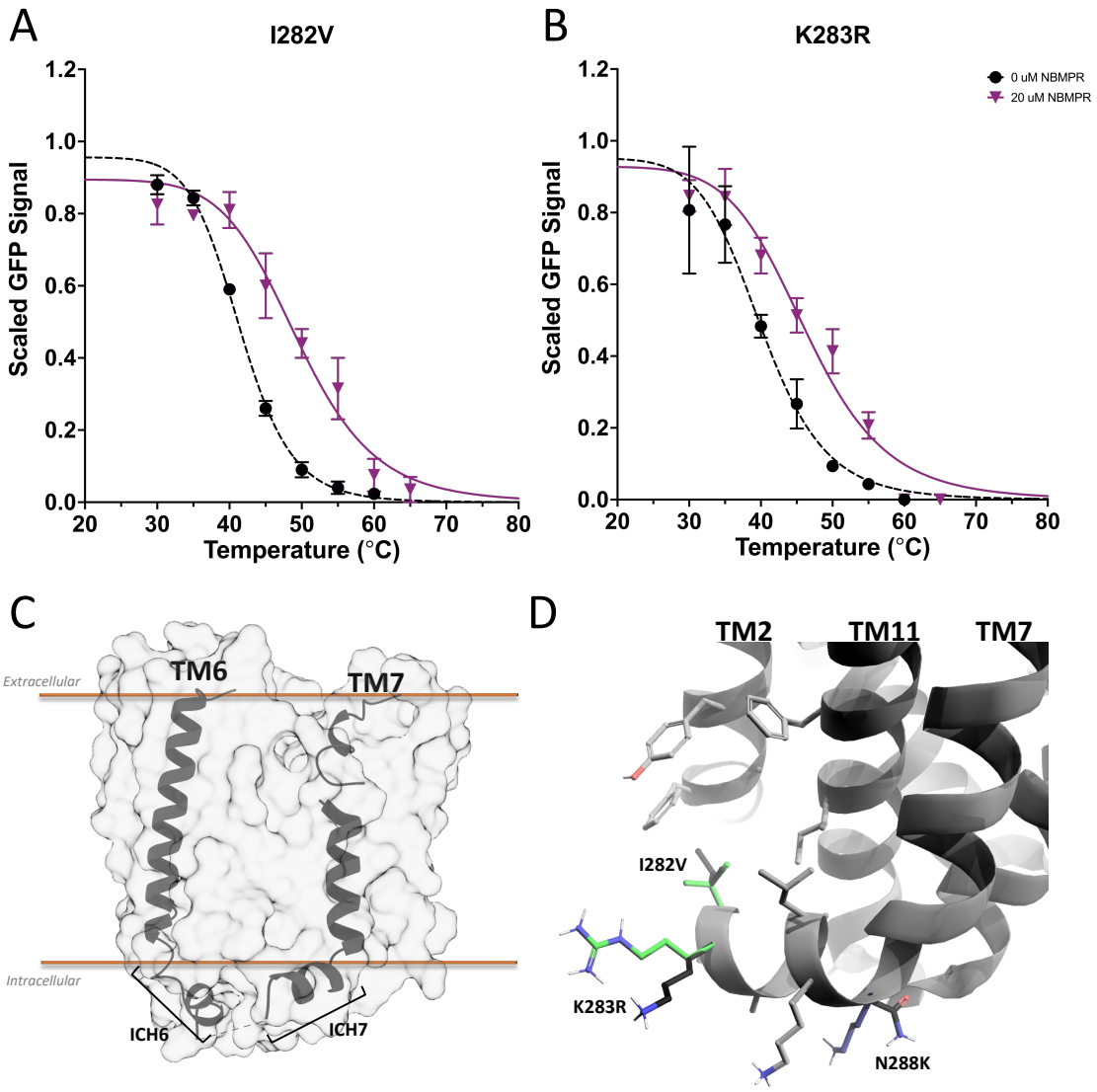


Figure 4.8 Investigation and rationalisation of variants I282V and K283R stabilisation of hENT1.

Data generation, fitting and error analysis performed as detailed in wild type. (A) Data shown for I282V *apo*-state is the average of n = 12 experiments, and I282V NBMPR-bound is the average of n = 3 experiments. (B) Data shown for K283R *apo*-state is the average of n = 12 experiments, and N30F NBMPR-bound is the average of n = 3 experiments. (C) A perpendicular view of TM6 and TM7. TM1-5 and TM8-11 removed for clarity. (D) A close-up perpendicular view of the intracellular region of the TM7 bridge of hENT1. The side chains of neighbouring native residues are shown in grey. Native residues are shown in black, and *in silico* mutagenesis of I282V and K283R are shown in green. The intracellular kink of TM7 is also the location of one of the three stabilising mutations (N288K, shown in blue) that was introduced by Wright & Lee to generate a crystallisable hENT1 construct[134].

4.2.4.3 Mutation of the ICL6

K263 and E264 are located at the ICL6. However, in the available structures of hENT1 (PDB: 6OB6 and 6OB7) residues 243-274, which contribute to ICL6, were deleted to generate a construct that was amenable to crystallisation[134] (Figure 4.5). Mutations K263A and E264A each stabilise the *apo*-state, Δ T_m of 1.0 ± 0.7 °C and 0.7 ± 0.7 °C, respectively (Table 4.1, Figure 4.2 and Figure

4.4A and B). As the ICL6 is predicted to be largely disordered[148] these variants may stabilise this region in the *apo*-state by reducing conformational flexibility. However, like I282V and K283R, each of these mutations had differing effects on stabilisation of the NBMPR-bound state. While E264A is no more stabilised by NBMPR than wild type, ΔΔ T_m -0.3 ± 1.5 °C, K263A is significantly stabilised, ΔΔT_m 5.0 ± 1.7 °C (p = 0.0001) (Table 4.2, Figure 4.4C and D, and Figure 4.9A and B). Additionally, the curve fit to the NBMPR-bound state data of K263A has a steeper hill slope than that of the *apo*-state, and other variant NBMPR-bound curves (Figure 4.9A). This suggests that K263A may unfold more cooperatively in the NBMPR-bound state. The replacement of the large, charged residue at position 263 with alanine may allow for other NBMPR-bound state stabilising interactions to take place.

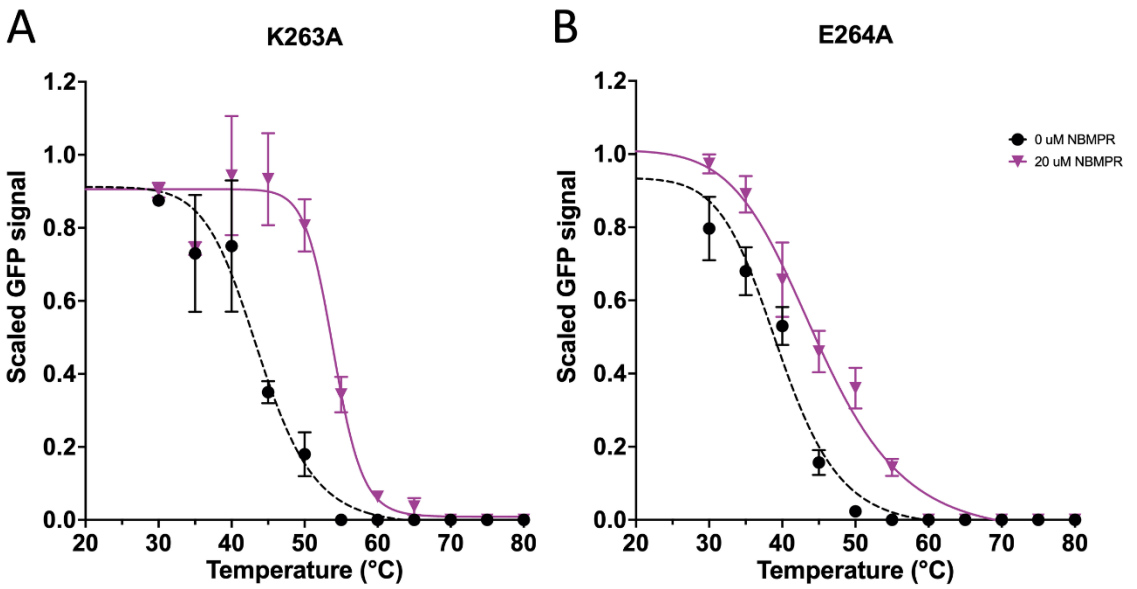


Figure 4.9 Investigation of variants K263A and E264A stabilisation of hENT1.

Data generation, fitting and error analysis performed as detailed for wild type. (A) Data shown for K263A *apo*-state is the average of n = 5 experiments, and K263A NBMPR-bound is the average of n = 3 experiments. (B) Data shown for E264A *apo*-state is the average of n = 7 experiments, and E264A NBMPR-bound is the average of n = 3 experiments.

4.2.4.4 Mutation of TM8

T336 is located at the extracellular region of TM8 and faces towards TM10 and the lipid bilayer (Figure 4.5 and Figure 4.10C). A number of residues on TM8 which face towards the central cavity have previously been shown to be important determinants in inhibitor sensitivity[143], with D341 and R345 specifically being shown to interact with the ribose moiety of NBMPR[134] (Figure 1.17 and Figure 4.10D). The mutation T336A has a destabilising effect on the *apo* state, Δ T_m -1.1 ± 0.6 °C, and shows no stabilisation in the presence of NBMPR, ΔΔ T_m -4.7 ± 1.1 °C (p = <0.0001) (Table 4.1, Table 4.2, Figure 4.2, Figure 4.4 and Figure 4.10).

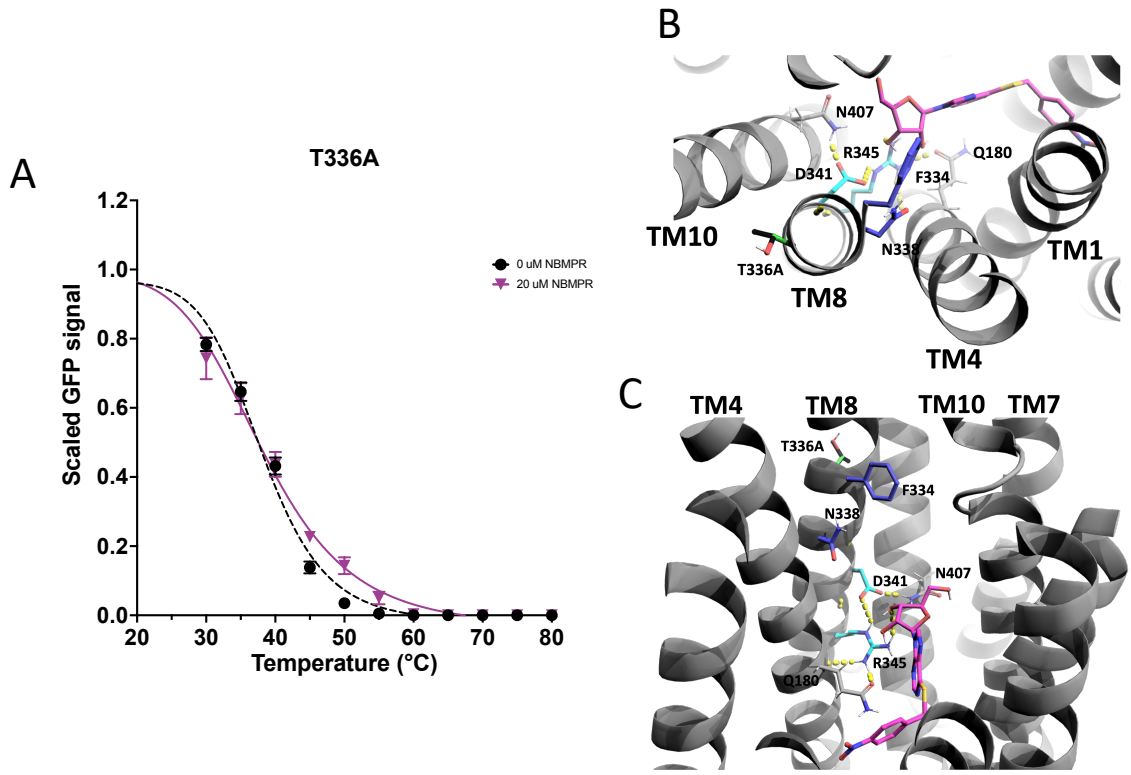


Figure 4.10 Investigation and rationalisation of variant T336A destabilisation of *apo*- and NBMPR bound hENT1.

Data generation, fitting and error analysis performed as detailed for wild type. (A) Data shown for T336A *apo*-state is the average of n = 12 experiments, and T336A NBMPR-bound is the average of n = 6 experiments. (B) A top-down and (C) A close-up perpendicular view of the central cavity of hENT1 (TM11 is removed in D for clarity). Side chains of residues involved in the surrounding of the ribose moiety of NBMPR (pink) are shown in cyan. Residues previously determined to be important determinants in inhibitor sensitivity are shown in blue. Native T336 is shown in black, and *in silico* mutagenesis of T336A is shown in green.

Due to the seeming absence of stabilisation by NBMPR, I set out to investigate the specific binding of NBMPR by T336A. Repeat small scale expressions were performed for wild type and T336A, and the cells harvested for a whole cell, single concentration radioligand binding assay and SDS-PAGE. Bound [3H]-NBMPR was normalised to hENT1-linked GFP signal, and then the normalised values were then scaled relative to wild type (Figure 4.11). These data suggest that the loss of stabilisation by NBMPR is a result of a specific reduction in binding, with T336A [3H]-NBMPR binding 0.14 times that of wild type.

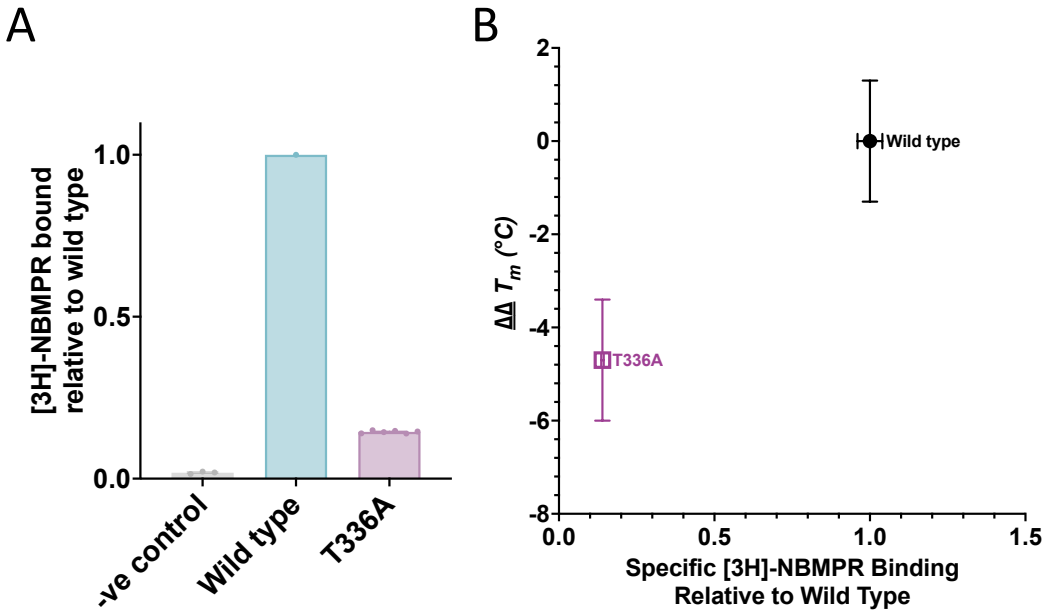


Figure 4.11 [3H]-NBMPR binding by wild type and T336A hENT1.

(A) Relative [3H]-NBMPR bound to wild type and T336A hENT1 following a single concentration radioligand binding assay. T336A and negative control results are normalized to wild type. (B) Scatter plot of hENT1 variant ΔΔ T_m versus amount of radiolabelled specific inhibitor [3H]-NBMPR bound in the membrane, relative to wild type. Error bars are representative of error as detailed in 2.4.4.1.

Mammalian orthologues of hENT1 feature a highly conserved polar residue (T/N) at the position equivalent to T336 of TM8 (Appendix Figure 4). Conversely, the equivalent residue in the NBMPR insensitive isoforms hENT2, hENT3 and hENT4 (and their mammalian orthologues) is a highly conserved hydrophobic residue (L/V). Therefore, the reduction in binding observed in T336A is possibly due to the exchange of a polar residue for the NBMPR-insensitive isoform (hENT2, hENT3 or hENT4)-like hydrophobic residue.

4.2.5 Comparing the mutations of the published hENT1 structures

In addition to screening all variants of hENT1 suggested by IMPROvER, I set out to screen the mutations detailed in the structure paper by Wright & Lee[134]. There is no comment on the degree of stabilisation offered by each of these mutations in the publication, either as single variants or in combination. Single mutations for L168F, P175A and N288K were successfully introduced to the hENT1-CTGH expression construct by quick change mutagenesis, and sequences were confirmed by Sanger sequencing (Eurofins). Ten-temperature challenges were performed on each of these variants, and bacmid DNA was extracted from the insoluble cell pellet to ensure that no-mishandling or cross contamination had occurred. The T_m of each variant was determined as detailed in 2.4.4.1 (Figure 4.12). All three variants showed stabilisation equal to, or significantly above that observed for all IMPROvER mutants. P175A

stabilised to a similar degree of the top stabilisers found using IMPROvER, with a ΔT_m of 1.5 ± 0.5 °C, which is the same as was observed in the top IMPROvER *apo*-state variant, G305A (4.2.4.2). P175 is located on TM5 and faces towards TM1, and the mutation to alanine here may improve helix packing. L168F and N288K stabilise the *apo*-state significantly more than all *apo*-state stabilising IMPROvER variants, with ΔT_m of 4.6 ± 0.5 °C (p = 0.016) and 6.7 ± 1.3 °C (p = <0.0001), respectively. L168F is located at the ICL4, which connects TM4 and TM5. This intracellular region of the N-terminal domain (TM1 and ICH6 specifically) is rich with aromatic residues (Figure 4.12D). Therefore, this mutation to phenylalanine may contribute to π-π stacking interactions in this region. N288K, like I282V and K283R (4.2.4.2), is located at the intracellular helix of TM7, ICH7, that bridges the connection to the ICL6. N288K faces towards the cytosol and the ICL10, which connects TM10 and TM11. This ICL10 features several charges residues. Therefore, the mutation from a polar residue to a charged residue at position 288 may allow for stabilising charged interactions to take place between the ICH7 and the ICL10 (Figure 4.12D).

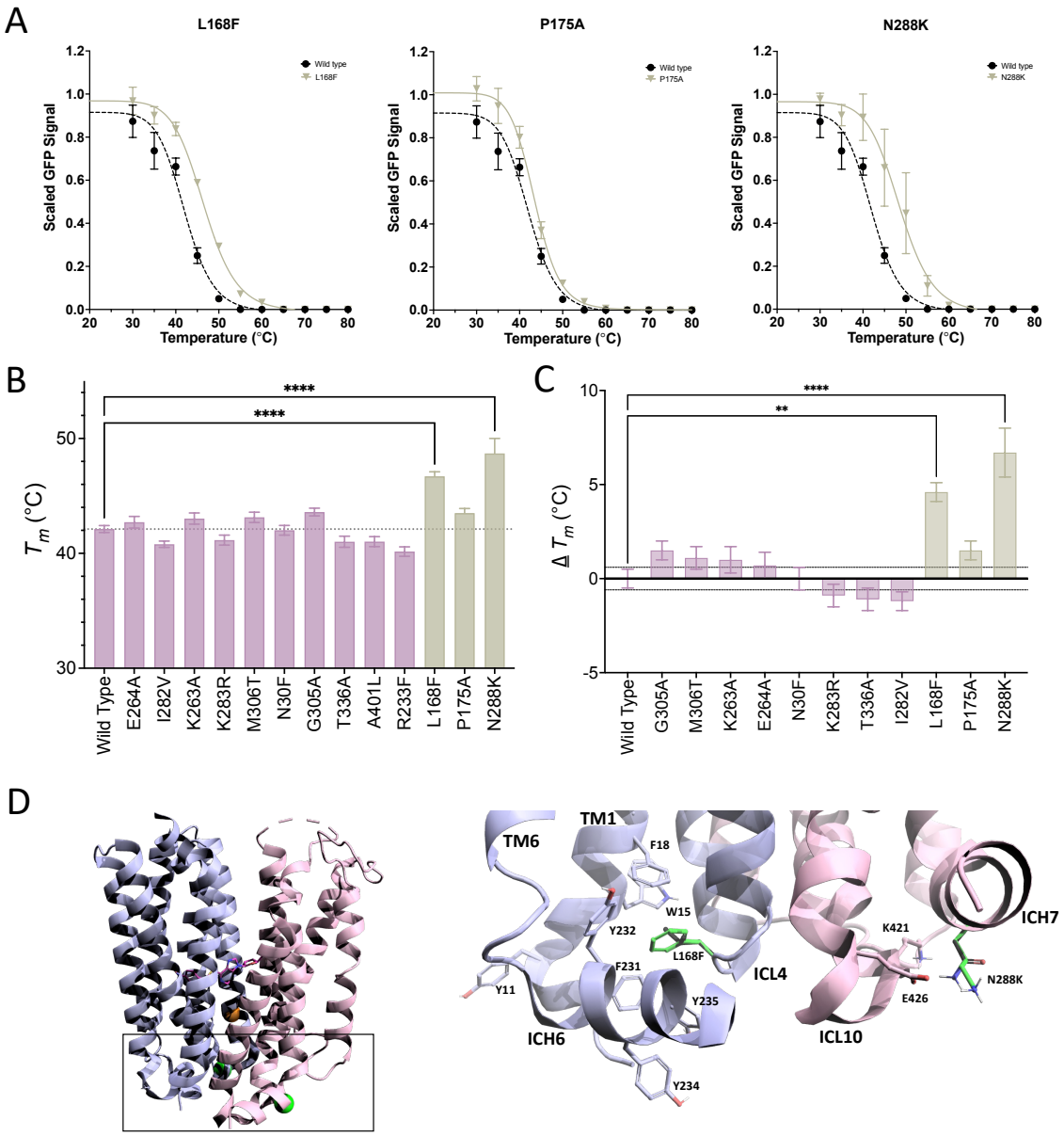


Figure 4.12 Ten-temperature stability assays of *apo*-variants (Wright & Lee[134]) of hENT1.

(A) Ten-temperature stability curves of each of the variants detailed by Wright & Lee[134] that contributed to the generation of an hENT1 construct that was amenable to crystallisation. (B) The comparison of these three variants (khaki) to the IMPROvER suggested variants (pink) demonstrates that L168F and N288K are significantly stabilising (****, p = <0.0001). (C) This significance is upheld when presented as ΔT_m values, with L168F p = 0.016 (**) and N288K p = < 0.0001 (****). (0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***) and <0.0001 (****)). (D) The locations of L168F (green), P175A (orange) and N288K (green) are mapped to the structure of hENT1 (PDB: 6OB6), with the N-terminal domain shown in light blue and the C-terminal domain shown in light pink. A cropped view of the intracellular domain of hENT1 is shown in the right panel, with wild type residues shown in black, and variants shown in green. Local aromatic resides are shown with stick representation in light blue on the N-terminal domain, and local charged residues are shown with stick representation in light pink on the C-terminal domain.

4.3 Discussion

4.3.1 Developing a new method for bacmid DNA extraction

As a result of the discovery of inconsistencies at the outset of this work, the continuation of the investigation of the thermostability of IMPROvER selected variants depended on all the experiments being repeated. At that point, only plasmid DNA was available for validation. As discussed in 4.1.2, baculovirus expression systems require multi-step processes, from plasmid DNA transfection to repeat baculovirus infections for increasing the viral titre before achieving adequate levels of protein expression. Therefore, without a means to validate bacmid DNA the workload would have been far more labour and reagent intensive. Moreover, there would still be multiple opportunities for handling errors to occur. Therefore, my initial goal was to explore methodologies available for the validation of bacmid DNA from transfected Sf9 cells.

There are commercially available specialist kits for the extraction of DNA from tissues and cells. However, there is no explicit data on the use of these kits for the extraction of bacmid DNA from transfected Sf9 cells. As of May 2022, the only reported data for bacmid DNA extraction from transfected Sf9 cell cultures is that of McCarthy & Romanowski, 2008[236]. However, this method requires the chloroform extraction of viral particles from the cell-free media.

Therefore, to eliminate the need for both toxic organic solvents and specialist extraction kits, I developed a protocol to make the commercially available kits for the isolation of plasmid DNA from bacterial cultures work for the isolation of bacmid DNA from Sf9 cells. This methodology allows for the validation of bacmid DNA at all steps in the Sf9 expression process. Furthermore, as I have shown that I successfully isolated, amplified, and sequenced DNA extracted from the insoluble pellet, this methodology allows for the extraction of DNA from an experimental 'waste' product. Therefore, this also eliminates the need for additional cell cultures and materials solely for the purposes of validating DNA. In addition to data validation, the ability to extract bacmid DNA at all steps allows for the confirmation of target integration, screening of cross-contamination and the identification of why expression levels have decreased[277].

4.3.2 Differential stabilisation of hENT1 states

4.3.2.1 The role of TM8 and T336A

In the *apo*-state, T336A is destabilising variant, but without any statistical significance (Table 4.1). However, in the NBMPR-bound state, T336A is a statistically significant destabilising variant

(p = <0.0001) (Figure 4.4). As detailed in 4.1.2, in this work, thermostability is assessed by relating the amount of protein that remains in solution following heat challenges and high-speed centrifugation to a 4 °C control. The amount of protein that remains in solution is determined by the in-gel intensity of hENT1-linked GFP signal. As discussed in 1.7.2.1, GFP is often used as a reporter protein as it can serve as a highly stable and sensitive indicator of protein folding, with a reported T_m of 78 °C. As GFP is expected to remain folded at temperatures significantly higher than hENT1, it is expected that the unfolding of the target protein occurs before GFP, and thus loss of GFP signal as a fusion protein is driven by the aggregation of the unfolded target, rather than the unfolding of GFP. Experiments using an alternate thermostable protein have shown that in this assay, all GFP signal is lost at temperatures above 70 °C (Appendix Figure 6). However, using the same assay, the apparent T_m of wild type hENT1 was determined to be 42.0 ± 0.3 °C. Therefore, as the GFP signal is lost at a significantly lower temperature as observed for GFP, this suggests that this is most likely representing the specific unfolding of hENT1.

Therefore, as the data analysed in gel represents the soluble, folded protein that is resistant to the temperature challenge, the loss of the stabilising effect from NBMPR in T336A is most likely because of a specific change in the binding, rather than simply being a result of misfolded protein. I set out to explore the ability of T336A to bind NBMPR by performing a single concentration radioligand binding assay, and these results (4.2.4.4) showed that the mutation T336A results in a significant reduction in the binding of NBMPR, with the T336A variant binding seven times less [3H]-NBMPR than wild type hENT1 (Figure 4.11). Therefore, this data further supports that the mutation T336A affects the ability of hENT1 to bind NBMPR.

As discussed in 4.2.4.4, it is possible that the reduction in NBMPR binding is due to the exchange of the polar threonine at position 336 to the hydrophobic alanine, with this exchange conferring a more NBMPR-insensitive isoform of hENT. However, as discussed in 1.6.3, NBMPR sensitivity is most likely mediated by the residue at position 154, and thus the availability of the deep hydrophobic pocket which the *p*-nitrobenzyl ring of NBMPR occupies upon binding. As this pocket is still present in the T336A variant, NBMPR may still be able to bind. Performing radioligand uptake assays would allow me to investigate the sensitivity of T336A to inhibition by NBMPR. However, owing to time constraints I have been unable to perform these.

T336 is located at the extracellular region of TM8 and faces towards TM10 and the lipid bilayer (Figure 4.5 and Figure 4.10B). Therefore, while I propose that residue 336 confers NBMPR affinity, it does not appear to participate in direct interactions with NBMPR. Several residues on TM8 have previously been shown to be important determinants in inhibitor sensitivity[142,143,278]

(1.6.2). In the structure of hENT1 with NBMPR bound (PDB: 6OB6) (Figure 4.10C) D341 and R345 were shown to participate in direct interactions with the ribose moiety of NBMPR[134]. D341 is exclusively conserved across mammalian ENTs, and residue 345 is a highly conserved positively charged residue (R/K) (Appendix Figure 4). As a nucleoside analogue, interactions with NBMPRs purine and ribose moiety are expected to represent interactions with endogenous nucleoside substrates, and thus interactions with the ribose moiety of nucleosides are predicted to also be mediated by D341 and R345. As discussed in 1.5.2.2, in addition to NBMPR sensitivity, hENT isoforms differ in their substrate selectivity. hENT1 has a higher affinity for nucleosides and hENT2 a higher affinity for nucleobases[90,91,99]. Therefore, although D341 and R345 are highly conserved in hENT1 and hENT2, the isoforms have differing affinities for the ribose moiety. Therefore, I propose that the residue at position 336 may mediate interactions with TM10 and/or the lipid bilayer, and thus may affect the ability of TM8 to support specific ribose interactions *via* D341 and R345. Consequently, the loss of the stabilisation by NBMPR for variant T336A may be due to a reduction in affinity for the ribose moiety (of NBMPR and endogenous substrates), rather than a specific insensitivity to NBMPR. As discussed, performing radioligand binding assays for wild type and T336A hENT1 would allow me to investigate the comparative kinetics of binding. Furthermore, it could be of interest to explore the effects of the mutation T336A on substrate kinetics, in addition to NBMPR kinetics and inhibition.

4.3.2.2 Towards understanding the role of TM7 and ICL6 in hENT1

In MFS transporters, the ICL6 bridges the connection between the N- and C-terminal domains through TM6 and TM7, respectively. TM7 plays a significant role in the mechanism of action in many MFS transporters. It is typically present as a discontinuous helix that undergoes rearrangements, such as partial unwinding at the extracellular region, during substrate binding and translocation[50,51,53,69,77]. This characteristic discontinuous helix is observed in hENT1 and supports the extracellular gating interactions between P308 and M33 of TM1 (1.6.1). However, as only outward-facing inhibitor bound structures of hENT1 are available, it is unknown what rearrangements TM7 may undergo in the transition from *apo*-state to substrate/inhibitor bound and how these rearrangements may influence the ICL6.

The data presented in this chapter suggests that the ICL6, ICH7 and TM7 play a role in the conformational stabilisation of hENT1. This is seen in both variants predicted by IMPROvER (G305A, M306T, K263A, E264A and I282V), and those used by Wright & Lee[134] in the generation of a construct of hENT1 that was amenable to crystallisation (L168F and N288K). Furthermore, the data presented in this chapter suggests that the IMPROvER variants provide differential

stabilisation of hENT1 states, with variants at the ICL6 (K263A and E264A) and TM7 (G305A and M306) stabilising the *apo*-state, and a single variant at the ICL6 (K263A) and a variant at the ICH7 (I282V) stabilising the NBMPR-bound state. I propose that these variants support interactions that contribute to the gating mechanisms at the intracellular face. However, without structures of the ICL, the mechanisms by which this is achieved remain unknown.

As discussed in 1.6.4.1 and 4.2.4.3 and, in the available structures of hENT1 (PDB: 6OB6 and 6OB7) residues 243-274 of the ICL6 were deleted to generate a construct that was amenable to crystallisation[134] (Figure 4.5 and Figure 4.8C). Therefore, the structure and orientation of the ICL6 in any state remains unknown. Previous NMR studies have suggested that the ICL6 is unstructured[146,147,149]. However, sequence analysis and computational structural predictions[155,158] suggest that, in addition to the helices at the intracellular domain of TM6 and TM7 (Figure 4.8C), there is a short helix at residues 243-256. Thus, suggesting that the ICL6 of hENT1 is highly α-helical.

As discussed in 1.6.5, despite recent advancements in the field, models produced using currently available computational methods have low to very low confidence in the prediction of this region (Figure 1.19). The structure and function of loop regions of MFS members is highly diverse, and thus highlights the difficulties in identifying suitable homology models for the prediction of loop regions in hENT1. For example, the MFS sugar porter subfamily, which includes the mammalian GLUTs[77] and their bacterial homologues XylE [67] and GlcP [64], feature an intracellular domain comprised of a series of short helices (the ICH domain) at the ICL6 (Appendix Figure 5B). This ICH domain directly interacts with TMs and is suggested to act as a latch which secures the closure of the intracellular gating domain in the outward-open conformation[64,66,67,77]. The minor intra-domain rearrangements of the ICH region are reportedly driven by rearrangements of TM7 upon substrate binding[63]. The di-/tripeptide transporter PepT2[184], and the plant nitrate transporter NRT1.1[279], also feature an ICL6 that is highly α-helical. However, here the helical loop extends away from the transporter (Appendix Figure 5C) and feature charged residues at the distal end which are suggested to help stabilise the domain on the intracellular side of the membrane. The ICL6 of hENT1 also contains several charged residues that may support interactions with the lipid-bilayer[147,184,279].

This diversity in the structure and function of MFS loops highlights the importance of experimentally determined structures of ENTs. Furthermore, as interactions with the lipid-bilayer may contribute to regulation and conformational stabilisation [7,17,21], the study of

membrane proteins in a native-like lipid containing environment is essential for understanding native structures and the molecular basis of their mechanism of action.

4.4 Summary

The effects of mutations discussed in this chapter support a role for the ICL6 in the intracellular gating mechanisms of hENT1[146,149]. However, the molecular basis by which this is achieved remains unknown and the mechanisms cannot be properly addressed. Thus, this work further highlights the need for experimentally determined full-length hENT1 structures, with the inclusion of key features such as the ICL6. Owing to the increase in stability above that of wild type in the presence of NBMPR, variants N30F, K263A and I282V (Table 4.2) could be good candidates for further investigations and may help in achieving this goal.

In addition, as discussed in 4.3.2.1, the work also supports a role for residue 336 in the indirect mediation of substrate selectivity. Therefore, while T336A is a destabilising variant of hENT1, and thus would not be suitable for structural studies, this variant could still provide opportunities to gain insights into hENT1 mechanism of action.

Chapter 5 The study of a thermostable homologue of hENT1

5.1 Introduction

At the outset of my project, the status of the work on hENT1 suggested that it was highly unstable and structural analysis would not be feasible. As discussed in 1.6, there are several different and complementary approaches that can be utilised for the optimisation of protein to produce a construct or variant that is more amenable to structural analysis. Three of the most successful approaches for improving the amenability of transporter proteins for structural analysis are protein engineering (1.7.2), the use of binding partners (1.7.4), and homology modelling (1.7.3). In turn each of these approaches have been explored. Binding partners and protein engineering have been discussed in detail in previous chapters.

The study of homologous proteins has provided the foundation for the understanding of the molecular basis of transport of several mammalian transporters[64,67,77,104,183,280-282]. The identification of homologues may facilitate the 'mass production' of a target by using simpler and cheaper expression systems. However, eukaryotic proteins typically require eukaryotic expression systems to produce functional protein[145,162,163]. ENTs are expressed exclusively by eukaryotes, and thus the options for large scale growths and 'mass production' of proteins are limited to yeast expression systems (1.7.1). Therefore, while identifying an ENT homologue amenable to yeast expression would prove advantageous, the identification of a homologue with improved stability would also prove beneficial, and may produce protein more amenable to processes required to achieve structural characterisation[173] .

5.1.1 Aims and strategy

The work presented in this chapter focused on the identification of a new, thermostable homologue of hENT1. To this end, I selected five putative homologues from diverse eukaryotic kingdoms for screening. As there are already several pre-existing studies for the characterisation of mammalian, plant, yeast, and parasitic protozoan homologues of hENT1, these classes were not explored any further in this study[131,142,283-290]. All five homologues were expressed with an N-terminal GFP tag to allow for visualisation and tracking throughout expression and purification. Furthermore, GFP based assays, as used in chapter 4, were utilised to assess thermostability. In addition, the in-gel presentations of all homologues were analysed for the presence of the suggested markers of instability that are seen in hENT1 (higher MW species and broad bands, 3.3.1). Once a thermostable homologue was identified large scale expression and

purification was performed for the purposes of establishing crystallisation trials. I also set out to use radioligand binding and thermostability assays to explore the sensitivity to the hENT1 inhibitor NBMPR, and to attempt to identify endogenous substrates.

5.2 Results

5.2.1 Screening of putative homologues of hENT1

5.2.1.1 Bioinformatic selection of candidates

I performed literature and bioinformatic searches to identify putative homologues of hENT1 for screening of expression and thermostability[291-298]. Species of interest were selected to provide a range of homologies. Putative homologues (21-60% identical as determined by BLAST®) to hENT1 from five species, each from distinct classes, were selected for screening (Table 5.1). All homologues were predicted to share the same conserved ENT topology, featuring an intracellular N-terminus, extracellular C-terminus and 11 transmembrane α-helices.

Table 5.1 Putative homologues of hENT1 selected for screening

Kingdom	Phylum	Class	Species		% identical to hENT1
Animalia	Chordata	Aves	*Gallus gallus*	*'Gg*ENT'	60 %
		Leptocardii	*Branchiostoma belcheri*	*'Bb*ENT'	39 %
		Mammalia	***Homo sapiens***	**hENT1**	**100 %**
	Tardigrada	Eutardigrada	*Ramazzottius varieornatus*	*'Rv*ENT'	27 %
Fungi	Ascomycota	Eurotiomycetes	*Byssochlamys spectabilis*	*'Bs*ENT'	26 %
Protozoa	Ciliophora	Oligohymenophorea	*Tetrahymena thermophila*	*'Tt*ENT'	21 %

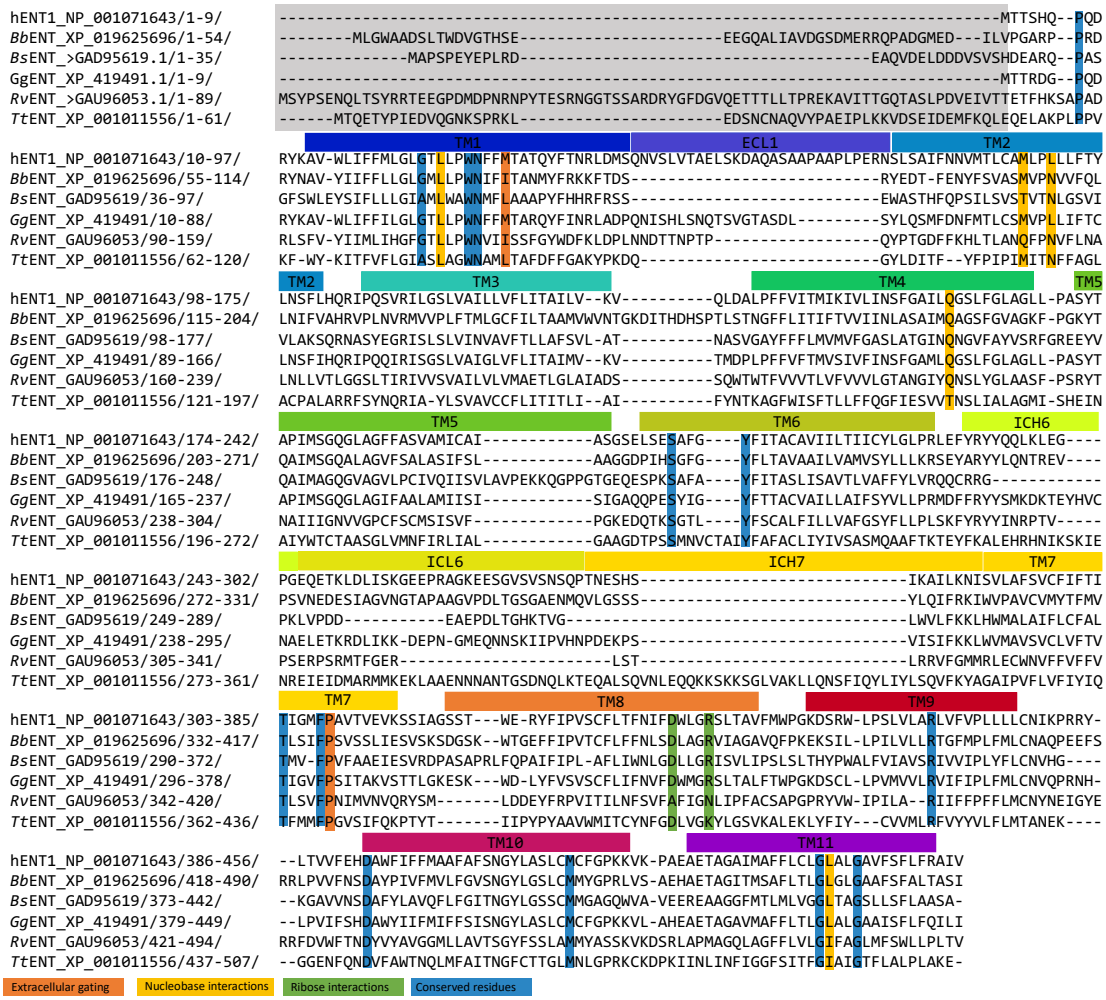


Figure 5.1 Sequence alignment of all putative homologues of hENT1.

The 11 transmembrane helices that are predicted for all homologues are annotated using a rainbow schematic. Any extended N-terminal regions that are predicted to be disordered are highlighted in grey. Equivalent residues to those involved in extracellular gating interactions in hENT1 are highlighted in orange, and those involved in interactions with the purine and ribose moiety of NBMPR are highlighted in yellow and green, respectively. Residues that are conserved in all putative homologues are highlighted in blue.

5.2.1.2 Construct design

5.2.1.2.1 Expression system

The screenings of all homologues were performed in parallel, in equal conditions. Therefore, as expression hosts most like that of the native host produce higher yields of functional protein and the selected homologues have diverse classes and kingdoms, I set out to express all homologues in Sf9 cells using a baculovirus mediated expression system. While post translational modifications in Sf9 cells could be more complex than some of the homologues 'need', the use of systems with less complex post translational modifications are likely be insufficient for some of the homologues included. Therefore, taking an approach where 'over

equipped' is preferred to 'under equipped', the use of Sf9 cells would hopefully increase the likelihood of the expression of functional protein for all homologues.

5.2.1.2.2 Sequence modification

As discussed in 1.7.2, there are several simple approaches that can be applied to construct design to improve protein expression and stability. I set out to explore the expression of all homologues in Sf9 cells. To control for codon usage, and thus improve translation efficiency and expression, all constructs were codon optimised for Sf9. In addition, as structural disorder gives rise to structural instability, sequences were analysed using Phobius[237] for the identification of large and disordered regions. Topological prediction identified that hENT1 features a 12 residue N-terminus that precedes the first TM. Therefore, following topological prediction of all homologues, all constructs were truncated at the consistently predicted disordered N-terminus to achieve the same 12 residue N-terminal extension. Sequences were also analysed to ensure that no typical signal or cleavable sequences were in the sequence that was to be truncated.

5.2.1.2.3 Location of protein tags

In-keeping with the optimised expression of hENT1 in Sf9 cells previously achieved by Dr. Harborne, all putative homologues of hENT1 were designed to be expressed with N-terminal tags. While C-terminal tags are frequently used in literature reports of expression and purification of ENTs[134,241,289], previous work determined that in-house expression and purification of hENT1 was improved by using N-terminal tags. Furthermore, Dr. Harborne determined that, in incidences where C-terminally tagged constructs were well expressed, the cleavage of tags for the elution of a tag-free construct was unsuccessful. Owing to the asymmetry of ENTs, the N- and C-termini are oriented differently, with the N-terminus towards the cytosol and the C-terminus towards the extracellular space. However, in the structures of hENT1 it appears that the C-terminus is possibly, or close to being, buried within the lipid bilayer[134]. Therefore, C-terminal tags are more likely less flexible and less accessible than N-terminal tags without the addition of further linker sequences. Similarly, the proximity to the membrane is possibly more likely to cause steric hinderance, and thus affect protein function. Therefore, N-terminal tags were the focus of the initial screens.

As all tags have the potential to affect protein folding and function, regardless of N-terminal or C-terminal location, the validation of expression and function for all constructs is essential. Furthermore, while the initial constructs investigated had N-terminal tags, cloning and construct design was done in a way that should alternate tags and C-terminal tagging be explored, this

would be easily achieved through simple sub-cloning protocols into alternate in-house pFastBac vectors.

5.2.1.2.4 Choice of tags

As discussed, the expression and purification of all putative homologues were investigated using N-terminal tags. The use of a His_6-tag has been used successfully for the purification of hENT1 (3.2.1). However, as the homogeneity of hENT1 is reduced following TEV-protease digestion for the elution of a tag free construct (3.2.1.3.3), I wanted to investigate the performance of an alternate protease, human rhinovirus (HRV) 3C protease. Furthermore, to allow for the visualisation and fluorescent analysis of expressed protein, I explored the use of a GFP tag as a reporter protein.

5.2.1.3 Expression in Sf9

All putative homologues were subcloned into a pFastBac™ donor plasmid with an expression cassette that provides N-terminal His_6-GFP tags, followed by a HRV protease recognition site (pFastBac™-NHGV). Insertion was confirmed by colony PCR and Sanger sequencing. Sequencing results determined that initial cloning of *Bb*ENT resulted in a frameshift mutation, and so was omitted from further studies. The remaining four homologue constructs were propagated in the DH10Bac™ *E. coli* strain and positive recombinant clones were identified by antibiotic selection and blue/white screening. Bacmids for *Bs*ENT, *Gg*ENT, *Rv*ENT and *Tt*ENT were isolated for transfection into adherent Sf9 cultures.

5.2.1.3.1 Expression and localisation

In baculovirus mediated expression, recombinant bacmid DNA is propagated in Sf9 cells following transfection through the secretion of recombinant baculovirus particles, which can infect further Sf9 cells. As the pFastBac™ expression cassette is under the control of a baculovirus specific polyhedrin promoter, baculovirus propagation coincides with recombinant gene expression. Therefore, following transfection, cells were observed, and successful transfection was confirmed by changes in cell morphology. Furthermore, owing to the use of GFP as a fluorescence reporter protein, expression and localisation was confirmed by fluorescence microscopy. All four homologues showed good levels of protein expression and positive GFP fluorescence, with correct plasma membrane localisation (Figure 5.2) and no accumulation of GFP fluorescence within the cytosol.

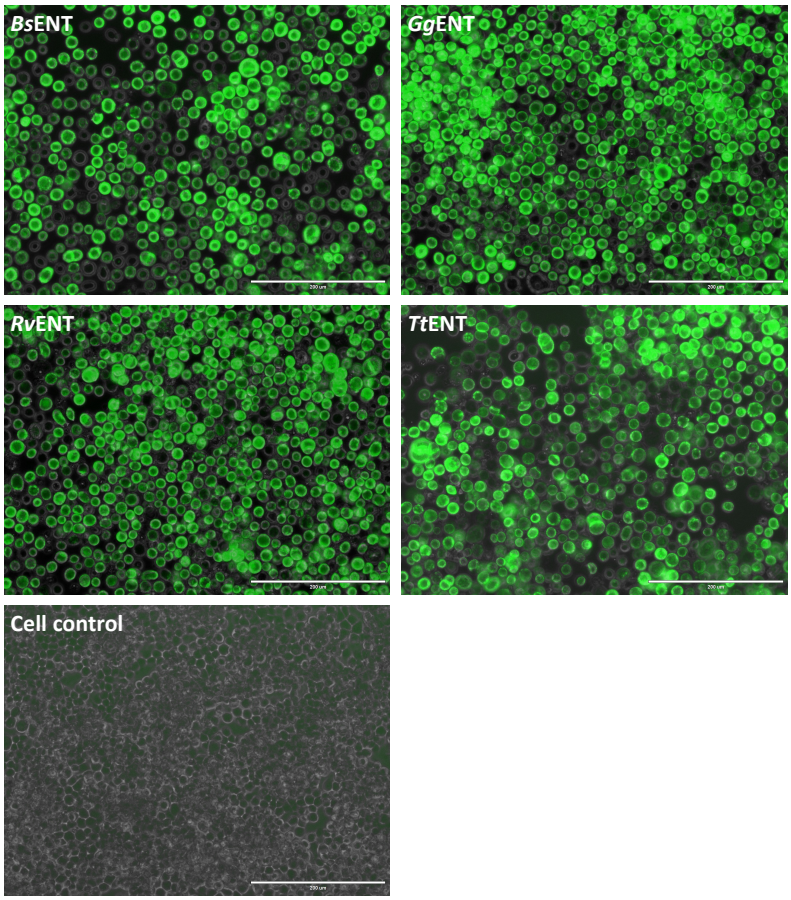


Figure 5.2 Sf9 cells expressing N-terminally GFP tagged homologues of hENT1.

40 x microscopy (EVOS) imaging with phase contrast and fluorescence overlay of Sf9 cells transfected with bacmids for hENT1 orthologues: *Bs*ENT, *Gg*ENT, *Rv*ENT and *Tt*ENT. Positive fluorescence confirms successful transfection of Sf9 cells with bacmid and construct expression. Furthermore, GFP presence in the plasma membrane confirms correct localisation. Cell control shows an un-transfected cell sample.

5.2.1.3.2 Expression testing

5.2.1.3.2.1 In gel presentation

As there were concerns about the quality and stability of hENT1 owing to the in-gel presentation (3.2.1.3), all homologues were analysed initially in SDS-PAGE to determine their in-gel presentation and compare that observed for each homologue to that typically observed for hENT1. Cells were harvested 120 hours post transfection and SDS-PAGE performed on whole cell samples, followed by fluorescence imaging and Western blot for the visualisation of N-terminal tags (Figure 5.3).

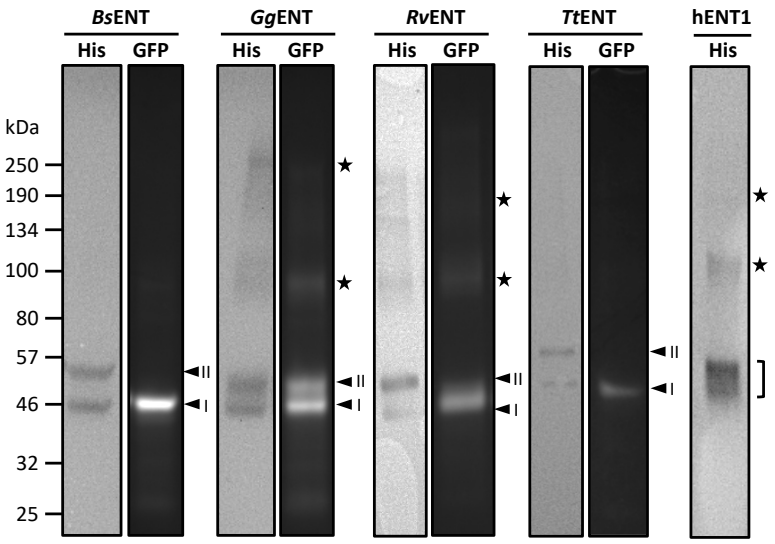


Figure 5.3 Expression tests on hENT1 homologues in Sf9 cells.

SDS-PAGE of whole Sf9 cells harvested 120 hours post transfection with anti-His-HRP Western blot and fluorescent imaging, for the visualisation of homologues with N-terminal His and GFP tags. An example image of hENT1 is included for reference. Bands that likely represent higher MW species are marked by a star where relevant. The monomeric broad band of hENT1 is marked by the bracket. The two species of monomeric ENTs are marked by an arrow, with 'I' indicating the lower MW species present in both fluorescent imaging and Western blot for all, and 'II' indicating the higher MW species that is present in all Western blots.

5.2.1.3.2.1.1 Double bands at the monomeric MW

All homologues present with two (presumed) monomeric bands in the anti-His Western blot, a lower band between 40-50 kDa, and an upper band between 50-70 kDa (Figure 5.3, bands I and II, respectively). Mixed MW monomers are also suspected in hENT1. However, the monomeric bands observed for the homologues have more definition than those observed in hENT1 (Figure 5.3, bracket). In hENT1 these 'mixed MW' monomeric bands are suspected to be aggregation, degradation, or sub-species. For all homologues, except *Gg*ENT, the bands at 50-70 kDa are absent from the fluorescent image (Figure 5.3, band II). As the His-tag cannot be intact without the GFP also being intact, this suggests that absence of fluorescent signal from the upper of these bands is due to signal quenching. As discussed in 1.7.2.1, GFP is a sensitive indicator of protein folding. Due to its stability in conditions encountered in protein expression, signal quenching is typically a result of protein denaturation[172]. Therefore, these higher MW monomeric bands are most likely misfolded protein. However, the retention of GFP signal in the higher MW band of the *Gg*ENT monomer suggests that the GFP is not denatured, and thus could instead be indicative of sub-species[146].

5.2.1.3.2.1.2 Higher MW bands

Further higher MW bands are present in both the anti-His Western blot and fluorescent image (Figure 5.3, stars), at ~90-100 kDa and ~250 kDa, which could indicate oligomeric species. The same pattern is observed in hENT1, and these higher MW/'oligomeric' species are presumed to be SDS-PAGE aggregates rather than true oligomers. Therefore, as the GFP signal for *Gg*ENT is present in both the higher MW monomer band and these presumed SDS-resistant aggregates (Figure 5.3, band II and stars), this suggests that although the *Gg*ENT-linked GFP is not denatured, the protein is likely aggregated. These presumed SDS-resistant aggregates are also seen in the *Rv*ENT Western blot and fluorescent image (Figure 5.3, stars). *Bs*ENT and *Tt*ENT both present with only monomeric species, and although both have two bands in the anti-His Western blot, the absence of the upper band in the fluorescent image suggests that the single lower MW monomeric bands (Figure 5.3, band I), represent stable, folded, and monomeric expression products.

5.2.1.3.3 Establishing a protocol for the large-scale expression of homologues of hENT1

Cell culture media that contained secreted recombinant baculovirus particles was harvested 72 hours post transfection (V0). This media was used to infect small scale suspension cultures for the purposes of increasing the viral titre (V1). Cells infected with V0 were observed for rate of proliferation and cell morphology. V1 was harvested 48 hours after the infected cells achieved proliferation arrest. Volumes of this higher viral titre media, V1, were then used to determine optimal volumes for the infection of cultures for expression. These optimal volumes were used to infect suspension cultures and to determine the optimal time for protein expression, and thus the target time point for harvesting the cells. The cell cultures for *Rv*ENT became contaminated during incubation and are consequently absent from this, and subsequent, analysis. While repeats of infection could have been performed to allow for comparison, as *Rv*ENT appears to aggregate it is not a favoured homologue. Therefore, repeat infections were not pursued. For *Bs*ENT, *Gg*ENT and *Tt*ENT, the infection of Sf9 suspension cell cultures at a density of 1 x10^6 cells/mL with V1 virus at a ratio of 1 in 4000 was optimal for expression of protein. As per 2.3.6, baculovirus infected insect cell (BIIC) stocks were prepared from these cultures for future large scale expression cultures.

As was observed in the initial in-gel analysis (Figure 5.3), *Bs*ENT has a single, defined, monomeric band at ~46 kDa in fluorescence imaging, with no other bands present (Figure 5.4A and B). However, a second monomeric band at ~55 kDa is present in the anti-His Western blot which is

presumed to be misfolded protein (Figure 5.4B). *Tt*ENT also shows a defined monomeric band at ~50 kDa in both gel images (Figure 5.4A and B), with a second misfolded band at ~60 kDa in the anti-His Western blot (Figure 5.4B). For *Tt*ENT, some lower MW bands are also observed in the fluorescence images for samples at +96 hours onwards, although these are absent from the anti-His Western blot, so are possibly other unrelated fluorescent proteins. However, lower MW bands are also present in both images for *Gg*ENT, predominantly in the +96 and +120 hours samples (Figure 5.4A and B). The presence of these bands in the anti-His Western blot suggest that the N-terminus remains intact. Therefore, these may represent C-terminally degraded protein. Again, as was observed in the initial in-gel analysis, *Gg*ENT also has two presumed monomeric bands, ~46 kDa and ~50 kDa, and further higher MW bands at ~100 kDa and 250 kDa in both fluorescent and anti-His Western blot (Figure 5.4A and B).

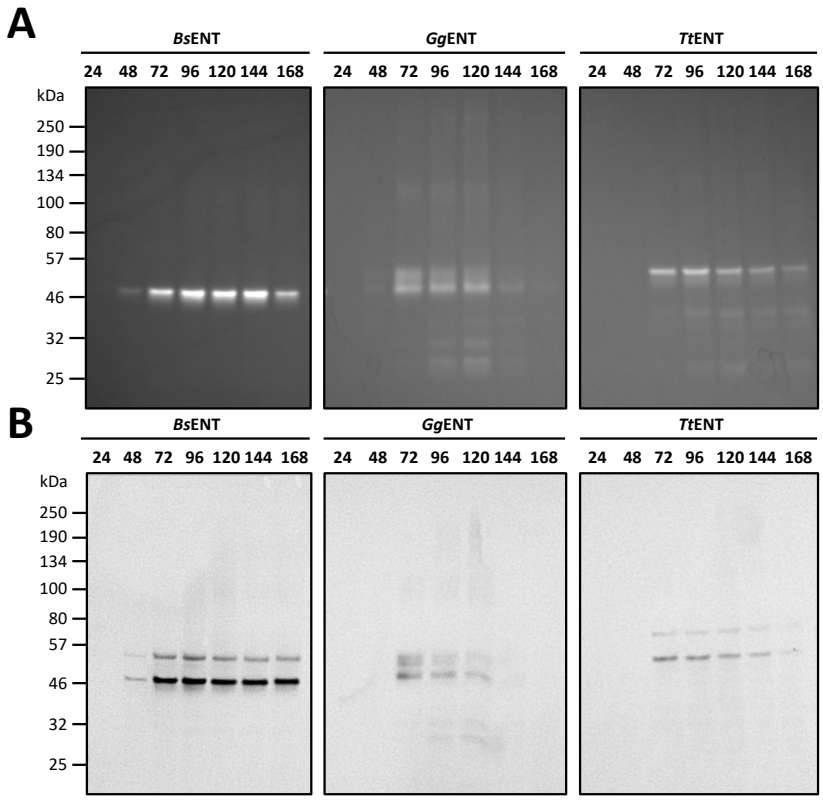


Figure 5.4 Determination the optimum expression of *Bs*ENT, *Gg*ENT and *Rv*ENT in Sf9 cells.

SDS-PAGE of whole Sf9 cell samples taken every 24 hours, from +24 to +168 hours, following a 1 in 4000 infection with V1 virus, with (A) fluorescent imaging followed by (B) anti-His Western blot.

SDS-PAGE analysis was performed with all samples prepared in a similar way (resuspended to the same OD_{600} and the same volume applied to gel). Therefore, the degree of signal observed in gel imaging suggests that there is a distinct difference in the level of expression between the homologues (Figure 5.4), with *Bs*ENT being the best expressed, and *Gg*ENT and *Tt*ENT being less well expressed. Furthermore, *Bs*ENT shows increasing levels of expression from +72 to +144 hours following infection with V1, and expression was still sustained at +168. However, In the +96 and +120 samples there appears to be some smearing in the Western blot, and there is possibly some prominent but diffuse banding observed at ~100 kDa. This may be an artefact of overloading or loading a whole cell sample on to the gel, or it could be small amounts of aggregation. In addition, for all homologues analysed, the cell morphology and health at 96 hours post-infection and later was very poor, and both *Gg*ENT and *Tt*ENT begin to show lower MW bands that could represent C-terminally degraded protein. Therefore, while +72 hours is maybe not the maximum expression for *Bs*ENT, to ensure cell health and protein quality was maintained, 72 hours post infection was determined as optimal incubation period for protein expression for all homologues.

5.2.1.4 Preliminary solubilisation screening

5.2.1.4.1 Selection of detergents for screening

While *Bs*ENT, *Gg*ENT and *Tt*ENT can be expressed and optimal expression in Sf9 has been established for the current constructs, expression is only one of the challenges in the study of membrane proteins, as discussed in 1.7. I performed literature searches and analysed the purification methods used in the structural characterisation of eukaryotic MFS transporters available to compile a list of detergents to be investigated for the solubilisation of each of the homologues. The most frequently used detergents are the maltosides DDM and DM[13,223], with the neopentyl glycol LMNG showing increasing popularity for MFS. LMNG has proven popular and successful with solved structures of GPCRs. Therefore, these detergents, along with some others from the classes, were explored for solubilisation of all homologues.

As extraction from the lipid bilayer is likely to be detrimental to membrane proteins, the reintroduction of lipids within a detergent system may prove beneficial for solubilised membrane protein function and stability (1.7.5). The reintroduction of lipids into protein-detergent complexes can be achieved though the inclusion of sterols, such as cholesterol hemi succinate (CHS) at solubilisation. The inclusion of CHS has previously contributed to thermostabilising hENT1 expressed in Sf9 cells, as detailed by Rehan & Jaakola[241]. Furthermore, detergents supplemented with CHS have proven increasingly successful for the solubilisation, purification and structural characterisation of numerous eukaryotic membrane proteins[167]. Therefore, I set out to screen detergents for the efficiency of solubilisation, with and without CHS supplementation. The zwitterionic detergent dodecyl-phosphocholine (FC12) was included as a positive control to approximate the upper solubility limit

5.2.1.4.2 In-gel presentations of solubilised protein

Following solubilisation and high-speed centrifugation for the removal of the insoluble fractions, SDS-PAGE was performed, and the gels imaged using anti-His Western blot. In the Western blot for *Bs*ENT all insoluble fractions (Figure 5.5, pellet) show the two bands that have been seen consistently in Western blot, ~46 and ~55 kDa. However, in all solubilised fractions (supernatant), the 55 kDa band is absent. Therefore, while fluorescent imaging of whole cell samples suggested that this band represents misfolded protein, this Western blot demonstrates that it is also insoluble. Therefore, solubilisation seemingly only extracts folded *Bs*ENT.

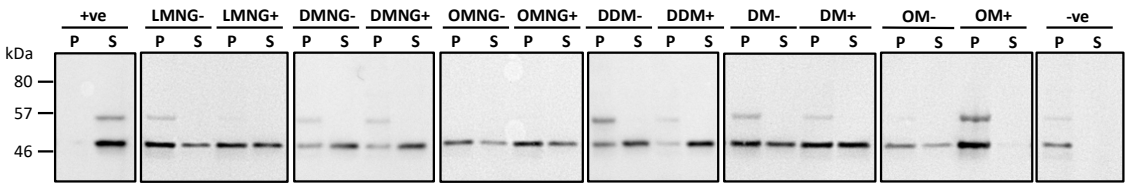


Figure 5.5 Detergent screen for the solubilisation of *Bs*ENT.

*Bs*ENT was solubilised with each of the detergents as detailed in Table 2, with (+) and without (-) CHS at 1 mg/mL. The negative control is cells that were incubated without detergent, and the positive control is cells that were incubated with the detergent Fos-choline 12. Fos-choline is a zwitterionic detergent that is highly successful at solubilising membrane proteins, but it also highly denaturing and is not a detergent that is suitable for purification.

Solubilisation efficiency was determined by relation of the density of the 46 kDa band in the supernatant to the same band in the pellet. This analysis determined that DDM and DMNG had the highest degree of solubilisation, with ~82% and 78% respectively (Figure 5.6). Detergent with CHS was shown consistently to have a reduced solubilisation efficiency relative to detergent only. Furthermore, the supplementation of n-octyl-β-D-maltopyranoside (OM) with CHS results in a near total loss of solubilisation. Cholesterol is the main sterol found in animals, but *Bs*ENT is fungal in origin, and the main sterol found in fungi is ergosterol. However, the structures only differ in the presence of additional double bonds on the B-ring and the acyl chain of ergosterol (Figure 5.7). Therefore, it may be that the reduced solubilisation efficiency observed is due to a lack of interactions between *Bs*ENT and sterols, whether CHS or ergosterol, and the inclusion of CHS simply reduces the availability of detergent for solubilisation. However, as the data presented here was only a single repeat, it may also be simply due to experimental variance. Additional repeats would allow me to determine whether this is a true reduction in solubilisation in the presence of CHS, or whether with and without CHS solubilise with a similar efficiency.

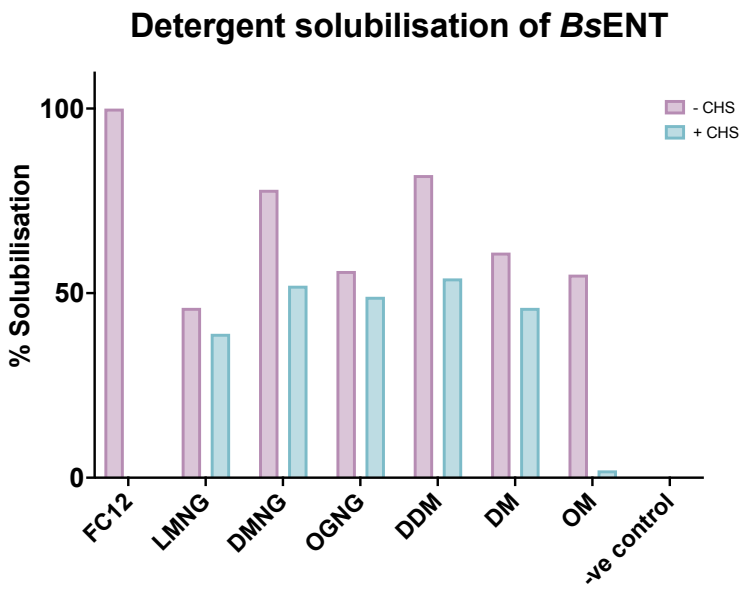


Figure 5.6 Efficiency of solubilisation of *Bs*ENT.

Densitometry analysis of the 46 kDa band in Western blot was used to relate the soluble fraction to the insoluble fraction for each condition. Data corresponds to a single repeat.

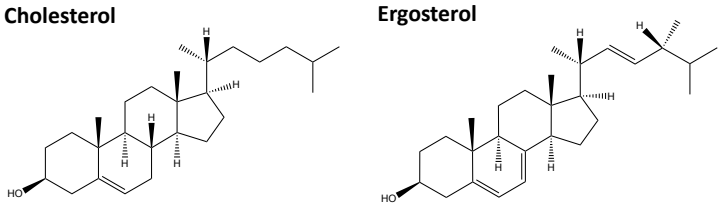


Figure 5.7 Chemical structures of sterols cholesterol and ergosterol.

Ergosterol (CAS: 57-87-4), the main sterol found in fungi differs from cholesterol (CAS: 57-88-5), the main sterol found in animals, by the presence of additional double bonds on the B-ring and on the acyl chain.

I attempted to investigate the solubilisation of the remaining homologues, *Gg*ENT, *Rv*ENT and *Tt*ENT. The experiments were to be repeated as performed for *Bs*ENT. However, as short chain detergents are typically more destabilising than their longer chain counterparts, and there are already concerns regarding the stability of the remaining homologues, the detergents octyl maltose neopentyl glycol (OGNG) and OM were excluded from analysis. In gel fluorescence analysis suggests that minimal solubilisation was achieved in all conditions for *Rv*ENT and *Tt*ENT. However, the sample of FC12 also suggests that no solubilisation was achieved in the positive control. It is possible that solubilisation is achieved in the positive control and other conditions, but protein has precipitated, and thus has not entered the gel. There is some solubilisation observed for *Gg*ENT. However, unlike *Bs*ENT, both bands for *Gg*ENT are solubilised. As both bands have been shown to retain GFP fluorescence, and thus may represent different sub-species rather than misfolded protein, the solubilisation of both bands may support this theory.

However, there is also a difference in the MW presentation of the upper band between the insoluble and soluble fraction, ~55 kDa and ~52 kDa, respectively, suggesting that solubilisation is causing some change in this upper band. Therefore, there are concerns about the quality of *Gg*ENT. In addition, in the FC12 positive control, there are no bands at the expected MW for the monomer, only a single band ~90-100 kDa. This suggests that, in FC12 at the very least, all solubilised protein has aggregated.

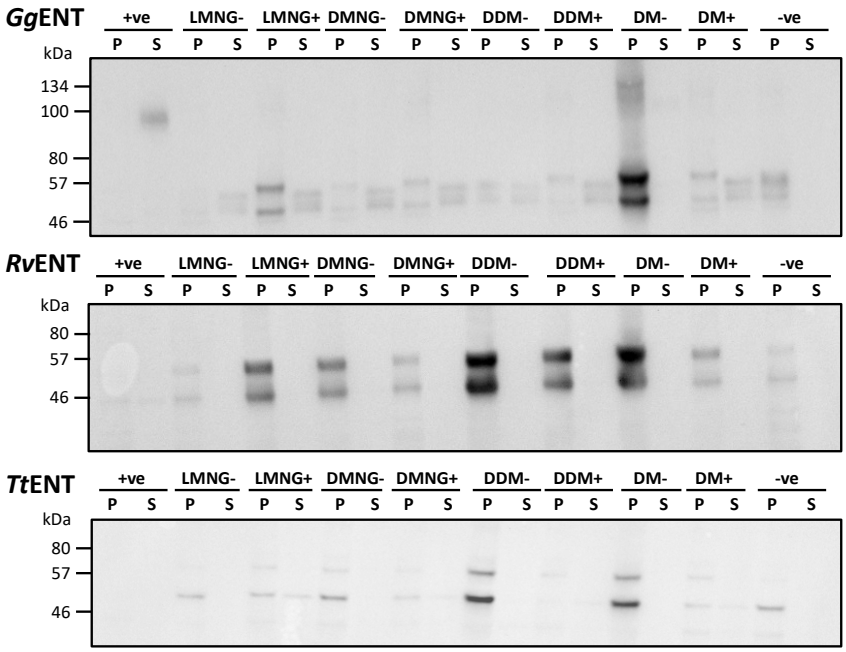


Figure 5.8 Detergent screen for the solubilisation of *Gg*ENT, *Rv*ENT and *Tt*ENT.

All homologues were solubilised with each of the detergents, with (+) and without (-) CHS at 1 mg/mL. The negative controls are cells that were incubated without detergent, and the positive controls are cells that were incubated with the detergent Fos-choline 12.

5.2.1.5 Summary of homologue screening

Owing to the consistent issues with presumed aggregation and precipitation, along with overall low expression, *Rv*ENT and *Tt*ENT were not explored any further. In addition, although it is well expressed, due to concerns over aggregation and the presence of a potential uncharacterised second sub-species at ~55 kDa, *Gg*ENT was also not explored any further. Conversely, the expression and stability of *Bs*ENT suggests that this homologue would be amenable to purification and hopefully, methods for structural characterisation. Not only does *Bs*ENT show more favourable characteristics during purification than the remaining homologues, but it also shows more favourable characteristics than hENT1 (Figure 5.3). Therefore, *Bs*ENT was taken forward for further optimisations and to explore large scale expression, purification, and characterisation.

5.2.2 Purification of *Bs*ENT

While DDM was determined to be the best detergent for solubilisation efficiency, this does not necessarily translate to this being the most stabilising detergent condition. Therefore, the thermostability of other conditions were analysed. As the inclusion of glycerol in the buffers for hENT1 reduced aggregation (3.2.4), and glycerol is known to stabilise proteins and reduce aggregation[247], I explored the influence of glycerol on thermostability. In addition, as *Bs*ENT is a putative homologue of hENT1, and hENT1 is thermostabilised by NBMPR (4.2.2.1.), I explored the influence of NBMPR on *Bs*ENT thermostability. The detergents OGNG and OM were excluded from thermostability analysis. While both detergents solubilised *Bs*ENT with better success than LMNG, they were not as successful as DDM, DM or DMNG. Furthermore, as short chain detergents are typically more destabilising than their longer chain counterparts, it was presumed that they would not perform as well. Therefore, in the interests of experimental efficiency, they were not explored any further.

5.2.2.1 Initial condition screening

Analysis of in-gel *Bs*ENT-linked GFP survival following a 10-temperature challenge (Figure 5.9A) determined that *Bs*ENT solubilised in 1% (w/v) DDM and hypertonic buffer (50 mM Tris-HCl pH 7.4, 3 mM DTT and protease inhibitor) has an apparent T_m of 58.4 ± 0.7 °C (Figure 5.9B). Conditions of interest were established to screen for the influence of glycerol and NBMPR (Figure 5.9C), and a single temperature challenge was performed. In each condition tested, surviving *Bs*ENT-linked GFP following an incubation at 60 °C was related to an on-ice control. The relative survival then allows for the identification of the most stabilising conditions. The single temperature challenge determined that DMNG without CHS, regardless of buffer conditions, is the most thermostabilising detergent. In all DMNG conditions without CHS, over 85% *Bs*ENT-linked GFP survived following the incubation at 60 °C. In all detergents, the conditions supplemented with NBMPR seem to show no increase to thermostability relative to the conditions without. Therefore, it is possible that NBMPR has low-to-no affinity for *Bs*ENT. In addition, supplementation with glycerol seems to have limited effect. However, for all detergents, the addition of CHS appears to be destabilising. This agrees with what I observed in the solubilisation efficiency, with CHS negatively affecting solubilisation. As discussed in 5.2.1.4.2, CHS may be a poor mimic of the fungal sterol, ergosterol. Therefore, *Bs*ENT may not interact with CHS, and CHS may be negatively affecting the availability of detergents, and thus may be destabilising the protein. All four DMNG conditions without CHS were taken forward and full 10-temperature challenges performed for the determination of specific condition T_m.

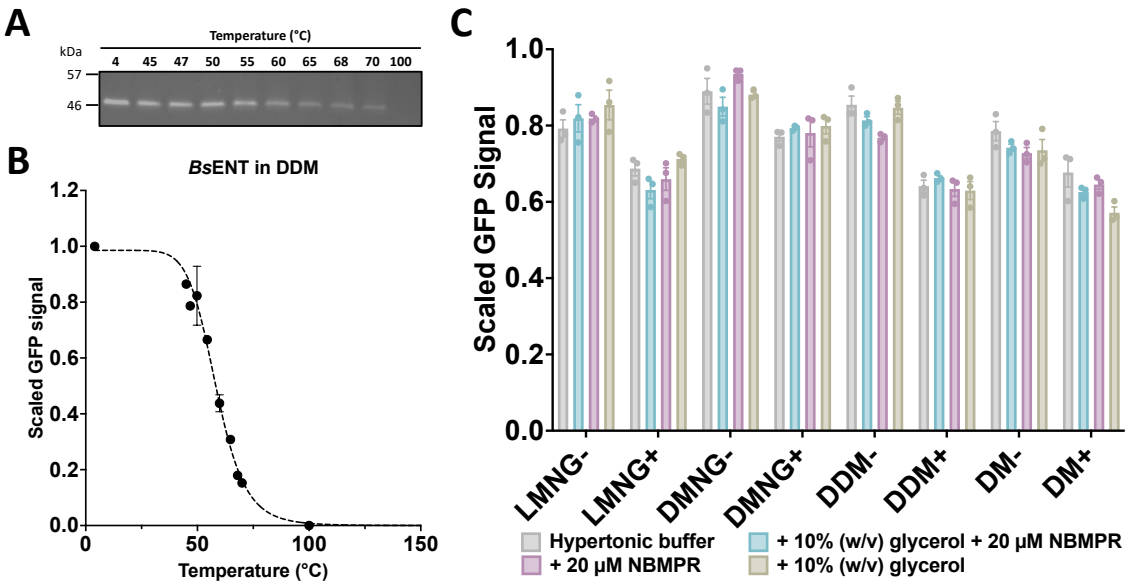


Figure 5.9 Preliminary detergent and buffer screen of *Bs*ENT.

(A) SDS-PAGE of *Bs*ENT following incubation at 10-temperature points, (4, 45, 47, 50, 55, 60, 65, 68, 70 and 75 °C) with fluorescence imaging for the visualisation of *Bs*ENT-linked GFP. (B) The intensity of the protein that remained in solution after the temperature challenge were quantified and data normalised to the 4 °C control. Data shown is the average of n = 3 experiments. Data were fit with a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting in GraphPad Prism 9.0. Error bars are representative of SEM. (C) Relative survival of *Bs*ENT-linked GFP signal following incubation at 60°C with a range of detergents with (+) and without (-) CHS. Data shown is the average of n = 3 experiments. Error bars represent SEM.

5.2.2.2 Optimal condition screening

Analysis of in-gel *Bs*ENT-linked GFP survival following a 10-temperature challenge determined that *Bs*ENT solubilised in 1% (w/v) DMNG and hypertonic buffer (50 mM Tris-HCl pH 7.4, 3 mM DTT and protease inhibitor) has a T_m of 67.6 ± 1.1 °C (Figure 5.10A and Figure 5.11). When supplemented with 20 µM NBMPR the apparent T_m is 70.4 ± 0.6 °C (Figure 5.10B); when supplemented with 20 µM NBMPR and 10% (w/v) glycerol the apparent T_m is 72.5 ± 1.4 °C (Figure 5.10C); and when supplemented with 10% (w/v) glycerol the apparent T_m is 75.7 ± 1.8 °C (Figure 5.10D).

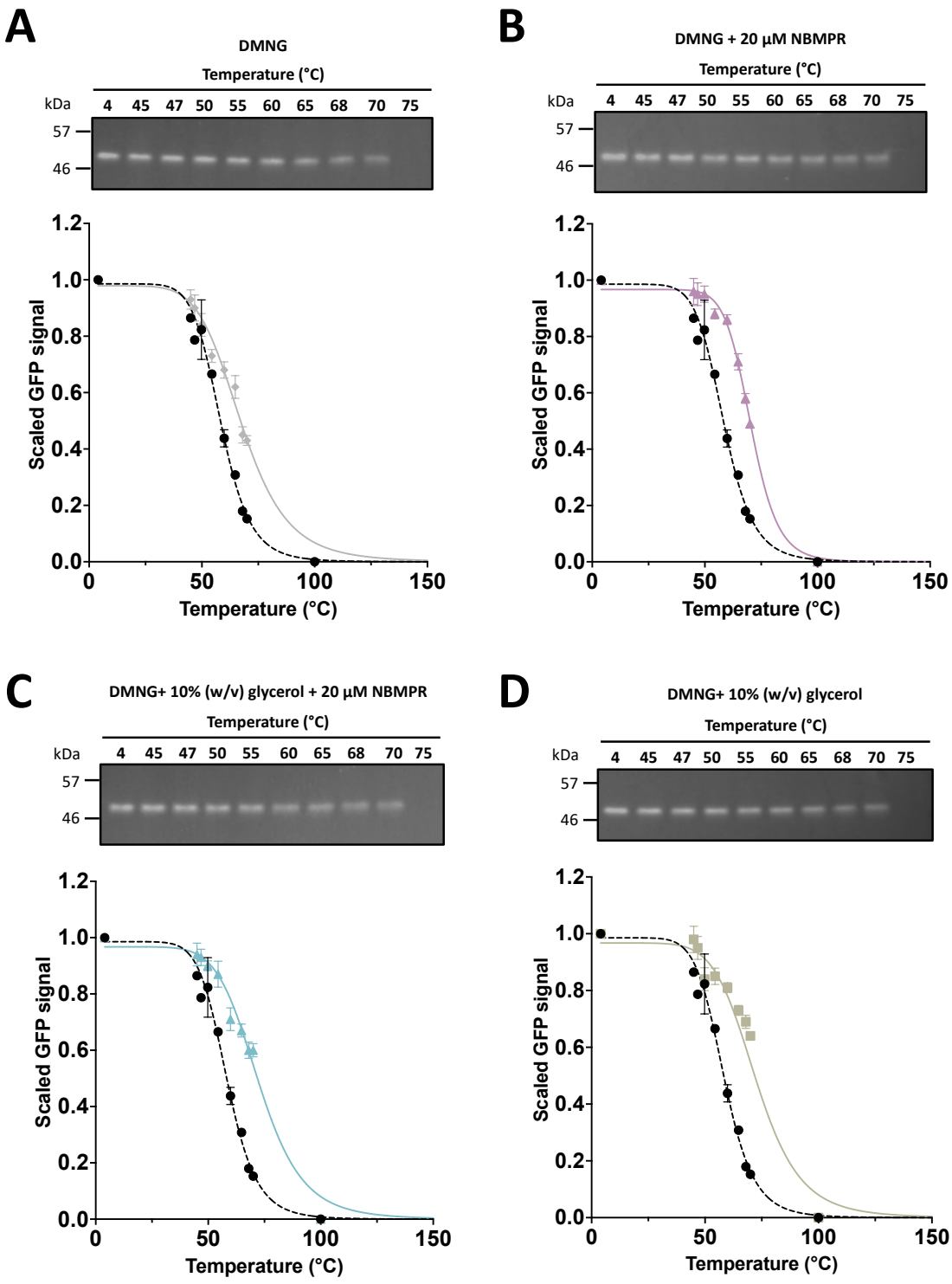


Figure 5.10 Ten-temperature stability curves of *Bs*ENT in differing solubilisation conditions.

SDS-PAGE of *Bs*ENT following incubation at 10-temperature points, (4, 45, 47, 50, 55, 60, 65, 68, 70 and 75 °C) with fluorescence imaging for the visualisation of *Bs*ENT-linked GFP. The intensity of the protein that remained in solution after the temperature challenge was quantified, and data normalised to the 4 °C control. Data shown is the average of n = 3 experiments. Data were fit with a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting in GraphPad Prism 9.0. Error bars represent SEM.

In this assay, T_m is defined as the temperature at which there is 50% of the *Bs*ENT-linked GFP signal remaining, relative to an on ice/4 °C control, with the data fit with a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting. However, the use of GFP as a fluorescent reporter for highly thermostable proteins has limitations that may contribute to inaccuracies in the data generated. If the reporter protein denatures at a temperature close to (or lower than) the target protein, then the loss of fluorescent signal cannot be assigned with any specificity. GFP itself has a T_m of 78 °C. Using this assay, all GFP signal is lost at temperatures 70-75 °C regardless of the condition, and thus is the upper limit of detection (Appendix Figure 6). A similar issue is discussed in the recent paper by Chatzikyriakidou *et al.*[299], where they found that in a different thermal shift assay, GFP is no longer fluorescent at temperatures >76 °C.

Furthermore, as GFP completely unfolds at 70-75°C, some degree of GFP specific unfolding and aggregation must also have already occurred at lower temperatures. In addition, an apparent high T_m also results in poor curve fitting. For example. all conditions with DMNG have 45-65% GFP signal surviving at 70 °C, before then dropping to 0% at 75 °C, and thus there is not a full distribution of data across the y-axis for the curve to be fit to. Therefore, for proteins with a high T_m, GFP directly contributes to inaccuracies in the determination of T_m at high temperatures, and the concerns with regards to specificity and accuracy of T_m would remain an issue regardless of detection methods used. The use of constructs free of GFP with analysis by alternate methods, such as DSF or CPM assay, would overcome these concerns about data quality. However, for the purposes of this investigation, considering the data in relative terms rather than as specific T_m values still provides insights into the overall thermostability of conditions. Therefore, ΔT_m is a more appropriate representation of the data than the specific T_m values (Figure 5.11).

In hypertonic buffer only, *Bs*ENT is significantly more thermostable in DMNG than in DDM (ΔT_m 9.2 ± 1.4 °C, p = 0.0035). When *Bs*ENT in DMNG is supplemented by 20 µM NBMPR this increases to ΔT_m 12.0 ± 0.9 °C (p = 0.0005). This is increased further to ΔT_m 14.1 ± 1.5 °C (p = 0.0001) when supplemented with 10% (w/v) glycerol in addition to 20 µM NBMPR. However, supplementing with 10% (w/v) glycerol alone is the most significantly stabilising with a ΔT_m of 17.3 ± 1.9 °C (p = <0.0001) (Figure 5.11). These results disagreed with my earlier findings, which suggested that *Bs*ENT may have low-to-no affinity for NBMPR, and thus was later investigated further (5.2.3.1).

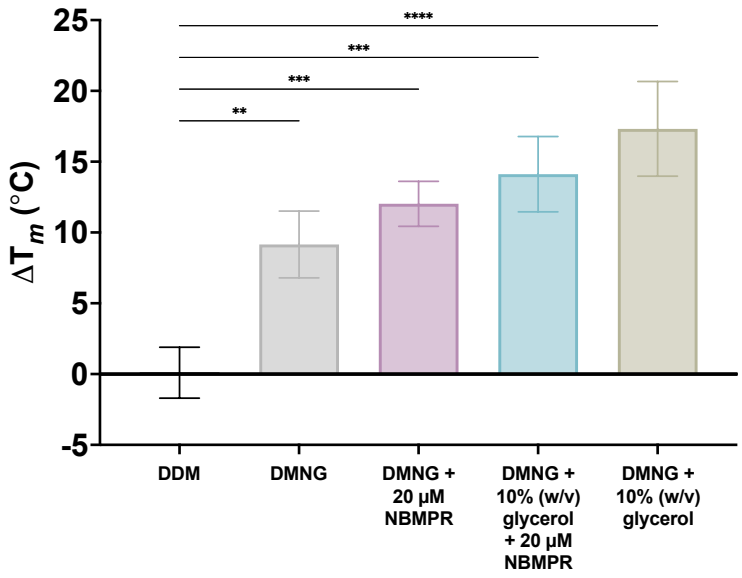


Figure 5.11 ΔT_m of NHGV-*Bs*ENT in differing conditions.

Δ T_m of *Bs*ENT, with each condition collected as an average of 3 repeats. Error bars are representative of error propagation as detailed in 2.3.2. Statistical analysis was performed using an ordinary one-way ANOVA with a follow up Dunnett's multiple comparisons test. The asterisks indicate the *p* value, with ** = <0.0021, *** = 0.0002 and **** = <0.0001.

5.2.2.3 Purification of *Bs*ENT from Sf9 cells

Large scale expression of NHGV-*Bs*ENT was achieved using BIIC infection of Sf9 cultures, and expression regularly achieved ~2.5 mg of *Bs*ENT/L of Sf9 culture at 1 x 10^6 cells/mL (as estimated from relative losses and final yields (Table 5.2)). The presence of the N-terminal GFP tag allows for fluorescent tracking of the protein throughout purification steps and informs the determination of purification efficiencies. Large scale solubilisation of *Bs*ENT in 1% (w/v) DMNG averages at 60-70%. While this is below the efficiency of solubilisation observed in the small-scale solubilisation trials (78%, Figure 5.6), this discrepancy could be attributed to differences in solubilisation conditions because of experimental scale up. Also, the initial solubilisation efficiency was determined using anti-His Western blot. All large-scale purification analysis uses in-gel fluorescence in place of Western blot, in addition to Coomassie staining. Therefore, there may also be differences in the relative amount of signal acquired.

Table 5.2 A comparison of the purification efficiency of hENT1 and *Bs*ENT, with estimated protein (mg/mL) per L of Sf9 culture.

	hENT1	***Bs*ENT**
Expressed in Sf9 (at 1.0 x 10^6 cells/mL) (mg/mL)	~1.0*	~2.5*
Solubilised (mg/mL)	0.35* (35%)	1.5 – 1.8* (60-70%)
NiNTA capture (mg/mL)	0.25* (70%)	1.0 – 1.4* (70-80%)
Digestion efficiency (mg/mL)	*undetermined*	1.0 – 1.3* (95%)
Eluted (mg/mL)	0.14 (55%)	0.5 – 0.8 (50-60%)
Heterogeneity	Low	Low

*concentration estimated by relation of the efficiency of each step (determined by relative in gel band intensity) to the final eluted concentration, as determined by $A280_{nm}$ and BCA assay

Overall, the two-step purification of *Bs*ENT performs well (Table 5.2, Figure 5.12). ~70% of solubilised protein binds to the resin and, as seen in the purification of hENT1, washing steps appear to remove proteins with non-specific or low-affinity binding to the Ni^{2+}-NTA resin, with no further proteins eluting in the final wash steps. HRV 3C protease digestion routinely achieves >95% efficiency, with digestion demonstrated by the shift in MW, with full length NHGV-*Bs*ENT at ~46 kDa and cleaved NHGV-tags at ~26 kDa, and as expected, no fluorescence is seen in the tag free *Bs*ENT at ~34 kDa. Following the elution of tag-free *Bs*ENT the resin is also washed with a high imidazole buffer, thus eluting any proteins with high affinity binding to the resin. The retention of all the undigested NHGV-*Bs*ENT on the resin following this wash suggests that this 'un-digestible' protein has precipitated and rendered the HRV 3C protease recognition sequence inaccessible. Following cleavage, *Bs*ENT can only be visualised in-gel by Coomassie staining. However, the efficiency of the washes and HRV 3C protease digestions means that *Bs*ENT elutes with a defined, homogeneous, and monomeric band at ~34 kDa, with average yields of 0.5 – 0.8 mg/L culture, as estimated by $A280_{nm}$ and BCA assay. This is a considerable improvement in both yield and protein quality in comparison to hENT1 (Table 5.2 and Figure 5.12B).

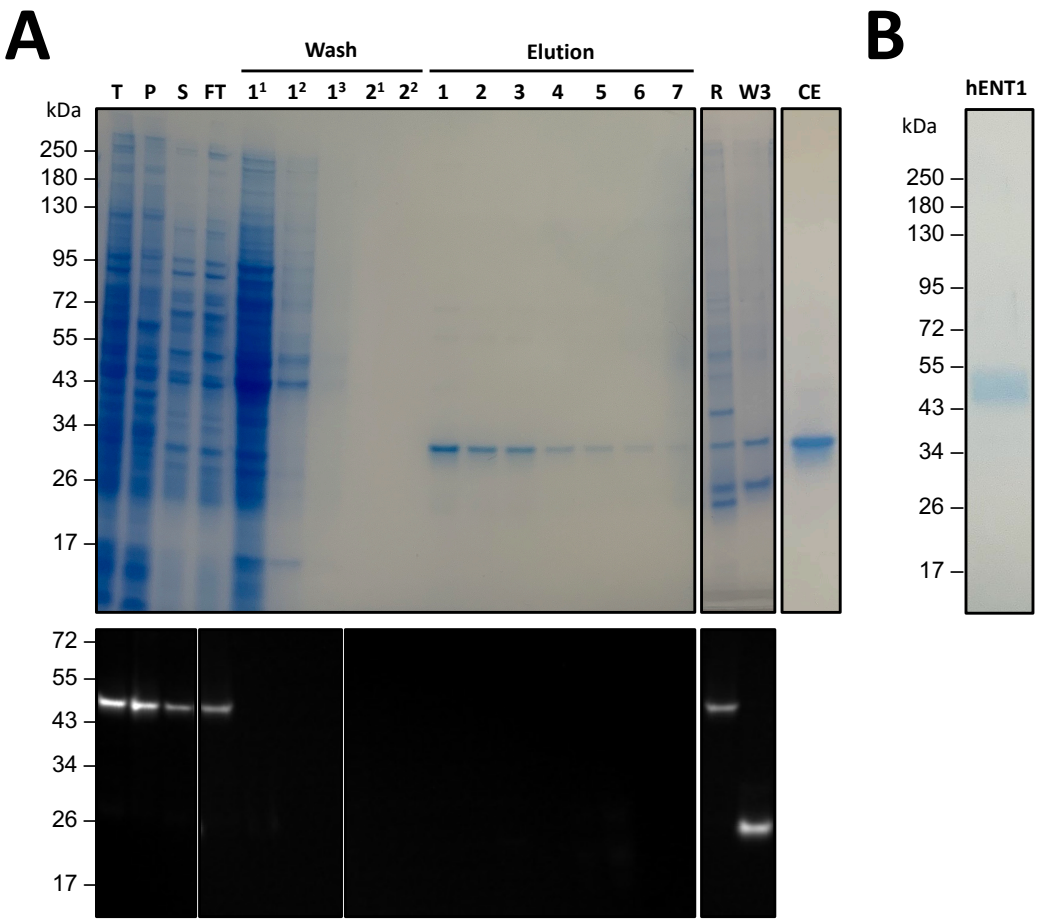


Figure 5.12 Representative purification of *Bs*ENT following a two-step purification protocol.

(A) SDS-PAGE of a two-step purification of *Bs*ENT with fluorescent imaging (bottom panel), followed by Coomassie staining (top panel). T = whole cell, total protein sample, P = pellet, S = supernatant, FT = Ni^{2+}-NTA resin flow though, wash 1 = high NaCl, low imidazole washing for elution of non-specific and low affinity binding, wash 2 = low NaCl, imidazole wash, for equilibration into protease digestion and elution buffer, elutions 1-7, R = resuspended resin following digestion elution and wash 3 (W3), W3 = high imidazole wash, CE = concentrated elutions (1-7). Fluorescent imaging demonstrates that NHGV-*Bs*ENT is not being eluted during the high salt and low imidazole, or low salt washes. HRV 3C digestion is demonstrated by the shift from 46 kDa to ~26 kDa. Some full length NHGV-*Bs*ENT remains on the resin. (B) Purified hENT1, as detailed in 3.2.3.2, as a reference.

While Ni^{2+}-NTA capture and HRV 3C digestion is performed with high efficiency, there are losses during the elution of tag-free *Bs*ENT with up to 50% of protein being lost on the resin. Structural predictions suggest that *Bs*ENT contains clusters of three and four histidine residues on the intracellular and extracellular surfaces, respectively (Figure 5.13). The residues that form these clusters are located on both the small intracellular/extracellular loops and the end of TMs. Therefore, it is possible that these residues were resulting in low-affinity binding to the resin. Furthermore, while <90% of this bound tag-free protein can be eluted with a high imidazole wash. it does so with the His_8 tag of the NHGV. I investigated SEC and dialysis for the recovery of the tag-free *Bs*ENT. However, owing to the similarity in the sizes of the NHGV and the tag-free *Bs*ENT, this was unsuccessful (data not shown). While the losses incurred throughout

purification highlight potential opportunities for purification optimisation, the final elution of highly homogenous *Bs*ENT with average yields of 0.5-1.0 mg /L culture are adequate for downstream processes.

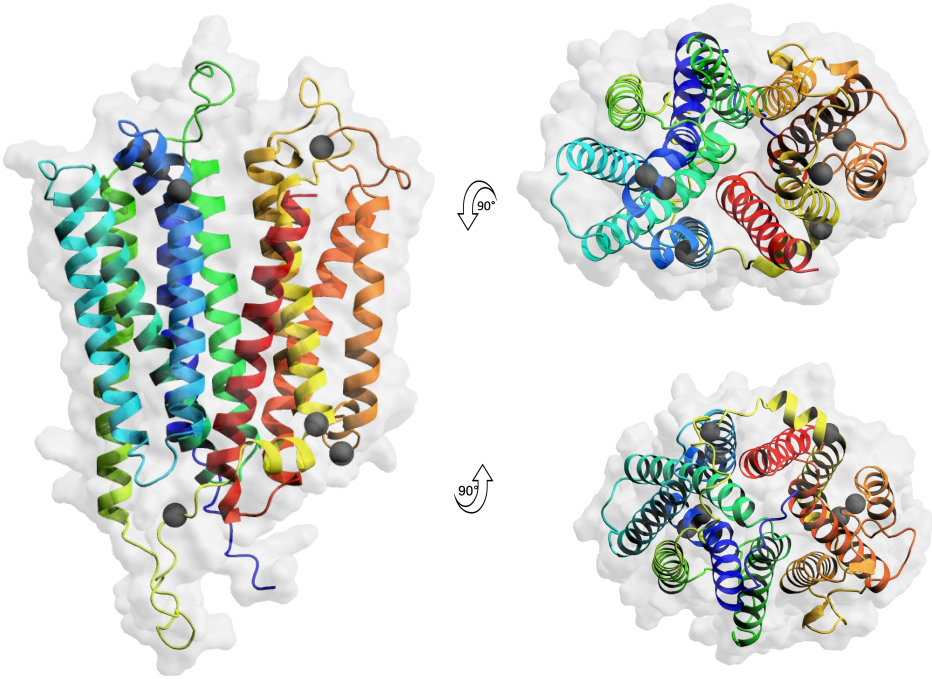


Figure 5.13 The locations of all histidine residues mapped to *Bs*ENT.

A structural prediction of *Bs*ENT was produced using AlphaFold[155]. Helices are shown with rainbow representation, with the N-terminal TM1 in blue and the C-terminal TM11 in red. All histidine's are shown as grey spheres.

5.2.2.4 Experimental troubleshooting

During the national lockdown that occurred because of the Covid-19 pandemic, the university also went into a temporary shutdown. Consequently, we had no access to the laboratories for nearly 6 months. Following the reopening of the laboratories, a significant number of difficulties were encountered, with previously successful and standardised experiments now failing or performing poorly. I observed this in the purification of *Bs*ENT, which suddenly began to incur significant losses due to increased protein precipitation at the point of elution. Furthermore, in some experiments simple attempts at solubilisation failed. After discussing the problems that I and others were encountering with Dr Postis, Dr Muench and Prof. Goldman, it was suspected that both the laboratory water supply and one of the two dishwashers for glassware cleaning were the sources of the problems. To confirm this, commercially available ultrapure water (ACROS Organics) was decanted into glassware off the shelf, shaken and set to stir. Significant frothing was observed (Figure 5.14A), and thus confirmed that detergent was still present in the

glassware following the wash cycle. In addition, I investigated the efficiency of solubilisation of NHGV-*Bs*ENT when using 25% (w/v) DDM stocks prepared with Milli-Q® H_2O obtained following the return to the lab after shutdown. I directly compared the results to those obtained by solubilising a biological replicate using 25% (w/v) DDM stocks prepared before the shutdown (Figure 5.14B). This data confirmed that the water quality was of concern and was likely the cause of protein precipitation.

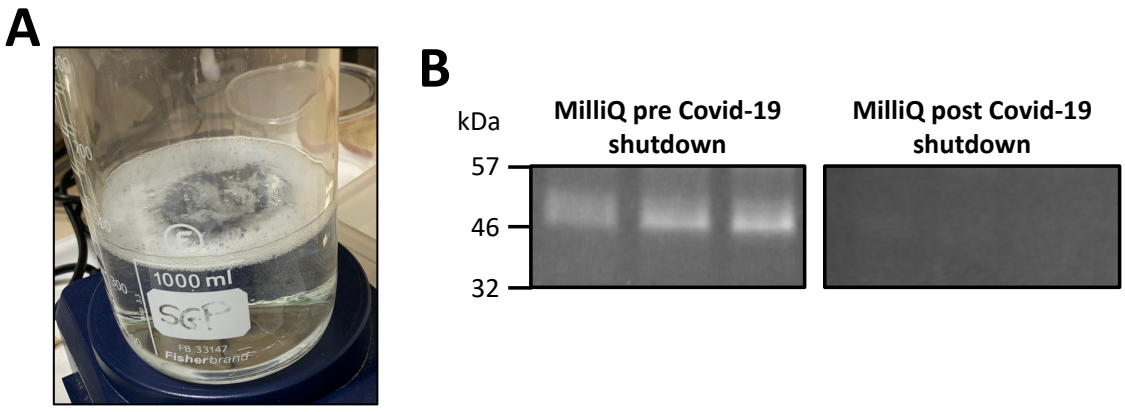


Figure 5.14 Investigations into dishwasher detergent contamination and water quality concerns following Covid-19 shutdown.

(A) **Commercial** ultrapure water in glassware off the shelf following shaking and stirring. (B) *Bs*ENT solubilised using DDM stocks that had been prepared with Milli-Q® water from before (left) and after (right) the Covid-19 shutdown.

For further and continued investigations into the effects of water quality and detergent contamination, small scale purifications of *Bs*ENT were performed as per 5.2.2.3, with the same cell pellet used for all conditions in each round of investigations. However, all reagents and buffers were prepared fresh with differing water sources and/or in different vessels, depending on the target of investigation, as detailed in Table 5.3. Therefore, aside from acids and bases used to pH buffers, the only variable in the condition was the water source or the vessel. Comparison of relative losses incurred during purification steps to the averages achieved before Covid-19, as detailed earlier in Table 5.2, allowed for continued monitoring of purification performance.

Losses incurred during these small-scale purifications were within the expected ranges (Table 5.2) at all steps except for protein elution. During the protein elution steps there were consistent issues with increased protein precipitation (Table 5.3). This was observed when using commercial ultrapure water in glassware, and when using Milli-Q® water in plasticware. Therefore, further confirming that both glassware and water independently were causing protein precipitation. The control condition of commercial water in plasticware consistently

performed as expected, with observed losses within the ranges of those seen previously. This confirms that the variation in the amount of precipitation is a true result of water/glassware quality, rather than experimental variation.

The experiments in March 2021 demonstrated that the concerns regarding water quality seemed to be specific to our location. The water obtained from Type I units from lab A and lab B, performed in line with the commercial ultrapure water. However, the water from our lab Type I unit performed poorly. Furthermore, the water from our lab Type I unit performed worse than the water obtained from Type II units from lab C and lab D. All Type I and Type II water systems were showing as functioning within the parameters and standards for those specific units (Table 5.3[a]), with no errors displayed. Furthermore, all units were within service periods and had recently had filters replaced. Therefore, while our Type I unit performed worse than a Type II unit experimentally, the unit was reportedly functioning appropriately, and thus whatever was contributing to the poor performance of the water was likely not a parameter that is measured as an ASTM standard for laboratory reagent water.

Table 5.3 Table of all conditions tested in small scale purifications of *Bs*ENT.

Date	Water source	Water standard[a]	Vessel	Target of investigation	Eluted	Precipitated
Pre Covid-19	Our laboratory MilliQ	Type I	Glassware	N/A, reference	50-60%	40-50%
Oct '20	Purchased Ultrapure	Type I	Plasticware	N/A, control	54%	46%
	Purchased Ultrapure	Type I	Glassware	Glassware	26%	74%
	Our laboratory MilliQ	Type I	Glassware	Glassware and water	15%	85%
Jan '21	Purchased Ultrapure	Type I	Plasticware	N/A, control	53%	47%
	Purchased Ultrapure	Type I	Glassware	Glassware	33%	67%
	Our laboratory MilliQ	Type I	Plasticware	Water	31%	69%
Mar '21	Purchased Ultrapure	Type I	Plasticware	N/A, control	52%	48%
	Our laboratory MilliQ	Type I	Plasticware	Water	35%	65%
	Alternate laboratory, A	Type I	Plasticware	Water	51%	49%
	Alternate laboratory, B	Type I	Plasticware	Water	52%	48%
	Alternate laboratory, C	Type II	Plasticware	Water	42%	58%
	Alternate laboratory, D	Type II	Plasticware	Water	40%	60%

a ASTM Standards for Laboratory Reagent Water (ASTM D1193-91)
Type I: resistivity >18 MΩ-cm, conductivity < 0.056 µS/cm, total organic carbon < 50 µg/L, Na < 1 µg/L, Cl < 1 µg/L, silica < 3 µg/L, pH at 25 °C N/A
Type II: resistivity >1 MΩ-cm, conductivity < 1 µS/cm, total organic carbon < 50 µg/L, sodium < 5 µg/L, chloride < 5 µg/L, silica < 3 µg/L, pH at 25 °C N/A
Type III: resistivity >4 MΩ-cm, conductivity < 0.25 µS/cm, total organic carbon < 200 µg/L, sodium < 10 µg/L, chloride < 10 µg/L, silica < 3 µg/L, pH at 25 °C N/A
Type IV: resistivity >0.2 MΩ-cm, conductivity < 5.0 µS/cm, total organic carbon N/A, sodium < 50 µg/L, chloride < 50 µg/L, silica N/A, pH at 25 °C 5.0 – 8.0

The issues regarding the dishwasher were resolved in February 2021 as the faulty unit was replaced. However, the issues regarding water quality are ongoing. No further small-scale experiments have been performed by me, but in September 2021 faculty wide updates to the water system were put in place. During this work further issues specific to our location were identified, and complaints about unusual or inexplicable experimental performance continue.

As a result of this work, any experiments that I performed from October 2020 onwards were done so using commercial ultrapure water and plasticware.

5.2.2.5 Expression of *Bs*ENT in yeast

As *Bs*ENT originates from a fungus, in an effort to further scale up the expression and do to so in a more native-like and cost-efficient system, I explored expression of *Bs*ENT in yeast, in both *S. cerevisiae* and *P. pastoris*.

5.2.2.5.1 S. cerevisiae

*Bs*ENT from the pFastBac-NHGV-*Bs*ENT was amplified and cloned into a pDDGFP2 expression vector. The recombinant plasmid was transformed into two different strains of *S. cerevisiae,* BJ1991 and FGY217. The pDDGFP2 yeast expression plasmid features an N-terminal His_6-tag, super folder-GFP (sfGFP) and TEV protease recognition sequence, to produce a N-terminally tagged His_6-sfGFP-TEV-*Bs*ENT (NHsfGT-*Bs*ENT), with expression under the inducible galactose promoter. Positive clones were confirmed using selection plates and suspension cultures were grown overnight with selection medium with 2% glucose. Stationary phase was consistently achieved in all overnight cultures. Therefore, to begin expression overnight cultures were diluted to an OD_{600} = 0.1 in selection medium with 0.1% glucose and grown to OD_{600} = 0.6 before being induced by the addition of 2% galactose.

Following induction cell densities were monitored until 22 hours post induction. Owing to the N-terminal GFP, cell fluorescence was also monitored, and whole cell samples imaged using phase and fluorescent microscopy. Induction with 2% galactose of the transformed cells (both BJ1991 and FGY217) at OD_{600} = 0.6 appeared to cause proliferation arrest, with minimal growth observed following induction. As cell growth in glucose media was as expected, it appeared to be the induction of cells that was causing issues. Therefore, I explored galactose induction at differing densities ranging from OD_{600} =0.2 to 1.0. However, did this not improve growth. In a period of 22 hours cell densities never surpassed OD_{600} = 6.0, nor was a plateau reached. GFP expression did appear to achieve a plateau (Figure 5.15). However, the signal overall was very low, and whole cell visualisation suggests that GFP which was expressed was likely in vacuoles[300] and not localised to membrane, vacuolar or plasma[300], and thus was not correctly localised (Figure 5.16). There was no difference observed between cell growth, GFP expression or localisation between BJ1991 and FGY217.

Owing to the plateau of GFP expression observed, this suggests that maximum possible expression was achieved. The upward trend of OD suggests that the cells achieve exponential growth following induction, without a lag phase, but did not reach a stable stationary phase. However, the rate of growth was extremely slow. As cells grow well in the uninduced sample and it was only post induction that they appeared to slow down, this suggests that it was the expression that was the cause of slow growth. GFP imaging of uninduced cell samples confirmed that there was no leaky expression. Therefore, it is possible that expression of NHsfGT-*Bs*ENT is toxic to *S. cerevisiae* and causes cellular arrest.

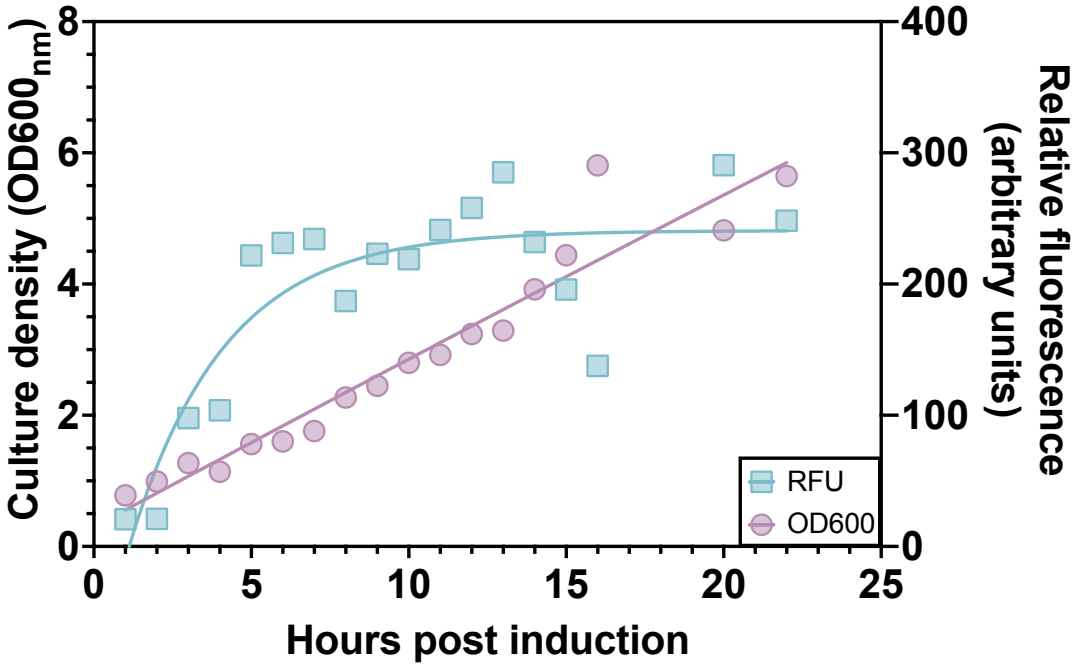


Figure 5.15 Growth and fluorescence curves of induced *S. cerevisiae*.

Growth of *S. cerevisiae* (FGY217) following galactose induction for the expression of NHsfGT-*Bs*ENT was monitored up until 21 hours post induction. Relative fluorescence of cell cultures is normalised to cell densities. Data are fit with a nonlinear growth curve with an exponential plateau. Data are representative of a single measurement.

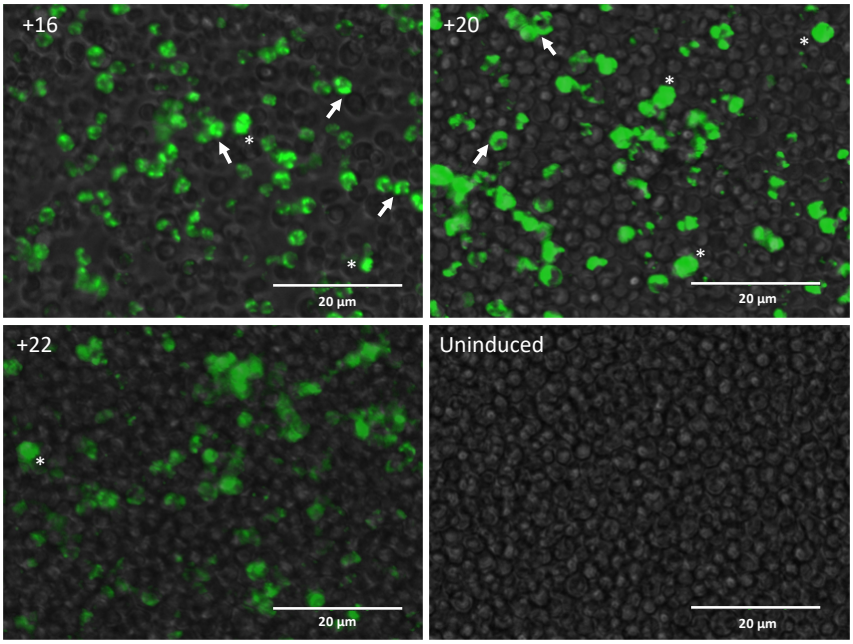


Figure 5.16 *S. cerevisiae* expression of NHsfGT-*Bs*ENT.

40x fluorescent and phase overlay of *S. cerevisiae* following induction with 2% galactose, at +16, +20 and +22 post induction. Uninduced cells are included as a negative control. GFP signal suggests that at +16 protein is being expressed, but it is not localised to the plasma membranes. Instead, it is likely in vacuoles (arrows). By +20 there are large amounts of GFP accumulated within the cytosol (*)

5.2.2.5.2 P. pastoris

NHGV-*Bs*ENT from the pFastBac-NHGV-*Bs*ENT was amplified and subcloned into a pPICZB expression vector, for the expression under the alcohol oxidase I promoter. The recombinant expression vector was then transformed into two different strains of *P. pastoris,* SMD1163 and SuperMan5. Both strains feature a His-selectable marker and disrupted genes for proteinase A and B. However, SuperMan5 also co-overexpresses several glycosyltransferases to produce glycoproteins with complex glycosylation. While this may not be necessary for *Bs*ENT, SuperMan5 was included out of interest. Positive clones were confirmed using selection plates and suspension cultures were grown to saturation. Expression was induced by the addition of 0.5% methanol and cells were monitored for GFP expression up until 12 hours post induction. Both SMD1163 and SuperMan5 produced several clones with high levels of GFP expression, with optimal expression at 7 hours post induction. However, using confocal microscopy for visualisation of localisation it appeared that GFP was contained within vacuoles and not localised to membranes[300], and thus was not correctly localised (Figure 5.17). There was no difference observed between GFP expression or localisation between SMD1163 and SuperMan5.

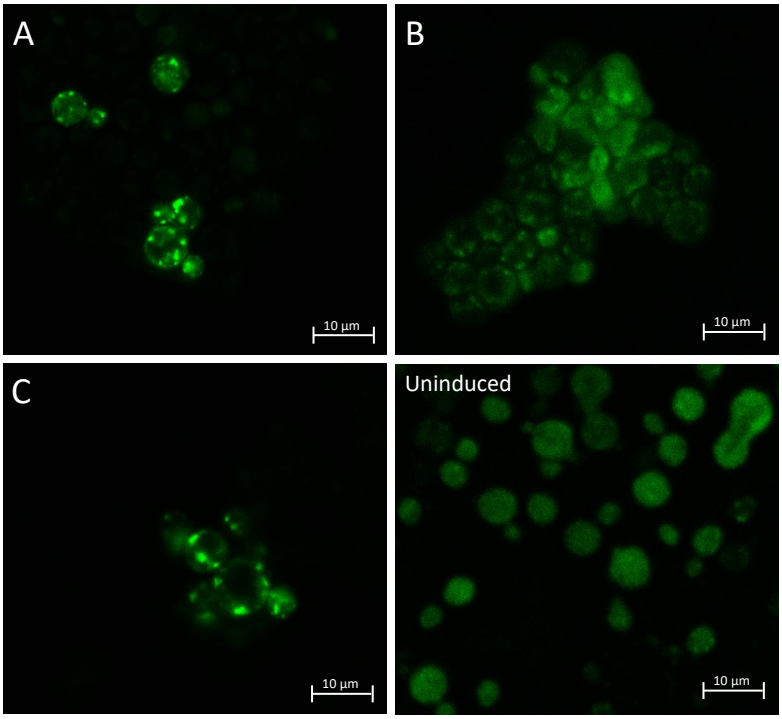


Figure 5.17 *P. pastoris* expression of NHGV-*Bs*ENT.

Panels A-C show confocal imaging of *P. pastoris* at 7 hours post-induction with 0.5% methanol. Panel D shows an uninduced cell control. The cells in panels (A) and (C) have some areas of strong GFP signal that suggests some GFP expression above that of the uninduced control. However, as with the expression in *S. cerevisiae* (Figure 5.16) all signal appears to be localised in vacuoles. The cells in panel (B) may possibly show diffuse cytosolic signal. However, this is minimal. Furthermore, both cytosolic and vacuolar localisation of GFP would indicate that expressed protein is not being localised to the membrane.

5.2.2.5.3 Summary of yeast expression of *Bs*ENT

Expression of *Bs*ENT in both *S. cerevisiae* and *P. pastoris* was unsuccessful. While expression in *P. pastoris* did not encounter the same issues with cell growth following induction that expression in *S. cerevisiae* did, both strains in both species encountered the same issues with incorrect localisation. In all conditions GFP signal was predominantly contained within the cytosol or the endoplasmic reticulum, thus was incorrectly localised. In all yeast expression the *Bs*ENT gene sequence had previously been codon optimised for expression in Sf9 cells. Therefore, it is possible that the issues encountered with expression in yeast were due to incompatible codon usage. Codon bias in *P. pastoris* has been shown to modulate transcription and translation efficiency, and protein folding and activity, with gene optimisation contributing to significantly higher yields of protein[164,301-303] . Therefore, expression in yeast would likely benefit greatly from codon optimisation. However, in the interest of time and as I was able to obtain high yields of homogenous *Bs*ENT from expression in Sf9 cells, I did not undertake any further explorations into expression in yeast.

5.2.3 Functional characterisation of *Bs*ENT

During solubilisation screening for *Bs*ENT it appeared as though for most of the detergent conditions tested using the single temperature challenge, the supplementation with 20 µM NBMPR provided no increase to thermostability (Figure 5.9C). Thus, suggesting that *Bs*ENT has low-to affinity for NBMPR. However, in the ten-temperature challenge for DMNG there were disagreements as to whether NBMPR provides any stabilisation. As discussed in 5.2.2.2, owing to concerns regarding the specificity of loss of signal in the assay, these apparent T_m are anyway, likely inaccurate. Therefore, I set out to explore the relationship between NBMPR and *Bs*ENT. This work pre-dates the publication of the structures of hENT1 with NBMPR bound by Wright & Lee[134]. Thus, the molecular basis by which NBMPR specifically inhibits hENT1 was unknown. Numerous publications have detailed residues that are critical for substrate selectivity and inhibitor sensitivity by hENT1. Of these residues, M89[138] (TM2) and G154[144] (TM4) appeared to be critical for conferring sensitivity to NBMPR, with the NBMPR insensitive hENT2 featuring a polar residue at each equivalent position (Figure 5.18). Sequence alignments suggest that while *Bs*ENT features a hydrophobic valine at the position equivalent to M89, like hENT2 it also features a polar residue at the position equivalent to G154. Therefore, I expected *Bs*ENT to have reduced affinity for NBMPR, relative to hENT1.

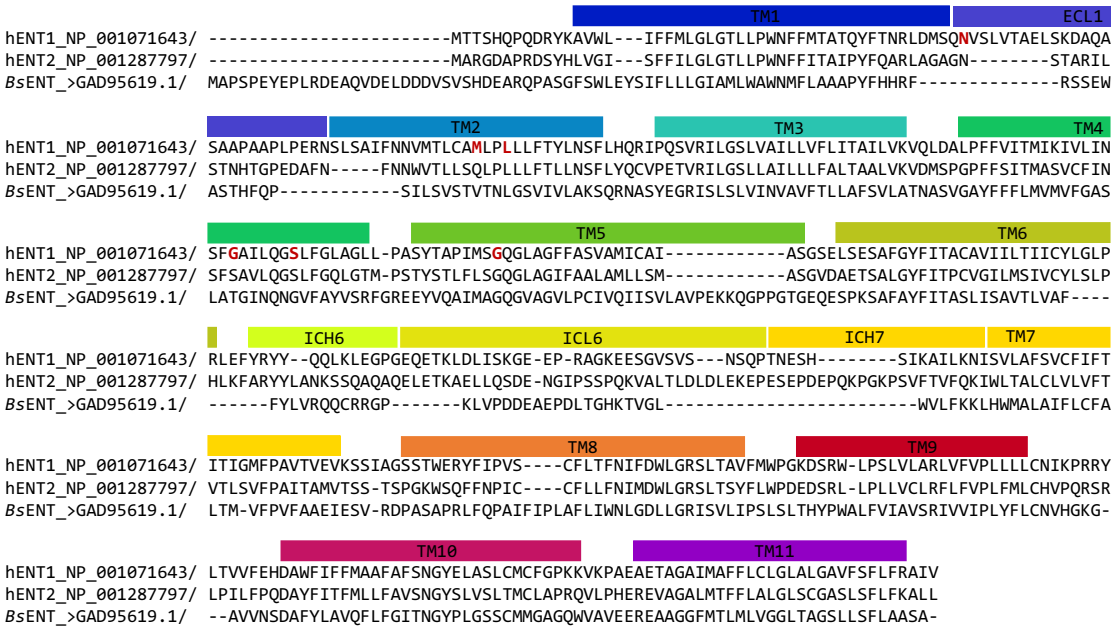


Figure 5.18 Sequence alignment of hENT1, hENT2 and *Bs*ENT.

Residues of hENT1 that have been shown to confer NMBPR sensitivity are highlighted in red[138,139,144,304,305]. The sequence for NBMPR sensitive hENT1 is aligned with that of the NBMPR insensitive hENT2, and *Bs*ENT. Topological predictions of helices are assigned with a rainbow schematic.

5.2.3.1 [3H]-NBMPR binding

I began my investigations into *Bs*ENT binding of NBMPR by performing a single concentration binding assay. Using a concentration of [3H]-NBMPR (61 nM) that achieves near equimolar binding in hENT1 (0.85 ± 0.10), I determined that *Bs*ENT binds 0.07 ± 0.01 pmol of [3H]-NBMPR per pmol of protein (Figure 5.19A). While this value is very low, it is above that of the negative control so could suggest there is some affinity. It may be that affinity of *Bs*ENT for NBMPR is low, and thus the concentration used in the single concentration assay was too low to see binding. Therefore, I then went on to perform a saturation binding assay to investigate whether B_{max} can be achieved, and if so, then to use the determined B_{max} and K_d values to determine a concentration of [3H]-NBMPR for use in future single concentration binding assays. However, at concentrations up to 1 µM, saturation was not achieved and B_{max} and K_d were unable to be determined (Figure 5.19B). Therefore, this suggests that *Bs*ENT is unable to bind NBMPR.

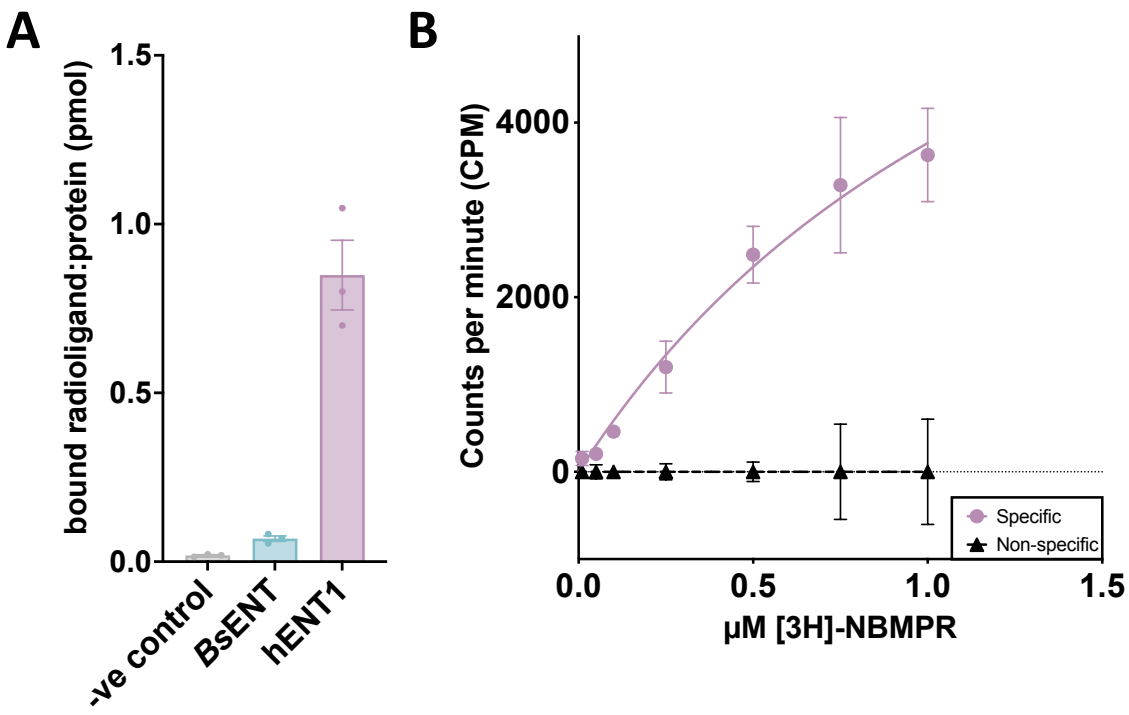


Figure 5.19 [3H]-NBMPR radioligand binding assays of *Bs*ENT.

(A) single concentration (61 nM) and (B) saturation binding assay for *Bs*ENT with [3H]-NBMPR. Saturation binding assay data are fit with a one site-specific binding curve, with the deduction of non-specific values for a baseline correction of data. Error bars are representative of SEM, n = 3.

As the structures of hENT1 with NBMPR bound have since been published this has allowed for modelling of a structure of *Bs*ENT using AlphaFold[155]. In addition, as the molecular basis of NBMPR binding and hENT1 specificity have now been detailed, analysis of the structure of *Bs*ENT

shows that it lacks the deep hydrophobic pocket that facilitates the binding of the *p*-nitrobenzyl moiety of NBMPR, as detailed in 1.6.3. Like the NBMPR insensitive hENT2, *Bs*ENT features a polar residue at the G154 equivalent position, T155, that likely occupies this pocket (Figure 5.20).

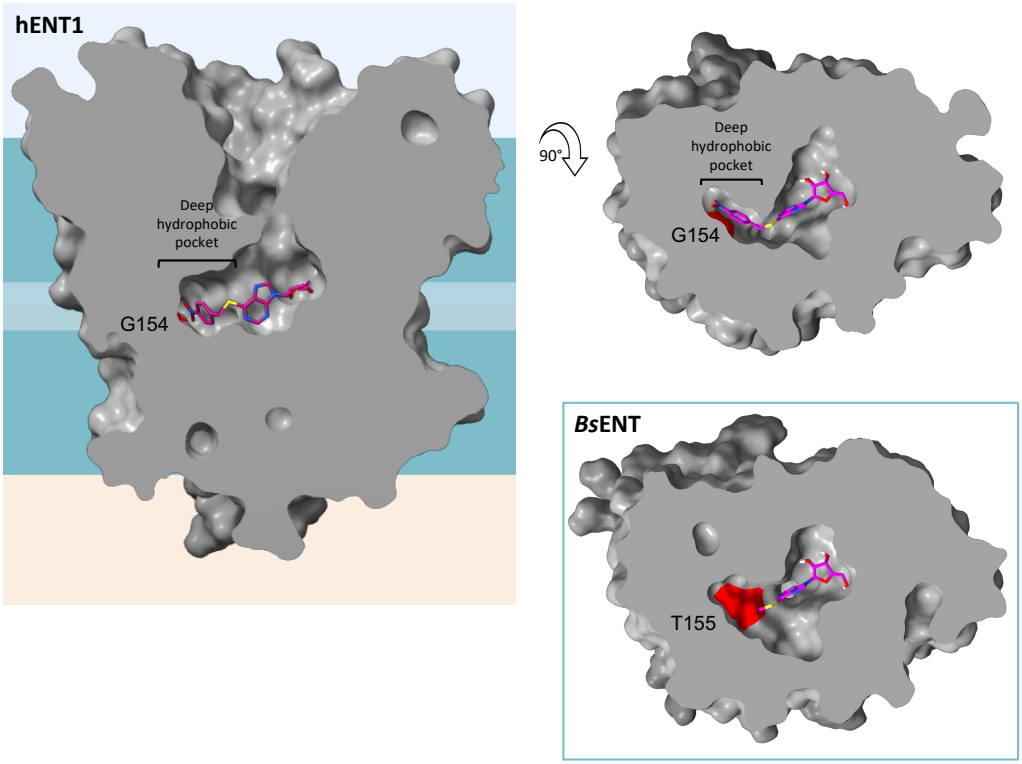


Figure 5.20 The hydrophobic pocket of hENT1 and NBMPR binding.

A perpendicular to the membrane, and top down, clipped view of the deep hydrophobic pocket of hENT1 with NBMPR bound (PDB: 6OB6)[134], with G154 highlighted in red. A structural prediction of *Bs*ENT using AlphaFold[155] is highlighted with the blue square. PDB: 6OB6 is aligned to the *Bs*ENT prediction and the NBMPR is shown in pink. The density of the residue at the position equivalent to G154, T155 is shown in red. The density of T155 occupies the space that is occupied by the *p*-nitrobenzyl moiety of the NBMPR in hENT1.

5.2.3.2 Identifying endogenous substrates

As *Bs*ENT has little or no affinity for NBMPR, I set out to identify endogenous substrates of *Bs*ENT. This would then be used to inform future transport assays and allow for the functional characterisation of *Bs*ENT. *Bs*ENT is 26% identical to hENT1 (Table 5.1), but it is also 24-25% identical to the remaining hENT isoforms. As each of the isoforms differ in their substrate specificity, including differing affinities for nucleobases as well as nucleosides, I set out to perform an initial narrow screen of purine and pyrimidine nucleosides and nucleobases.

I decided to use the in-gel ten-temperature GFP survival assay to screen the compounds. As was seen for hENT1 and NBMPR and has been demonstrated in a number of other thermal shift

assays for the identification of ligands of MFS and SLC members[299,306], binding interactions are expected to translate to thermostabilising effects. However, as discussed in 5.2.2.2, when using the most thermostabilising solubilisation conditions for *Bs*ENT, GFP survival cannot be accurately determined at high temperatures. Therefore, to reduce the likelihood of reaching the upper limit of the assay, the assay was performed in a lesser thermostabilising condition (resuspended in 1 x PBS and 1 x protease inhibitor and solubilised in 1% (w/v) DDM) (2.8.7)). This allowed for the best fit of the data (Figure 5.21), and as the data is analysed in relative terms (ΔT_m) this was adequate for the purposes of this investigation.

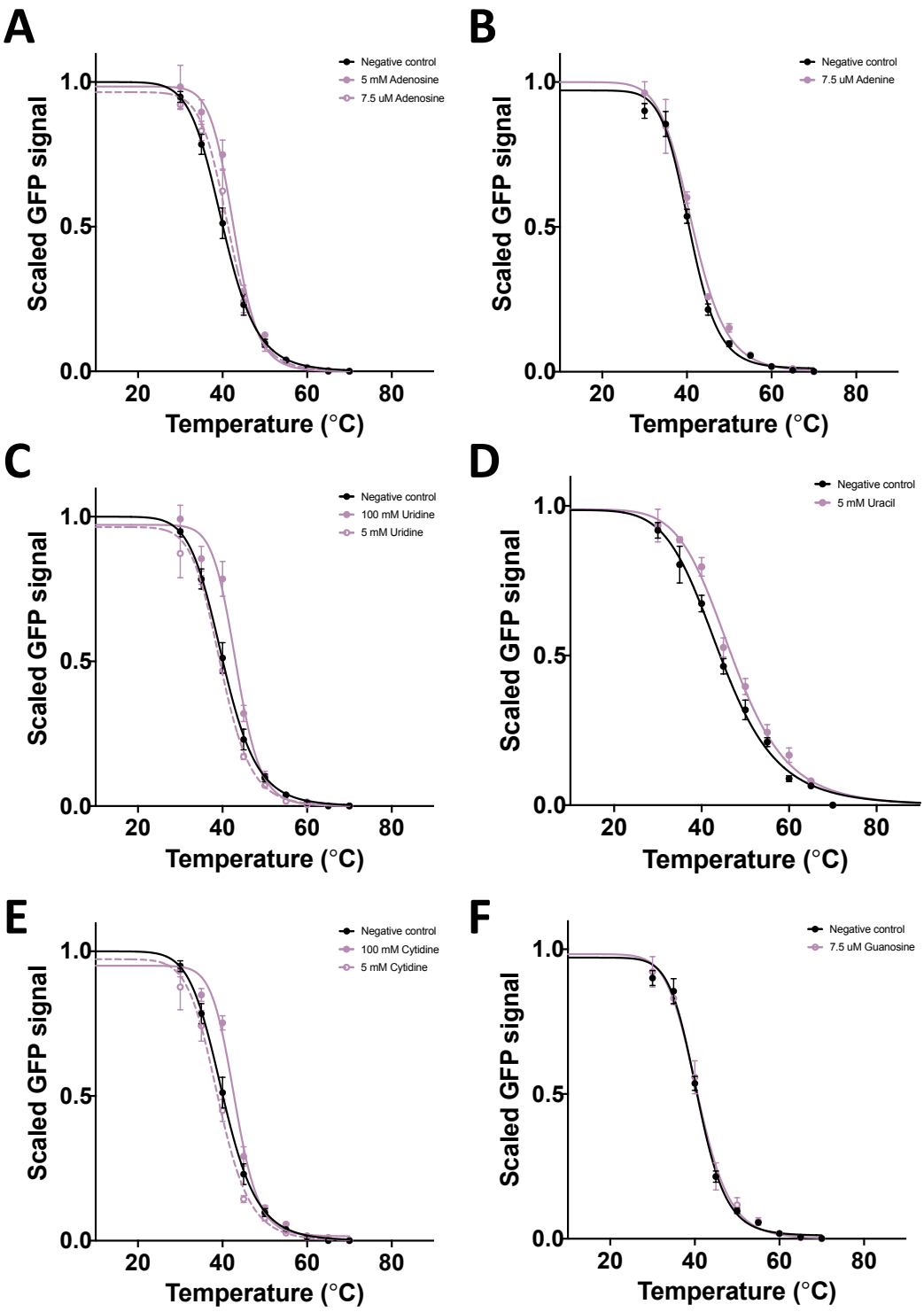


Figure 5.21 Ten-temperature stability curves for *Bs*ENT in the presence of putative endogenous substrates.

Protein that remained in solution after a ten-temperature challenge in the presence of (A) adenosine, (B) adenine, (C) uridine, (D) uracil, (E) cytidine or (F) guanosine was quantified using in-gel GFP fluorescence. Data was normalised to the 4 °C control and were fit with a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting in GraphPad Prism 9.0. Error bars are representative of SEM. Data shown is an average of n = 3 experiments.

At a concentration of 7.5 µM, there is no stabilising effect in the presence of guanosine, ΔT_m = 0.1 ± 0.5 °C, and there is a negligible effect in the presence of adenine, ΔT_m = 0.7 ± 0.5°C. However, at 7.5 µM adenosine has a ΔT_m of 1.6 ± 0.5 °C. This stabilising effect is further increased to ΔT_m = 2.7 ± 0.5 when the concentration of adenosine is increased to 5 mM. At 5 mM uridine and cytidine have no stabilising effect, with ΔT_m = -0.4 ± 0.5 °C and ΔT_m = -1.0 ± 0.6 °C, respectively. It takes 20 times as much uridine and cytidine to achieve a stabilising effect similar to that of 5 mM adenosine, with 100 mM concentrations of uridine and cytidine having ΔT_m of 3.3 ± 0.5 °C and 3.0 ± 0.5 °C, respectively. However, at 5 mM the nucleobase uracil has a ΔT_m of 2.3 ± 0.9 °C. In both the conditions analysed in the presence of 5mM NaOH, 5 mM uracil and the negative control, there is a distinct difference in the curve slope of the data. The slope here is less steep than in all other conditions analysed (hill slope values of ~-7 vs -10 to -15 in other conditions). This suggests that the 5 mM NaOH, and the subsequent increase in pH to 8.2, possibly affects the co-operativity of protein unfolding. Overall, this analysis of this data suggests that *Bs*ENT has a distinct preference for adenosine and appears to prefer nucleosides over nucleobases. There also seems to be some preference for uracil. However, owing to the difference observed in the slope of the curve this would benefit from further investigations. Unfortunately, owing to time constraints I have not had the opportunity to pursue any further investigations into the possible substrates of *Bs*ENT.

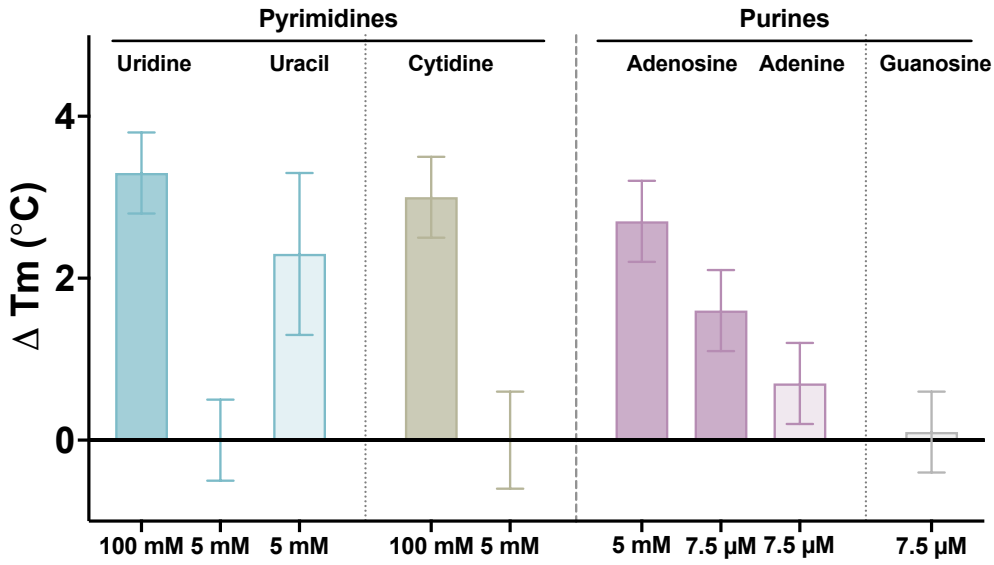


Figure 5.22 ΔT_m values for all putative substrates screened.

Δ T_m of *Bs*ENT in the presence of putative substrates, relative to the appropriate negative control (as shown in Figure 5.21). Data shown is an average of n = 3 repeats. Error bars are representative of error propagation as detailed in 2.3.2.

5.2.4 Nanobodies as conformational selectors and crystallisation chaperones for *Bs*ENT

As nanobodies have been identified as valuable tools in the structural characterisation of challenging membrane targets (1.7.4.1), I set out to select for nanobodies against *Bs*ENT. However, as discussed in 3.3.3, owing to the difficulties encountered with the yeast display system for in-house selection of synthetic nanobodies for membrane proteins I opted to pursue camelid affinity matured nanobodies. I applied for a second Instruct-ERIC funded project with the Steyaert Lab for the generation of nanobodies against *Bs*ENT. I provided the Steyaert Lab with purified *Bs*ENT for camelid immunisation, with the intention of performing the screening and selection of nanobodies myself during a visit to VIB-VUB. *Bs*ENT was solubilised in DMNG, following the protocols as described in 2.8.2. As per the protocols described by Pardon *et al*[187] camelid immunisations were performed and a diverse, affinity matured nanobody library was prepared from camelid blood samples. As discussed in 3.2.6.2, for optimal selection of nanobodies with desired properties, it is recommended that parallel selections in differing conditions and using different methods are performed[187]. Therefore, as with hENT1, I exchanged *Bs*ENT from DMNG micelles into A8-35. Use of a 1:2 mass ratio was found to be the optimal ratio for exchange, with 96% *Bs*ENT recovered following exchange and detergent removal (Figure 5.23). *Bs*ENT in DMNG and A8-35 were labelled with the thiol-cleavable amine-reactive biotinylating agent, EZ-Link™ Sulfo-NHS-SS-Biotin (Thermo Fisher Scientific), and a HABA/Avidin assay[234] determined that *Bs*ENT was successfully labelled with the EZ-Link™ Sulfo-NHS-SS-Biotin in a 1:1 ratio.

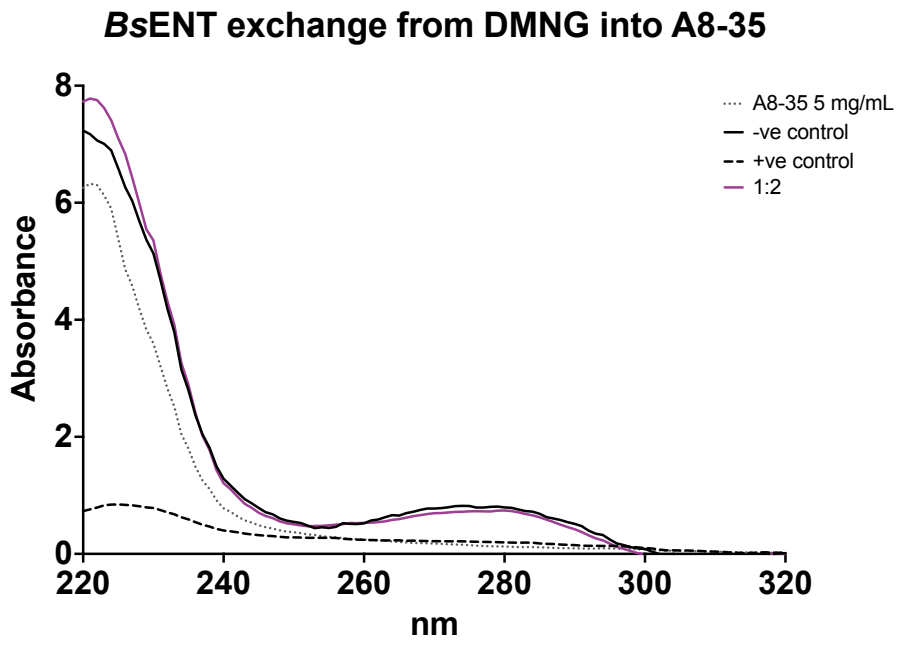


Figure 5.23 Absorbance trace of *Bs*ENT exchanged from DMNG into A8-35.

A8-35 control samples at 5 show contributions to absorbance at ~225-230$_{nm}$ only, with no contributions at 280$_{nm}$. The negative protein control was incubated without A8-35 or biobeads. Exchange into A8-35 results in an increase in the absorbance at ~225-230$_{nm}$ relative to the negative control. Similarly, precipitation of protein, and thereby efficiency of exchange, results in a reduction in the absorbance at 280$_{nm}$. The positive protein control was incubated with biobeads only. This results in the precipitation of all protein, as shown by the loss of absorbance at 220$_{nm}$ and 280$_{nm}$.

I was due to perform selections during a visit to the Steyaert Lab at the VIB-VUB, Belgium. However, due to restrictions that arose because of Covid-19, this visit was unable to take place. The Steyaert Lab generously agreed to perform the selections on my behalf, and I provided them with *Bs*ENT samples in both DMNG and A8-35, with and without the cleavable biotin linker. The immune library was cloned for phage display, and selection and enrichment of target specific nanobodies was performed by phage panning using solid phase coating for biotin-free *Bs*ENT and neutravidin capture for biotinylated *Bs*ENT. 274 individual clones from the enriched sub-library were isolated during solid phase and neutravidin capture selection (136 and 138 clones each, respectively). 76 *Bs*ENT specific clones were found positive (34 clones on the solid phase screening and 42 clones on the neutravidin capture screening). Sequencing determined that, owing to high similarity in the CDR3 sequence, these 76 clones represented 29 families. The plasmids for 44 of the clones from these 29 families were provided to me by Steyaert Lab for expression of nanobodies in *E. coli* WK6 cells. 14 of the 44 nanobodies belong to one family, and thus likely bind the same target epitope. Furthermore, these 14 nanobodies were selected for in all four conditions. Therefore, as a consistent positive result, they have a high chance of being

*Bs*ENT specific binders. Unfortunately, owing to time constraints I have not had the opportunity to express and validate these nanobodies.

5.2.5 Crystallisation of *Bs*ENT

I set up preliminary crystallisation trials for *Bs*ENT using vapour diffusion and LCP crystallisation methods. For LCP, I purified *Bs*ENT in DMNG and obtained a final yield of 200 µL at 5.4 mg/mL. To reach concentrations sufficient for LCP crystallisation, *Bs*ENT was enriched in the mesophase following the cubicon method[238]. The final concentration of protein once diluted with the host lipid, monoolein, for LCP formation was estimated to be 25 mg/mL. Crystallisation screens were prepared in 384 conditions, using the commercially available MemMeso, MemTrans, MemSys, MemStart and MemGold screens. However, no crystal growth was observed.

For vapour diffusion, a crystallisation screen was prepared in 96 conditions using the MemGold2 screen with purified *Bs*ENT in DMNG at a concentration of 16.4 mg/mL. 14 crystals were harvested from 4 conditions after 34 days. All 14 crystals were shot using the I24 Microfocus beam. Four crystals from one condition (33% PEG 1000, 0.5 M MgCl2, 0.02 LiCl and 0.02 M Glycine pH 10; Figure 5.24) diffracted, but 7.3 Å was the highest resolution achieved. I have not had an opportunity to establish any further crystallisation trials.

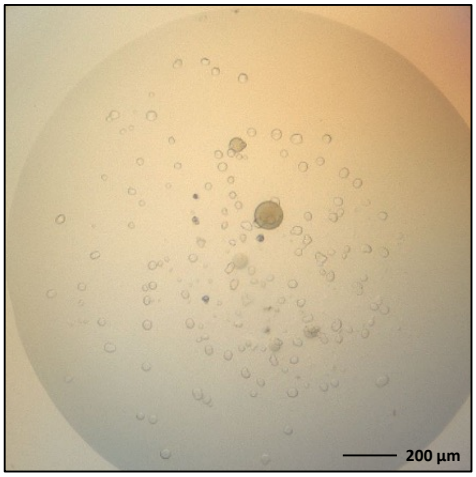


Figure 5.24 Vapor diffusion crystallisation of *Bs*ENT.

Globular crystals (20-40 µm) were obtained after seeding *Bs*ENT (at 16.4 mg/mL) in well D11 of the MemGold2 screen (33% PEG 1000, 0.5 M MgCl2, 0.02 LiCl and 0.02 M Glycine pH 10). Crystals were imaged and harvested at +34 days.

5.3 Discussion

5.3.1 *Bs*ENT as a candidate for homology modelling for hENT1

Overall, at all stages of expression and purification from Sf9 cells, *Bs*ENT outperformed hENT1 (Table 5.2), as well as all the other homologues screened (5.2.1.5). Presentation in-gel suggests that *Bs*ENT is significantly more stable than hENT1 (Figure 5.12). During purification, hENT1 consistently presents with broad and undefined bands and higher MW species that are likely aggregation and/or degradation. Despite efforts towards optimisation and further stabilisation of the current hENT1 constructs, this cannot be improved upon. Therefore, if similar in-gel presentation was observed in any of the homologues screened, they were no longer pursued (5.2.1.5). However, *Bs*ENT consistently presents as a single, defined monomeric band at all points after solubilisation. There is a second monomeric band that is observed in the initial whole cell sample but owing to its absence in fluorescence imaging and subsequent removal with the insoluble fraction, this is most likely unfolded and insoluble protein (Figure 5.5).

During expression, *Bs*ENT is estimated to achieve ~2.5 mg protein per litre of Sf9 culture, which is 2.5 times that as estimated for hENT1. Furthermore, solubilisation of *Bs*ENT in 1% (w/v) DMNG is nearly twice as efficient as that achieved for hENT1 in 1% (w/v) LMNG (60-70% vs ~35%, respectively). The subsequent purification steps are comparable in performance (Table 5.2). However, typical final yields of *Bs*ENT are 3-5 times higher than that obtained for hENT1, with an average of 0.65 mg/L vs 0.14 mg/L, respectively. Therefore, it is the initial steps of expression and purification that seemingly contribute most significantly to the higher yields obtained for *Bs*ENT. Furthermore, the higher yields, increased stability, and high homogeneity of purified *Bs*ENT make it more amenable to downstream processes, such as structural characterisation. In my first vapour diffusion crystallisation trials I obtained protein crystals that diffracted to 7.8 Å. Therefore, *Bs*ENT is amenable to crystallisation and would benefit greatly from further repeat and optimised crystallisation trials. In addition, selection against camelid affinity matured nanobodies for specific *Bs*ENT binders identified 76 positive nanobody clones. While validation of these nanobodies has not yet been performed, the identification of *Bs*ENT binders with high affinity and/or conformational selection would only serve to increase the amenability of *Bs*ENT to crystallisation, and hopefully, structural characterisation. Therefore, I propose that *Bs*ENT is a promising candidate for homology studies for hENT1.

5.3.2 *Bs*ENT as a potential therapeutic target

In my initial search for putative thermostable homologues, I sought out literature on reported thermotolerant species. As ENTs are expressed exclusively in eukaryotes, canonical thermotolerant species from eubacteria and archaea, which can grow in conditions near or above 100 °C[297] were not suitable for homology searches. However, reviews of thermophilic fungi led to me identifying *Byssochlamys spectabilis*[297,307,308] as a species of interest.

B. spectabilis is the sexual, ascospore producing stage (teleomorph) of the ascomycete *Paecilomyces variotii. B. spectabilis* is reportedly a common mold found in soils, plants and animals, and it is frequently encountered in products that have been heat-treated, such as pasteurised fruit juices[307]. It is reportedly a fast-growing fungus that is one of the most heat-resistant known[309]. Furthermore, *B. spectabilis* can grow in environments with low oxygen levels and in the presence of preservatives such as formaldehyde[292,310]. The genome for *B. spectabilis* was first published in 2014[292], and was generated with the intent of investigating formaldehyde resistance. This genome was predicted to feature 8,877 open reading frames and proteins were assigned a predicted function. Several nucleoside and nucleobase transporters were identified. However, '*Bs*ENT' is the only proposed ENT within the genome. Proteins purified directly from *B. spectabilis* have already been explored for the purposes of obtaining protein yields with purity, functionality and stability that are of use for bioindustry, with consistent reports of broad pH tolerance and high thermostability[310-313]. However, there are no reports of structural studies on any proteins from *B. spectabilis* or *P. variotii*.

In addition to *B. spectabilis* and *P.variotii* causing significant economic costs as a result of food spoilage[307,309], the species produces the mycotoxin viriditoxin, and consequently is also associated with many types of human infections, such as pneumonia, peritonitis and endocarditis[314,315]. Therefore, while initial reviews of *P.variotii* and *B. spectabilis* considered it primarily as a contaminant, it is being increasingly recognised as a pathogen[316] and is listed as an emerging causative agent of opportunistic fungal diseases in immunocompromised hosts[317,318]. However, *P. variotii* is highly resistant to the clinically relevant antifungal therapeutic voriconazole[319-321]. Therefore, while studies of *Bs*ENT could provide insight for homology modelling of hENT1, determining structures of *Bs*ENT and gaining insights into the mechanism of action specific to *Bs*ENT may allow for the development of fungal ENT inhibitors that could be of significant therapeutic and economic benefit.

5.3.3 Functional characterisation of *Bs*ENT

As discussed in 5.2.3, *Bs*ENT appears to be NBMPR-insensitive (Figure 5.19), possibly prefers nucleosides over nucleobases, has a strong preference for adenosine, and possibly has a preference for uracil (Figure 5.22). Therefore, I feel confident that the protein assigned as a putative ENT[292], is a true nucleoside transporter. However, further investigations into the relationship with substrates need to be explored for true functional characterisation. Furthermore, as discussed in 5.2.3.2, *Bs*ENT is 26% and 25% identical to hENT1 and hENT2. However, it is also 24% identical to the pH dependent hENT3[90,111,115]. Owing to the pH dependence of hENT3, it is suggested to be a proton-dependent transporter. Sequence analysis predicts that *Bs*ENT may be localised to membranes of endosome/vacuole compartments. This localisation is observed in the well characterised, presumed hENT3 homologue from *S. cerevisiae*, function unknown 26[107,109,290,322] (FUN26). At present, the function of hENT3 remains poorly understood. However, the clinical relevance of hENT3 has been explored extensively, and mutations and dysfunction have been shown to contribute to lysosomal storage disorders[114,116-118]. Therefore, *Bs*ENT may also prove to be a valuable homology model for gaining insights into the function and mechanism of action of hENT3. A goal during my project had been to reconstitute purified *Bs*ENT into liposomes and perform radioligand uptake assays. However, owing to limited time availability and significant delays because of Covid-19, I was not able to accomplish this. Transport assays will be crucial for exploring the true activity of *Bs*ENT, and hopefully the initial substrate screen presented in this work would provide a good starting point for further substrate screening.

5.3.4 Future work

5.3.4.1 Alternatives to GFP-linked thermostability screens

Using a similar detergent and solubilisation screening methodology as was used for *Bs*ENT (5.2.2.1), Dr. Harborne had previously determined that solubilisation of hENT1 with 1% (w/v) LMNG and 1 mg/mL was the optimal condition for thermostability, with a reported apparent T_m of ~42 °C (data not available). As discussed in 5.2.2.2, in all conditions tested for *Bs*ENT in DMNG, the upper limit of the GFP-linked thermostability assay was reached (70-75 °C). While reaching the upper limit renders any determined specific T_m values inaccurate, the fact that the upper limit was reached suggests high thermostability. However, I have not performed assays in parallel for *Bs*ENT and hENT1 in their respective optimal conditions. Therefore, I cannot perform any quantitative comparison of thermostability.

Use of a *Bs*ENT construct free of GFP would overcome the limitations of a GFP linked assay, as discussed in 5.2.2.2, and would allow for specific determination of T_m. This could be achieved easily by using the tag-free purified *Bs*ENT sample obtained following purification. However, as this protein is free of tags, analysis would need to be performed using alternate methods, such as differential scanning fluorimetry (DSF). Dr. Harborne had previously explored the use of the DSF using the thiol reactive N-[4-(7-diethylamino- 4-methyl-3-coumarinyl)phenyl] maleimide (CPM) dye for the specific detection of cysteines during protein unfolding[323] for the determination of hENT1 T_m. However, the signal to noise ratio was poor and no specific data could be determined (data not available). NanoDSF is alternative method that utilises the intrinsic fluorescence of tryptophan to measure protein unfolding in place of reactive dyes. This method would be an ideal approach for further investigations into the thermostability of *Bs*ENT and hENT1, as it would allow for use of existing, well-expressed constructs, without the need for further protein modifications. Furthermore, this method has already been used by other members of the group with success [324]. However, without further optimisation of the purification of a tag free hENT1, further stability analysis would still only be able to compare the stable *Bs*ENT to an unstable hENT1 (3.3.1).

5.3.4.2 Functional characterisation of *Bs*ENT

In addition to characterising substrate specificity of transport, as an extremotolerant species that is resistant to preservatives, heat treatment, variable pH ranges, and low oxygen conditions, it would be interesting to explore the influence of a number of these factors on the activity of *Bs*ENT. Furthermore, as discussed in 5.3.3, as *Bs*ENT may function as a pH and possibly proton dependent transporter, like hENT3[90,111,115] and FUN26[107,290,322], exploring the influence of pH and protons would be critical for true functional characterisation.

As discussed in 1.2.3, local lipid bilayer factors such as lateral pressure, membrane fluidity and packing, surface charge distribution, and the segregation of microdomains have been shown to contribute towards membrane protein structure and function[19-22]. In addition, protein function has been shown to be influenced by specific lipid-protein interactions[17,18] . However, as a highly thermotolerant species it is possible that the lipid membranes of *B. spectabilis* undergo homeoviscous adaptation as a means of maintaining membrane fluidity[15]. Therefore, it could be interesting to explore the relationship between *Bs*ENT, lipids and temperature. During the solubilisation screens, the inclusion of the sterol CHS seemed to negatively affect the solubilisation efficiency and thermostability of all detergents tested (Figure 5.6 and Figure 5.9). However, I am unsure if this was simply because of a negative impact on detergent availability,

and thus solubilisation efficiency. CHS, a cholesterol mimic, differs from the main sterol found in fungal lipids, ergosterol. Cholesterol and ergosterol are structurally very similar, only differing in the presence of additional double bonds on the B-ring and the acyl-chain (Figure 5.7). As discussed in 5.3.2, *P*. *variotii* (*B. spectabilis*) is highly resistant to the clinically relevant antifungal therapeutic voriconazole. Voriconazole is an inhibitor of the cytochrome P-450-dependent 14α-sterol demethylase, which results in the depletion of ergosterol and causes the accumulation of sterol precursors, and subsequently alters the structure and function of the plasma membrane[325,326]. Therefore, *Bs*ENT may not interact with ergosterol (or CHS). As the thermostability assay is determined by relative survival post solubilisation, the negative result there seems to be a specific effect of the CHS. Thus, it could be interesting to address the relationship of *Bs*ENT with ergosterol, sterols as a whole, and other lipids.

Chapter 6 Final conclusions

6.1 Towards the structural characterisation of ENTs in diverse conformations

Human equilibrative nucleoside transporters (ENTs) represent a major pharmaceutical target for cardiac, cancer and viral therapies, and understanding the molecular basis for transport is crucial for the development of improved therapeutics through structure-based drug design (1.5.2.3).

The first structures of hENT1 in distinct, outward-facing conformations were published by Wright & Lee[134] in 2019 and have contributed to significant advancements in our understanding of hENTs. However, there are still large gaps in our knowledge of their mechanism of action. ENTs have been proposed to utilise an alternating access mechanism, similar to that of MFS, in which transporters occupy one of two major conformations, outward-facing or inward-facing (1.4.2.2). However, the understanding of the molecular basis of transport requires the insight into extended conformational states, or at the very least, the opposing major conformation[51,53,58]. Furthermore, ENTs lack features of MFS that are essential for function, suggesting that they may use a different transport mechanism (1.6.5). Therefore, this highlights the need for experimentally determined structures of ENTs in diverse conformational states.

In Chapter 3 I found that the purification of NHT-hENT1 from Sf9 cells suffered from significant losses, and the final protein yields were unstable and heterogenous (3.3.1). This instability and overall poor quality meant that not only were the final yields not viable for structural studies, but they also significantly hindered any attempts to investigate and further optimise protein purification. Therefore, in Chapter 5 I explored putative thermostable homologues of hENT1 to identify alternate candidates for structural characterisation of ENTs, and thus homology modelling for hENT1. In this work I identified an ENT from *Byssochlamys spectabilis* that can be highly expressed in Sf9 cell cultures and can be obtained with homogenous and monodisperse protein yields following a simple two-step purification protocol (5.2.2.3). At all stages of expression and purification *Bs*ENT outperforms hENT1, and typical final yields of *Bs*ENT are on average 4 times higher than those obtained for hENT1 (5.3.1). Furthermore, initial crystallisation trials suggest that following the methodologies outlined in this work, *Bs*ENT is amenable to structural characterisation. Therefore, *Bs*ENT is a promising candidate for structural characterisation, and thus homology modelling of hENT1. However, only preliminary trials were established. Therefore, there are significant opportunities for the optimisation of crystallisation conditions (5.2.5). In addition, *Bs*ENT was amenable to the selection of camelid affinity matured

nanobodies (5.2.4). As nanobodies can serve as conformation selectors and crystallisation chaperones, I believe that the nanobodies selected against *Bs*ENT have the potential to serve as valuable tools in the structural characterisation of an ENT in diverse conformation (5.3.1).

6.2 Towards the structural characterisation of full-length ENTs

Defined mechanisms of alternating access in MFS transporters suggests that the transition from apo to substrate bound state is mediated by subtle rearrangements[18,51,58]. Loop regions of ENTs have been demonstrated to be crucial for function and may also modulate their actions by subtle rearrangements[148-150,153]. However, the ECL1 and the ICL6, are absent from both structures of hENT1. The ICL6 is absent due to its deletion contributing to a construct that was amenable to crystallisation, and the ECL1 is absent due to presumed poor resolution of the electron density of this region because of disorder. Consequently, insights into the molecular basis for the role of ECL1 and ICL6 are lacking. Despite recent advancements in the field, models produced using currently available computational methods have low to very low confidence in the prediction of these regions (Figure 1.19). As discussed in 1.6.5, a major limiting factor in the confident structural prediction of ENTs is the lack of suitable homology models. The distinct, significant differences between ENTs and MFS transporters, as demonstrated in the X-ray structures of hENT1[134] (1.6.1) gives rise to poorly fitted and low confidence models. This is further highlighted in the diversity in the structure and function of loop regions within MFS subfamilies (4.3.2.2). Therefore, the structural characterisation of hENT1, inclusive of loop regions essential for function, is key for our understanding of hENT1 mechanism of action.

As discussed in 3.3.5, I propose that ECL1 may contribute to the stabilisation of hENT1 and may be mediated by the N-linked glycosylation of N48. In addition, in Chapter 4 I identified variants at the ICL6 (4.2.4.3) and the TM7 (4.2.4.2) that stabilise the *apo*-state of hENT1. Furthermore, I identified that K263A (ICL6) and I282V (TM7) stabilise the inhibitor bound state (Figure 4.4). I propose that these variants support interactions that contribute to gating at the intracellular face of the NBMPR-bound state, as in the sugar porters. However, without structures of the ICL6, the mechanisms by which this is achieved remain unknown.

In May 2022 Wu *et al*.[327] published a paper in which they detailed molecular dynamic simulations of outward-facing conformations of hENT1, with the binding of adenosine and dilazep. The simulations were performed in a 3:1 POPE:POPG lipid bilayer using a full length hENT1 model obtained from SWISS-MODEL[327,328], and the simulations modelled the transitions from outward-open to outward-occluded and to fully occluded. These simulations found that the most

significant conformational differences were found in the ECL1 and ICL6. The ECL1 was observed to participate in substrate regulated extracellular gating, with different conformations observed in different states. In an apo-state simulation the ECL1 was observed to have an open-close movement, in which the loop transitions from extending away from the extracellular surface to associating with the extracellular surface and obstructing the pore of the central cavity. Upon adenosine binding hENT1 was more inclined to be in the closed conformation, with the ECL1 packed in a more compact conformation to prevent adenosine dissociation. Conversely, upon dilazep binding the ECL1 remained in the open conformation. They observed that in all states, the ICL6 folds up and contacts the intracellular side of the membrane. However, no further details of the mechanism or lipid interactions were given. As discussed in 1.2.3, interactions with the lipid-bilayer may contribute to regulation and conformational stabilisation[7,17,21]. Therefore, the study of membrane proteins in a native-like lipid containing environment is essential for understanding native structures and the molecular basis of their mechanism of action [184,279,329-331].

In addition to loop motions, Wu *et al*.[327] observed that the major conformational changes in the transition from outward-open to occluded states occur at the extracellular side. In addition, in the *apo*-state, the motion magnitudes for all TMs were relatively small, except for in TM8. TM8 was observed to have a motion direction towards the intracellular side of hENT1 and are suggested to represent an inwards contraction. These motions increased in the adenosine-bound simulation, and the inwards contractions are observed in TM1, TM2 and TM4, in addition to TM8. Thus, Wu *et al*.[327] propose that this supports the rocker-switch mechanism of action of hENT1 suggested by Wright *et al.*[134], where large-scale inwards contractions correspond to the transition from an outward open to a metastable occluded state (Figure 1.8), and contribute to the inter-domain motions between the N-terminal and C-terminal domains that lead to the alternating access mechanism. The role of TM8 in substrate selectivity has been well established[142,143,278]. However, all previously studied residues are oriented towards the central cavity. In Chapter 4 I proposed that residue 336 of TM8 may contribute to the indirect mediation of substrate selectivity. Therefore, in the absence of further structures of ENTs, there are still considerable insights that can be gained from functional and *in silico* studies.

6.3 Significance of this work

Overall, I believe that the findings of my research have the potential to provide new insights into extended conformational states of ENTs and to elucidate the molecular basis of ENT mechanism of action. This information would subsequently assist in the structure-based design of new and

important therapeutics used in anti-cancer and anti-viral treatments, in addition to important therapeutic modulators of purinergic signalling pathways.

References

1 Voet, D. & Voet, J. G. in *Biochemistry* (John Wiley & Sons, INC, 2011).

2 Alberts, B., Johnson, A. & Lewis, J. in *Mol Biol Cell* (New York: Garland Science, 2002).

3 Vinothkumar, K. R. & Henderson, R. *Structures of membrane proteins*. Vol. 43 (2010).

4 Luckey, M. *Membrane structural biology: with biochemical and biophysical foundations*. 2nd edn, (Cambridge University Press, 2014).

5 Andersen, O. S. & Koeppe, R. E. Bilayer thickness and membrane protein function: An energetic perspective. *Annu Rev Biophys* **36**, 107-130, doi:10.1146/annurev.biophys.36.040306.132643 (2007).

6 Harayama, T. & Riezman, H. Understanding the diversity of membrane lipid composition. *Nat Rev Mol Cell Biol* **19**, 281-296, doi:10.1038/nrm.2017.138 (2018).

7 Phillips, R., Ursell, T., Wiggins, P. & Sens, P. Emerging roles for lipids in shaping membrane-protein function. *Nature* **459**, 379-385, doi:10.1038/nature08147 (2009).

8 van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* **9**, 112-124, doi:10.1038/nrm2330 (2008).

9 Nastou, K. C., Tsaousis, G. N. & Iconomidou, V. A. PerMemDB: A database for eukaryotic peripheral membrane proteins. *Biochim Biophys Acta Biomembr* **1862**, 183076, doi:10.1016/j.bbamem.2019.183076 (2020).

10 Kinoshita, T. & Fujita, M. Thematic review series: Glycosylphosphatidylinositol (GPI) anchors: Biochemistry and cell biology biosynthesis of GPI-anchored proteins: special emphasis on GPI lipid remodeling. *J Lipid Res* **57**, 6-24, doi:10.1194/jlr.R063313 (2016).

11 Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. Monotopic membrane proteins join the fold. *Trends Biochem Sci* **44**, 7-20, doi:10.1016/j.tibs.2018.09.013 (2019).

12 White, S. H. Biophysical dissection of membrane proteins. *Nature* **459**, 344-346, doi:10.1038/nature08142 (2009).

13 White, S. *Membrane Protein Structure Database*, <http://blanco.biomol.uci.edu/mpstruc/> (

14 Jensen, M. & Mouritsen, O. G. Lipids do influence protein function - The hydrophobic matching hypothesis revisited. *Biochim Biophys Acta Biomembr* **1666**, 205-226, doi:10.1016/j.bbamem.2004.06.009 (2004).

15 Ernst, R., Ejsing, C. S. & Antonny, B. Homeoviscous adaptation and the regulation of membrane lipids. *J Mol Biol* **428**, 4776-4791, doi:10.1016/j.jmb.2016.08.013 (2016).

16 Kalli, Antreas C. & Sansom, Mark S. P. Interactions of peripheral proteins with model membranes as viewed by molecular dynamics simulations. *Biochemical Society Transactions* **42**, 1418-1424, doi:10.1042/BST20140144 (2014).

17 Valiyaveetil, F., I., Zhou, Y. & MacKinnon, R. Lipids in the structure, folding and function of the KcsA K^+ channel. *Biochemistry* **41**, 10771-10777 (2002).

18 Drew, D., North, R. A., Nagarathinam, K. & Tanabe, M. Structures and general transport mechanisms by the major facilitator superfamily (MFS). *Chem Rev* **121**, 5289-5335, doi:10.1021/acs.chemrev.0c00983 (2021).

19 Syeda, R. *et al.* Piezo1 channels are inherently mechanosensitive. *Cell Rep* **17**, 1739-1746, doi:10.1016/j.celrep.2016.10.033 (2016).

20 Escribá, P. V. *et al.* Membranes: a meeting point for lipids, proteins and therapies. *J Cell Mol Med* **12**, 829-875, doi:10.1111/j.1582-4934.2008.00281.x (2008).

21 Wang, B. *et al.* Pocket delipidation induced by membrane tension or modification leads to a structurally analogous mechanosensitive channel state. *Structure* **30**, 608-622 e605, doi:10.1016/j.str.2021.12.004 (2022).

22 Pyle, E. *et al.* Structural lipids enable the formation of functional oligomers of the eukaryotic purine symporter UapA. *Cell Chem Biol* **25**, 840-848 e844, doi:10.1016/j.chembiol.2018.03.011 (2018).

23 McLaughlin, S. & Murray, D. Plasma membrane phosphoinositide organization by protein electrostatics. *Nature* **438**, 605-611, doi:10.1038/nature04398 (2005).

24 Michailidis, I. E. *et al.* Phosphatidylinositol-4,5-bisphosphate regulates epidermal growth factor receptor activation. *Pflugers Arch - Eur J Physiol* **461**, 387-397, doi:10.1007/s00424-010-0904-3 (2011).

25 Hansen, S. B., Tao, X. & MacKinnon, R. Structural basis of PIP2 activation of the classical inward rectifier K^+ channel Kir2.2. *Nature* **477**, 495-498, doi:10.1038/nature10370 (2011).

26 Zheng, H., Liu, W., Anderson, L. Y. & Jiang, Q.-X. Lipid-dependent gating of a voltage-gated potassium channel. *Nature Communications* **2**, 250, doi:10.1038/ncomms1254 (2011).

27 Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 - 1260411-1260419, doi:10.1126/science.1260419 (2015).

28 Overington, J. P., Al-Lazikani, B. & Hopkins, A., L. How many drug targets are there? *Nat Rev Drug Discov* **5**, 993-996, doi:10.1038/nrd2199 (2006).

29 Sood, V. D. & Gross, A. W. in *Advances in Membrane Proteins: Building, Signaling and Malfunction* (ed Yu Cao) 49-65 (Springer Singapore, 2019).

30 Yin, H. & Flynn, A. D. Drugging membrane protein interactions. *Annu Rev Biomed Eng* **18**, 51-76, doi:10.1146/annurev-bioeng-092115-025322 (2016).

31 Sriram, K. & Insel, P. A. G protein-coupled receptors as targets for approved drugs: How many targets and how many drugs? *Mol Pharmacol* **93**, 251-258, doi:10.1124/mol.117.111062 (2018).

32 Saier, M. H., Jr., Reddy, V. S., Tamang, D. G. & Vastermark, A. The transporter classification database. *Nucleic Acids Res* **42**, 251-258, doi:10.1093/nar/gkt1097 (2014).

33 Gulezian, E. *et al.* Membrane protein production and formulation for drug discovery. *Trends Pharmacol Sci* **42**, 657-674, doi:10.1016/j.tips.2021.05.006 (2021).

34 César-Razquin, A. *et al.* A call for systematic research on solute carriers. *Cell* **162**, 478-487, doi:10.1016/j.cell.2015.07.022 (2015).

35 Colas, C., Ung, P. M.-U. & Schlessinger, A. SLC transporters: structure, function, and drug discovery. *MedChemComm* **7**, 1069-1081, doi:10.1039/C6MD00005C (2016).

36 Lin, L., Yee, S. W., Kim, R. B. & Giacomini, K. M. SLC transporters as therapeutic targets: Emerging opportunities. *Nat Rev Drug Discov* **14**, 543-560, doi:10.1038/nrd4626 (2015).

37 Rask-Andersen, M., Almen, M. S. & Schioth, H. B. Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov* **10**, 579-590, doi:10.1038/nrd3478 (2011).

38 Rask-Andersen, M., Masuram, S. & Schioth, H. B. The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu Rev Pharmacol* **54**, 9-26, doi:10.1146/annurev-pharmtox-011613-135943 (2014).

39 Henderson, P. J. F. in *Encyclopedia of Biophysics* (ed Gordon C. K. Robers) 1496-1504 (Springer, 2013).

40 Forrest, L. R., Kramer, R. & Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim Biophys Acta* **1807**, 167-188, doi:10.1016/j.bbabio.2010.10.014 (2011).

41 Klingenberg, M. Transport viewed as a catalytic process. *Biochimie* **89**, 1042-1048, doi:10.1016/j.biochi.2007.02.010 (2007).

42 Perland, E. & Fredriksson, R. Classification systems of secondary active transporters. *Trends Pharmacol Sci* **38**, 305-315, doi:10.1016/j.tips.2016.11.008 (2017).

43 Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: The power to change. *Nat Rev Mol Cell Biol* **10**, 218-227, doi:10.1038/nrm2646 (2009).

44 Wilkens, S. Structure and mechanism of ABC transporters. *F1000Prime Rep* **7**, doi:10.12703/P7-14 (2015).

45 Verhalen, B. *et al.* Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. *Nature* **543**, 738-741, doi:10.1038/nature21414 (2017).

46 Kim, J. *et al.* Subnanometre-resolution electron cryomicroscopy structure of a heterodimeric ABC exporter. *Nature* **517**, 396-400, doi:10.1038/nature13872 (2015).

47 Kopcho, N., Chang, G. & Komives, E. A. Dynamics of ABC Transporter P-glycoprotein in Three Conformational States. *Sci Rep* **9**, 15092, doi:10.1038/s41598-019-50578-2 (2019).

48 Timachi, M. H. *et al.* Exploring conformational equilibria of a heterodimeric ABC transporter. *Elife* **6**, doi:10.7554/eLife.20236 (2017).

49 Sobczak, I. & Lolkema, J. S. Structural and mechanistic diversity of secondary transporters. *Curr Opin Microbiol* **8**, 161-167, doi:10.1016/j.mib.2005.02.005 (2005).

50 Shi, Y. Common folds and transport mechanisms of secondary active transporters. *Annu Rev Biophys* **42**, 51-72, doi:10.1146/annurev-biophys-083012-130429 (2013).

51 Quistgaard, E. M., Low, C., Guettou, F. & Nordlund, P. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat Rev Mol Cell Biol* **17**, 123-132, doi:10.1038/nrm.2015.25 (2016).

52 Kobylka, J., Kuth, M. S., Muller, R. T., Geertsma, E. R. & Pos, K. M. AcrB: a mean, keen, drug efflux machine. *Ann N Y Acad Sci* **1459**, 38-68, doi:10.1111/nyas.14239 (2020).

53 Zhang, X. C., Zhao, Y., Heng, J. & Jiang, D. Energy coupling mechanisms of MFS transporters. *Protein Sci* **24**, 1560-1579, doi:10.1002/pro.2759 (2015).

54 Hediger, M. A. *Genomic Transorter Database - SLC series*, <http://slc.bioparadigms.org/> (

55 Hediger, M. A., Clémençon, B., Burrier, R. E. & Bruford, E. A. The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Mol Aspects Med* **34**, 95-107, doi:10.1016/j.mam.2012.12.009 (2013).

56 Perland, E., Bagchi, S., Klaesson, A. & Fredriksson, R. Characteristics of 29 novel atypical solute carriers of major facilitator superfamily type: Evolutionary conservation, predicted structure and neuronal co-expression. *Open Biol J* **7**, 170142-170142, doi:10.1098/rsob.170142 (2017).

57 Pao, S. S., Paulsen, I. T. & Saier, M. H. Major facilitator superfamily. *Microbiol Mol Biol Rev* **62**, 1-34 (1998).

58 Law, C. J., Maloney, P. C. & Wang, D.-N. Ins and outs of major facilitator superfamily antiporters. *Annu Rev Microbiol* **62**, 289-305, doi:10.1146/annurev.micro.61.080706.093329 (2008).

59 Madej, M. G. in *Methods in Enzymology* Vol. 557 *Membrane Proteins—Engineering, Purification and Crystallization* Ch. 24, 521-549 (Elsevier Inc., 2015).

60 Fluman, N. & Bibi, E. Bacterial multidrug transport through the lens of the major facilitator superfamily. *Biochim Biophys Acta* **1794**, 738-747, doi:10.1016/j.bbapap.2008.11.020 (2009).

61 Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610-615, doi:10.1126/science.1088196 (2003).

62 Custódio, T. F., Paulsen, P. A., Frain, K. M. & Pedersen, B. P. Structural comparison of GLUT1 to GLUT3 reveal transport regulation mechanism in sugar porter family. *Life Sci Alliance* **4**, doi:10.26508/LSA.202000858 (2021).

63 Deng, D. *et al.* Crystal structure of the human glucose transporter GLUT1. *Nature* **510**, 121-125, doi:10.1038/nature13306 (2014).

64 Iancu, C. V., Zamoon, J., Sang, B. W., Aleshin, A. & Choe, J. Y. Crystal structure of a glucose/H^+ symporter and its mechanism of action. *Proc Natl Acad Sci U S A* **110**, 17862-17867, doi:10.1073/pnas.1311485110 (2013).

65 Mueckler, M. & Thorens, B. The SLC2 (GLUT) family of membrane transporters. *Mol Aspects Med* **34**, 121-138, doi:10.1016/j.mam.2012.07.001 (2013).

66 Nomura, N. *et al.* Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature* **526**, 397-401, doi:10.1038/nature14909 (2015).

67 Sun, L. *et al.* Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature* **490**, 361-366, doi:10.1038/nature11524 (2012).

68 Radestock, S. & Forrest, L. R. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *J Mol Biol* **407**, 698-715, doi:10.1016/j.jmb.2011.02.008 (2011).

69 Yan, N. Structural biology of the major facilitator superfamily transporters. *Annu Rev Biophys* **44**, 257-283, doi:10.1146/annurev-biophys-060414-033901 (2015).

70 Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem Sci* **38**, 151-159, doi:10.1016/j.tibs.2013.01.003 (2013).

71 Jessen-Marshall, A. E., Paul, N. J. & Brooker, R. J. The conserved motif, GXXX(D/E)(R/K)XG[X](R/K)(R/K), in hydrophilic loop 2/3 of the lactose permease. *J Biol Chem* **270**, 16251-16257, doi:10.1074/jbc.270.27.16251 (1995).

72 Henderson, P. J. F. & Baldwin, S. A. This is about the in and the out. *Nat Struct Mol Biol* **20**, 654-655, doi:10.1038/nsmb.2604 (2013).

73 Solcan, N. *et al.* Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J* **31**, 3411-3421, doi:10.1038/emboj.2012.157 (2012).

74 Taniguchi, R. *et al.* Outward-and inward-facing structures of a putative bacterial transition-metal transporter with homology to ferroportin. *Nat Commun* **6**, doi:10.1038/ncomms9545 (2015).

75 Zheng, H., Wisedchaisri, G. & Gonen, T. Crystal structure of a nitrate/nitrite exchanger. *Nature* **497**, 647-651, doi:10.1038/nature12139 (2013).

76 Tetsch, L. & Jung, K. How are signals transduced across the cytoplasmic membrane? Transport proteins as transmitter of information. *Amino Acids* **37**, 467-477, doi:10.1007/s00726-009-0235-x (2009).

77 Deng, D. *et al.* Molecular basis of ligand recognition and transport by glucose transporters. *Nature* **526**, 391-396, doi:10.1038/nature14655 (2015).

78 Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347-355, doi:10.1038/nature08143 (2009).

79 Pedata, F. *et al.* Purinergic signalling in brain ischemia. *Neuropharmacology* **104**, 105-130, doi:10.1016/j.neuropharm.2015.11.007 (2016).

80 Tocco, A., Pinson, B., Thiébaud, P., Thézé, N. & Massé, K. Comparative genomic and expression analysis of the adenosine signaling pathway members in *Xenopus*. *Purinergic Signal* **11**, 59-77, doi:10.1007/s11302-014-9431-6 (2015).

81 Dunwiddie, T. V. & Masino, S. a. The role and regulation of adenosine. *Annu Rev Pharmacol Toxicol* **41**, 145-174, doi:10.1146/annurev.neuro.24.1.31 (2001).

82 Cheng, R. K. Y. *et al.* Structures of human A_1 and A_2A adenosine receptors with xanthines reveal determinants of selectivity. *Structure* **25**, 1275-1285.e1274, doi:10.1016/j.str.2017.06.012 (2017).

83 Fredholm, B., B., Chen, J.-F., Cunha, R., A., Svenningsson, P. & Vaugeois, J.-M. Adenosine and brain function. *Int Rev Neurobiol* **63**, 191-270, doi:10.1016/S0074-7742(05)63007-3 (2005).

84 Lebon, G. *et al.* Agonist-bound adenosine A_2A receptor structures reveal common features of GPCR activation. *Nature* **474**, 521-525, doi:10.1038/nature10136 (2011).

85 Lu, Q. *et al.* Sustained adenosine exposure causes lung endothelial barrier dysfunction *via* nucleoside transporter–mediated signaling. *Am J Respir Cell Mol Biol* **47**, 604-613, doi:10.1165/rcmb.2012-0012OC (2012).

86 Baldwin, S. A., Mackey, J. R., Cass, C. E. & Young, J. D. Nucleoside transporters: Molecular biology and implications for therapeutic development. *Mol Med Today* **5**, 216-224, doi:10.1016/S1357-4310(99)01459-8 (1999).

87 King, A. E., Ackley, M. A., Cass, C. E., Young, J. D. & Baldwin, S. A. Nucleoside transporters: from scavengers to novel therapeutic targets. *Trends Pharmacol Sci* **27**, 416-425, doi:10.1016/j.tips.2006.06.004 (2006).

88 Molina-Arcas, M., Trigueros-Motos, L., Casado, F. J. & Pastor-Anglada, M. Physiological and pharmacological roles of nucleoside transporter proteins. *Nucleosides Nucleotides Nucleic Acids* **27**, 769-778, doi:10.1080/15257770802145819 (2008).

89 Dos Santos-Rodrigues, A., Grañé-Boladeras, N., Bicket, A. & Coe, I. R. Nucleoside transporters in the purinome. *Neurochem Int* **73**, 229-237, doi:10.1016/j.neuint.2014.03.014 (2014).

90 Young, J. D., Yao, S. Y. M., Sun, L., Cass, C. E. & Baldwin, S. a. Human equilibrative nucleoside transporter (ENT) family of nucleoside and nucleobase transporter proteins. *Xenobiotica* **38**, 995-1021, doi:10.1080/00498250801927427 (2008).

91 Young, J. D., Yao, S. Y. M. M., Baldwin, J. M., Cass, C. E. & Baldwin, S. A. The human concentrative and equilibrative nucleoside transporter families, SLC28 and SLC29. *Mol Aspects Med* **34**, 529-547, doi:10.1016/j.mam.2012.05.007 (2013).

92 Ye, J. & van den Berg, B. Crystal structure of the bacterial nucleoside transporter Tsx. *EMBO J* **23**, 3187-3195, doi:10.1038/sj.emboj.7600330 (2004).

93 Wang, C. *et al.* Molecular basis for substrate recognition by the bacterial nucleoside transporter NupG. *J Biol Chem* **296**, 100479, doi:10.1016/j.jbc.2021.100479 (2021).

94 Patching, S. G. *et al.* The nucleoside transport proteins, NupC and NupG, from *Escherichia coli*: specific structural motifs necessary for the binding of ligands. *Org Biomol Chem*, 462-470, doi:10.1039/B414739A (2005).

95 Yao, S. Y. M., Ng, A. M. L., Cass, C. E., Baldwin, S. A. & Young, J. D. Nucleobase transport by human equilibrative nucleoside transporter 1 (hENT1). *J Biol Chem* **286**, 32552-32562, doi:10.1074/jbc.M111.236117 (2011).

96 Yao, S. Y. M. *et al.* Functional and molecular characterization of nucleobase transport by recombinant human and rat equilibrative nucleoside transporters 1 and 2: Chimeric constructs reveal a role for the ENT2 helix 5-6 region in nucleobase translocation. *J Biol Chem* **277**, 24938-24948, doi:10.1074/jbc.M200966200 (2002).

97 Vastermark, A., Wollwage, S., Houle, M. E., Rio, R. & Saier, M. H., Jr. Expansion of the APC superfamily of secondary carriers. *Proteins* **82**, 2797-2811, doi:10.1002/prot.24643 (2014).

98 Gray, J. H., Owen, R. P. & Giacomini, K. M. The concentrative nucleoside transporter family, SLC28. *Pflugers Arch* **447**, 728-734, doi:10.1007/s00424-003-1107-y (2004).

99 Baldwin, S. A. *et al.* The equilibrative nucleoside transporter family, SLC29. *Pflugers Arch* **447**, 735-743, doi:10.1007/s00424-003-1103-2 (2004).

100 Löffler, M., Morote-Garcia, J. C., Eltzschig, S. A., Coe, I. R. & Eltzschig, H. K. Physiological roles of vascular nucleoside transporters. *Arterioscler Thromb Vasc Biol* **27**, 1004-1013, doi:10.1161/ATVBAHA.106.126714 (2007).

101 Arimany-Nardi, C. *et al.* Identification and characterization of a secondary sodium-binding site and the main selectivity determinants in the human concentrative nucleoside transporter 3. *Mol Pharm* **14**, 1980-1987, doi:10.1021/acs.molpharmaceut.7b00085 (2017).

102 Hirschi, M., Johnson, Z. L. & Lee, S.-Y. Visualizing multistep elevator-like transitions of a nucleoside transporter. *Nature* **545**, 66-70, doi:10.1038/nature22057 (2017).

103 Johnson, Z. L., Cheong, C. G. & Lee, S. Y. Crystal structure of a concentrative nucleoside transporter from *Vibrio cholerae* at 2.4 Å. *Nature* **483**, 489-493, doi:10.1038/nature10882 (2012).

104 Stecula, A., Schlessinger, A., Giacomini, K. M. & Sali, A. Human concentrative nucleoside transporter 3 (hCNT3, SLC28A3) forms a cyclic homotrimer. *Biochemistry* **56**, 3475-3483, doi:10.1021/acs.biochem.7b00339 (2017).

105 Zhou, Y. *et al.* Cryo-EM structure of the human concentrative nucleoside transporter CNT3. *PLoS Biol* **18**, e3000790, doi:10.1371/journal.pbio.3000790 (2020).

106 Boswell-Casteel, R. C. & Hays, F. A. Equilibrative nucleoside transporters—A review. *Nucleosides Nucleotides Nucleic Acids* **36**, 7-30, doi:10.1080/15257770.2016.1210805 (2017).

107 Boswell-Casteel, R. C., Johnson, J. M., Duggan, K. D. & Hays, F. A. Function unknown number 26 (Fun26) protein from *Saccharomyces Cerevisiae* is a nucleoside selective integral membrane transporter. *Biophys J* **106**, 369a-369a, doi:http://dx.doi.org/10.1016/j.bpj.2013.11.2089 (2014).

108 Young, J. D. The SLC28 (CNT) and SLC29 (ENT) nucleoside transporter families: a 30-year collaborative odyssey. *Biochem Soc Trans* **44**, 869-876, doi:10.1042/BST20160038 (2016).

109 Griffiths, M. *et al.* Cloning of a human nucleoside transporter implicated in the cellular uptake of adenosine and chemotherapeutic drugs. *Nat Med* **3**, 89-93, doi:10.1038/nm0197-89 (1997).

110 Griffith, M. T. *et al.* Molecular cloning and characterization of a nitrobenzylthioinosine-insensitive (ei) equilibrative nucleoside transporter from human placenta. *Biochem J* **328**, 739-743, doi:10.1042/bj3280739 (1997).

111 Baldwin, S. A. *et al.* Functional characterization of novel human and mouse equilibrative nucleoside transporters (hENT3 and mENT3) located in intracellular membranes. *J Biol Chem* **280**, 15880-15887, doi:10.1074/jbc.M414337200 (2005).

112 Barnes, K. *et al.* Distribution and functional characterization of equilibrative nucleoside transporter-4, a novel cardiac adenosine transporter activated at acidic pH. *Circ Res* **99**, 510-519, doi:10.1161/01.RES.0000238359.18495.42 (2006).

113 Sundaram, M. *et al.* Topology of a human equilibrative, nitrobenzylthioinosine (NBMPR) -sensitive nucleoside transporter (hENT1) implicated in the cellular uptake of adenosine and anti-cancer drugs. *J Biol Chem* **276**, 45270-45275, doi:10.1074/jbc.M107169200 (2001).

114 Hsu, C.-L. *et al.* Equilibrative nucleoside transporter 3 deficiency perturbs lysosome function and macrophage homeostasis. *Science* **335**, 89-92, doi:10.1126/science.1213682 (2012).

115 Rahman, M. F., Askwith, C. & Govindarajan, R. Molecular determinants of acidic pH-dependent transport of human equilibrative nucleoside transporter 3. *J Biol Chem* **292**, 14775-14785, doi:10.1074/jbc.M117.787952 (2017).

116 Kang, N. *et al.* Human equilibrative nucleoside transporter-3 (hENT3) spectrum disorder mutations impair nucleoside transport, protein localization, and stability. *J Biol Chem* **285**, 28343-28352, doi:10.1074/jbc.M110.109199 (2010).

117 Liu, B. *et al.* Equilibrative nucleoside transporter 3 depletion in β-cells impairs mitochondrial function and promotes apoptosis: Relationship to pigmented

hypertrichotic dermatosis with insulin-dependent diabetes. *Biochim Biophys Acta Mol Basis Dis* **1852**, 2086-2095, doi:10.1016/j.bbadis.2015.07.002 (2015).

118 Spiegel, R. *et al.* Expanding the clinical spectrum of SLC29A3 gene defects. *Eur J Med Genet* **53**, 309-313, doi:10.1016/j.ejmg.2010.06.012 (2010).

119 Zhou, M., Duan, H., Engel, K., Xia, L. & Wang, J. Adenosine transport by plasma membrane monoamine transporter: Reinvestigation and comparison with organic cations. *Drug Metab Dispos* **38**, 1798-1805, doi:10.1124/dmd.110.032987 (2010).

120 Cunha, R., A. Adenosine as a neuromodulator and as a homeostatic regulator in the nervous system: different roles, different sources and different receptors. *Neurochem Int* **38**, 107-125, doi:10.1016/S0197-0186(00)00034-6 (2001).

121 Deguchi, H. *et al.* Dilazep, an antiplatelet agent, inhibits tissue factor expression in endothelial cells and monocytes. *Blood* **90**, 2345-2356 (1997).

122 Dale, N. & Frenguelli, B. G. Release of adenosine and ATP during ischemia and epilepsy. *Curr Neuropharmacol* **7**, 160-179, doi:10.2174/157015909789152146 [doi] (2009).

123 Hyde, R. J., Cass, C. E., Young, J. D. & Baldwin, S. A. The ENT family of eukaryote nucleoside and nucleobase transporters: Recent advances in the investigation of structure/function relationships and the identification of novel isoforms. *Mol Membr Biol* **18**, 53-63, doi:10.1080/09687680110033765 (2001).

124 Molina-Arcas, M., Casado, F. & Pastor-Anglada, M. Nucleoside transporter proteins. *Curr Vasc Pharmacol* **7**, 426-434, doi:10.2174/157016109789043892 (2009).

125 Verdaguer, H., Saurí, T. & Macarulla, T. Predictive and prognostic biomarkers in personalized gastrointestinal cancer treatment. *J Gastrointest Oncol* **8**, 405-417, doi:10.21037/jgo.2016.11.15 (2017).

126 Loosen, S. H., Neumann, U. P., Trautwein, C., Roderburg, C. & Luedde, T. Current and future biomarkers for pancreatic adenocarcinoma. *Tumor Biol* **39**, 101042831769223-101042831769223, doi:10.1177/1010428317692231 (2017).

127 Pastor-Anglada, M. & Pérez-Torras, S. Nucleoside transporter proteins as biomarkers of drug responsiveness and drug targets. *Front Pharmacol* **6**, doi:10.3389/fphar.2015.00013 (2015).

128 Tavolari, S. *et al.* Membrane human equilibrative nucleoside transporter 1 is associated with a high proliferation rate and worse survival in resected intrahepatic cholangiocarcinoma patients not receiving adjuvant treatments. *Eur J Cancer* **106**, 160-170, doi:10.1016/j.ejca.2018.11.005 (2019).

129 Deniskin, R., Frame, I. J., Sosa, Y. & Akabas, M. H. Targeting the *Plasmodium vivax* equilibrative nucleoside transporter 1 (PvENT1) for antimalarial drug development. *Int J Parasitol Drugs Drug Resist* **6**, 1-11, doi:10.1016/j.ijpddr.2015.11.003 (2016).

130 Frame, I. J., Deniskin, R., Arora, A. & Akabas, M. H. Purine import into malaria parasites as a target for antimalarial drug development. *Ann N Y Acad Sci* **1342**, 19-28, doi:10.1111/nyas.12568 (2015).

131 Arora, A. *et al.* Substrate and inhibitor specificity of the *Plasmodium berghei* equilibrative nucleoside transporter type 1. *Mol Pharmacol* **89**, 678-685, doi:10.1124/mol.115.101386 (2016).

132 Lapponi, M. J., Rivero, C. W., Zinni, M. A., Britos, C. N. & Trelles, J. A. New developments in nucleoside analogues biosynthesis: A review. *J Mol Catal B Enzym* **133**, 218-233, doi:10.1016/j.molcatb.2016.08.015 (2016).

133 Rehan, S., Ashok, Y., Nanekar, R. & Jaakola, V.-P. Thermodynamics and kinetics of inhibitor binding to human equilibrative nucleoside transporter subtype-1. *Biochem Pharmacol* **98**, 681-689, doi:10.1016/j.bcp.2015.09.019 (2015).

134 Wright, N. J. & Lee, S. Y. Structures of human ENT1 in complex with adenosine reuptake inhibitors. *Nat Struct Mol Biol* **26**, 599-606, doi:10.1038/s41594-019-0245-7 (2019).

135 Visser, F. *et al.* Mutation of residue 33 of human equilibrative nucleoside transporters 1 and 2 alters sensitivity to inhibition of transport by dilazep and dipyridamole. *J Biol Chem* **277**, 395-401, doi:10.1074/jbc.M105324200 (2002).

136 Visser, F. *et al.* Residue 33 of human equilibrative nucleoside transporter 2 is a functionally important component of both the dipyridamole and nucleoside binding sites. *Mol Pharmacol* **67**, 1291-1298, doi:10.1124/mol.104.005884 (2005).

137 Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* **503**, 85-90, doi:10.1038/nature12533 (2013).

138 Endres, C. J. Residues met89 and ser160 in the human equilibrative nucleoside transporter 1 affect Its affinity for adenosine, guanosine, S6-(4-nitrobenzyl)-mercaptopurine riboside, and dipyridamole. *Mol Pharmacol* **67**, 837-844, doi:10.1124/mol.104.008102 (2004).

139 Endres, C. J., Sengupta, D. J. & Unadkat, J. D. Mutation of leucine-92 selectively reduces the apparent affinity of inosine, guanosine, NBMPR [S6-(4-nitrobenzyl)-mercaptopurine riboside] and dilazep for the human equilibrative nucleoside transporter, hENT1. *Biochem J* **380**, 131-137, doi:10.1042/bj20031880 (2004).

140 Paproski, R. J. *et al.* Mutation of Trp 29 of human equilibrative nucleoside transporter 1 alters affinity for coronary vasodilator drugs and nucleoside selectivity. *Biochem J* **414**, 291-300, doi:10.1042/BJ20080074 (2008).

141 Visser, F., Baldwin, S. A., Isaac, R. E., Young, J. D. & Cass, C. E. Identification and mutational analysis of amino acid residues involved in dipyridamole interactions with human and *caenorhabditis elegans* equilibrative nucleoside transporters. *J Biol Chem* **280**, 11025-11034, doi:10.1074/jbc.M410348200 (2005).

142 Arastu-Kapur, S., Ford, E., Ullman, B. & Carter, N. S. Functional analysis of an inosine-guanosine transporter from *Leishmania donovani*: The role of conserved

residues, aspartate 389 and arginine 393. *J Biol Chem* **278**, 33327-33333, doi:10.1074/jbc.M305141200 (2003).

143 Visser, F. *et al.* Residues 334 and 338 in transmembrane segment 8 of human equilibrative nucleoside transporter 1 are important determinants of inhibitor sensitivity, protein folding, and catalytic turnover. *J Biol Chem* **282**, 14148-14157, doi:10.1074/jbc.M701735200 (2007).

144 SenGupta, D. J. & Unadkat, J. D. Glycine 154 of the equilibrative nucleoside transporter, hENT1, is important for nucleoside transport and for conferring sensitivity to the inhibitors nitrobenzylthioinosine, dipyridamole, and dilazep. *Biochem Pharmacol* **67**, 453-458, doi:10.1016/j.bcp.2003.09.018 (2004).

145 Bicket, A. & Coe, I. R. N-linked glycosylation of N48 is required for equilibrative nucleoside transporter 1 (ENT1) function. *Biosci Rep* **36**, e00376-e00376, doi:10.1042/BSR20160063 (2016).

146 Aseervatham, J., Tran, L., Machaca, K. & Boudker, O. The role of flexible loops in folding, trafficking and activity of equilibrative nucleoside transporters. *PLoS ONE* **10**, doi:10.1371/journal.pone.0136779 (2015).

147 Reyes, G., Nivillac, N. M. I., Chalsev, M. & Coe, I. R. Analysis of recombinant tagged equilibrative nucleoside transporter 1 (ENT1) expressed in *E. coli*. *Biochem Cell Biol* **89**, 246-255, doi:10.1139/O10-155 (2011).

148 Hughes, S. J., Cravetchi, X., Vilas, G. & Hammond, J. R. Adenosine A1 receptor activation modulates human equilibrative nucleoside transporter 1 (hENT1) activity *via* PKC-mediated phosphorylation of serine-281. *Cell Signal* **27**, 1008-1018, doi:10.1016/j.cellsig.2015.02.023 (2015).

149 Reyes, G. *et al.* The equilibrative nucleoside transporter (ENT1) can be phosphorylated at multiple sites by PKC and PKA. *Mol Membr Biol* **28**, 412-426, doi:10.3109/09687688.2011.604861 (2011).

150 Handa, M. *et al.* Cloning of a novel isoform of the mouse NBMPR-sensitive equilibrative nucleoside transporter (ENT1) lacking a putative phosphorylation site. *Gene* **262**, 301-307, doi:10.1016/S0378-1119(00)00555-2 (2001).

151 Conseil, G., Rothnie, A. J., Deeley, R. G. & Cole, S. P. C. Multiple roles of charged amino acids in cytoplasmic loop 7 for expression and function of the multidrug and organic anion transporter MRP1 (ABCC1). *Mol Pharmacol* **75**, 397-406, doi:10.1124/mol.108.052860 (2009).

152 Fusca, T. *et al.* Single point mutations in the small cytoplasmic loop of ACA8, a plasma membrane $Ca2^{+}$-ATPase of *Arabidopsis thaliana*, generate partially deregulated pumps. *J Biol Chem* **284**, 30881-30888, doi:10.1074/jbc.M109.006148 (2009).

153 Choi, D.-S. *et al.* The type 1 equilibrative nucleoside transporter regulates ethanol intoxication and preference. *Nat Neurosci* **7**, 855-861, doi:10.1038/nn1288 (2004).

154 Mailliard, W. S. & Diamond, I. Recent advances in the neurobiology of alcoholism: The role of adenosine. *Pharmacol Ther* **101**, 39-46, doi:10.1016/j.pharmthera.2003.10.002 (2004).

155 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

156 Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735-1742, doi:10.1016/j.str.2013.08.005 (2013).

157 Yang, J. *et al.* The I-TASSER suite: Protein structure and function prediction. *Nat Methods* **12**, 7-8, doi:10.1038/nmeth.3213 (2014).

158 Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-D444, doi:10.1093/nar/gkab1061 (2022).

159 Hardy, D., Bill, R. M., Jawhari, A. & Rothnie, A. J. Overcoming bottlenecks in the membrane protein structural biology pipeline. *Biochem Soc Trans* **44**, 838-844, doi:10.1042/BST20160049 (2016).

160 Lyons, J. A., Shahsavar, A., Paulsen, P. A., Pedersen, B. P. & Nissen, P. Expression strategies for structural studies of eukaryotic membrane proteins. *Curr Opin Struct Biol* **38**, 137-144, doi:10.1016/j.sbi.2016.06.011 (2016).

161 Scott, D. J., Kummer, L., Tremmel, D. & Plückthun, A. Stabilizing membrane proteins through protein engineering. *Curr Opin Chem Biol* **17**, 427-435, doi:10.1016/j.cbpa.2013.04.002 (2013).

162 Andréll, J. & Tate, C. G. Overexpression of membrane proteins in mammalian cells for structural studies. *Mol Membr Biol* **30**, 52-63, doi:10.3109/09687688.2012.703703 (2013).

163 Tate, C. G. Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett* **504**, 94-98, doi:10.1016/S0014-5793(01)02711-9 (2001).

164 Byrne, B. *Pichia pastoris* as an expression host for membrane protein structural biology. *Curr Opin Struct Biol* **32**, 9-17, doi:10.1016/j.sbi.2015.01.005 (2015).

165 Parker, J. L. & Newstead, S. Method to increase the yield of eukaryotic membrane protein expression in *saccharomyces cerevisiae* for structural and functional studies. *Protein Sci* **23**, 1309-1314, doi:10.1002/pro.2507 (2014).

166 Kapoor, K. *et al.* Mechanism of inhibition of human glucose transporter GLUT1 is conserved between cytochalasin B and phenylalanine amides. *Proc Natl Acad Sci U S A* **113**, 4711-4716, doi:10.1073/pnas.1603735113 (2016).

167 He, Y., Wang, K. & Yan, N. The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein Cell* **5**, 658-672, doi:10.1007/s13238-014-0086-4 (2014).

168 Tate, C. G. in *Heterologous Expression of Membrane Proteins: Methods and Protocols* (ed Isabelle Mus-Veteau) 187-203 (Humana Press, 2010).

169 Goehring, A. *et al.* Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat Protoc* **9**, 2574-2585, doi:10.1038/nprot.2014.173 (2014).

170 Chun, E. *et al.* Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20**, 967-976, doi:10.1016/j.str.2012.04.010 (2012).

171 Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805 (1994).

172 Stepanenko, O. V., Verkhusha, V. V., Kuznetsova, I. M., Uversky, V. N. & Turoverov, K. K. Fluorescent proteins as biomarkers and biosensors: Throwing color lights on molecular and cellular processes. *Curr Protein Pept Sci* **9**, 338-369, doi:10.2174/138920308785132668 (2008).

173 Tate, C. G. Identifying thermostabilizing mutations in membrane proteins by bioinformatics. *Biophys J* **109**, 1307-1308, doi:10.1016/j.bpj.2015.08.020 (2015).

174 Tate, C. G. A crystal clear solution for determining G-protein-coupled receptor structures. *Trends Biochem Sci* **37**, 343-352, doi:10.1016/j.tibs.2012.06.003 (2012).

175 Magnani, F. *et al.* A mutagenesis and screening strategy to generate optimally thermostabilized membrane proteins for structural studies. *Nat Protoc* **11**, 1554-1571, doi:10.1038/nprot.2016.088 (2016).

176 Xiang, J. *et al.* Successful strategies to determine high-resolution structures of GPCRs. *Trends Pharmacol Sci* **37**, 1055-1069, doi:10.1016/j.tips.2016.09.009 (2016).

177 Heydenreich, F. M., Vuckovic, Z., Matkovic, M. & Veprintsev, D. B. Stabilization of G protein-coupled receptors by point mutations. *Front Pharmacol* **6**, doi:10.3389/fphar.2015.00082 (2015).

178 Vaidehi, N., Grisshammer, R. & Tate, C. G. How can mutations thermostabilize G-protein-coupled receptors? *Trends Pharmacol Sci* **37**, 37-46, doi:10.1016/j.tips.2015.09.005 (2016).

179 Sauer, D. B., Karpowich, N. K., Song, J. M. & Wang, D. N. Rapid bioinformatic identification of thermostabilizing mutations. *Biophys J* **109**, 1420-1428, doi:10.1016/j.bpj.2015.07.026 (2015).

180 Harborne, S. P. D. *et al.* IMPROvER: The integral membrane protein stability selector. *Sci Rep* **10**, doi:10.1038/s41598-020-71744-x (2020).

181 Bhattacharya, S., Lee, S., Grisshammer, R., Tate, C. G. & Vaidehi, N. Rapid computational prediction of thermostabilizing mutations for G protein-coupled receptors. *J Chem Theory Comput* **10**, 5149-5160, doi:10.1021/ct500616v (2014).

182 Popov, P. *et al.* Computational design of thermostabilizing point mutations for G protein-coupled receptors. *eLife* **7**, doi:10.7554/eLife.34729 (2018).

183 Latek, D. Rosetta Broker for membrane protein structure prediction: concentrative nucleoside transporter 3 and corticotropin-releasing factor

receptor 1 test cases. *BMC Struct Biol* **17**, 8-8, doi:10.1186/s12900-017-0078-8 (2017).

184 Parker, J. L. *et al.* Cryo-EM structure of PepT2 reveals structural basis for proton-coupled peptide and prodrug transport in mammals. *Sci Adv* **7**, eabh3355 (2021).

185 Arbabi-Ghahroudi, M. Camelid single-domain antibodies: Historical perspective and future outlook. *Front Immunol* **8**, 1-8, doi:10.3389/fimmu.2017.01589 (2017).

186 McMahon, C. *et al.* Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat Struct Mol Biol* **25**, 289-296, doi:10.1038/s41594-018-0028-6 (2018).

187 Pardon, E. *et al.* A general protocol for the generation of nanobodies for structural biology. *Nat Protoc* **9**, 674-693, doi:10.1038/nprot.2014.039 (2014).

188 Zimmermann, I. *et al.* Synthetic single domain antibodies for the conformational trapping of membrane proteins. *eLife* **7**, e34317-e34317, doi:10.7554/eLife.34317 (2018).

189 Zimmermann, I. *et al.* Generation of synthetic nanobodies against delicate proteins. *Nat Protoc* **15**, 1707-1741, doi:10.1038/s41596-020-0304-x (2020).

190 Doshi, R. *et al.* In vitro nanobody discovery for integral membrane protein targets. *Sci Rep* **4**, 1-8, doi:10.1038/srep06760 (2014).

191 Manglik, A., Kobilka, B. K. & Steyaert, J. Nanobodies to study G protein–coupled receptor structure and function. *Annu Rev Pharmacol Toxicol* **57**, 19-37, doi:10.1146/annurev-pharmtox-010716-104710 (2017).

192 Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101-106, doi:10.1038/nature12735 (2013).

193 Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the β2-adrenoceptor. *Nature* **469**, 175-181, doi:10.1038/nature09648 (2011).

194 Uchanski, T. *et al.* Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nat Methods* **18**, 60-68, doi:10.1038/s41592-020-01001-6 (2021).

195 Roy, A. in *Methods in Enzymology* Vol. 557 *Membrane Proteins—Engineering, Purification and Crystallization* Ch. 3, 45-56 (Elsevier Inc., 2015).

196 Urner, L. H. *et al.* Modular detergents tailor the purification and structural analysis of membrane proteins including G-protein coupled receptors. *Nat Commun* **11**, 564, doi:10.1038/s41467-020-14424-8 (2020).

197 Bertheleme, N. *et al.* Unlocking the secrets of the gatekeeper: Methods for stabilizing and crystallizing GPCRs. *Biochim Biophys Acta Biomembr* **1828**, 2583-2591, doi:10.1016/j.bbamem.2013.07.013 (2013).

198 Loll, P. J. Membrane proteins, detergents and crystals: What is the state of the art? *Acta Crystallogr F Struct Biol Commun* **70**, 1576-1583, doi:10.1107/S2053230X14025035 (2014).

199 Parker, J. L. & Newstead, S. in *The Next Generation in Membrane Protein Structure Determination. Advances in Experimental Medicine and Biology* Vol. 922 (ed Isabel Moraes) 61-72 (Springer New York LLC, 2016).

200 Wiener, M. C. A pedestrian guide to membrane protein crystallization. *Methods* **34**, 364-372, doi:10.1016/j.ymeth.2004.03.025 (2004).

201 Oliver, R. C. *et al.* Dependence of micelle size and shape on detergent alkyl chain length and head group. *PLoS One* **8**, e62488, doi:10.1371/journal.pone.0062488 (2013).

202 Poulos, S., Morgan, J. L. W., Zimmer, J. & Faham, S. in *Methods in Enzymology* Vol. 557 *Membrane Proteins—Engineering, Purification and Crystallization* Ch. 19, 393-416 (Elsevier Inc., 2015).

203 Rigaud, J.-L., Pitard, B. & Levy, D. Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *Biochim Biophys Acta* **1231**, 223-246 (1995).

204 Durr, U. H., Soong, R. & Ramamoorthy, A. When detergent meets bilayer: birth and coming of age of lipid bicelles. *Prog Nucl Magn Reson Spectrosc* **69**, 1-22, doi:10.1016/j.pnmrs.2013.01.001 (2013).

205 Ratkeviciute, G., Cooper, B. F. & Knowles, T. J. Methods for the solubilisation of membrane proteins: the micelle-aneous world of membrane protein solubilisation. *Biochem Soc Trans* **49**, 1763-1777, doi:10.1042/BST20210181 (2021).

206 Efremov, R. G., Gatsogiannis, C. & Raunser, S. in *Methods in Enzymology* Vol. 594 *A Structure-Function Toolbox for Membrane Transporter and Channels* Ch. 1, 1-31 (Elsevier, 2017).

207 Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett* **2**, 853-856, doi:10.1021/nl025623k (2002).

208 Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed self-assembly of nonodisperse phospholipid bilayer nanodiscs with controlled size. *J Am Chem Soc* **126**, 3477-3487, doi:10.1021/ja0393574 (2004).

209 Billesbølle, C. B. *et al.* Structure of hepcidin-bound ferroportin reveals iron homeostatic mechanisms. *Nature* **586**, 807-811, doi:10.1038/s41586-020-2668-z (2020).

210 Pan, Y. *et al.* Structural basis of ion transport and inhibition in ferroportin. *Nat Commun* **11**, 5686, doi:10.1038/s41467-020-19458-6 (2020).

211 Cater, R. J. *et al.* Structural basis of omega-3 fatty acid transport across the blood-brain barrier. *Nature* **595**, 315-319, doi:10.1038/s41586-021-03650-9 (2021).

212 Carlson, M. L. *et al.* The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution. *eLife* **7**, doi:10.7554/eLife.34085 (2018).

213 Le Bon, C., Marconnet, A., Masscheleyn, S., Popot, J. L. & Zoonens, M. Folding and stabilizing membrane proteins in amphipol A8-35. *Methods* **147**, 95-105, doi:10.1016/j.ymeth.2018.04.012 (2018).

214 Pocanschi, C. L. *et al.* Amphipathic polymers: Tools to fold integral membrane proteins to their active form. *Biochemistry* **45**, 13954-13961, doi:10.1021/bi0616706 (2006).

215 Gohon, Y. *et al.* Partial specific volume and solvent interactions of amphipol A8-35. *Anal Biochem* **334**, 318-334, doi:10.1016/j.ab.2004.07.033 (2004).

216 Dawaliby, R. *et al.* Allosteric regulation of G protein–coupled receptor activity by phospholipids. *Nat Chem Biol* **12**, 35-41, doi:doi:10.1038/nchembio.1960 (2016).

217 Hardy, D., Desuzinges Mandon, E., Rothnie, A. J. & Jawhari, A. The yin and yang of solubilization and stabilization for wild-type and full-length membrane protein. *Methods* **147**, 118-125, doi:10.1016/j.ymeth.2018.02.017 (2018).

218 Dorr, J. M. *et al.* The styrene-maleic acid copolymer: a versatile tool in membrane research. *Eur Biophys J* **45**, 3-21, doi:10.1007/s00249-015-1093-y (2016).

219 Gulamhussein, A. A. *et al.* Examining the stability of membrane proteins within SMALPs. *Eur Polym J* **112**, 120-125, doi:10.1016/j.eurpolymj.2018.12.008 (2019).

220 Gulamhussein, A. A., Uddin, R., Tighe, B. J., Poyner, D. R. & Rothnie, A. J. A comparison of SMA (styrene maleic acid) and DIBMA (di-isobutylene maleic acid) for membrane protein purification. *Biochim Biophys Acta Biomembr* **1862**, 183281, doi:10.1016/j.bbamem.2020.183281 (2020).

221 Hawkins, O. P. *et al.* Membrane protein extraction and purification using partially-esterified SMA polymers. *Biochim Biophys Acta Biomembr* **1863**, 183758, doi:10.1016/j.bbamem.2021.183758 (2021).

222 Lee, S. C. *et al.* A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nat Protoc* **11**, 1149-1162, doi:10.1038/nprot.2016.070 (2016).

223 *Protein Data Bank*, <https://www.rcsb.org/> (

224 Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152-156, doi:10.1038/s41586-020-2829-0 (2020).

225 Merk, A. *et al.* Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell* **165**, 1698-1707, doi:10.1016/j.cell.2016.05.040. (2016).

226 Earl, L. A., Falconieri, V., Milne, J. L. S. & Subramaniam, S. Cryo-EM: beyond the microscope. *Curr Opin Struct Biol* **46**, 71-78, doi:10.1016/j.sbi.2017.06.002 (2017).

227 Renaud, J. P. *et al.* Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat Rev Drug Discov* **17**, 471-492, doi:10.1038/nrd.2018.77 (2018).

228 Kwan, T. O. C., Axford, D. & Moraes, I. Membrane protein crystallography in the era of modern structural biology. *Biochem Soc Trans* **48**, 2505-2524, doi:10.1042/BST20200066 (2020).

229 Kang, H. J., Lee, C. & Drew, D. Breaking the barriers in membrane protein crystallography. *Int J Biochem Cell Biol* **45**, 636-644, doi:10.1016/j.biocel.2012.12.018 (2013).

230 Kermani, A. A. A guide to membrane protein X-ray crystallography. *FEBS J* **288**, 5788-5804 (2021).

231 Nollert, P. Lipidic cubic phases as matrices for membrane protein crystallization. *Methods* **34**, 348-353, doi:10.1016/j.ymeth.2004.03.030 (2004).

232 Caffrey, M. A comprehensive review of the lipid cubic phase or *in meso* method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallogr F Struct Biol Commun* **71**, 3-18, doi:10.1107/S2053230X14026843 (2015).

233 Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* (ed John M. Walker) 571-607 (Humana Press, 2005).

234 G-Biosciences. *How much biotin is coupled to my protein? The HABA/Avidin assay*, <https://info.gbiosciences.com/blog/how-much-biotin-is-coupled-to-my-protein-the-haba/avidin-assay> (2017).

235 Baginski, E. S., Foà, P. P. & Zak, B. Microdetermination of inorganic phosphate, phospholipids, and total phosphorus in biological materials. *Clin Chem* **13**, 326-332 (1967).

236 McCarthy, C. B. & Romanowski, V. A simplified method for the extraction of baculoviral DNA for PCR analysis: a practical application. *J Virol Methods* **148**, 286-290, doi:10.1016/j.jviromet.2007.11.004 (2008).

237 Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*, gkac240, doi:10.1093/nar/gkac240 (2022).

238 Ma, P. *et al.* The cubicon method for concentrating membrane proteins in the cubic mesophase. *Nat Protoc* **12**, 1745-1762, doi:10.1038/nprot.2017.057 (2017).

239 Gietz, R. D. Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol Biol* **1163**, 33-44, doi:10.1007/978-1-4939-0799-1_4 (2014).

240 Wu, S. & Letchworth, G. J. High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. *BioTechniques* **36**, 152-154, doi:10.2144/04361DD02 (2004).

241 Rehan, S. & Jaakola, V.-P. Expression, purification and functional characterization of human equilibrative nucleoside transporter subtype-1 (hENT1) protein from Sf9 insect cells. *Protein Expr Purif* **114**, 99-107, doi:10.1016/j.pep.2015.07.003 (2015).

242 Crichton, P. G., Harding, M., Ruprecht, J. J., Lee, Y. & Kunji, E. R. S. Lipid, detergent, and Coomassie blue G-250 affect the migration of small membrane proteins in blue native gels. *J Biol Chem* **288**, 22163-22173, doi:10.1074/jbc.M113.484329 (2013).

243 Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci U S A* **106**, 1760-1765, doi:10.1073/pnas.0813167106 (2009).

244 Rehan, S., Paavilainen, V. O. & Jaakola, V.-P. Functional reconstitution of human equilibrative nucleoside transporter-1 into styrene maleic acid co-polymer lipid

particles. *Biochim Biophys Acta Biomembr* **1859**, 1059-1065, doi:10.1016/j.bbamem.2017.02.017 (2017).

245 Rasmussen, S. G. F. *et al.* Crystal structure of the β2-adrenergic receptor-Gs protein complex. *Nature* **477**, 549-557, doi:10.1038/nature10361 (2011).

246 Structural Genomics, C. *et al.* Protein production and purification. *Nat Methods* **5**, 135-146, doi:10.1038/nmeth.f.202 (2008).

247 Vagenende, V., Yap, M. G. & Trout, B. L. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry* **48**, 11084-11096, doi:10.1021/bi900649t (2009).

248 Stetsenko, A. & Guskov, A. An overview of the top ten detergents used for membrane protein crystallization. *Crystals* **7**, doi:10.3390/cryst7070197 (2017).

249 Breyton, C. *et al.* Assemblies of lauryl maltose neopentyl glycol (LMNG) and LMNG-solubilized membrane proteins. *Biochim Biophys Acta Biomembr* **1861**, 939-957, doi:10.1016/j.bbamem.2019.02.003 (2019).

250 Chaptal, V. *et al.* Quantification of detergents complexed with membrane proteins. *Sci Rep* **7**, 41751, doi:10.1038/srep41751 (2017).

251 Gohon, Y. *et al.* Well-defined nanoparticles formed by hydrophobic assembly of a short and polydisperse random terpolymer, Amphipol A8-35. *Langmuir* **22**, 1281-1290, doi:10.1021/la052243g (2006).

252 Zoonens, M., Zito, F., Martinez, K. L. & Popot, J.-L. in *Membrane Proteins Production for Structural Analysis* Ch. 7, 173-203 (2014).

253 Zoonens, M. & Popot, J. L. Amphipols for each season. *J Membr Biol* **247**, 759-796, doi:10.1007/s00232-014-9666-8 (2014).

254 Popot, J. L. *et al.* Amphipols from A to Z. *Annu Rev Biophys* **40**, 379-408, doi:10.1146/annurev-biophys-042910-155219 (2011).

255 Jeffries, C. M. *et al.* Small-angle X-ray and neutron scattering. *Nat Rev Methods Primers* **1**, 70, doi:10.1038/s43586-021-00064-9 (2021).

256 Meisburger, S. P., Thomas, W. C., Watkins, M. B. & Ando, N. X-ray scattering studies of protein structural dynamics. *Chem Rev* **117**, 7615-7672, doi:10.1021/acs.chemrev.6b00790 (2017).

257 Graewert, M. A. *et al.* Adding size exclusion chromatography (SEC) and light scattering (LS) devices to obtain high-quality small angle X-ray scattering (SAXS) data. *Crystals* **10**, doi:10.3390/cryst10110975 (2020).

258 Hutin, S., Brennich, M., Maillot, B. & Round, A. Online ion-exchange chromatography for small-angle X-ray scattering. *Acta Crystallogr D Struct Biol* **72**, 1090-1099, doi:doi:10.1107/S2059798316012833 (2016).

259 Gabel, F. in *Biological Small Angle Scattering: Techniques, Strategies and Tips* (eds Barnali Chaudhuri, Inés G. Muñoz, Shuo Qian, & Volker S. Urban) 201-214 (Springer Singapore, 2017).

260 Le Roy, A. *et al.* in *Methods in Enzymology* Vol. 562 (ed James L. Cole) Ch. Chapter Twelve, 257-286 (Academic Press, 2015).

261 Maric, S. *et al.* Time-resolved small-angle neutron scattering as a probe for the dynamics of lipid exchange between human lipoproteins and naturally derived membranes. *Sci Rep* **9**, 7591, doi:10.1038/s41598-019-43713-6 (2019).

262 Ebel, C., Breyton, C. & Martel, A. in *Biophysics of Membrane Proteins: Methods and Protocols* (eds Vincent L. G. Postis & Adrian Goldman) 147-175 (Springer US, 2020).

263 Dias Mirandela, G. *et al.* Merging in-solution X-ray and neutron scattering data allows fine structural analysis of membrane–protein detergent complexes. *J Phys Chem Lett* **9**, 3910-3914, doi:10.1021/acs.jpclett.8b01598 (2018).

264 Birch, J. *et al.* Changes in membrane protein structural biology. *Biology* **9** (2020).

265 Dhandapani, G., Wachtel, E., Das, I., Sheves, M. & Patchornik, G. Purification of antibody fragments via interaction with detergent micellar aggregates. *Sci Rep* **11**, 11697, doi:10.1038/s41598-021-90966-1 (2021).

266 Nizak, C., Moutel, S., Goud, B. & Perez, F. in *Methods in Enzymology* Vol. 403 *GTPases Regulating Membrane Targeting and Fusion* 135-153 (2005).

267 Shi, X. & Jarvis, D. L. Protein N-glycosylation in the baculovirus-insect cell system. *Curr Drug Targets* **8**, 1116-1125, doi:10.2174/138945007782151360 (2007).

268 Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc Natl Acad Sci U S A* **99**, 13419-13424, doi:10.1073/pnas.212519299 (2002).

269 Chang, V. T. *et al.* Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15**, 267-273, doi:10.1016/j.str.2007.01.011 (2007).

270 Mesters, J. R. & Hilgenfeld, R. Protein glycosylation, sweet to crystal growth? *Cryst Growth Des* **7**, 2251-2253, doi:10.1021/cg7006843 (2007).

271 Coleman, J. A., Green, E. M. & Gouaux, E. X-ray structures and mechanism of the human serotonin transporter. *Nature* **532**, 334-339, doi:10.1038/nature17629 (2016).

272 Fay, J. F. *et al.* Cryo-EM visualization of an active high open probability CFTR anion channel. *Biochemistry* **57**, 6234-6246, doi:10.1021/acs.biochem.8b00763 (2018).

273 Huang, Y. *et al.* Structural basis for the Mg^{2+} recognition and regulation of the CorC Mg^{2+} transporter. *Sci Adv* **7**, eabe6140, doi:10.1126/sciadv.abe6140.

274 Hopf, T. A. *et al.* Mutation effects predicted from sequence co-variation. *Nature Biotechnology* **35**, 128-135, doi:10.1038/nbt.3769 (2017).

275 Hopf, T. A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, e03430, doi:10.7554/eLife.03430 (2014).

276 Ashok, Y., Nanekar, R. & Jaakola, V.-P. Defining thermostability of membrane proteins by western blotting. *Protein Eng Des Sel* **28**, 539-542, doi:10.1093/protein/gzv049 (2018).

277 Garretson, T. A., Shang, H., Schulz, A. K., Donohue, B. V. & Cheng, X. W. Expression- and genomic-level changes during passage of four baculoviruses derived from bacmids in permissive insect cell lines. *Virus Res* **256**, 117-124, doi:10.1016/j.virusres.2018.08.009 (2018).

278 Kim, S. R. *et al.* Thirty Novel Genetic Variations in the SLC29A1 Gene Encoding Human Equilibrative Nucleoside Transporter 1 (hENT1). *Drug Metab Pharmacokinet* **21**, 248-256, doi:10.2133/dmpk.21.248 (2006).

279 Parker, J. L. & Newstead, S. Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1. *Nature* **507**, 68-72, doi:10.1038/nature13116 (2014).

280 Ehrnstorfer, I. A., Geertsma, E. R., Pardon, E., Steyaert, J. & Dutzler, R. Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport. *Nat Struct Mol Biol* **21**, 990-996, doi:10.1038/nsmb.2904 (2014).

281 Newstead, S. *et al.* Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J* **30**, 417-426, doi:10.1038/emboj.2010.309 (2011).

282 Singh, S. K. & Pal, A. *Biophysical approaches to the study of LeuT, a prokaryotic homolog of neurotransmitter sodium symporters*. 1 edn, Vol. 557 (Elsevier Inc., 2015).

283 Reyes, G. *et al.* Characterization of mammalian equilibrative nucleoside transporters (ENTs) by mass spectrometry. *Protein Expr Purif* **73**, 1-9, doi:10.1016/j.pep.2010.04.008 (2010).

284 Arastu-Kapur, S., Arendt, C. S., Purnat, T., Carter, N. S. & Ullman, B. Second-site suppression of a nonfunctional mutation within the *Leishmania donovani* inosine-guanosine transporter. *J Biol Chem* **280**, 2213-2219, doi:10.1074/jbc.M408224200 (2005).

285 Valdés, R., Arastu-Kapur, S., Landfear, S. M. & Shinde, U. An ab initio structural model of a nucleoside permease predicts functionally important residues. *J Biol Chem* **284**, 19067-19076, doi:10.1074/jbc.M109.017947 (2009).

286 Valdés, R., Elferich, J., Shinde, U. & Landfear, S. M. Identification of the intracellular gate for a member of the equilibrative nucleoside transporter (ENT) family. *J Biol Chem* **289**, 8799-8809, doi:10.1074/jbc.M113.546960 (2014).

287 Valdés, R., Shinde, U. & Landfear, S. M. Cysteine cross-linking defines the extracellular gate for the *Leishmania donovani* nucleoside transporter 1.1 (LdNT1.1). *J Biol Chem* **287**, 44036-44045, doi:10.1074/jbc.M112.414433 (2012).

288 Vasudevan, G., Ullman, B. & Landfear, S. M. Point mutations in a nucleoside transporter gene from *Leishmania donovani* confer drug resistance and alter substrate selectivity. *Proc Natl Acad Sci U S A* **98**, 6092-6097, doi:10.1073/pnas.101537298 (2001).

289 Girke, C. *et al.* High yield expression and purification of equilibrative nucleoside transporter 7 (ENT7) from *Arabidopsis thaliana*. *Biochim Biophys Acta Gen Subj* **1850**, 1921-1929, doi:10.1016/j.bbagen.2015.06.003 (2015).

290 Boswell-Casteel, R. C. *et al.* FUN26 (function unknown now 26) protein from *Saccharomyces cerevisiae* is a broad selectivity, high affinity, nucleoside and nucleobase transporter. *J Biol Chem* **289**, 24440-24451, doi:10.1074/jbc.M114.553503 (2014).

291 Li, G., Yang, X., Shu, Z. H., Chen, X. Y. & Wang, Y. Q. Consecutive spawnings of chinese amphioxus, *Branchiostoma belcheri*, in captivity. *PLoS ONE* **7**, doi:10.1371/journal.pone.0050838 (2012).

292 Oka, T., Ekino, K., Fukuda, K. & Nomura, Y. Draft genome sequence of the formaldehyde-resistant fungus *Byssochlamys spectabilis* No. 5 (anamorph *Paecilomyces variotii* No. 5) (NBRC109023). *Genome Announc* **2**, doi:10.1128/genomea.01162-13 (2014).

293 *Kumamushi genome project*, <http://kumamushi.org/> (

294 Duhra, J. K., Fan, L., Gill, V. S., Lee, J. J. & Yeon, J. The effect of temperature on the population size of *Tetrahymena thermophila*. *The Expedition* **3** (2014).

295 Eisen, J. A. *et al.* Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* **4**, e286, doi:10.1371/journal.pbio.0040286 (2006).

296 Hashimoto, T. *et al.* Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nat Commun* **7**, 12808, doi:10.1038/ncomms12808 (2016).

297 Maheshwari, R., Bharadwaj, G. & Bhat, M. K. Thermophilic fungi: Their phyiology and enzymes. *Microbiol Mol Biol Rev* **64**, 461-488, doi:10.1128/MMBR.64.3.461-488.2000 (2000).

298 Zhong, J. *et al.* Complete mitochondrial genomes defining two distinct lancelet species in the West Pacific Ocean. *Mar Biol Res* **5**, 278-285, doi:10.1080/17451000802430817 (2009).

299 Chatzikyriakidou, Y., Ahn, D. H., Nji, E. & Drew, D. The GFP thermal shift assay for screening ligand and lipid interactions to solute carrier transporters. *Nat Protoc* **16**, 5357-5376, doi:10.1038/s41596-021-00619-w (2021).

300 Drew, D. *et al.* GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nature Protocols* **3**, 784-798, doi:10.1038/nprot.2008.44 (2008).

301 Bai, J. *et al.* A gene optimization strategy that enhances production of fully functional P-glycoprotein in *Pichia pastoris*. *PLoS One* **6**, e22577, doi:10.1371/journal.pone.0022577 (2011).

302 Xu, Y. *et al.* Codon usage bias regulates gene expression and protein conformation in yeast expression system *P. pastoris*. *Microb Cell Fact* **20**, 91, doi:10.1186/s12934-021-01580-9 (2021).

303 Karbalaei, M., Rezaee, S. A. & Farsiani, H. *Pichia pastoris*: A highly successful expression system for optimal synthesis of heterologous proteins. *J Cell Physiol* **235**, 5867-5881, doi:10.1002/jcp.29583 (2020).

304 SenGupta, D. J. *et al.* A single glycine mutation in the equilibrative nucleoside transporter gene, hENT1, alters nucleoside transport activity and sensitivity to nitrobenzylthioinosine. *Biochemistry* **41**, 1512-1519, doi:10.1021/bi015833w (2002).

305 Vickers, M. F. *et al.* Functional production and reconstitution of the human equilibrative nucleoside transporter (hENT1) in *Saccharomyces cerevisiae*. Interaction of inhibitors of nucleoside transport with recombinant hENT1 and a glycosylation-defective derivative (hENT1/N48Q). *Biochem J* **339 (Pt 1**, 21-32, doi:10.1042/0264-6021:3390021 (1999).

306 Majd, H. *et al.* Screening of candidate substrates and coupling ions of transporters by thermostability shift assays. *eLife* **7**, e38821, doi:10.7554/eLife.38821 (2018).

307 Houbraken, J., Varga, J., Rico-Munoz, E., Johnson, S. & Samson, R. A. Sexual reproduction as the cause of heat resistance in the food spoilage fungus *Byssochlamys spectabilis* (anamorph *Paecilomyces variotii*). *Appl Environ Microbiol* **74**, 1613-1619, doi:10.1128/AEM.01761-07 (2008).

308 Samson, R. A., Houbraken, J., Varga, J. & Frisvad, J. C. Polyphasic taxonomy of the heat resistant ascomycete genus *Byssochlamys* and its *Paecilomyces* anamorphs. *Persoonia* **22**, 14-27, doi:10.3767/003158509X418925 (2009).

309 Tournas, V. Heat-resistant fungi of importance to the food and beverage industry. *Crit Rev Microbiol* **20**, 243-263, doi:10.3109/10408419409113558 (1994).

310 Kondo, T., Morikawa, Y. & Hayashi, N. Purification and characterization of alcohol oxidase from *Paecilomyces variotii* isolated as a formaldehyde-resistant fungus. *Appl Microbiol Biotechnol* **77**, 995-1002, doi:10.1007/s00253-007-1237-9 (2008).

311 Battestin, V. & Macedo, G. A. Purification and biochemical characterization of tannase from a newly isolated strain of *Paecilomyces Variotii*. *Food Biotechnol* **21**, 207-216, doi:10.1080/08905430701533588 (2007).

312 Job, J., Sukumaran, R. K. & Jayachandran, K. Production of a highly glucose tolerant β-glucosidase by *Paecilomyces variotii* MG3: optimization of fermentation conditions using Plackett–Burman and Box–Behnken experimental designs. *World J Microbiol Biotechnol* **26**, 1385-1391, doi:10.1007/s11274-010-0311-0 (2010).

313 Michelin, M. *et al.* Purification and biochemical characterization of a thermostable extracellular glucoamylase produced by the thermotolerant fungus *Paecilomyces variotii*. *J Ind Microbiol Biotechnol* **35**, 17-25, doi:10.1007/s10295-007-0261-1 (2008).

314 Bellanger, A. P. *et al. Paecilomyces variotii* fungemia in a patient with lymphoma needing liver transplant. *Mycopathologia* **182**, 761-765, doi:10.1007/s11046-017-0131-y (2017).

315 Salle, V. *et al. Paecilomyces variotii* fungemia in a patient with multiple myeloma: case report and literature review. *J Infect* **51**, e93-95, doi:10.1016/j.jinf.2004.09.006 (2005).

316 Feldman, R., Cockerham, L., Buchan, B. W., Lu, Z. & Huang, A. M. Treatment of *Paecilomyces variotii* pneumonia with posaconazole: case report and literature review. *Mycoses* **59**, 746-750, doi:10.1111/myc.12525 (2016).

317 Houbraken, J., Verweij, P. E., Rijs, A. J., Borman, A. M. & Samson, R. A. Identification of *Paecilomyces variotii* in clinical samples and settings. *J Clin Microbiol* **48**, 2754-2761, doi:10.1128/JCM.00764-10 (2010).

318 Walsh, T. J. *et al.* Infections due to emerging and uncommon medically important fungal pathogens. *Clin Microbiol Infect* **10**, 48-66, doi:10.1111/j.1470-9465.2004.00839.x (2004).

319 Chamilos, G. & Kontoyiannis, D. P. Voriconazole-resistant disseminated *Paecilomyces variotii* infection in a neutropenic patient with leukaemia on voriconazole prophylaxis. *J Infect* **51**, e225-228, doi:10.1016/j.jinf.2005.02.005 (2005).

320 Drogari-Apiranthitou, M. *et al.* In vitro antifungal susceptibility of filamentous fungi causing rare infections: synergy testing of amphotericin B, posaconazole and anidulafungin in pairs. *J Antimicrob Chemother* **67**, 1937-1940, doi:10.1093/jac/dks137 (2012).

321 Castelli, M. V. *et al.* Susceptibility testing and molecular classification of *Paecilomyces* spp. *Antimicrob Agents Chemother* **52**, 2926-2928, doi:10.1128/AAC.00538-08 (2008).

322 Vickers, M. F., Yao, S. Y., Baldwin, S. A., Young, J. D. & Cass, C. E. Nucleoside transporter proteins of *Saccharomyces cerevisiae*. Demonstration of a transporter (FUI1) with high uridine selectivity in plasma membranes and a transporter (FUN26) with broad nucleoside selectivity in intracellular membranes. *J Biol Chem* **275**, 25931-25938, doi:10.1074/jbc.M000239200 (2000).

323 Alexandrov, A. I., Mileni, M., Chien, E. Y., Hanson, M. A. & Stevens, R. C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16**, 351-359, doi:10.1016/j.str.2008.02.004 (2008).

324 Cecchetti, C. *et al.* A novel high-throughput screen for identifying lipids that stabilise membrane proteins in detergent based solution. *PLoS One* **16**, e0254118, doi:10.1371/journal.pone.0254118 (2021).

325 Rodrigues, M. L. The multifunctional fungal Ergosterol. *mBio* **9**, doi:10.1128/mBio.01755-18 (2018).

326 Ghannoum Mahmoud, A. & Rice Louis, B. Antifungal agents: Mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. *Clin Microbiol Rev* **12**, 501-517, doi:10.1128/CMR.12.4.501 (1999).

327 Wu, Z. *et al.* Insight into the nucleoside transport and inhibition of human ENT1. *Curr Res Struct Biol* **4**, 192-205, doi:10.1016/j.crstbi.2022.05.005 (2022).

328 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).

329 Chakrapani, S. in *Methods in Enzymology* Vol. 557 279-306 (Elsevier Inc., 2015).

330 Fowler, P. W. *et al.* Gating topology of the proton-coupled oligopeptide symporters. *Structure* **23**, 290-301, doi:10.1016/j.str.2014.12.012 (2015).

331 Kapsalis, C. *et al.* Allosteric activation of an ion channel triggered by modification of mechanosensitive nano-pockets. *Nat Commun* **10**, 4619, doi:10.1038/s41467-019-12591-x (2019).

Appendix

Chapter 2 supplementary information

Appendix Table 1 Protein sequences used in this work, with NCBI accession numbers and protein parameters

Sequence source	FASTA protein sequence	Protein parameters
*Branchiostoma belcheri; 'Bb*ENT' >XP_019625696.1 PREDICTED: ENT1-like isoform X1	PGARPPRDRYNAVYIIFFLLGLGMLLPWNIFITANMYFRKKFTDSRYEDTFENYFSVASMVPNVV FQLLNIFVAHRVPLNVRMVVPLFTMLGCFILTAAMVWVNTGKDITHDHSPTLSTNGFFLITIFTV VIINLASAIMQAGSFGVAGKFPGKYTQAIMSGQALAGVFSALASIFSLAAGGDPIHSGFGYFLTA VAAILVAMVSYLLLKRSEYARYYLQNTRVPSVNEDESIAGVNGTAPAAGVPDLTGSGAENMQVL GSSSYLQIFRKIWVPAVCVMYTFMVTLSIFPSVSSLIESVSKSDGSKWTGEFFIPVTCFLFFNLSDL AGRVIAGAVQFPKEKSILLPILVLLRTGFMPLFMLCNAQPEEFSRRLPVVFNSDAYPIVFMVLFGV SNGYLGSLCMMYGPRLVSAEHAETAGITMSAFLTLGLGLGAAFSFALTASI	**Number of amino acids:** 443 **Molecular weight:** 48207.52 **Theoretical pI:** 8.58 **Identical to hENT1:** 39.1% **Ext. coefficient:** 45840 **Abs 0.1% (=1 g/l):** 0.951
*Byssochlamys spectabilis; 'Bs*ENT' >GAD95619.1 nucleoside transporter family	EARQPASGFSWLEYSIFLLLGIAMLWAWNMFLAAAPYFHHRFRSSEWASTHFQPSILSVSTVTN LGSVIVLAKSQRNASYEGRISLSLVINVAVFTLLAFSVLATNASVGAYFFFLMVMVFGASLATGIN QNGVFAYVSRFGREEYVQAIMAGQGVAGVLPCIVQIISVLAVPEKKQGPPGTGEQESPKSAFAY FITASLISAVTLVAFFYLVRQQCRRGPKLVPDDEAEPDLTGHKTVGLWVLFKKLHWMALAIFLCFA LTMVFPVFAAEIESVRDPASAPRLFQPAIFIPLAFLIWNLGDLLGRISVLIPSLSLTHYPWALFVIAVS RIVVIPLYFLCNVHGKGAVVNSDAFYLAVQFLFGITNGYLGSSCMMGAGQWVAVEEREAAGGF MTLMLVGGLTAGSLLSFLAASA	**Number of amino acids:** 414 **Molecular weight:** 44958.74 **Theoretical pI:** 8.30 **Identical to hENT1:** 26.2% **Ext. coefficient:** 67380 **Abs 0.1% (=1 g/l):** 1.499
*Gallus gallus; 'Gg*ENT' >XP_419491.1 PREDICTED: equilibrative nucleoside transporter 1	MTTRDGPQDRYKAVWLIFFILGLGTLLPWNFFMTARQYFINRLADPQNISHLSNQTSVGTASDL SYLQSMFDNFMTLCSMVPLLIFTCLNSFIHQRIPQQIRISGSLVAIGLVFLITAIMVKVTMDPLPFF VFTMVSIVFINSFGAMLQGSLFGLAGLLPASYTAPIMSGQGLAGIFAALAMIISISIGAQQPESYIG YFTTACVAILLAIFSYVLLPRMDFFRYYSMKDKTEYHVCNAELETKRDLIKKDEPNGMEQNNSKII PVHNPDEKPSVISIFKKLWVMAVSVCLVFTVTIGVFPSITAKVSTTLGKESKWDLYFVSVSCFLIFN VFDWMGRSLTALFTWPGKDSCLLPVMVVLRVIFIPLFMLCNVQPRNHLPVIFSHDAWYIIFMIF FSISNGYLASLCMCFGPKKVLAHEAETAGAVMAFFLTLGLALGAAISFLFQILI	**Number of amino acids:** 448 **Molecular weight:** 50058.43 **Theoretical pI:** 8.45 **Identical to hENT1:** 60.4% **Ext. coefficient:** 57870 **Abs 0.1% (=1 g/l):** 1.153
hENT1 (SLC29A1) NP_001071643.1 Uniprot: Q99808	MTTSHQPQDRYKAVWLIFFMLGLGTLLPWNFFMTATQYFTNRLDMSQNVSLVTAELSKDAQA SAAPAAPLPERNSLSAIFNNVMTLCAMLPLLLFTYLNSFLHQRIPQSVRILGSLVAILLVFLITAILVK VQLDALPFFVITMIKIVLINSFGAILQGSLFGLAGLLPASYTAPIMSGQGLAGFFASVAMICAIASG SELSESAFGYFITACAVIILTIICYLGLPRLEFYRYYQQLKLEGPGEQETKLDLISKGEEPRAGKEESGV SVSNSQPTNESHSIKAILKNISVLAFSVCFIFTITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLTF NIFDWLGRSLTAVFMWPGKDSRWLPSLVLARLVFVPLLLLCNIKPRRYLTVVFEHDAWFIFFMAA FAFSNGYLASLCMCFGPKKVKPAEAETAGAIMAFFLCLGLALGAVFSFLFRAIV	**Number of amino acids:** 456 **Molecular weight:** 50219.45 **Theoretical pI:** 8.62 **Ext. coefficient:** 56380 **Abs 0.1% (=1 g/l):** 1.123
*Ramazzottius varieornatus; 'Rv*ENT' >GAU96053.1 hypothetical protein RvY_07553	FHKSAPADRLSFVYIIMLIHGFGTLLPWNVIISSFGYWDFKLDPLNNDTTNPTPQYPTGDFFKHLT LANQFPNVFLNALNLLVTLGGSLTIRIVVSVAILVLVMAETLGLAIADSSQWTWTFVVVTLVFVVV LGTANGIYQNSLYGLAASFPSRYTNAIIIGNVVGPCFSCMSISVFPGKEDQTKSGTLYFSCALFILLV AFGSYFLLPLSKFYRYYINRPTVPSERPSRMTFGERLSTLRRVFGMMRLECWNVFFVFFVTLSVF PNIMVNVQRYSMLDDEYFRPVITILNFSVFAFIGNLIPFACSAPGPRYVWIPILARIIFFPFFLMCN YNEIGYERRFDVWFTNDYVYAVGGMLLAVTSGYFSSLAMMYASSKVKDSRLAPMAGQLAGFF LVLGIFAGLMFSWLLPLTV	**Number of amino acids:** 413 **Molecular weight:** 46565.77 **Theoretical pI:** 8.94 **Identical to hENT1:** 26.8% **Ext. coefficient:** 73800 **Abs 0.1% (=1 g/l):** 1.585
*Tetrahymena thermophile; 'Tt*ENT' >XP_001011556.1 equilibrative nucleoside transporter family protein	ELAKPLPPVKFWYKITFVFLGIASLAGWNAMLTAFDFFGAKYPKDQGYLDITFYFPIPIMITNFFA GLACPALARRFSYNQRIAYLSVAVCCFLITITLIAIFYNTKAGFWISFTLLFFQGFIESVVTNSLIALAG MISHEINAIYWTCTAASGLVMNFIRLIALGAAGDTPSSMNVCTAIYFAFACLIYIVSASMQAAFTK TEYFKALEHRHNIKSKIENREIEIDMARMMKEKLAAENNNANTGSDNQLKTEQALSQVNLEQQ KKSKKSGLVAKLLQNSFIQYLIYLSQVFKYAGAIPVFLVFIYIQTFMMFPGVSIFQKPTYTIIPYPYAA VWMITCYNFGDLVGKYLGSVKALEKLYFIYCVVMLRFVYYVLFLMTANEKGGENFQNDVFAWT NQLMFAITNGFCTTGLMNLGPRKCKDPKIINLINFIGGFSITFGIAIGTFLALPLAKE	**Number of amino acids**: 455 **Molecular weight**: 51236.46 **Theoretical pI:** 9.08 **Identical to hENT1:** 20.7% **Ext. coefficient:** 68760 **Abs 0.1% (=1 g/l):** 1.342

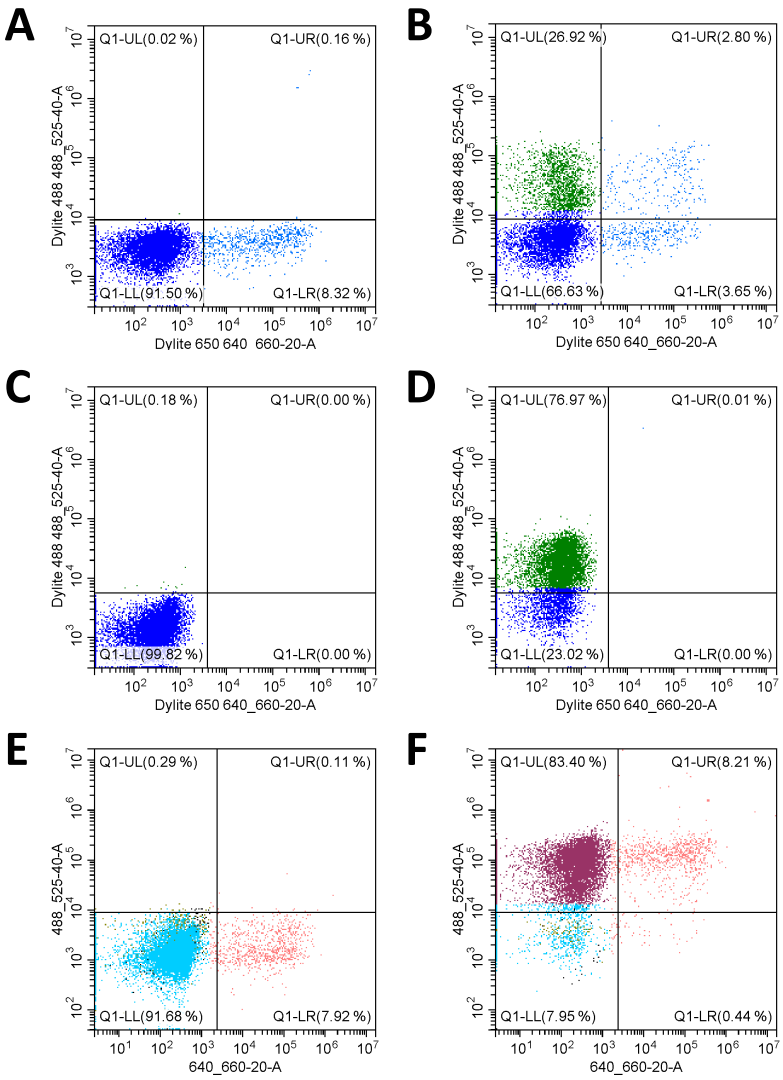
Appendix Table 2 Parameters for hENT1 selected nanobodies

Nanobody	MW	Theoretical pI	Ext. coefficient	Abs 0.1%
CA14400	14344.98	8.01	19940	1.390
CA14401	15011.67	8.00	28420	1.893
CA14402	13866.37	6.76	16960	1.223
CA14404	13779.39	6.38	19940	1.447
CA14405	14881.37	6.31	21680	1.457

Chapter 3 supplementary information

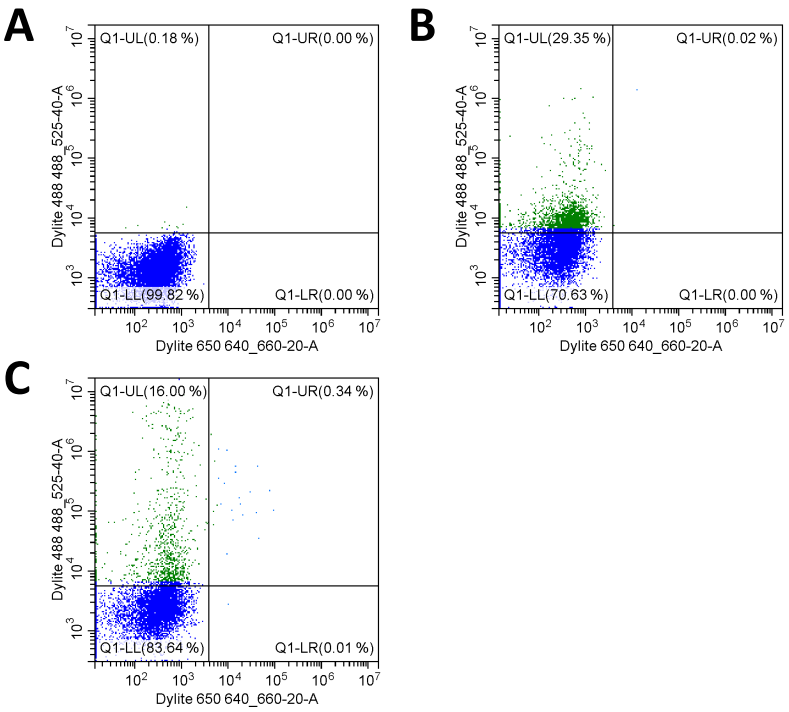
Appendix Table 3 Theoretical amounts of hENT1 based on percentage losses determined by anti-hENT1 Western blot and the concentration of the final yield of hENT1, as estimated by $A280_{nm}$.

	Purification					
mg/L hENT1	**1**	**2**	**3**	**4**	**5**	**6**
Total	0.91	0.93	1.06	0.94	1.03	0.90
Solubilised	0.25	0.20	0.38	0.25	0.34	0.26
Applied to resin	0.14	0.14	0.22	0.16	0.23	0.15
Applied to SEC	0.06	0.05	0.10	0.06	0.10	0.07
Yield	0.03	0.03	0.05	0.03	0.05	0.04



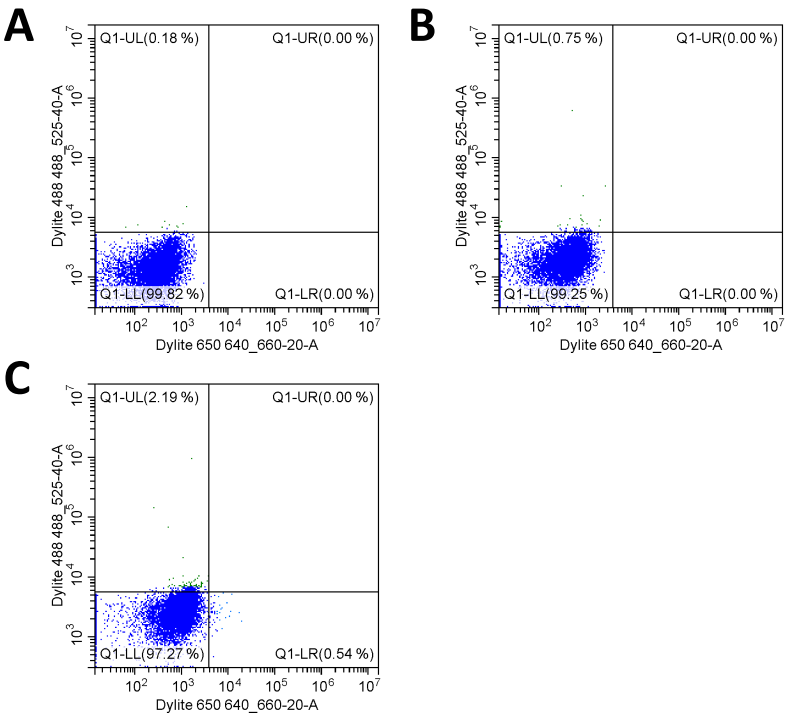
Appendix Figure 1 Flow cytometry analysis of the yeast display of a synthetic nanobody library and the membrane protein MaTap, an MFS transporter, in DDM.

(A and B) A: Induced cells following incubation with anti-HA labelled with DyLight 650 represent nanobody expression, which is demonstrated by the increase in population within Q1-LR. B: Induced cells following incubation with anti-HA labelled with DyLight 650 and MaTap labelled with DyLight 488, represent non-specific binding, as demonstrated by increase in the population within Q1-UL. (C and D) C: An induced but unlabelled cell sample is a negative control, allows for the establishment of the quadrants for analysis of specific intensities. D: Induced cells following blocking with 1% (w/v) BSA and incubation with MaTap labelled with DyLight 488 determine non-specific binding, as demonstrated by increase to the population within Q1-UL. There is a significant overall shift in population upwards, suggesting an overall increase in the intensity of DyLight 488 signal across the total cell population. (E and F) E: Induced cells following incubation with anti-HA labelled with DyLight 650 determine nanobody expression, as demonstrated by the increase in population within Q1-LR. F: induced cells following blocking with 3 x CMC DDM and incubation with MaTap labelled with DyLight 488, determine non-specific binding. There is a significant overall shift in population upwards, suggesting an overall increase in the intensity of DyLight 488 signal across the total cell population. In both conditions with blocking steps, with BSA and DDM, significant non-specific binding is observed.



Appendix Figure 2 Flow cytometry analysis of the yeast display of a synthetic nanobody library and the membrane bound pyrophosphatase, TmPPase, in DDM.

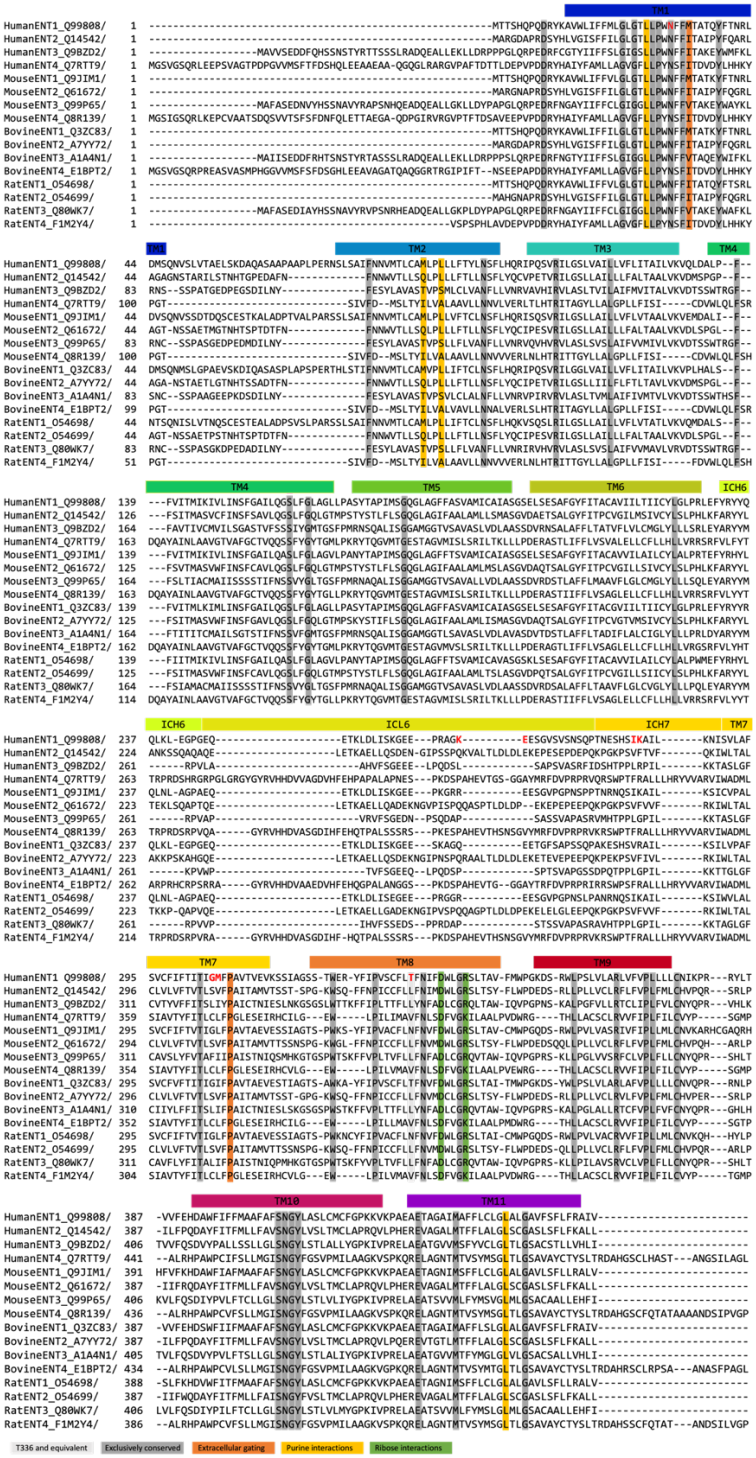
(A) An induced but unlabelled cell sample as a negative control allows for the establishment of the quadrants for analysis of specific intensities. (B) Induced cells following blocking with 1% (w/v) BSA and incubation with TmPPase labelled with DyLight 488, determine non-specific binding, as demonstrated by increase to the population within Q1-UL. (C) Induced cells following blocking with 3 x CMC DDM and incubation with TmPPase labelled with DyLight 488, determine non-specific binding, as demonstrated by increase to the population within Q1-UL. In both conditions with blocking steps, with BSA and DDM, significant non-specific binding is observed.



Appendix Figure 3 Flow cytometry analysis of the yeast display of a synthetic nanobody library and the soluble protein alcohol dehydrogenase (ADH).

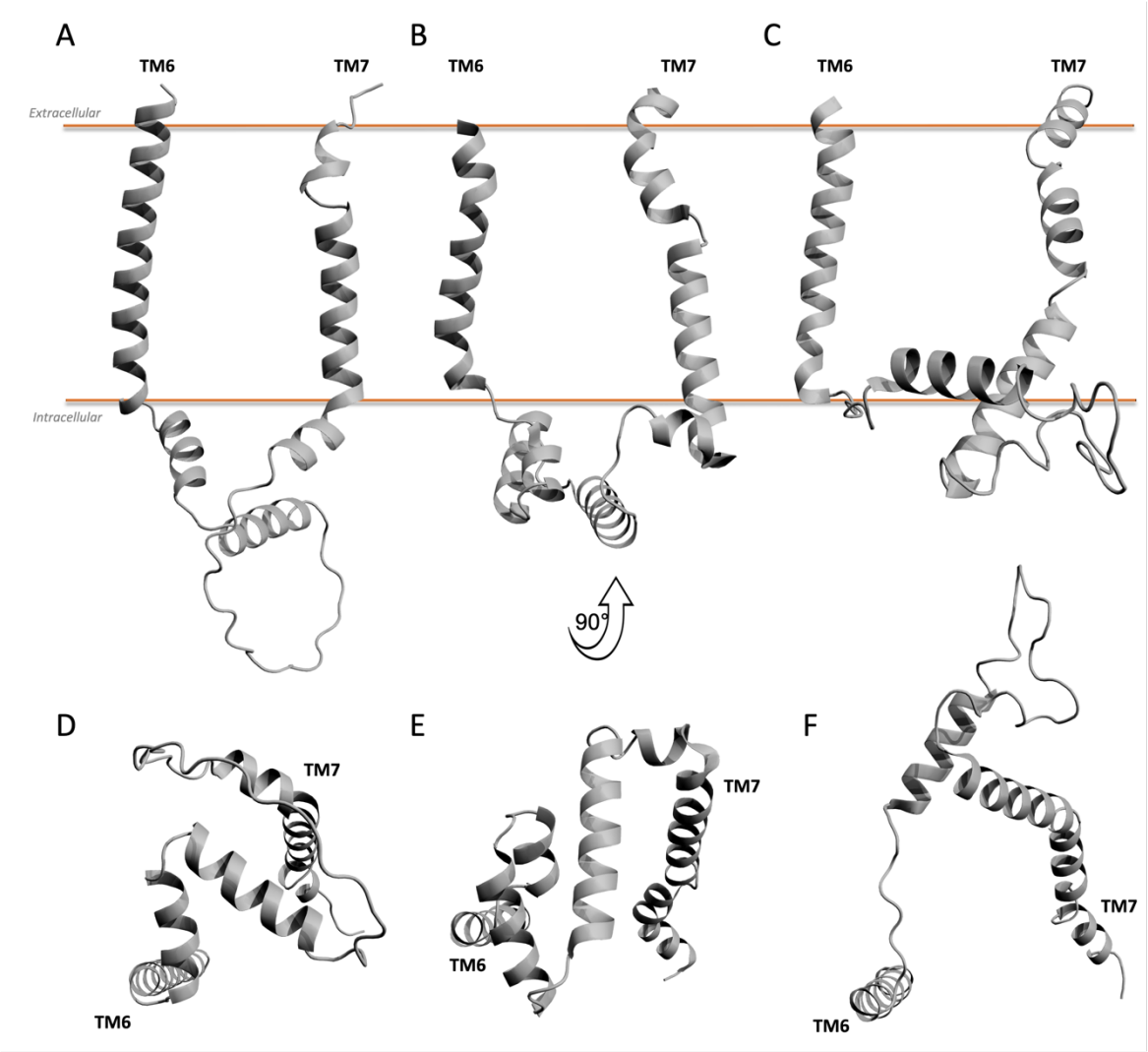
(A) Induced but unlabelled cell sample as a negative control allows for the establishment of the quadrants for analysis of specific intensities. (B) Induced cells following blocking with 1% (w/v) BSA and incubation with ADH labelled with DyLight 488 determine non-specific binding. (C) Induced cells following blocking with 3 x CMC DDM and incubation with ADH labelled with DyLight 488 determine non-specific binding. There is minimal non-specific binding observed in any of the ADH conditions, including those blocked with BSA and DDM.

Chapter 4 supplementary information



Appendix Figure 4 Multiple sequence alignment of select mammalian ENTs with UniProt identifiers.

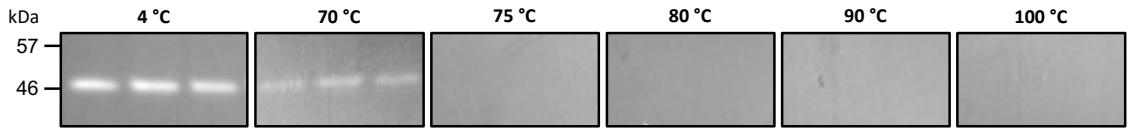
Residues mutated in this study are displayed in red, T336 and equivalent residues are highlighted in light grey. Residues that are exclusively conserved are highlighted in dark grey. Residues involved in extracellular gating interactions are highlighted in orange. Residues discussed involved in interactions with the purine and ribose moiety of NBMPR are highlighted in yellow and green, respectively.



Appendix Figure 5 Representations of differing ICL6 domains of MFS members.

Representative structures of the TM6, ICL and TM7 of (A, D) Alphafold [155] predicted full-length wild type hENT1, (B, E) the mammalian sugar transporters, Glut3 (PDB: 4ZW9) and (C, F) the mammalian peptide transporter PepT2 (PDB: 7NQK).

Chapter 5 supplementary information



Appendix Figure 6 The survival of *Bs*ENT linked GFP signal at >70 °C.

SDS-PAGE of NHGV-*Bs*ENT with fluorescence imagining following temperature challenge >70 °C (2.4.4).