Structural, Functional and Computational Studies on the SLC7 Family of Amino Acid Transporters



Dimitrios Kolokouris

Department of Biochemistry

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy.

Brasenose College – Hilary 2023

Copyright © 2023 Dimitrios Kolokouris

All rights reserved.

At the time of submission, this D.Phil thesis has a word count of 49,929 and therefore is in line with the Medical Sciences Board permitted word limit (exclusive of bibliography, appendices, diagrams, and tables).

The *viva voce* examination was held on 08.02.2023 by Prof. Mark S.P. Sansom (University of Oxford) and Prof. Manuel Palacín (University of Barcelona). A copy of this thesis has been deposited in the Bodleian Library, Oxford.

Αρχάς είναι των όλων ατόμους και κενόν, τα δ' άλλα πάντα νενομίσθαι.

The beginning of everything is atoms and void, and everything else is perception.

— Democritus, 470-370 BC

Abstract

Amino acids are central components of the cell as protein building blocks, nitrogen sources and metabolic and cell signalling regulators. As such amino acid concentration in the cytosol and cellular compartments must be regulated. The SLC7 family of amino acid transporters provides a central gateway of amino acids to the cell. There is increasing understanding of mammalian SLC7 members, primarily, through structural and functional studies on prokaryotic and, recently, some human homologues. However, a member of the cationic amino acid transporter (CAT) subfamily remains of unknown function, prompting us to uncover its structure-function relationship. This thesis reports the first cryoEM structure of a eukaryotic SLC7A4 homologue, from *Arabidopsis thaliana* (*At*CAT4), closely related to the mammalian SLC7A4 by conserving a key acidic residue, and other CATs responsible for the transport of L-arginine, L-lysine, and L-ornithine into the cell. *At*CAT4 structural elucidation required the selection and generation of a nanomolar affinity sybody binder. The structure was solved in the apo outward-open state and showed *At*CAT4 has an APC fold with two additional TM helices previously unobserved, likely involved in homodimerisation. Structural comparisons revealed an intricate electrostatic gating mechanism, which rearranges upon substrate binding, uncovering early steps of the transport cycle. *At*CAT4 binding assays and MD simulation showed recognition of L-arginine, L-lysine, and L-ornithine with differential preferences for each. Human SLC7A4 however did not transport cationic amino acids, but preliminary data support a role for L-leucine transport. Structural analyses and mutagenesis suggested the root of the homologue discrepancy to be two bulky, hydrophobic residues absent from *At*CAT4 and other CATs. The current structure serves as a template to further understand mammalian cationic amino acid transport via the SLC7 family. *At*CAT4 was solved with a steroid lipid bound, similar to other SLC7s. With multiscale-MD and *in vitro* transport assays we suggest two cholesterol binding sites for HATs which stabilise transporter SLC7-SLC3 heterodimerisation and conformational state.

Acknowledgements

First and foremost, I would like to sincerely thank my supervisors, Prof. Simon Newstead and Prof. Philip Biggin. For the past four years, I experienced a great journey of exploration. For every step of the process my supervisors were overlooking my shoulder, making sure careful steps are taken in my training, research, and this thesis; all parts of the meticulous scientific training they have provided me. When pedagogical, they allowed me the freedom to explore my ideas as an independent researcher and provisioned me with all the resources necessary. For their guidance and continuous support, I am truly grateful. In particular, I would like to thank Simon Newstead for his excellent supervision, ideas and chance to work in his lab. His insight into membrane transport biology has been instrumental in the content of this thesis.

I would like to extend special gratitude to Dr Joanne Parker and Dr Takafumi Kato, whose expertise in membrane protein biochemistry and Cryo-EM, training, and fruitful discussions and advice were indispensable in the realisation of this thesis. Furthermore, many thanks to the members of the Newstead and Biggin groups, past and present, for having me as a group member and for making the time I spent with them enjoyable. I am especially grateful to my colleagues and friends Mark, Feng, Oli, and Gurdeep for their support and advice throughout my thesis. It was a great pleasure sharing my D.Phil days with you.

Many thanks go to my mentor, Prof. Antonios Kolocouris (National and Kapodistrian University of Athens), who has always altruistically supported my dreams and best interests, in the ups and downs that escort academic life.

I would also like to say a heartfelt thanks to my partner, Iris, who has been by my side throughout my studies as an endless source of support and encouragement – no matter the distance.

Finally, I reserve my deepest expression of gratitude for my parents, Despina

and Konstantinos Kolokouris, and my brother, Spiros, who have always undoubtedly expressed faith in my abilities, encouraged me to pursue my dreams and extended our home overseas to always include me during my days away from Greece. It is for my family that I have completed this thesis.

Many people made my time in Oxford memorable through their friendship and support. This includes my dear friends Giorgos P., Manos, Dimitris, Maria, Giorgos K., Antonia, and Nikolas; I will always cherish the great times we shared in Oxford. Their great friendship gave me a feeling of belonging allowing me to thrive in both my research and personal life. I wish to thank my friends from home for their support and encouragement throughout my studies.

Dimitrios Kolokouris
Oxford, January 2023

Contents

Abstract v

Acknowledgements vii

1 Introduction 1

1.1 Nutrient Transport into the Cell 1

1.2 Amino Acid Transporters of the APC Superfamily 5

1.2.1 Topology, Architecture & Fold Features 6

1.2.2 Binding Site 9

1.2.3 Transport Mechanism 11

1.3 SLC7 Family of Amino Acid Transporters 13

1.3.1 Prokaryotic SLC7 Homologues 15

1.3.2 Heteromeric Amino Acid Transporters (HATs) 18

1.3.3 Cationic Amino Acid Transporters (CATs) 23

1.3.4 Plant CAT Transporters 26

1.4 Transporter Interactions with Lipids 28

1.4.1 Mechanisms of Protein Modulation by Lipids 28

1.4.2 SLC7 Interactions with Cholesterols 35

1.5 Novel Methods in Membrane Protein Structural Biology 38

1.5.1 Electron Cryomicroscopy & Single Particle Analysis 38

1.5.2 Nanobodies as Structural Biology Aids 40

1.6 Hypotheses & Aims 45

2 Experimental & Computational Methods 51

2.1 Cloning & Mutagenesis 52

2.1.1 Yeast Expression Vectors 52

2.1.2 Mammalian & Oocyte Expression Vectors 55

2.1.3 Sybody Vectors 56

2.2 Protein Expression & Purification 57

2.2.1 Membrane Protein Expression in S. cerevisiae 57

2.2.2 Membrane Protein Expression in HEK293F cells 58

2.2.3 Membrane Fraction Preparation 59

2.2.4 Fluorescence Size Exclusion Chromatography 60

2.2.5 Membrane Protein Purification 61

2.2.6 Sybody Expression & Purification 63

2.3 Membrane Protein Characterisation 64

2.3.1 SEC-MALS 64

2.3.2 Circular Dichroism Spectroscopy (CD) 64

2.3.3 Differential Scanning Fluorimetry (DSF) 65
2.3.4 Native-PAGE Analysis 66
2.3.5 Native Mass Spectrometry (nMS) 66
2.3.6 Xenopus laevis Oocyte Two Electrode Voltage-Clamp 67
2.3.7 Cell-based Radioactive Substrate Transport Assays 68
2.4 Sybody Screening & Generation 69
2.4.1 AtCAT4 Biotinylation 69
2.4.2 Sybody Screening 70
2.4.3 Phage Display 72
2.4.4 ELISA Screening 74
2.4.5 Biolayer Interferometry (BLI) 75
2.5 CryoEM Structure Determination 76
2.5.1 Sample Preparation & Data Acquisition 76
2.5.2 Data Processing & Model Building 77
2.6 Molecular Dynamics for Substrate Interactions with SLC7s 78
2.6.1 System Preparation & Amino Acid-Bound MD Simulations 79
2.6.2 End-Point MMPBSA Calculations 82
2.7 Multiscale Molecular Dynamics Simulations of Lipid Interactions 84
2.7.1 CryoEM Structural Models & Coarse-graining Models 84
2.7.2 Coarse-grained Simulations 85
2.7.3 Potential of Mean Force (PMF) Calculations 86
2.7.4 Atomistic Simulations 87
2.7.5 Binding Site Detection and Analysis 88

3 SLC7A4 Homologue Screening for Structural Studies 91

3.1 Homologue Selection & Screening Pipeline 91
3.1.1 Expression Tests 96
3.1.2 Detergent Solubilisation & FSEC Monodispersity 98
3.1.3 Trial Purifications & Construct Characterisation 103
3.2 The Arabidospis thaliana SLC7A4 (AtCAT4) 114
3.2.1 Characterisation & Quality Control of AtCAT4 115
3.3 Crystallisation Trials 123
3.3.1 Lipidic Cubic Phase (LCP) Crystallisation Trials 123
3.3.2 Vapour Diffusion (VD) Crystallisation Trials 125
3.4 Chapter Conclusions 126

4 Synthetic Nanobody Generation for AtCAT4 129

4.1 Sybodies: Synthetic Nanobodies 129
4.2 Selection Against Lipidated CAT4 133
4.2.1 Sample Preparation for Sybody Selection 133
4.2.2 Ribosome Display 139
4.2.3 Phage Display 140
4.2.4 ELISA Screening 141
4.2.5 Sybody Sequence Analysis 144

4.2.6 Screening Sybodies for Purification and Biochemical Stability 147
4.3 Analysis of Syb Binding to AtCAT4 154
4.3.1 Syb Affinity Pull-downs 154
4.3.2 Biolayer Interferometry Analysis of Binding 155
4.3.3 Gel-filtration Co-elution Analysis 159
4.4 Chapter Conclusions 162

5 Structure, Dynamics & Characterisation of AtCAT4 163
5.1 CryoEM Pipeline for CAT4 164
5.1.1 Sample Preparation & Grid Optimisation for SPA 165
5.1.2 Single Particle Analysis of CAT4 169
5.2 CAT4 Structural Analysis 172
5.2.1 CAT4-SybB5 Interface 174
5.2.2 Conformational State & Comparison to Other APC Transporters . 178
5.2.3 TM Helices 11&12 and IL5 – Functional Implications 188
5.3 Substrate Recognition in the Binding Site 193
5.3.1 Structural Insights Into Substrate Specificity via the Binding Site. 193
5.3.2 Substrate Binding Assay 196
5.3.3 MD Simulations of CAT4-CAA and Comparison to xCT-Cystine.. 211
5.4 Progress Towards Functional Characterisation of SLC7A4 217
5.4.1 Electrophysiological Characterisastion 218
5.4.2 Whole-Cell Radioactive Uptake Assays 221
5.5 Chapter Conclusions 228

6 Cholesterol Interactions with SLC7 Transporters 231
6.1 Investigation & Comparison of Cholesterol Sites on HATs 232
6.1.1 Computational Pipeline 232
6.1.2 Cholesterol Enrichment in the 4F2hc/LAT1 Vicinity in Binary & Complex Membranes 232
6.1.3 Identification of Cholesterol Binding Sites in Human HATs 235
6.1.4 Energetics of Cholesterol Sites in HATs by US/PMF Calculations 244
6.1.5 Atomistic Refinement Analysis of HAT Cholesterol Sites 248
6.2 Characterisation of LAT1 Cholesterol Sites 251
6.2.1 Cholesterol Concentration Effects on Red & Green Site Occupancies 251
6.2.2 Computational Mutagenesis of LAT1 Red & Green Sites 254
6.2.3 Experimental Testing of LAT1 Red & Green Sites by Transport Assays 257
6.3 Functional Implications & Relations to other SLC7s 260
6.3.1 Cholesterol-coordinating Residues Implicated in Disease 260
6.3.2 Comparison of AtCAT4 and LAT1 Cholesterol Sites 262
6.4 Chapter Conclusions 264
6.5 Limitations & Future Directions 266

7 Final Remarks & Future Perspectives. 269

7.1 Why Was SLC7A4 Studied? 270
7.2 Is SLC7A4 an Amino Acid Transporter? 271
7.3 Binding Site Modularity in SLC7s 274
7.4 Cholesterol Interactions with SLC7s 275

Bibliography 279

A Appendix 315
Chapter 3 Supplementary Information 315
Chapter 4 Supplementary Information 319
Chapter 5 Supplementary Information 323
Chapter 6 Supplementary Information 331

B Appendix 333
Molecular Dynamics Theory And Methods 333

CHAPTER 1

Introduction

1.1 Nutrient Transport into the Cell

α-amino acids are a chemically broad category of molecules. An important subcategory consists the building blocks of proteins, known as proteinogenic amino acids. Amino acids also serve as metabolic precursors of many important nitrogen-containing compounds, notably heme, physiologically active amines, nucleotides, and nucleotide coenzymes (Voet and Voet, 2011). Emerging research suggests that amino acids are also essential in cell survival, growth and proliferation by playing the role of messengers in signalling pathways (Goberdhan *et al.*, 2005; Sancak *et al.*, 2008; W. Y. Lin *et al.*, 2015; Chantranupong *et al.*, 2016; Fan *et al.*, 2016; Lawrence and Zoncu, 2019).

Eukaryotic cells are compartmentalised by membranes that form boundaries between organelles. However, the hydrophobic membranes are impermeable to most ionic and polar substances and are thus transported within cells and organelles by dedicated proteins via multiple mechanisms (Figure 1.1). These are integral membrane proteins and are involved in the transport of crucial compounds such as nutrients, inorganic ions, amino acids and drugs (Hediger *et al.*, 2013). Consistent with their functionality, they are ubiquitously expressed, e.g. in the

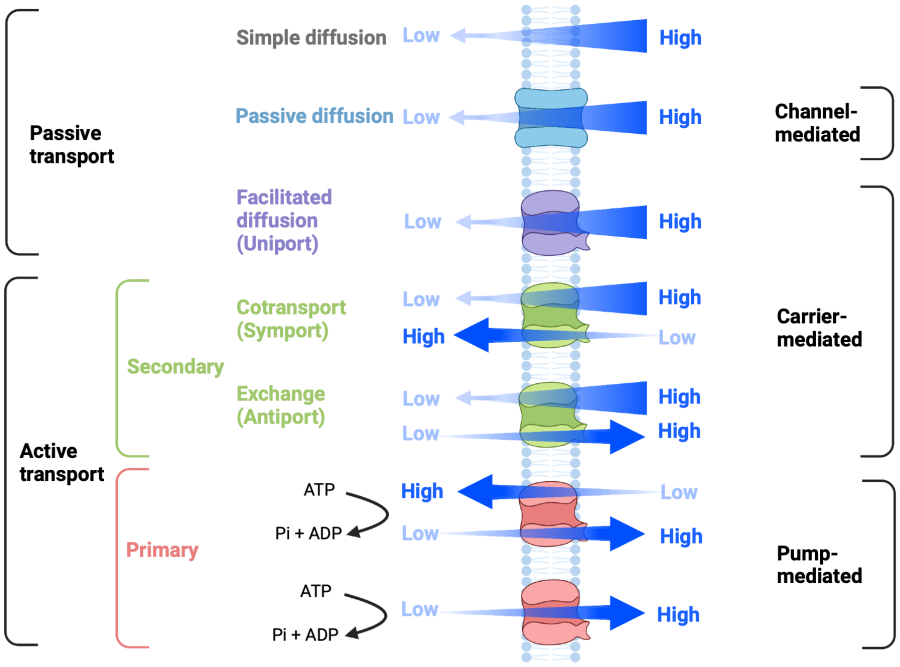


Figure 1.1 Cartoon illustration of molecule transport mechanisms through membranes. Molecules involve inorganic (e.g., ions) or organic substrates that are charged or a have a high non-zero electric dipole moment. Through the hydrophilic lining pore of membrane channels, molecules diffuse via passive diffusion mechanism. On the other hand, transporters translocate their substrate from one side of the membrane to the other as a result of a conformational change that the substrates induce. This can be done either with a requirement of energy expenditure against the substrate electrochemical gradient (active transport) or down the substrate electrochemical gradient (facilitated transport). Blue arrows indicate the transport of a solute; arrow width shows the electrochemical gradient.

epithelia of organs such as the liver, intestine, kidney, blood-brain barrier, and the brain and are abundantly present in subcellular locations (David T. Thwaites and Anderson, 2011; L. Lin *et al.*, 2015). The proteins involved in transport are broadly categorised as either *channels* or *transporters*. In their open state, channels are open to both environments at the same time, allowing ions or molecules (e.g., water – aquaporin) to passively diffuse. In contrast, transporters – i.e. solute carriers, ABC transporters and P-type ATPases – differ from channels

(Kotyk, 1997), by having a binding site which they can expose to one side of the membrane at a time (inward-facing versus outward-facing states), and may:

- facilitate diffusion in an exergonic process ($\Delta G<0$) by lowering the transition state energy for substrates to cross the membrane. Substrates flow from high concentration to low concentration to reach equilibrium (uniporter solute carriers; SLC transporters).
- transport substrates from low to high concentration, coupled to a sufficiently exergonic process to make the transport favourable. The energy source can be ATP hydrolysis (primary active transport; ABC transporters and P-type ATPases) (Borst and Oude Elferink, 2002). In secondary active transport there is no direct coupling of ATP, but instead the potential difference created by pumping ions out of the cell by P-type pumps is exploited (symporter and antiporter solute carriers; SLC transporters) (L. Lin *et al.*, 2015; Bosshart and Fotiadis, 2019) (Figure 1.1).

SoLute Carrier (SLC) superfamily transporters function as uniporters, symporters, or antiporters (Figure 1.1). The superfamily includes over 400 members organized into 66 families (Hediger *et al.*, 2004; Perland and Fredriksson, 2017) and transport a diverse range of substrates such as inorganic ions, amino acids, neurotransmitters, sugars, purines, and fatty acids, among others. It is therefore not surprising that SLCs are vital for maintaining homeostasis in the body and individual cells, and their genetic polymorphisms are associated with several diseases, such as amyotrophic lateral sclerosis (ALS), Alzheimer's disease, and schizophrenia, indicating a wealth of opportunities for therapeutic intervention (Colas, Ung and Schlessinger, 2016). SLC members have recently been identified as

playing a major role in drug pharmacokinetics (Giacomini and Huang, 2022), while others were known drug targets, such as monoamine transporters (SERT, DAT, NET) in mental disorders and SGLTs in diabetes (L. Lin *et al.*, 2015). Some constitute paths for drug absorption and excretion, such as the oligopeptide transporter PepT (Minhas and Newstead, 2019) and system yLAT1 (Fotiadis, Kanai and Palacín, 2013). Compared to other gene families of similar stature, SLCs are relatively understudied, with more than 30% reported with unknown function – the latter also termed *orphans* (César-Razquin *et al.*, 2015; Perland and Fredriksson, 2017), even though this class of transporters has been known for decades. ABC transporter research on the other hand has progressed throughout the years (Theodoulou and Kerr, 2015).

With membrane protein structural biology efforts flourishing within the past fifteen years (Newby *et al.*, 2009; Moraes *et al.*, 2014; Parker and Newstead, 2016; Cheng, 2018), incredible progress has been made with high-resolution structures of human SLC transporters and homologues. Even though they transport a diverse set of solutes, most transporters in the SLC superfamily have similar folds, implying commonalities in transport mechanism (Luckey, 2014). Two main fold categories have been identified:

- the major facilitator superfamily (MFS) with characteristic examples including the mitochondrial ADP/ATP carrier as well as the *E. coli* glycerol-3-phosphate transporter GlpT (Huang *et al.*, 2003) and lactose permease LacY (Abramson *et al.*, 2003), and
- the LeuT superfamily, also known as the Amino Acid-Polyamine-Organocation (APC) superfamily, the archetypal member of which is the well-known prokaryotic leucine transporter, LeuT, from *Aquifex*

aeolicus (Yamashita *et al.*, 2005a).

1.2 Amino Acid Transporters of the APC Superfamily

Analyses estimate that over 25% of the human SLCs have amino acids as their primary substrate, most of which belong to the APC Superfamily (APCS) (Schiöth *et al.*, 2013a). APCS members do not only transport amino acids and metabolic derivatives (e.g. serotonin, GABA, α-ketoglutarate), but also a myriad of other substrates. Examples are the sodium-glucose cotransporter family (SLC5), including SGLT2 (Niu *et al.*, 2021) – a major type-II diabetes target – and the electroneutral cation-coupled Cl cotransporter family (SLC12; CCC) which are involved in active Cl^- absorption in the kidney and are targets for diuretics (Arroyo, Kahle and Gamba, 2013). In epithelial cells, SLC36 transporters drive amino acid absorption at the apical membrane of intestinal epithelial cells via H^+-coupling (David T Thwaites and Anderson, 2011). Some APC transporters are "transceptors", i.e. transporter-substrate complexes that transduce signals to the inside of the cell in a receptor-like fashion (Holsbeeks *et al.*, 2004; Hundal and Taylor, 2009).

Exemplary carriers with the APC fold that transport α-amino acids and metabolic derivatives are members of the SLC6, 7, 32, 36 and 38 families (Closs *et al.*, 2006; Abramson and Wright, 2009; Schweikhard and Ziegler, 2012; Fotiadis, Kanai and Palacín, 2013; Schiöth *et al.*, 2013a; Bosshart and Fotiadis, 2019) and their homologues across species. A list of the APCS family members with nomenclature and common substrates is shown below (Table 1.1).

TC family	SLC family	Substrates	Coupling mechanism
AE	SLC4	HCO_3^-, Cl^-	Na^+/H^+-dependent (Exch., Symp.)
SSS	SLC5	sugars, choline, vitamins, myoinositol, anions	Na^+
NSS	SLC6	neurotransmitters, **amino acids**, osmolytes	Na^+, Cl^-
APC	SLC7	**amino acids**	H^+ (Exch, Unip)
NRAMP	SLC11	divalent metal cations (e.g. Mg^{2+},Fe^{2+},Mn^{2+})	H^+, Symp
CCC	SLC12	Na^+, K^+	Cl^-
NCS2	SLC23	nucleobases, purines, pyrimidines	H^+, Na^+
SulP	SLC26	sulphate, chloride	H^+, HCO_3^-,
HAAAP	SLC32	**amino acids**	H^+
AAAP	SLC36,38	**amino acids**, auxin	H^+, Na^+
BCCT		betaine, carnitine, choline	H^+, Na^+ (Exch, Uni)
AGCS		alanine, glycine	H^+, Na^+
NCS1		purines, pyrimidines and metabolites	H^+
PAAP		alanine	Na^+
LIVCS		aliphatic amino acids (Leu, Ile, Val)	H^+, Na^+
CstA		peptides	–
KUP		K^+	H^+
BenE		small aromatic compounds	H^+

Table 1.1 APC superfamily transporter families named with their Transporter Classification (TC) and their respective mammalian SLC numbering (where attributed). Substrates and coupling mechanism are shown (Exch = exchanger, Unip = uniporter, Symp = symporter). Information from *https://tcdb.org/* and *https://www.bioparadigms.org/slc/intro.htm*.

1.2.1 Topology, Architecture & Fold Features

In the late 2000s, with the developments of membrane transporter X-ray crystallography, such as sugar-based detergents – e.g. DDM (Vanaken *et al.*, 1986), LMNG (Chae *et al.*, 2010) – and lipidic cubic phase crystallisation (Landau and Rosenbusch, 1996; Caffrey, 2008), considerable insights into the structure and mechanism of APCS members was gained. In 2005, the structure of LeuT, a broad spectrum amino acid transporter and a member of the NSS

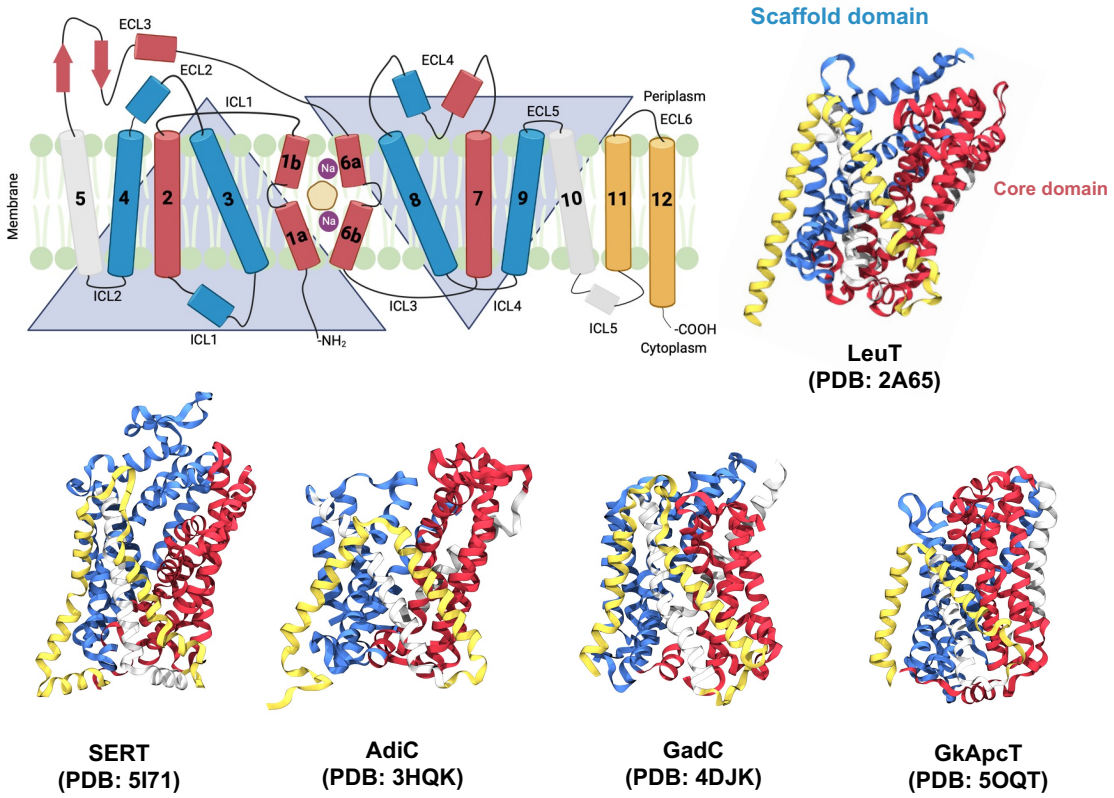


Figure 1.2 Transporters that share the LeuT fold consist of 10 TMs assembled into two inverted structural repeats (N-terminal repeat TMs 1–5, C-terminal repeat TMs 6–10 also known as the "5 + 5 inverted repeats"). TMs 1 and 6 are discontinuous. The short α-helices in TM1 and TM6 are linked by an unwound polypeptide stretch. The alternating mechanism of substrate transport governing LeuT-fold transporters is facilitated by the orchestrated movement between the core (in red) and the scaffold domain (hash subdomain in blue, TMs 5 & 10 in white). In yellow are TM11 and TM12 helices. Shown structures: LeuT (PDB ID 2A65), serotonin transporter SERT (PDB ID 5I71), the L-arginine/agmatine transporter AdiC (PDB ID 3HQK), glutamate-GABA antiporter GadC (PDB ID 4DJK) and the *G. kaustophilus* ApcT (*Gk*ApcT; PDB 5OQT).

family (Table 1.1), represented the first structural information of an APC family, uncovering the characteristic LeuT fold, for the first time (Figure 1.2). The high-resolution 1.65 Å of the outward-facing state allowed for an unambiguous assignment of the position of two coupling Na^+ ions, which are bound in the center of the transporter (Yamashita *et al.*, 2005a). Characteristically, this fold has two distinct bundles referred to as the scaffold domain (TM3, TM4, TM5, TM8, TM9, TM10) and the core/bundle domain (TM1, TM2, TM6, TM7).

Even with 20–25% overall sequence identity, the LeuT structure served as a prokaryotic template to understand the mammalian neurotransmitter sodium symporters (NSS; SLC6) (Singh, 2008) and model the mechanism underlying NSS function and pharmacology (Caplan *et al.*, 2008; Celik *et al.*, 2008; Noskov and Roux, 2008; Shi *et al.*, 2008; Shaikh and Tajkhorshid, 2010; Zhao *et al.*, 2010; Jørgensen *et al.*, 2017). SLC6s include the monoamine transporters, SERT (serotonin transporter; SLC6A4) and DAT (dopamine transporter; SLC6A3), which carry neurotransmitterers and are linked to neurological disorders such as anxiety, major depression, post-traumatic stress disorder, obsessive compulsive disorder (Kristensen *et al.*, 2011; Bröer and Gether, 2012).

Structures of sequence unrelated APC superfamily transporters revealed the same highly conserved fold which was exemplified by LeuT. The fold is characterised by two inverted structural repeats of five transmembrane helices related pseudosymmetrically (Krishnamurthy, Piscitelli and Gouaux, 2009; Boudker and Verdon, 2010). This suggested a gene duplication origin for APCS members, similar to how the prokaryotic semiSWEETs evolved to the mammalian SWEET transporters (Keller, Ziegler and Schneider, 2014; Feng and Frommer, 2015), suggesting that monomers dimerised in an antiparallel manner to form the functional protein (Keller, Ziegler and Schneider, 2014). Most APCS members contain 12 TM helices, although exceptions with 10, 11 and 14 have been recorded (Västermark and Saier, 2014). The shared architecture between APCS transporters (Figure 1.2) suggests shared mechanistic concepts involving substrate binding and transport.

1.2.2 Binding Site

The diversity of substrates – in size, shape, hydrophobicity, polarity, and charge – and their transport coupling mechanisms – Na^+ and Cl^- symport, H^+ or K^+ symport/antiport, or symport/antiport with other solutes – showcase the versatility of the APC fold (Rudnick, 2011). What is more, APCS transporters have low sequence identity yet produce the same overall fold (Edwards *et al.*, 2018). However, different families and even family members achieve differential substrate preferences, binding affinities and transport kinetics (Rodríguez-Banqueri *et al.*, 2016; Bartoccioni *et al.*, 2019; Errasti-Murugarren *et al.*, 2019a). The key in substrate transport selectivity originates from key residues in the binding site and translocation pathway while the fold permits a transition cycle with conserved mechanism (Schweikhard and Ziegler, 2012; Bosshart and Fotiadis, 2019). Indeed, many APC-fold transporters exhibit a conserved binding environment such that substrates sit in the same general locale (**Figure 1.3**) (Yamashita *et al.*, 2005a; Gao *et al.*, 2009a; Penmatsa, Wang and Gouaux, 2013; Coleman, Green and Gouaux, 2016; K. E. J. Jungnickel, Parker and Newstead, 2018; Lee *et al.*, 2019a; Focht *et al.*, 2021), termed the S1 site or primary binding site.

The topology diagram (**Figure 1.2**) and the substrate-bound structures of APC members (**Figure 1.3**) show unwound helical segments in TM1 and TM6 which form the ion and substrate binding sites (Yamashita *et al.*, 2005b; Penmatsa and Gouaux, 2014). The reasoning behind substrates being coordinated by unwound, unstructured residues is that they can expose their backbone groups to coordinate substrates via H-bonding, observed in substrate bound structures. In amino acid transporters TM1 coordinates the $-COO^-$ group and TM6 the $-NH_3^+$

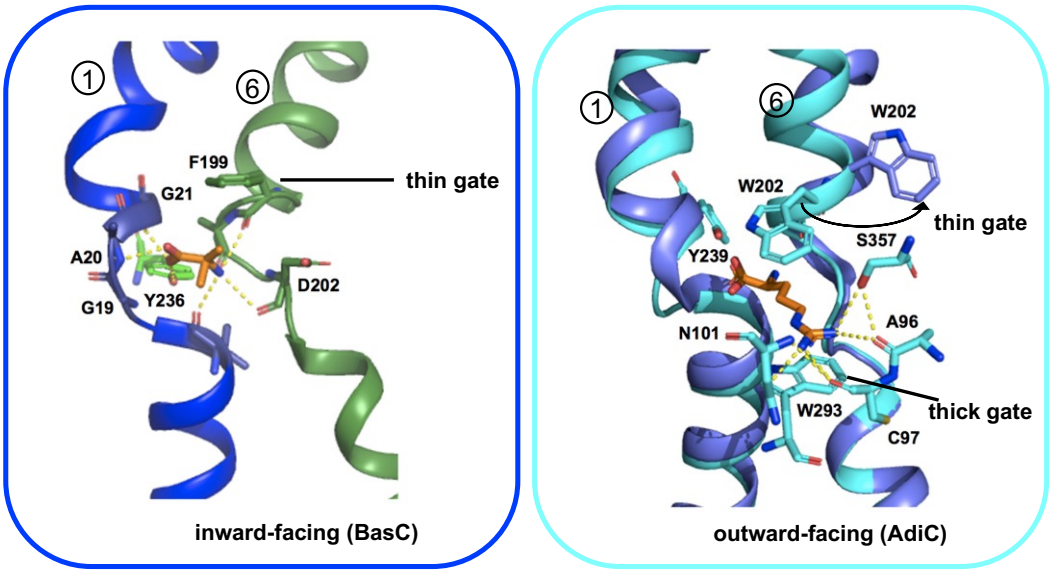


Figure 1.3 Substrate-binding site design in APC transporters. (A) View of the bacterial alanine–serine–cysteine exchanger substrate-binding site (PDB ID: 6F2W). H-bonds shown as dashed lines between TM1-6 and the substrate. (B) Superposed AdiC structures in outward-open apo (PDB ID: 3J4I) and outward-occluded L-arginine-bound (PDB ID 3L1L). H-bonds show the interactions specific for the L-arginine side chain (dashed lines). Trp 202, the external thin gate, swings away from the binding site in the outward-open conformation, allowing substrate release. Figure adapted from (Nicolàs-Aragó et al., 2021).

group of the substrate (**Figure 1.3**). On the other hand, they provide flexibility for conformational changes between inward and outward-facing states.

So how is specificity achieved if there are general, redundant interactions present? If the substrate has a side chain with specific chemistry, residues from surrounding TMs contribute auxiliary interactions, such as AdiC A96, C97, S357 (shown for L-arginine in **Figure 1.3**). The opposite is also true, i.e., residues can hinder substrate binding. Comparison between the binding site of LeuT and another prokaryotic Na^+-coupled SLC6 homologue, MhsT, showed that bulky residues in S1 prevented LeuT from binding L-tryptophan, MhsT's substrate. One can also change bulky residues to modulate steric hindrance such that LeuT and ApcT could transport non-native substrates (Piscitelli and Gouaux, 2012; K. E.

J. Jungnickel, Parker and Newstead, 2018). For example, TM3 contains non-conserved residues equivalent to LeuT's V104 (Livesay *et al.*, 2007) that are involved in forming the substrate binding site (volume and side chain interactions). Suitable site-mutations at this site directly affect substrate side-chain specificity (Edwards *et al.*, 2018). The V104-equivalent in mouse LAT2 (SLC7A8) is N133. Substituting the latter with the equivalent Ser from LAT1 (SLC7A5) increases the transport of LAT1-specific substrate 3,3-diiodo-L-thyronine (Hinz *et al.*, 2015). The binding site also secures the substrate in place by a semi-conserved aromatic thin gate (Figure 1.3) (Nicolàs-Aragó *et al.*, 2021). Mutating this to an aliphatic residue generates transport deficient mutants (Piscitelli and Gouaux, 2012). Collectively, the subtle balance of the H-bond donors/acceptors, the electrostatic composition of the substrate binding site and the size restrictions determined by S1-lining residues define the substrate specificity of APCS members.

1.2.3 Transport Mechanism

Multiple structures of LeuT in different conformations allowed for a molecular insight to the "alternating-access" mechanism, originally proposed to involve cycles of phosphorylation/dephosphorylation inducing conformational changes to capture the substrate on one side of the membrane and release it on the other (Jardetzky, 1966). How is the alternating access mechanism, achieved by APCS members? The intertwined 5+5 TM repeats form two distinct domains, the core/bundle (TMs 1, 2, 6, 7) and the scaffold domains (TMs 3, 4, 5, 8, 9, 10) (Figure 1.2). The core TMs move as a bundle against the rigid scaffold domain to expose the binding site to the intra- or extracellular milieu via the "rocking-

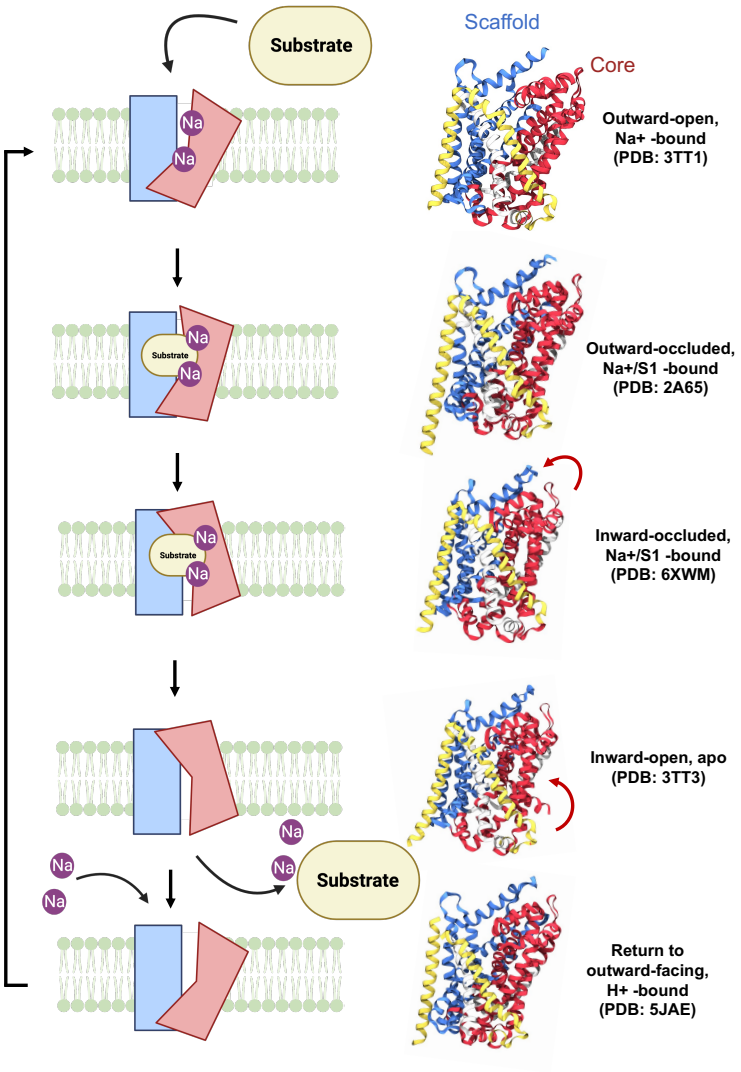


Figure 1.4 Cartoon illustration of the alternate access mechanism of transport of an APC-fold SLC transporter. On the right, cartoon representations of LeuT structures corresponding to each cycle step are shown (coloured as in **Figure 1.2**). In the outward-open state, Na^+ ions (purple) enter and bind close to the orthosteric S1 site. Next a substrate molecule binds, and the transporter isomerises to the outward-occluded conformation. Then, LeuT transitions towards the inward-occluded state to finally adopt an inward-open conformation after releasing its substrate. In this state, a counter-ion is thought to bind (H^+) to promote isomerisation back to the outward-open state, and the cycle repeats.

bundle" mechanism. The mechanism is shown for LeuT in **Figure 1.4**. The conformational changes are induced upon binding of the substrate and ion, which

by an induced-fit mechanism cause the transporter to shift to an occluded state, wherein the binding site is isolated from both sides of the membrane. This intermediate is resolved into an open state and the substrate is released before the cycle resets (Forrest *et al.*, 2008). In LeuT, the mobile bundle domain (TMs 1, 2, 6, and 7) pivots against the rigid hash domain (TMs 3, 4, 8, and 9) (Forrest *et al.*, 2008) with TM1α displaced as much as 45 °, creating access to the substrate binding pocket from the intracellular side (Krishnamurthy and Gouaux, 2012). However, it must be noted that slight mechanistic differences between APCS members have been identified. One such example is the *Microbacterium liquefaciens* Na^+-hydantoin transporter, Mhp1, suggested to transition by rigidly moving the hash domain against the bundle domain, while TMs 5 and 10 act as auxiliary gating elements (Shimamura *et al.*, 2010). On the other hand, AdiC involves the pivoting of both the bundle and the hash domains, while TMs 5 and 10 present minor conformational changes (Kowalczyk *et al.*, 2011a). The common LeuT fold therefore appears to represent an archetype of transporters that, by rather small structural modifications, has evolutionarily developed into distinct transporter families with similar substrate translocation mechanisms for different physiological conditions and substrates (Höglund *et al.*, 2011; Västermark and Saier, 2014; Focht *et al.*, 2021).

1.3 SLC7 Family of Amino Acid Transporters

The SLC7 family of transporters are Na^+-independent amino acid transporters within the APC superfamily, very similar in structure to the Na^+-coupled NSS family (e.g. LeuT, SERT). Most of SLC7s function as uniporters and exchangers (Closs, Gräf, *et al.*, 1997; Vékony *et al.*, 2001; Wolf *et al.*, 2002; Verrey, Ellen

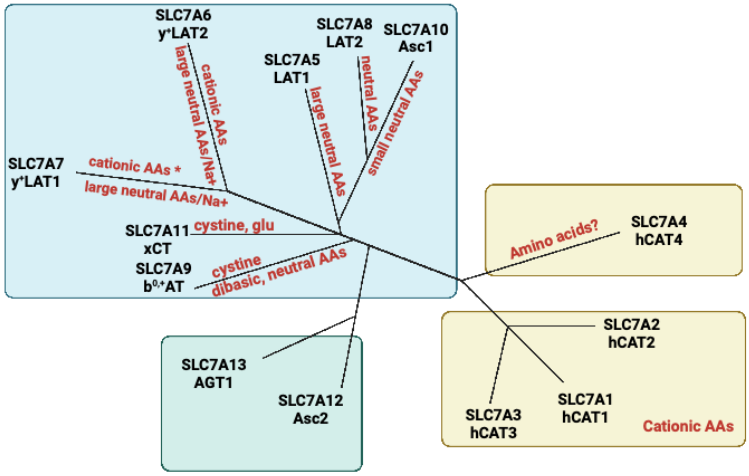


Figure 1.5 Phylogenetic tree of the SLC7 family proteins. The SLC7 family is composed of two subfamilies formed by the cationic amino acid transporters (CAT, SLC7A1–4) and the heteromeric amino acid transporters [HAT light chains, SLC7A5–11 and neutral amino acid Ala, Ser Cys preferring transporter-2 (Asc-2) and aspartate/glutamate transporter-1 (AGT1)]. SLC7A4 as well as Asc-2 and AGT1 differ slightly from the other subfamily members.

I Closs, *et al.*, 2004; Closs *et al.*, 2006). Fourteen mammalian members belong in the SLC7 family. Based on their evolutionary relationships SLC7s further divide into (a) the cationic amino acid transporters (CATs), and (b) the heteromeric amino acid transporter (HATs), the latter being heterodimers consisting of an SLC7 transporter and an SLC3 protein (**Figure 1.5**).

The increasing amino acid metabolism research has showcased SLC7s as a central player in cell metabolic regulation and as a main gateway for amino acids in the cell. SLC7 functional perturbations are linked to metabolic disorders, such as lysinuric protein intolerance, various forms of cancer and cystinurias (**Table 1.2**) (Verrey, Ellen I. Closs, *et al.*, 2004; Schweikhard and Ziegler, 2012; Colas, Ung and Schlessinger, 2016; Kandasamy *et al.*, 2018a).

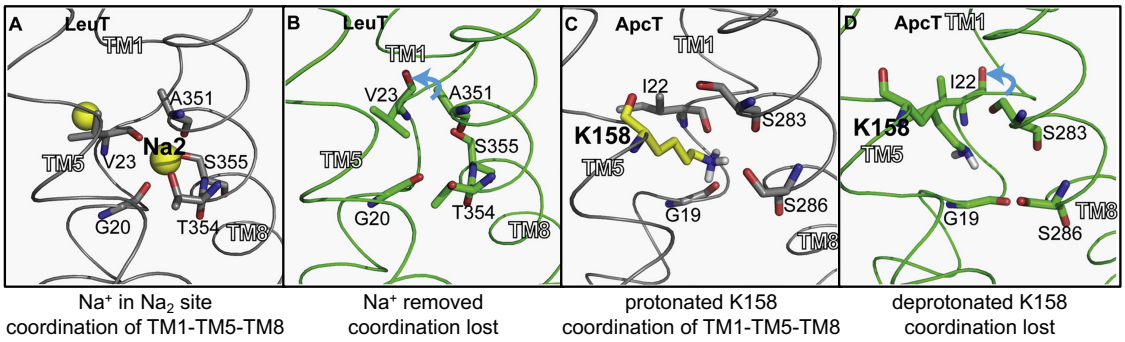


Figure 1.6 Configurational changes in the Na2 site of LeuT in the presence (A) and absence (B) of Na^+, and the corresponding vicinity of K158 of ApcT with K158 protonated (C) and deprotonated (D). Snapshots from unbiased MD simulations of LeuT and ApcT generated by (Shi and Harel, 2010).

1.3.1 Prokaryotic SLC7 Homologues

The SLC7 family members are mammalian homologues of the prokaryotic amino acid transporters ApcT, GadC, and AdiC (**Figure 1.2**). ApcT is a distant homologue to CATs and HATs similar to how LeuT is for SLC6 transporters. The first mechanistic model was derived from the crystal structure of ApcT from *Methanocaldococcus jannaschii*, a H^+-coupled broad-specificity amino acid transporter within the APC family, at 2.35 Å resolution (Shaffer *et al.*, 2009). *Mj*ApcT contains 12 transmembrane helices, with the first 10 consisting of an inverted structural repeat of 5 transmembrane helices like LeuT.

In Na^+-dependent SLC6 members, a Na^+ in the "Na2 site" is coordinated by TM5 residues and bridges the unwound region of TM1 (core domain) with TM8 residues (scaffold domain) resulting in increased substrate binding affinity and conformational transitions (Loo *et al.*, 1999; Kazmier *et al.*, 2014; Tavoulari *et al.*, 2016). The ApcT structure revealed that the protonable ammonium moeity of K158 in TM5 is in an equivalent position to the Na2 site Na^+ in LeuT. MD simulations revealed that ApcT's TM5 K158 also coordinates TM helices 1

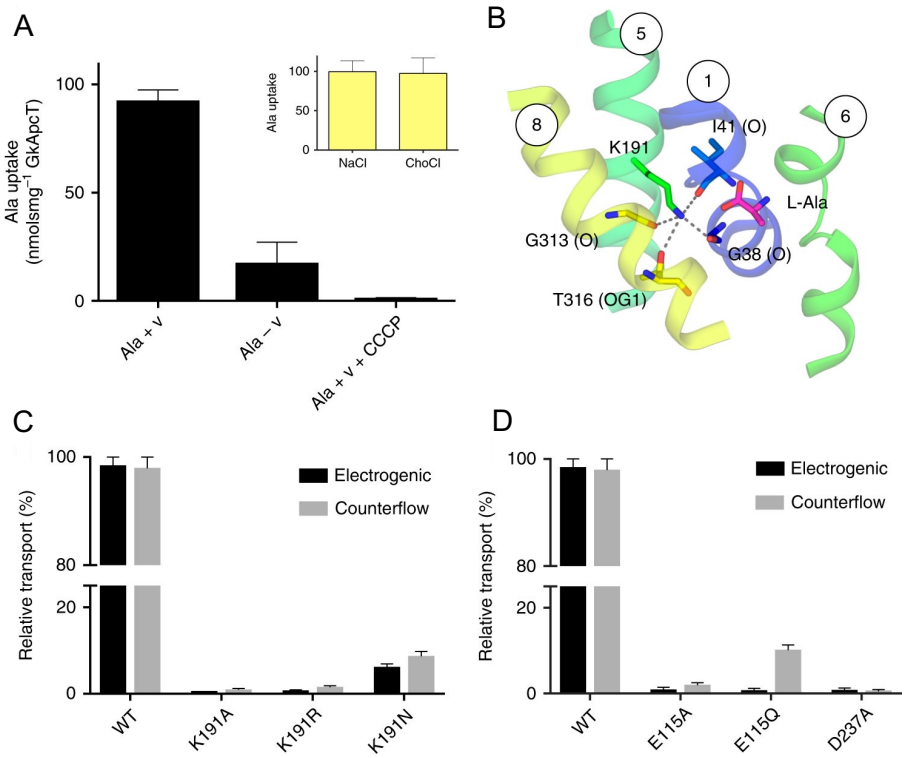


Figure 1.7 Functional characterisation of proton coupling residues in GkApcT. (A) H^+-coupled uptake assays showing that accumulation of L-Ala is driven by an inwardly directed H^+ electrochemical gradient generated using valinomycin (+v) and an inwardly directed pH gradient (−v). Transport was abolished in the presence of the proton ionophore CCCP. Inset—transport is independent of sodium. (B) View of K191 in the crystal structure, highlighting the interactions made to TMs 1 and 8. The bound L-Ala ligand is shown in magenta. (C) Mutational analysis of K191 using electrogenic and counterflow accumulation of [^{3}H]-L-Ala. (D) Mutational analysis of E115 and D237 using electrogenic and counterflow accumulation assays. Figure reproduced from (K. E. J. Jungnickel, Parker and Newstead, 2018).

and 8, demonstrating common principles among H^+ and Na^+-coupled amino acid transporters (**Figure 1.6**) (Shi and Harel, 2010).

Since *Mj*ApcT was shown to have pH-dependent uptake of L-alanine and other amino acids, it was suggested that K158 was also the H^+-coupling residue since a K158A mutation abrogated transport (Shaffer *et al.*, 2009). However the H^+-coupling role was not strictly proven, since the authors investigated the mutant only under an inwardly-directed pH gradient and not a H^+ electrochemical

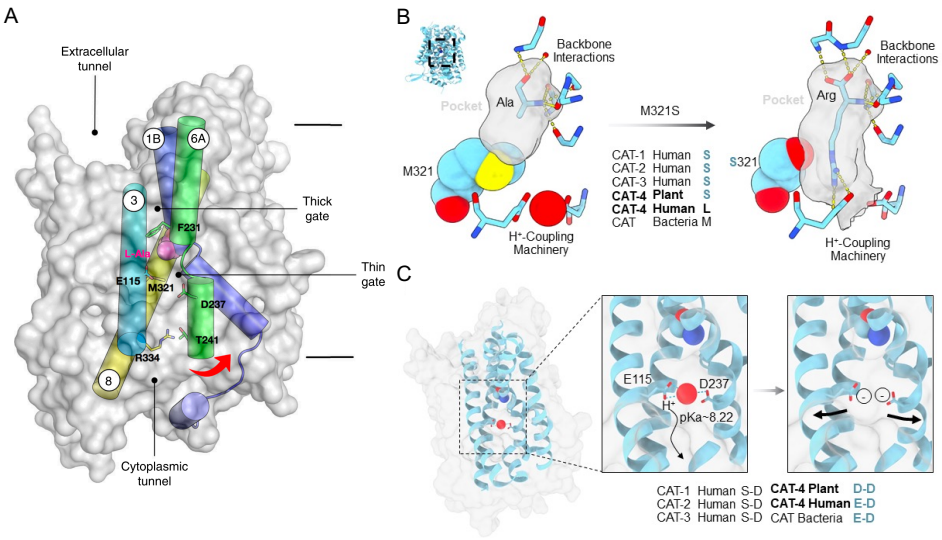


Figure 1.8 Molecular interactions governing the H^+-coupled ApcT transporter important for this thesis. (A) GkApcT surface representation with helices TM1, 3 and 6 involved in the substrate recognition are shown. (B) Binding interactions with L-alanine substrate are via backbone interactions with TM1 and TM6 residues. M321 sterically hinders the binding of larger substrates. Humanisation of *Gk*ApcT with the M321S mutation permits binding of hCAT substrate L-arginine. (C) Deprotonation of H^+-coupled E115 leads to repulsive forces between two negative charged residues (TM3 E115 – TM6 D237), causing opening of transporter and intracellular release of substrate.

gradient. The stringent H^+-coupling investigation was performed with another H^+-coupled ApcT L-alanine transporter, from thermophilic *Geobacillus kaustophilus* which shared structural similarities with *Mj*ApcT (K. E. J. Jungnickel, Parker and Newstead, 2018). It was shown that TM5 K191 (pKa = 7.3) indeed may undergo (de)protonation during the cycle but it is not directly coupled to the H^+ electrochemical gradient, since a K191N mutant managed to show electrogenic transport. Rather a TM3 Glu residue, E115 (pKa = 8.2), was identified as the H^+-coupling site since E115A was transport dead, but E115Q showed L-alanine transport but not electrogenic activity (Figure 1.7). The deprotonation leads to electrostatic repelling with a vicinial D237 (pKa = 5.5) leading to TM3-TM6 repulsion and substrate release (Figure 1.8-B). Perhaps the most important finding from the *Gk*ApcT study was that although the WT

transporter has no transport activity for cationic amino acids, a simple point mutation that reliefs steric hindrance, M321S, converted GkApcT to an L-arginine transporter making it into the first model for mammalian CATs (Figure 1.8-B,C). The M321S *Gk*ApcT is relevant to this thesis and will be discussed in many occasions.

Similar to ApcT, the inward-open structure of BasC – a bacterial Ala-Ser-Cys exchanger with 26–28% sequence identity with human HAT light subunits (Bartoccioni *et al.*, 2019) – conserved TM5 Lys interactions with a TM1 Gly by H-bonds. TM1-TM5 interactions are present in both inward-open and inward-occluded states (Shaffer *et al.*, 2009; K. E. J. Jungnickel, Parker and Newstead, 2018; Errasti-Murugarren *et al.*, 2019a). Upon substrate binding the TM1 Gly shifts by ~2 Å allowing TM5 K158 to H-bond with a TM8 Ser and create the extended TM1–TM5-TM8 bridge and stabilise the occluded state (Errasti-Murugarren *et al.*, 2019a). Loss of Na^+ in Na^+-coupled (SLC6s) or deprotonation of TM5 Lys (SLC7s) is essential to create the space needed for the release of substrate to the cytoplasm by both smFRET (Terry *et al.*, 2018) and unbiased MD (Shi and Harel, 2010). This prokaryotic mechanism of transport explains the rocking-bundle mechanism for the SLC7 family, wherein tilting of the bundle domain (TM 1, 2, 6 and 7) over the hash domain (TM 3, 4, 8, 9) rationalises the major inward-to-outward transitions (Forrest and Rudnick, 2009).

1.3.2 Heteromeric Amino Acid Transporters (HATs)

The uptake of dietary amino acids is largely mediated by heteromeric amino

acid transporters (HATs). This subfamily includes the classical system L transporters – i.e. L-leucine preferring transporters – that transport large neutral amino acids including branched and aromatic amino acids (Christensen, 1990). HATs consist of an SLC7 transporter domain and an SLC3 glycoprotein domain that heterodimerise, additionally linked via an extracellular disulfide bridge. The SLC3 family of heavy subunits, has two members; SLC3A1 (rBAT) – associates with SLC7A9 and SLC7A13 – and SLC3A2 (4F2hc or CD98hc) – associates with SLC7A5-8 and SLC7A10-11. This dimerisation offers both structural stability to the SLC7 subunit and trafficking from the endoplasmic reticulum (ER) to the plasma membrane (Nakamura *et al.*, 1999; Verrey, Ellen I Closs, *et al.*, 2004; Napolitano *et al.*, 2015; Y. Lee *et al.*, 2022). A study suggested that 4F2hc is necessary for transporter functionality in proteoliposomes (Yan *et al.*, 2019a) while others suggested the SLC7 subunit alone recovers the full transport kinetics, K_M and V_{max} (Reig *et al.*, 2002; Kantipudi and Fotiadis, 2021).

Physiologically the HATs transport a wide array of amino acids primarily as exchangers (Table 1.2) with different selectivity and ion dependence. System-L transporters, 4F2hc/LAT1 and 4F2hc/LAT2, transport large neutral and aromatic amino acids without ion coupling. System-y^+L transporters, 4F2hc/y^+LAT1 and 4F2hc/y^+LAT2, exchange large neutral amino acids for cationic amino acids but only couple the former to Na^+ having a requirement for a +1 net flux. System Asc transporters, 4F2hc/Asc1 and Asc2, transport small neutral amino acids and system $b^{0,+}$ (rBAT/$b^{0,+}$AT) exchanges cationic amino acids (and L-cystine) for neutral amino acids. System x_c^- selectively exchanges anionic L-cystine for L-glutamate, providing L-cystine as a building block of the antioxidant glutathionine.

Human gene name	Protein name	Predominant substrates	Transporter type/coupling ions[a]	Physiological Functions	Link to disease
SLC7A1	CAT1	Cationic L-amino acids	F[b] (non-obligatory E) System y^+	L-Arg transport into primary human T lymphocytes and T-cell activation	(↓) Hypertension
SLC7A2	CAT2A/B isoforms	Cationic L-amino acids	F System y^+	L-Arg transport, NO synthesis, insulin secretion, macrophage activation, regulation of inflammatory processes of the GI tract	(↓) Colonic inflammation
SLC7A3	CAT3	Cationic L-amino acids	F System y^+	L-Arg uptake in primary cortical neurons, NMDA receptor activation on the mTOR pathway and neuronal process formation	(↓) Autism spectrum disorder
SLC7A4	CAT4(?)	**unknown**	O		Unknown (suggested cancer?)
SLC7A5	LAT1	Large neutral L-amino acids, T_3, T_4, L-DOPA, BCH	E (similar intra- and extracellular selectivities, lower intracellular apparent affinity) System L	L-Leu transport, mTOR activation, pro-inflammatory cytokine production, pancreatic β-cell signalling and function	(↑) cancer (↑)rheumatoid arthritis
SLC7A6	y^+LAT2	Na^+ indep.: cationic amino acids: Na+/large neutral amino acids	E (preferentially intracellular cationic amino acid against extracellular neutral amino acid/Na^+) System y^+L	L-Arg transport in astrocytes, NO synthesis, podocyte development and function	(↑) hepatocarcinogenesis
SLC7A7	y^+LAT1	Na^+ indep.: cationic amino acids: Na+/large neutral L-amino acids	E (preferentially intracellular cationic amino acid against extracellular neutral amino acid/Na^+) System y^+L	Intestinal absorption and renal reabsorption of the CAAs L-Orn, L-Arg and L-Lys	(↓) Lysinuric protein intolerance
SLC7A8	LAT2	Neutral L-amino acids, T_3, T_4, BCH	E (similar intra- and extracellular selectivities, lower intracellular apparent affinity) System L	Flux of amino acids across cell barriers, amino acid renal reabsorption, optimal hearing function, platelet reactivity, clopidogrel response	(↓) age-related hearing loss, clopidogrel liver/kidney injury
SLC7A9	$b^{0,+}$AT	Cationic amino acids, large neutral amino acids, L-cystine	E (preferentially extracellular cationic amino acid and cysteine against intracellular neutral amino acid) System $b^{0,+}$	Transport of L-cystine, maintenance of glomerular and renal tubular function	(↓) cystinuria
SLC7A10	Asc-1	Small neutral amino acids	Preferentially E System Asc	L-Ala, L-Gly, L-Ser, L-Cys transport, glycinergic inhibitory function in the central nervous system	(↓) Hyperekplexia, stiff person syndrome
SLC7A11	xCT	L-cystine (anionic form), L-glutamate	E (preferentially extracellular cystine against intracellular glutamate) System x_c^-	L-cystine uptake, glutathione (GSH) synthesis	Various cancers
SLC7A12	Asc-2	L-glycine, L-alanine	System Asc		
SLC7A13	AGT-1	L-aspartate, L-glutamate, L-cystine	E		
SLC7A14		unknown	O	Retinal development and visual function	(↓) Retinitis pigmentosa

Table 1.2 Members of the mammalian SLC7 family (Fotiadis, Kanai and Palacín, 2013; Kandasamy *et al.*, 2018a).

a C: cotransporter, E: exchanger, F: facilitated transporter, O: orphan transporter

b CAT1 is strongly trans-stimulated and works better in the exchange mode, CAT2B and CAT3 are moderately trans-stimulated, and CAT2A is completely independent of the presence of trans-substrate.

The regulation of amino acid concentration in the cell by these systems has connected them with many different pathophysiological disease states (Table 1.2). Among these, LAT1 has attracted the most interest due to its prominent role in stimulating cancer-cell growth through the uptake of L-leucine (Yanagida *et al.*, 2001). Cancer cell growth has high demands in nutrients, including essential amino acids and LAT1 overexpression prospers amino acid-hungry tumour cell growth (Fuchs and Bode, 2005; Nawashiro *et al.*, 2006; Kaira *et al.*, 2008; Kobayashi *et al.*, 2008; Liang *et al.*, 2011; Furuya *et al.*, 2012; Betsunoh *et al.*, 2013). A more direct connection is that L-leucine activates mechanistic Target of Rapamycin Complex 1 (mTORC1), a lysosomal proliferation signalling hub that regulates protein translation and cell growth (Saxton and Sabatini, 2017). The importance of LAT1 extends to supplying thyroid hormones to developing fetuses and transporting L-DOPA and other neurological drugs (gabapentin, melphalan) as well as alkylating drugs (L-metasarcolysin and D,L-NAM) across the blood–brain barrier (Kanai *et al.*, 1998; del Amo, Urtti and Yliperttula, 2008; Braun *et al.*, 2011) Therefore, LAT1 is believed to be an important drug target for cancer treatment (Kim *et al.*, 2006; Nakanishi and Tamai, 2011; Marshall *et al.*, 2016) and promises that other SLC7 members will be recognised as drug targets. Aside to the traditional orthosteric inhibition strategies, a recent strategy has taken advantage of the water network found in the distal S2 binding site (Yan, Li, *et al.*, 2020), to covalently inhibit by nucleophilic attack of C407 in distal S2 binding site (Figure 1.9-C) to a 1,2,3-dithiazoles via a water-stabilised trisulfide intermediate (Prejanò *et al.*, 2021).

Within the time of this thesis, HAT structures were successionally reported

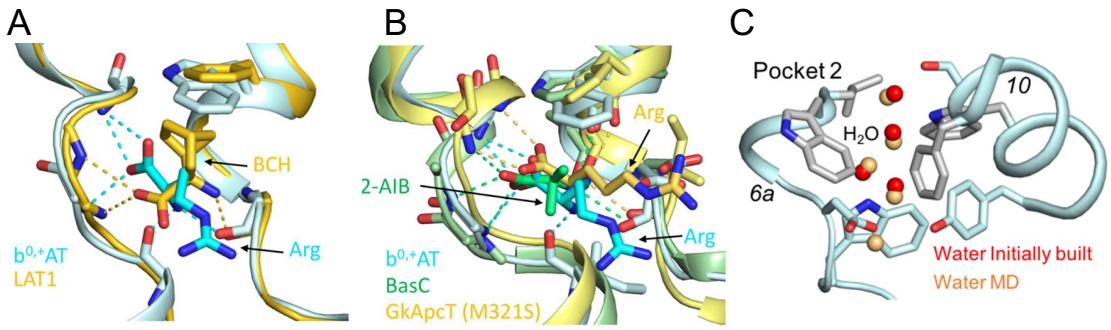


Figure 1.9 (A,B) Comparison of the binding modes of amino acids and derivatives (BCH, 2-AIB) in the canonical APC binding site between prokaryotic homologues, $b^{0,+}$AT and LAT1. (C) The water network located at the distal S2 binding site; network convergence via cryoEM and MD simulation. Adapted from (Yan, Li, *et al.*, 2020).

via CryoEM. The structure of the light subunit has the APC fold while the heavy subunit consists of a single pass TM helix followed by a bulky globular extracellular domain. The structures of apo and L-arginine bound heteromeric $b^{0,+}$AT/rBAT and dimer of heteromer $b^{0,+}$AT/rBAT have been reported in the inward open state (Wu *et al.*, 2020; Yan, Li, *et al.*, 2020; Y. Lee *et al.*, 2022) along with apo and inhibitor-bound inward-open LAT1/4F2hc (**Figure 1.9-A**) (Lee *et al.*, 2019b; Yan *et al.*, 2019a). Recently, LAT1/4F2hc was also solved bound to 3,3-diiodo-L-tyrosine or inhibitors in an outward-facing state (Yan *et al.*, 2021). LAT2/4F2hc was solved with cryoEM in the inward-open state both as apo (Rodriguez *et al.*, 2021) and in complex with L-leucine and L-trypthophan (Yan, Zhou, *et al.*, 2020) similar to outward facing structures of the SLC6s. The mean global $FSC_{0.143}$ resolution for these structures is ~3.3 Å. This suggests that mammalian SLC7 transporters share the rocking-bundle transport mechanism without having the same coupling mechanism.

The transporters displayed the same substrate binding mode explained in the prokaryotic homologue section (**Figure 1.9-B**). For example, L-arginine binding in M321S *Gk*ApcT shows the same backbone interactions with $b^{0,+}$AT, with only

slight modification in the guanidinium group orientation. The recent structures of SLC6 human SERT (Coleman, Green and Gouaux, 2016; Coleman *et al.*, 2019; Yang and Gouaux, 2021), SLC7 human LAT1 and $b^{0,+}$AT confirmed the predictions originally inferred from the LeuT structures. Although there are commonalities between members there are member-specific diversities that require extensive structural research of different conformations (Colas, 2020).

1.3.3 Cationic Amino Acid Transporters (CATs)

The cationic amino acid transporter branch of SLC7s (CATs) consists of four genes including SLC7A1, SLC7A2, SLC7A3 or otherwise named CAT1, CAT2, CAT3. CAT1-3 transport L-arginine, L-lysine, and L-ornithine. CAT1 additionally transports L-histidine at pH 5.5. The fourth gene associated with the family, member SLC7A4, remains orphan to date – i.e., of no known substrate. Since SLC7A4 is an orphan without demonstrated transport activity for cationic amino acids (CAAs), the name "CAT4" might be inappropriate (Fotiadis, Kanai and Palacín, 2013). The CAT family members are shown in (Table 1.2). CATs, unlike most APC members, are predicted to have 14 TM helices (12+2) (Figure 1.10) (Verrey, Ellen I. Closs, *et al.*, 2004) and share relatively high sequence identity between their members (Wolf *et al.*, 2002).

CATs mechanistically work as facilitators (uniporters) in a Na^+-independent and pH-independent manner and display varying affinities and K_M/V_{max} kinetics for CAAs (Table 1.3). Only CAT-2B has shown pH-dependency in the form of decreased transport with decreasing pH (Closs, Basha, *et al.*, 1997). No competition of transport was observed with D-Arg, GABA, L-serine or L-glutamate,

	Substrate	K_M (mM)	V_{max} (nmol substrate/oocyte/h)
CAT1	L-arginine	0.14 ± 0.03	1.7 ± 0.20
CAT2A	L-arginine	3.63 ± 0.27	5.3 ± 3.10
CAT2B	L-arginine	0.53 ± 0.21	2.6 ± 1.40
CAT3	L-arginine	0.45 ± 0.13	1.4 ± 0.30
	L-lysine	0.65 ± 0.23	1.4 ± 0.70
	L-ornithine	0.91 ± 0.26	2.5 ± 0.80

Table 1.3 Transport kinetics of human CAT members 1, 2A, 2B and 3. Values shown as means ± standard error of the mean from repeat radioactive uptake measurements reported in *Xenopus* oocytes. Data published in (Closs, Gräf, *et al.*, 1997; Vékony *et al.*, 2001).

suggesting the transporters are selective for cationic amino acids and stereospecific (Kim *et al.*, 1991; Ito and Groudine, 1997). CAT2 exists as two splice variants, A and B which differ in their transport kinetics (CAT2A has 10-fold lower affinity for L-arginine) due to differences in an intracellular loop 42 residues, in a remote position to the binding site. Therefore, differences are not due to affinities for CAAs. The structure and function analysis of the prokaryotic homologue *Gk*ApcT suggests the reason is the substitution of a CAT2B TM8 E396 to R396 in CAT2A changing TM8 configuration with respect to vicinal TM6β and TM3, involved in the translocation cycle (K. E. J. Jungnickel, Parker and Newstead, 2018).

CATs in Immune Cells

Other than providing CAAs for protein synthesis, there is a dictated role for L-arginine in cell metabolism. L-arginine catabolism involves four enzymes: nitric oxide synthase (NOS1, NOS2, NOS3), arginase (Arg1, Arg2), arginine decarboxylase (ADC), and arginine glycine amidinotransferase (AGAT). The three NOS

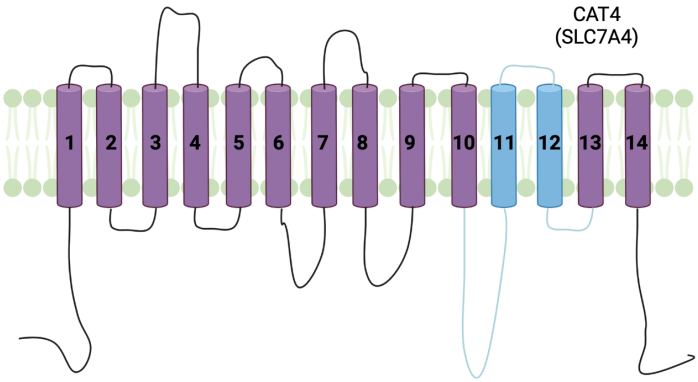


Figure 1.10 Eukaryotic cationic amino acid transporter topology diagram. CATs are the only members of the SLC7 family with two extra TMs compared to the family consensus, which is 12 TMs. The highlighted in light blue portion represents this additional amino acid stretch. The vacuolar side of the membrane corresponds to the extracellular side in related plasma membrane transporters. CAT4 is thought to localise on the lysosome.

isozymes produce NO and L-citrulline from L-arginine (Bogdan, 2001). NO released by macrophages acts as a cytokine and induces inflammation, vasodilation and downstream humoral immunity (Tripathi *et al.*, 2007). Arginase, expressed in immune cells, hydrolyses L-arginine into urea and L-ornithine as part of the urea cycle. L-ornithine is subsequently converted by ornithine decarboxylase to bioactive polyamines (putrescine, spermidine, and spermine) which are involved in proliferation, inflammation response and downregulation of iNOS production (Munder, 2009). Albeit the arginase and NOS arginine metabolism pathways were thought separate, recently it was found that LACC1 bridges proinflammatory NOS2 with polyamine immunometabolism in inflammatory macrophages, by converting L-citrulline to L-ornithine and isocyanic acid (Wei *et al.*, 2022). CATs provide high-level regulation of L-arginine concentrations in the tightly regulated L-arginine immunometabolism pathways.

CATs and mTORC signalling

Similar to L-leucine, L-arginine plays an important role in cellular homeostasis

and metabolism by activating the mTORC1 kinase by specific sensor proteins (Chantranupong *et al.*, 2016; Saxton *et al.*, 2016; Rabanal-Ruiz and Korolchuk, 2018). A specific lysosomal L-arginine sensor is the APC fold SLC38A9, acting as a transceptor (Rebsamen *et al.*, 2015; Lei *et al.*, 2018). A similar role as lysosomal transceptors regulating cell growth is known for other APCS members, the SLC36 PAT1 and PAT4 (Goberdhan *et al.*, 2005; Heublein *et al.*, 2010; Fan *et al.*, 2016). L-arginine concentration, regulated by CATs, is directly associated with the growth regulation pathway.

1.3.4 Plant CAT Transporters

Determining the intracellular concentration of L-arginine and other cationic amino acids is crucial in understanding key aspects of cell physiology (Morris, 2016). Recently a plant CAT homologue, the *S. lycopersicum* CAT2, was shown to transport L-arginine in response to pH gradients (Cosco *et al.*, 2019a). E142 in *S. lycopersicum* CAT2 is equivalent to the prokaryotic H^+-coupling site E115 suggesting the plant CAT to be H^+-coupled. Plant CATs may have diverged from the mammalian counterparts to rely on pH. This is not surprising as CATs are evolutionary clustered with H^+-coupled SLC36 amino acid symporters (Schweikhard and Ziegler, 2012; Perland and Fredriksson, 2017).

A. thaliana amino acid permeases (AAPs) serve as passageways for nutrients inside the plant cell. Arabidopsis has a cluster of nine CAT proteins, which share a defined similarity to the mammalian CAT genes (Y.-H. Su, Frommer and Ludewig, 2004) (Figure 1.11). CAT4, the closest Arabidopsis homologue to mammalian SLC7A4, is known to primarily localise in the tonoplast (Carter *et al.*,

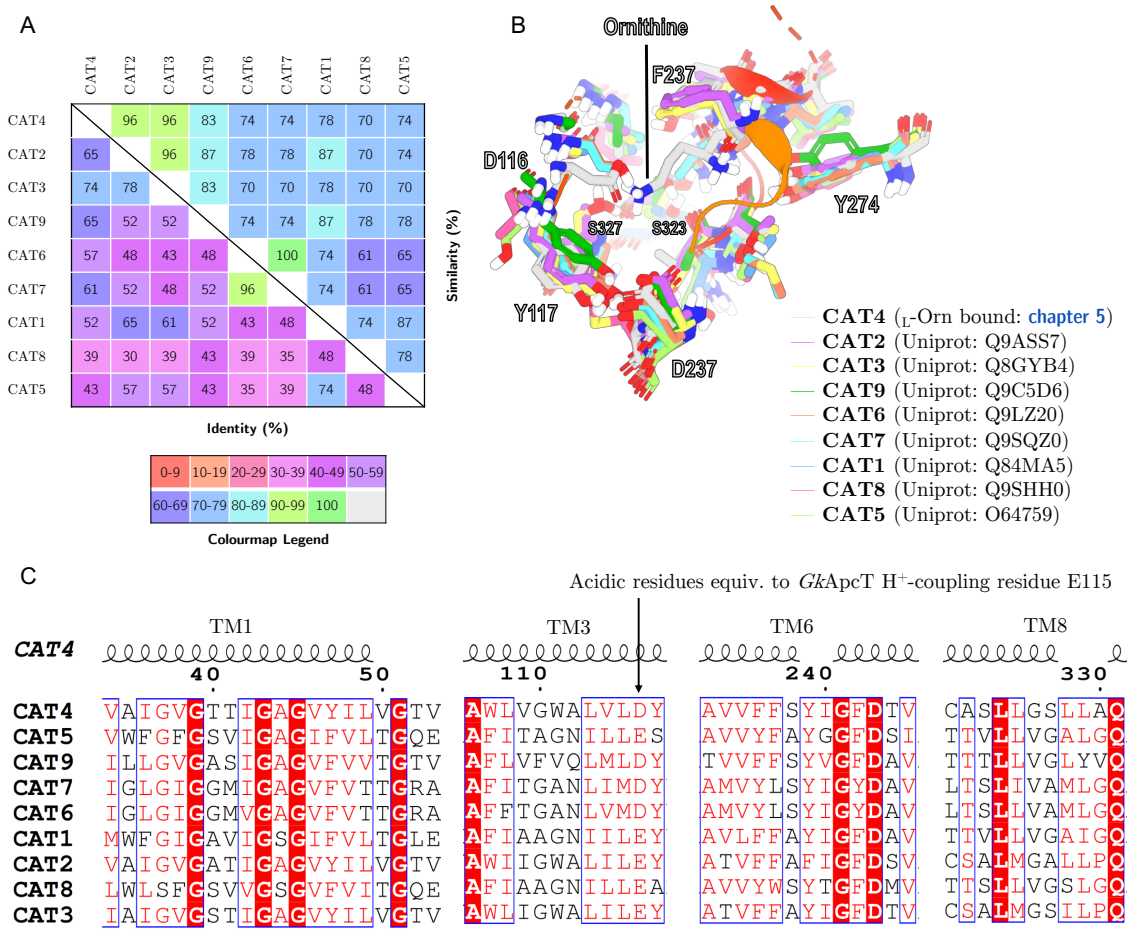


Figure 1.11 Comparison of Arabidopsis CAT family transporters. (A) Identity and similarity heat map of the binding site residues. (B) Structural superposition of CAT binding sites with L-Orn occupying the substrate pocket (from **chapter 5**) (TM6 residues hidden for clarity). (C) Multiple sequence alignment of CATs key residues lining the substrate binding site coloured in ESPript3.0 format (Robert and Gouet, 2014).

2004), the plant equivalent of mammalian lysosomes and yeast vacuoles, but it has also been detected in the ER. Additionally, CAT4 shares the largest binding site similarity with the L-arginine transporter CAT2, shown to be pH-dependent in *S. lycopersicum*. While CAT8 and CAT9 are localised partially in the tonoplast, only CAT2 and CAT4 are primarily localised in the tonoplast. CAT1, CAT5 and CAT6 localise in the plasma membrane, CAT3 in the ER and CAT9 is mainly vesicular (Carter *et al.*, 2004; Y. H. Su, Frommer and Ludewig, 2004; Yang *et al.*, 2014; Yang, Stierhof and Ludewig, 2015). With the localisation of

mammalian SLC7 members 1-3 known to be primarily in the plasma membrane, the lack of such information for mammalian SLC7A4 may hypothetically suggest an evolutionary relationship with *At*CAT4 and lysosomal localisation.

1.4 Transporter Interactions with Lipids

The presentation of APCS transporters above has focused on their physiology and structure-based mechanism of function. However, there is growing evidence of cross-talk between membrane transporters and their membrane environment that modulates their functional behaviour.

1.4.1 Mechanisms of Protein Modulation by Lipids

Membranes consist of dynamic lipid molecules that self-associate, diffuse at different rates, rotate about their lateral axis, and flip-flop spontaneously. Moreover, membranes composition differs between cells/tissues and subcellular organelles. For example, the neuronal plasma membrane (PM) has high percentages of cholesterol and anionic lipids (Fitzner *et al.*, 2020) which in turn play an essential role on neuronal APC transporter functionality – from conformational selectivity to stability and oligomerization (Scanlon, Williams and Schloss, 2001; Anderluh *et al.*, 2017). On the other hand intracellular organelle membranes are thinner compared to the PM because they are enriched in short-chain lipids (Mitra *et al.*, 2004). Yeast utilise these thickness gradients in their ER membrane as a filtering mechanism for trafficking proteins with long TM helices (Prasad, Sliwa-Gonzalez and Barral, 2020). The dynamics of the membranes are well known to affect the dynamics of the membrane proteins.

Biophysical Mechanisms of Protein Modulation

Membrane transporter function can be modulated by the combination of different forces from the membrane (Figure 1.12). A bilayer has a local thickness which is determined by the length and saturation of its lipid chains. "Hydrophobic mismatch" between the protein hydrophobic belt and the lipid hydrophobic moieties deforms the bilayer planar geometry to compensate the entropically unfavourable mismatch. This can in turn (a) apply compression forces to proteins inducing conformational and functional transitions (Cornelius, 2001; Botelho *et al.*, 2006; Andersen and Koeppe, 2007), (b) induce formation of protein dimers or higher-order clusters to cover the mismatched interface (Periole, Zeppelin and Schiøtt, 2018; Chadda *et al.*, 2021), (c) alleviate mismatch by lipid leaflet flipping (Falzone *et al.*, 2022). Membranes are also asymmetrical causing local thickening or thinning modulating curvature which can also be sensed by membrane proteins (Aimon *et al.*, 2014). Membrane proteins are also affected by membrane viscosity, impacting their lateral diffusion. Viscosity depends on lipid species, cholesterol and protein concentration, and temperature (homeoviscous adaptation) (Sinensky, 1974). Cells can apply mechanical stimuli via membrane stretching, compression or localised forces to modulate membrane proteins. An example are the mechanosensitive PIEZO1 and PIEZO2 channels who's gating is directly coupled to the deformation of the surrounding membrane (Coste *et al.*, 2010). There are no known cases of mechanosensation in solute carriers.

Cholesterol-induced lipid ordering can conformationally modulate the activity

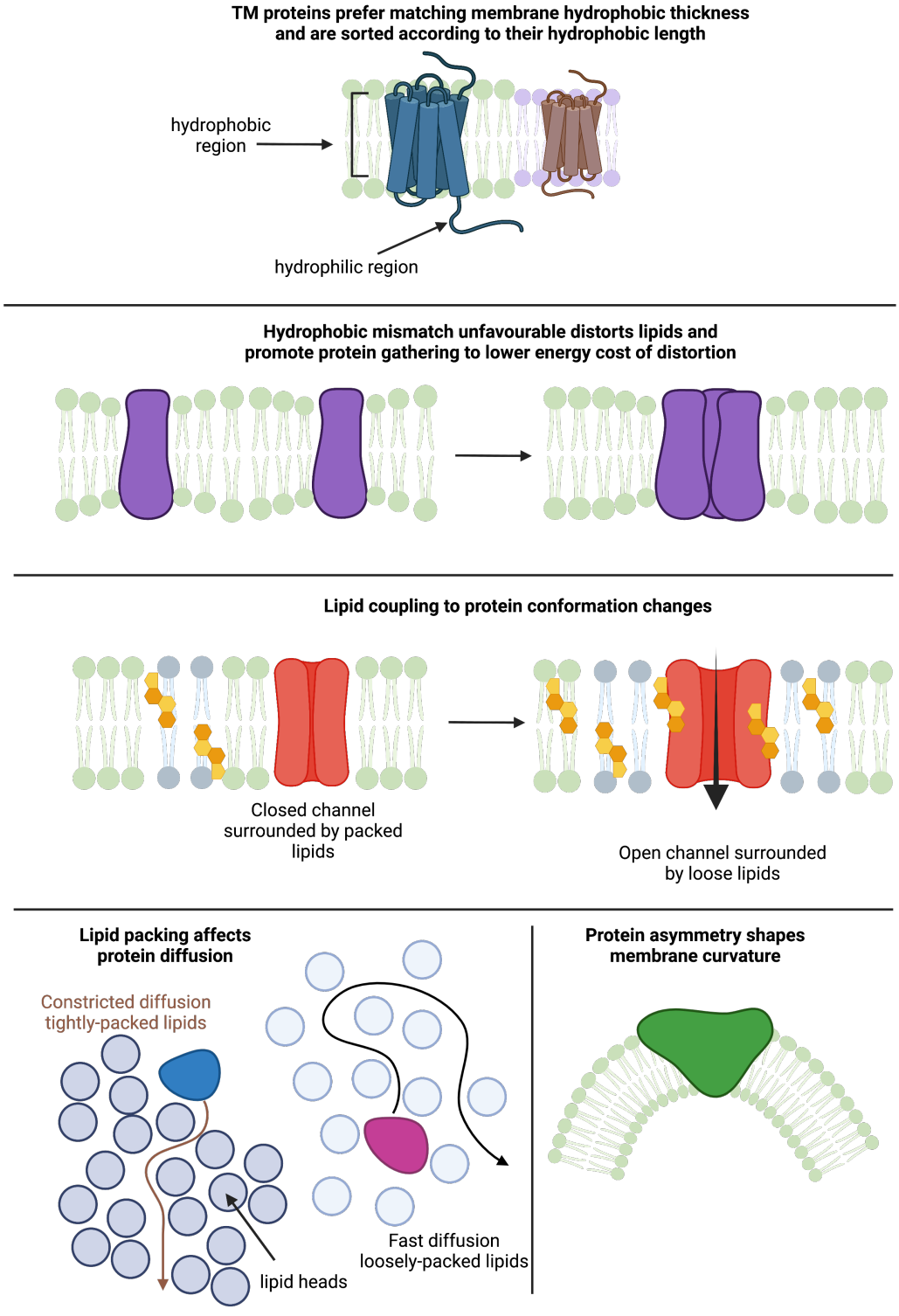


Figure 1.12 Mechanisms of protein modulation by the membrane. Schematic illustration based on mechanisms reviewed by Levental and Lyman, 2022.

of membrane proteins. For example, the adenosine $A2_A$ receptor in nanodiscs is

conformationally driven to the active state in relatively low cholesterol concentrations ($< 13\%$). The cholesterols (1-5 per leaflet) increased lipid ordering (35-40 per leaflet) but with only transient weak cholesterol interactions with $A2_AR$ by ^{19}F-NMR (Huang *et al.*, 2022). Cholesterols can also form microdomains, known as lipid-rafts, due to liquid-liquid separation; self-associated cholesterols and sphingolipids in proximity with saturated phospholipids (Sezgin *et al.*, 2017). Lipid-rafts consisting of liquid-ordered (L_o) and liquid-disordered (L_d) phases gather membrane proteins, affecting their trafficking and localisation (Schachter *et al.*, 2022).

Specific Chemically-Defined Protein-Lipid Interactions

Since membrane proteins and lipids diffuse along the two-dimensional membrane, they are bound to interact. These protein-lipid interactions can be highly specific, with the lipid interacting with membrane-facing residues (Contreras *et al.*, 2012). Protein-lipid interactions can be:

i. Annular; lipid sites lack specificity mediating the interaction of a hydrophobic transmembrane (TM) protein segment with the bulk membrane by forming a ring-like first solvation shell around the protein (Marsh, 2008) or, higher-order solvation shells (Gonen *et al.*, 2005).

ii. Non-annular; interacting lipids are typically embedded deeper within protein sites and are involved in regulation of protein function (Corradi, Besian I. Sejdiu, *et al.*, 2019).

Lipids can act as allosteric ligands, binding to specific protein pockets with ligand-like affinity presenting a novel strategy for lipophilic drug development (Payandeh and Volgraf, 2021). Specific lipids such as phosphatidylinositol 4,5-

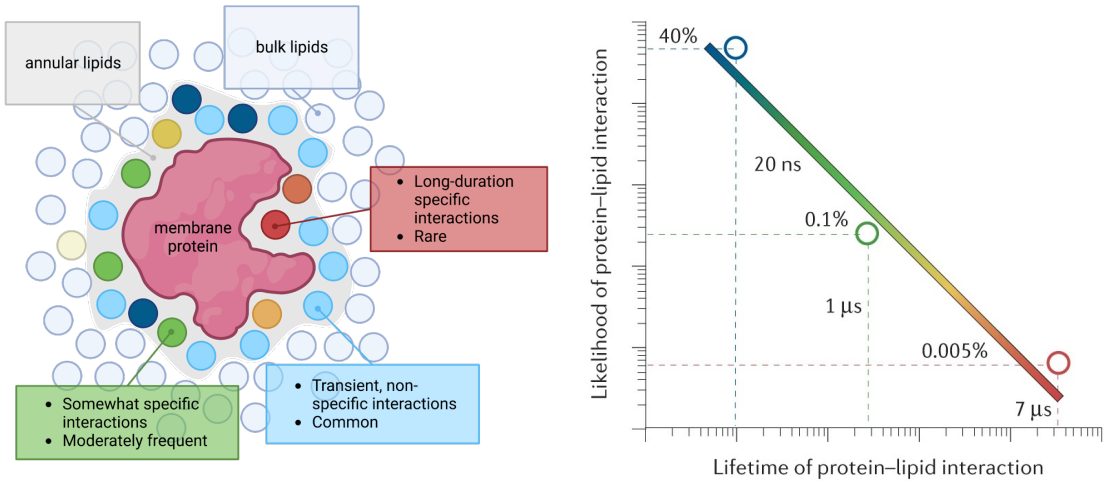


Figure 1.13. (Left panel) Cartoon illustration of membrane protein top-view and a shell of annular lipids (circles shaded grey) engulfing the protein. The surrounding bulk lipids (light-blue) exchange with fast rates. The exchange rates in the vicinity of the protein are slower due to longer-lived interactions (red circles). (Right panel) Correlation of the change of protein-lipids interaction time and the likelihood of the interaction. The bar scale is coloured according to lifetimes from red (long duration) to blue (short duration). The lipids on the left panel are coloured according to the bar scale on the right panel. Correlation plot from (Levental and Lyman, 2022).

bisphosphate (PIP2) or cardiolipin, usually have charged or polar headgroups to interact through electrostatic protein pockets, while protein pockets may drive lipid specificity by surface complementarity and charged/hydrophobic of residues (Corey *et al.*, 2021; Song *et al.*, 2022). Meta-analysis of MD simulations of protein-lipid binding revealed that the more specific an interaction is, the longer its lifetime and the lower the likelihood of it occurring (Figure 1.13) (Levental and Lyman, 2022).

Methods to Study Protein-Lipid Binding

Membrane protein-lipid interactions are often challenging to study as techniques are typically restricted to a relatively narrow spatio-temporal range (Levental and Lyman, 2022). On the structural front, X-ray crystallography of

detergent-engulfed proteins might show densities corresponding to lipids with high-affinity that survive the purification procedure (Long *et al.*, 2007). Similarly, protein-lipid complexes have been identified determined by electron crystallisation (Raunser and Walz, 2009), wherein the protein is reconstituted into a continuous 2D lipid bilayer to form crystals. Recent developments in cryoEM have permitted identification of densities corresponding to lipids in both detergent-engulfed protein and nanodisc reconstitutions (Bayburt, Grinkova and Sligar, 2002; Denisov *et al.*, 2004) – discoidal lipid bilayers with a diameter of 8-16 nm that are stable in solution by two encircling amphipathic helical protein belts. However, when lipid densities are found they may be misinterpreted as strongly bound specific lipid molecules. Whether these densities correspond to lipids with functional relevance is debated and functional investigation of the lipid site is necessary (Neumann *et al.*, 2022). Magic-angle-spinning (MAS) solid-state NMR (SSNMR) is suited for protein–lipid interactions in chemically-defined proteoliposomes (Elkins *et al.*, 2017), especially for [^{13}C]-cholesterols (Elkins, Sergeyev and Hong, 2018).

Native mass spectrometry (nMS) has proven to be highly beneficial to studying protein-lipid interactions (Bolla *et al.*, 2019) showing binding specificities for different lipids, such as LeuT phosphatidylglycerol-promoted dimerisation (Gupta *et al.*, 2017). Another method gaining popularity is the use of "clickable" photoactivatable lipid derivatives, such as cholesterol derivatised with a reactive diazirine group that can be UV-activated to chemically crosslink with the target protein (Hu *et al.*, 2020; Krishnan *et al.*, 2021). The complex can then be analysed by SDS-PAGE or mass spectrometry (Hulce *et al.*, 2013). Other methods include Förster resonance energy transfer (FRET) (Loura, Prieto and

Fernandes, 2010; Contreras *et al.*, 2012) and fluorescence cross-correlation spectroscopy (FCCS) (Betaneli, Petrov and Schwille, 2012).

Transport uptake assays for solute carriers and ion flux for channels can take the investigation to the final step, i.e. functional modulation by lipid binding, either in a whole-cell assay setup, or by proteoliposome reconstitution (Cosco *et al.*, 2020) – vesicles that contain the purified membrane proteins. The latter can benchmark specific lipid compositions and by extension associate conformation and function with lipid identity (Parker and Newstead, 2017; van 't Klooster *et al.*, 2020).

MD simulation has a unique advantage in providing chemical resolution visual detail and quantitative dynamics of protein-lipid microstates and ensemble statistics (Corey *et al.*, 2019; Corradi, Besian I Sejdiu, *et al.*, 2019; Marrink *et al.*, 2019). Coarse-grained representations of membrane lipids and proteins – in which groups of atoms are represented as single particles – is an effective way to alleviate computational demand and allow longer, biologically relevant simulation times (Marrink and Tieleman, 2013) to allow multiple lipids to sample the protein surface for binding. The coarse-grained approximations can be reconciled by conversion to finer atomistic resolution. Computational studies using multiscale representations, have been recognised for their role in unravelling the complex mechanisms of lipid specificity (Stansfeld and Sansom, 2011) and their impact on membrane transporter functionality (Laursen *et al.*, 2018; Zeppelin *et al.*, 2018). However MDs are limited by model accuracy, i.e. force field (Alessandri *et al.*, 2019), and simulation limitations, i.e. simulation spatiotemporal restrictions and adequate system sampling.

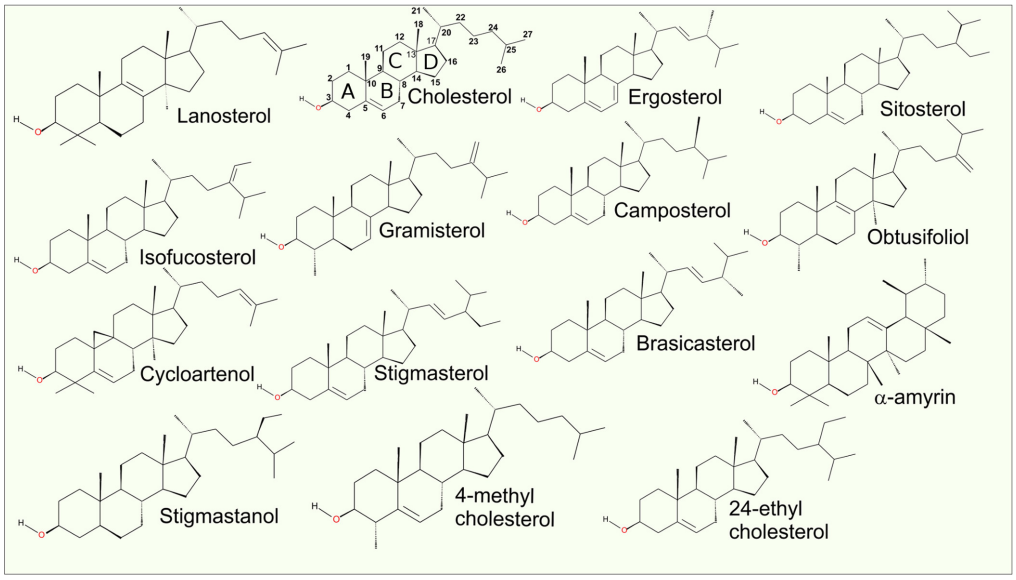


Figure 1.14 Structures of common sterols, including mammalian cholesterol, found across species in animals, plants, and fungi. Figure adapted (Enkavi *et al.*, 2019).

1.4.2 SLC7 Interactions with Cholesterols

Cholesterol comprises 20-40% of eukaryotic membranes (Mouritsen and Zuckermann, 2004). Due to its smaller molecular weight compared to other lipid classes, this percentage translates to almost half of the lipid molecules in human cell PMs being cholesterols (Das *et al.*, 2014). The highest concentration is in the PM reaching up to 50% mol and is lower in intracellular membranes and almost negligible in mitochondria (Enkavi *et al.*, 2019).

Cholesterol is vital to regulation of membrane fluidity – by promoting lipid ordering effects, participating in lipid-lipid interactions and exists in liquid-ordered lipid rafts (Allen, Halverson-Tamboli and Rasenick, 2007). Although they lack cholesterol, prokaryotes reserve a similar steroid molecule class, known as the hopanoids, which functions similarly in regulating membrane fluidity (Sáenz *et al.*, 2012). Despite the commonly held belief that plants do not contain cho-

lesterol, they contain phytosterols. Phytosterol composition is 14% mol cholesterol in arabidopsis, making cholesterol the second most abundant after stigmasterol (Valitova, Sulkarnayeva and Minibayeva, 2016). Cholesterol is present in fungal membranes, although the principal sterol is ergosterol (80 mol %) (Enkavi *et al.*, 2019).

Cholesterol interacts with membrane proteins both via protein-specific interactions (Grouleff *et al.*, 2015; Corradi *et al.*, 2018), or by modulating membrane protein properties such as lateral diffusion and mobility (Kwik *et al.*, 2003; Guigas and Weiss, 2016; Levental, Levental and Heberle, 2020). APC fold transporter structures report cholesterol densities in specific sites, namely; CHOL1 and CHOL2 sites on the drosophila Na^+-coupled DAT (Penmatsa, Wang and Gouaux, 2013; Wang, Penmatsa and Gouaux, 2015); CHOL1 CHOL2 and CHOL3 sites on the human Na^+-coupled SERT [cryoEM on SERT-nanodisc (Yang *et al.*, 2022), X-ray (Coleman, Green and Gouaux, 2016), cryoEM (Coleman *et al.*, 2019)]; one site on the lysosomal transceptor SLC38A9 (Castellano *et al.*, 2017a); three sites on the human SLC12A2 cation-chloride Na^+:K^+:$2Cl^-$ cotransporter NKCC1 (Neumann *et al.*, 2022). Cholesterol regulates neurotransmission by modulating DAT (Jones, Zhen and Reith, 2012) and SERT K_M and V_{max} of serotonin transport (Laursen *et al.*, 2018). Insight from coarse-grained MD simulation revealed six potential cholesterol binding sites in SERT, (Ferraro *et al.*, 2016), including five common sites between different monoamine transporters (Zeppelin *et al.*, 2018) (**Figure 1.15**).

Similar implications have been proposed for SLC7 transporters. The uptake

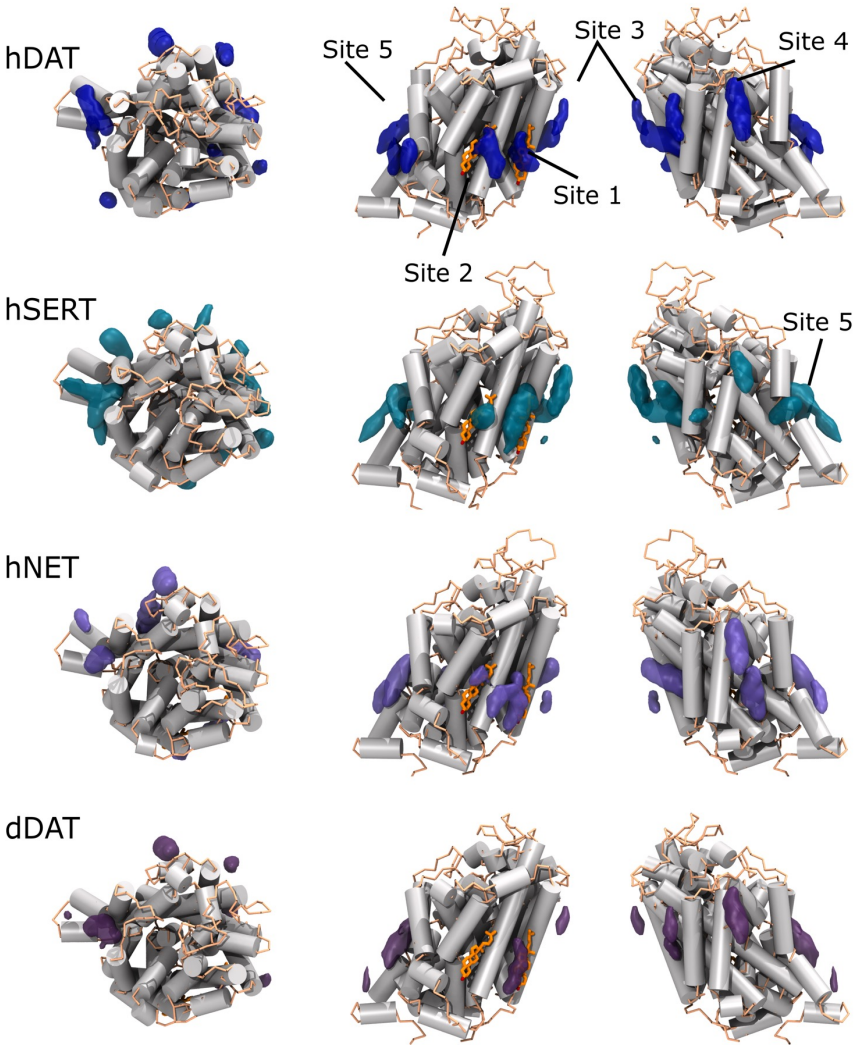


Figure 1.15 Cholesterol interaction sites in SLC6s from CG simulations with 20% membrane cholesterol. Cholesterol sites occupancy maps are shown. The maps are drawn at occupancy values at least three times higher than bulk values. Cholesterols structurally resolved are shown in orange sticks. Adapted from (Zeppelin *et al.*, 2018).

of ^{3}[H]-L-DOPA by 4F2hc/LAT1 -expressing HEK293 showed reduction of V_{max}, but not K_M after extracting membrane cholesterol with MβCD – a cholesterol extractor (Dickens *et al.*, 2017). The observation suggests reduced maximal transport activity without discernible change in substrate affinity implying changes in the translocation cycle. These K_M-V_{max} results from cells were verified in liposome reconstituted 4F2hc/LAT1 (Cosco *et al.*, 2020). Surprisingly, the external affinity of LAT1 for substrate was unaffected by cholesterol whereas the

internal affinity strongly increased, suggesting that cholesterol stabilises the inward-open LAT1 conformation (Cosco *et al.*, 2020), in contrast to the outward-facing conformation preferred by cholesterol in SERT (Laursen *et al.*, 2018; Zeppelin *et al.*, 2018). L-leucine uptake by 4F2hc/LAT1 in liposomes was also increased with cholesterol supplementation in another study (Lee *et al.*, 2019a).

To enrich the picture, cryoEM structures of human HAT transporters, rBAT/$b^{0,+}$AT, 4F2hc/LAT1, 4F2hc/xCT, are accompanied by co-purified steroid-like densities in their EM maps (see **chapter 6**; **Figure 6.6**), which have mostly been assigned to cholesterol or cholesteryl hemisuccinate, the latter being a detergent additive in purifications to mimic cholesterol and increase stability. Visualisation of these sites showed that the cholesterol/CHS pockets are the amalgamation of residues belonging to adjacent TM helices. This precludes a cholesterol interacting domain mechanism, CRAC/CARC (Fantini and Barrantes, 2013), and suggests that the APC architecture drives cholesterol binding, perhaps by creating interhelical grooves common in the SLC7 fold. Surface complementarity for cholesterol binding has been demonstrated for GPCRs (Taghon *et al.*, 2021), enhancing the hypothesis. Unlike other lipids, such as the anionic cardiolipin (Corey *et al.*, 2021), molecular binding determinants have not been suggested for cholesterol, since it is a small, flat molecule that lacks net charge.

1.5 Novel Methods in Membrane Protein Structural Biology

1.5.1 Electron Cryomicroscopy & Single Particle Analysis

The principal physical underpinning for cryoEM is electron scattering. After

the incident plane wave is transmitted through the sample, the scattered electron wave signal produces images on electron detectors that are 2D projections of the 3D structures of biological macromolecules. The information about the projected structure is contained primarily in the spatial variations of the electron wave phase. Many projections need to be merged to recover information in the third dimension. After data have been merged, it seems possible to reconstruct a 3D density map of the object by computing a 3D inverse Fourier transform. In practise the 3D reconstruction step is not as simple, and sophisticated algorithms are used to obtain the 3D reconstruction (Glaeser, Nogales and Chiu, 2021).

Over the past 40 years, electron cryo-microscopy (cryoEM) has advanced into one of the most powerful in molecular biology, not least due to its success in membrane protein structural biology. Its advantage is the lack of conditions for crystallisation and large quantities of extremely pure protein samples whilst allowing analysis of nanodisc reconstitutions (Kühlbrandt, 2022). However, the technique suffered from poor signal-to-noise of individual micrographs making alignments of smaller proteins difficult (Kühlbrandt, 2022). The recent advent of direct electron detectors (e.g., Gatan K2 and Falcon-II), improved signal-to-noise and the detective quantum efficiency, leading to the "resolution revolution". New detectors are highly sensitive and with fast readouts so the electron dose (i.e. 20-50 $e/Å^2$) can be spread over multiple movie frames, allowing on-the-fly correction of beam-induced movement (Zheng *et al.*, 2017) and increasing map clarity. As methodological and computational bottlenecks are being resolved, the protein sample's quality will pose the challenge in structure determination (Kühlbrandt, 2022).

Size and domain rigidity of the protein are important to generate 2D projections (class averages) clear enough for 3D reconstruction. This is especially the case for small membrane proteins below ~ 60 kDa (Herzik, Wu and Lander, 2019). The emerging technology of nanobody binders pushes the limit of membrane proteins to a lower value, shown recently for the 55 kDa H^+-coupled folate transporter in complex with a nanobody (Parker, Deme, Kuteyi, *et al.*, 2021). Such is also the case of SLC7 CATs (without a major solvent-exposed domain) according to their topology diagram (Figure 1.10). The deep-learning algorithm, Alphafold2 (Jumper *et al.*, 2021), can help predict transporter structures and inform about the need for imaging fiducials.

1.5.2 Nanobodies as Structural Biology Aids

1.5.2.1 Introduction to Nanobodies

Antibodies (Ab) are a class of glycoproteins that are secreted by immune cells and are designed to be selective binders for their protein targets via specific protein-protein interactions. Their common role is to bind xenobiotic antigens and mark them for degradation. This selectivity is a result of panning the antigen through a vast library of antibodies. A biotechnological revolution sparked from this ability, i.e., protein scientists could employ this strategy to generate nanobodies targeting proteins of interest *in vitro*. Example applications in molecular and cellular biology techniques that leverage antigen-antibody technology are the enzyme-linked immunosorbent assay (ELISA), western blotting, immunoprecipitation, immunohistochemistry, and flow cytometry.

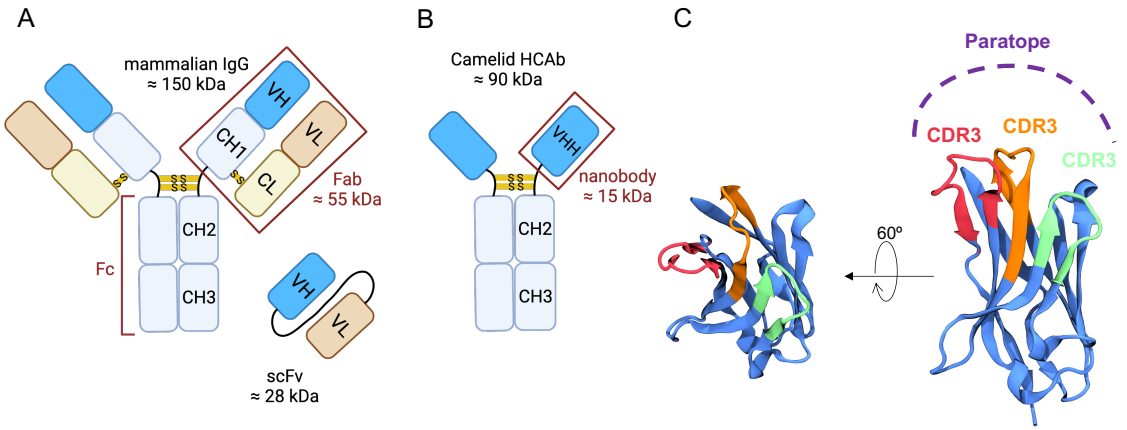


Figure 1.16 The structural organisation of mammalian and camelid antibodies. Schematic of a (A) mammalian Ab and a (B) camelid HCAb. The V_HH domain is a nanobody (in rectangle), with the characteristic beta sheet Greek-motif and hypervariable CDRs 1-3 coloured in green/orange/red (C).

Ab's consist of four polypeptides, two heavy chains and two light chains each composed of smaller domains (V_H, C_H1-3 and V_L, C_L). In 3D, 3 globular domains are formed: 2 identical Fab fragments (each having V_L, V_H, C_L, and C_H1) and 1 crystalline fragment F_C (2× C_H2, 2× C_H3). The V_{L+H} domains make up the "variable" region (F_V), with the V_H domain being the "hypervariable" complementarity-determining regions (CDRs 1-3) (**Figure 1.16-A**). The variable region can be separately produced and still retain binding specificity to the antigen, known as scF_V. Camelid heavy-chain antibodies (HCAb) do not possess a light chain (**Figure 1.16-B**). Their specificity is determined by a single V_HH. This domain, truncated and expressed on its own is known as a nanobody (Nb, ~15kDa).

Nanobodies have high-affinities (10^{-6}-10^{-9} M) for the selected target. A Nb has 3 CDRs that compose the paratope, the molecular surface that selectively binds the target's epitope (**Figure 1.16-C**). Nbs compensate for the loss of the light chain primarily by diversification in CDR backbone conformations (Sircar et al., 2011;

Mitchell and Colwell, 2018) especially in CDR3 which has more sequence diversity and length variability. CDR3 has more configurational degrees of freedom than the other CDRs and improves shape-complementarity to the target (Mitchell and Colwell, 2018). The paratope mediates H-bonds, salt bridges, π-π and π-cation stacking, van der Waals interactions whilst simultaneously providing surface-complementarity to the target. Commonly, a protruding CDR surface is inserted deeply into membrane protein cavities, either aqueous or hydrophobic pockets (Rasmussen, Choi, *et al.*, 2011; Löbel *et al.*, 2022).

1.5.2.2 Nanobodies as CryoEM Imaging Fiducials

The recent membrane protein structural biology literature provides a myriad of examples in where structural determination of a resistant target protein was afforded with the use of molecular chaperones like nanobodies (Bräuer *et al.*, 2019; Löbel *et al.*, 2022) and derivative technologies such as NabFab, legobodies, (pro)macrobodies, megabodies (Pardon *et al.*, 2014; Bloch *et al.*, 2021; Uchański *et al.*, 2021; Wu and Rapoport, 2021; Botte *et al.*, 2022). For example, molecular chaperones can be used in membrane protein crystallography to either provide or enhance pre-existing crystal contacts, or even rigidify conformations of a flexible target, thus driving crystallisation (**Figure 1.17**) (Rasmussen, DeVree, *et al.*, 2011).

In cryoEM the molecular chaperone helps with conformational homogenisation and can be utilised as an imaging fiducial. The binder acts as a guide feature against which to "*in silico* purify" the dataset; aiding image alignment, supplementing particle size, and breaking any target-intrinsic pseudo-symmetry (Wu et al., 2012; Hofmann et al., 2019). This technology has proved fruitful in resolving

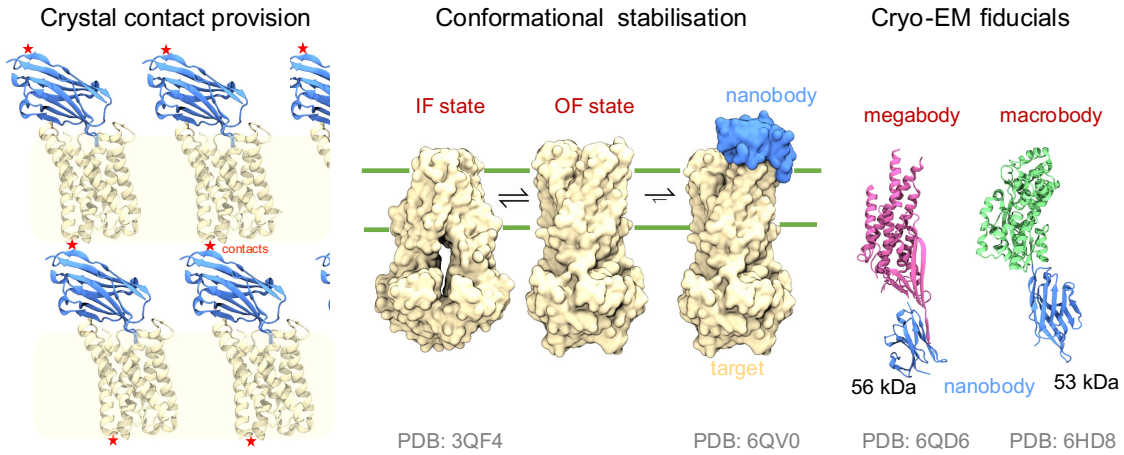


Figure 1.17 Examples of nanobody applications in modern membrane protein structural biology as: (A) crystal contact mediator for the KDEL receptor, (B) conformational-selector of the outward-open state of the ABC exporter TM287/288. (C) Examples of cryoEM fiducial marker utilisation of a nanobody covalently linked with cHopQ or MBP.

a number of small membrane proteins by cryoEM, including the serotonin transporter SERT [70 kDa, 3.6 Å, (Coleman et al., 2019)] and complexes such as the thioredoxin-fused $A2_A$ adrenergic receptor coupled to mini-Gs-β1γ2 [135 kDa, 3.8 Å, (García-Nafría et al., 2018)].

1.5.2.3 Nanobody Generation Methods

Camelid Immunisation

The method of choice for Nb generation is by immunisation of camelids with target membrane protein for the pre-enrichment of B-cells encoding HCAb's, and downstream isolation of peripheral blood lymphocytes. A target-customised V_HH mRNA library is made (Pardon *et al.*, 2014) and is subsequently screened and ranked against purified target with directed evolution methods, most commonly phage or yeast display (McMahon *et al.*, 2018; Zimmermann *et al.*, 2018). After several rounds of affinity selection against the target (panning), hits are identified via ELISA. Camelid nanobodies are relatively easy to generate diverse sets for high-throughput selection, express and purify to high yields from multiple

hosts, are generally stable ($T_{m,app}$ [60–80] °C), can have up to nanomolar affinities for the target, and can be used for other applications (e.g. fluorescent imaging, cross-linking etc.) (Pardon *et al.*, 2014). Despite their benefits, the generation of Nbs is an animal-involving process, requiring special access with ethical considerations. An immunisation cycle is lengthy (typically 6-8 weeks) with multiple injections rounds followed by lymphocyte preparation and construction of the phage library (Pardon *et al.*, 2014). In addition, the injected protein molecule must be exceptionally stable at the high temperatures (38 °C) and withstand immune defence. This is specifically challenging for membrane proteins as they must be correctly reconstituted in proteoliposomes.

Synthetic Nanobodies for *In-vitro* Generation

In an effort to solve the immunisation drawbacks related to nanobody camelid generation, synthetic nanobodies (Syb) were recently developed (Zimmermann *et al.*, 2018, 2020a). Sybs can be generated *in vitro* with detergent-purified protein, which is inexpedient in camelid immunisation as the detergent may dissociate and the protein will unfold (Pardon *et al.*, 2014). In the Syb pipeline, the immunisation process is replaced by a 1-2 day ribosome display of a rich mRNA library (Zimmermann *et al.*, 2018) and can be completed within 3 weeks. Perhaps the most important advantage is the short time required for Syb selection against a target until reaching the panning stage. The Syb library is advantageous by being in the form of an mRNA library which is exceptionally rich in sequence space ($>10^{12}$ unique binders); compensating for the lack of affinity maturated binders (natively in camelid immune system). Other *in vitro* libraries exist but are limited to 10^{8}-10^{10} library members (Yan *et al.*, 2014; Moutel *et al.*, 2016;

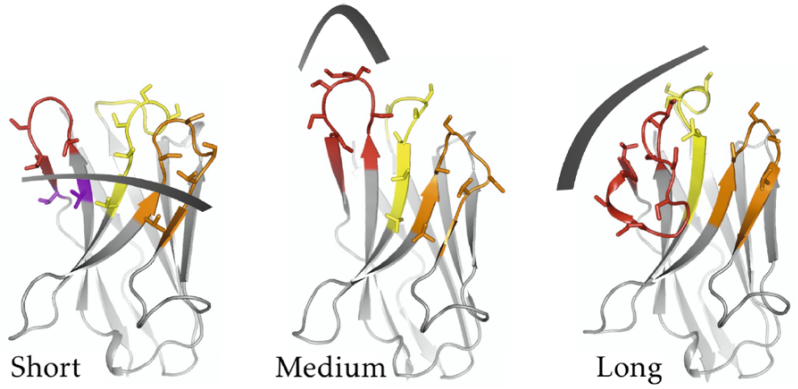


Figure 1.18 The structural templates of the three Syb libraries used in this thesis. The short CDR3 library creates a concave surface, the medium CDR3 length creates a protruding loop and a long CDR3 creates a convex surface. Figure adapted from (Zimmermann *et al.*, 2018).

McMahon *et al.*, 2018). Additionally, Sybs are designed to have "concave", "convex" and "loop" geometries, based on the paratope binding surface ("short", "long" and "medium" respectively), to reproduce the three different binder modes of target-Nb observed in the PDB (Figure 1.18). Sybs have been successfully used as alternatives to cameloid Nb in crystallisation (Bräuer *et al.*, 2019; Löbel *et al.*, 2022) and cryoEM (Custódio *et al.*, 2020).

1.6 Hypotheses & Aims

SLC7 structural research has started to explain HAT mechanism of transport (§1.3). However, the structural basis of CAT function from a eukaryotic homologue was lacking. Comparison of SLC7A4 Alphafold2 prediction with prokaryotic structures suggests similarities with the bacterial transporter (Figure 1.19-A,B). However, *Gk*ApcT contains 12 TM helices while all eukaryotic CATs are known to have 14 TMs, further raising the need for a eukaryotic structure to provide the structural basis for this difference.

Functionally SLC7A4 remains uncharacterised, but it is unlikely that it is without transport function (Wolf *et al.*, 2002). *SLC7A4* mRNA is expressed in the testes, placenta and the brain, while the gene is deleted in velocardiofacial syndrome (Sperandeo *et al.*, 1998). SLC7A4 with a GFP-tag localises in the plasma membrane in *X. laevis* oocytes, but showed no transport activity for the tested substrates (twelve α-amino acids incl. cationic and polyamines) (Wolf *et al.*, 2002). The authors did not test all possible substrate and pH combinations leaving possibility for amino acid transport. No gene knockout studies have been reported for SLC7A4 and thus no phenotypical information is available.

Intriguingly, multiple sequence alignment of *Gk*ApcT with CATs provided an interesting observation; the H^+-coupling E115 residue is mutated to Ser in pH-independent hCATs 1-3, yet the acidic residue is conserved in human SLC7A4 (**Figure 1.19-C,D**). This hypothesises that the orphan transporter SLC7A4 might be H^+-coupled. If this is the case, SLC7A4 would be the first H^+-coupled amino acid transporter within the SLC7 family.

Among the pertinent questions remaining unanswered at the inception of my D.Phil were:

- How are eukaryotic SLC7 CAT structures related to the prokaryotic homologue?
- What is the molecular function of the additional TM helices predicted for eukaryotic CATs which are absent in HATs?
- Is SLC7A4 an amino acid transporter?
- What is the substrate for SLC7A4?
- Is SLC7A4 a H^+-coupled transporter?

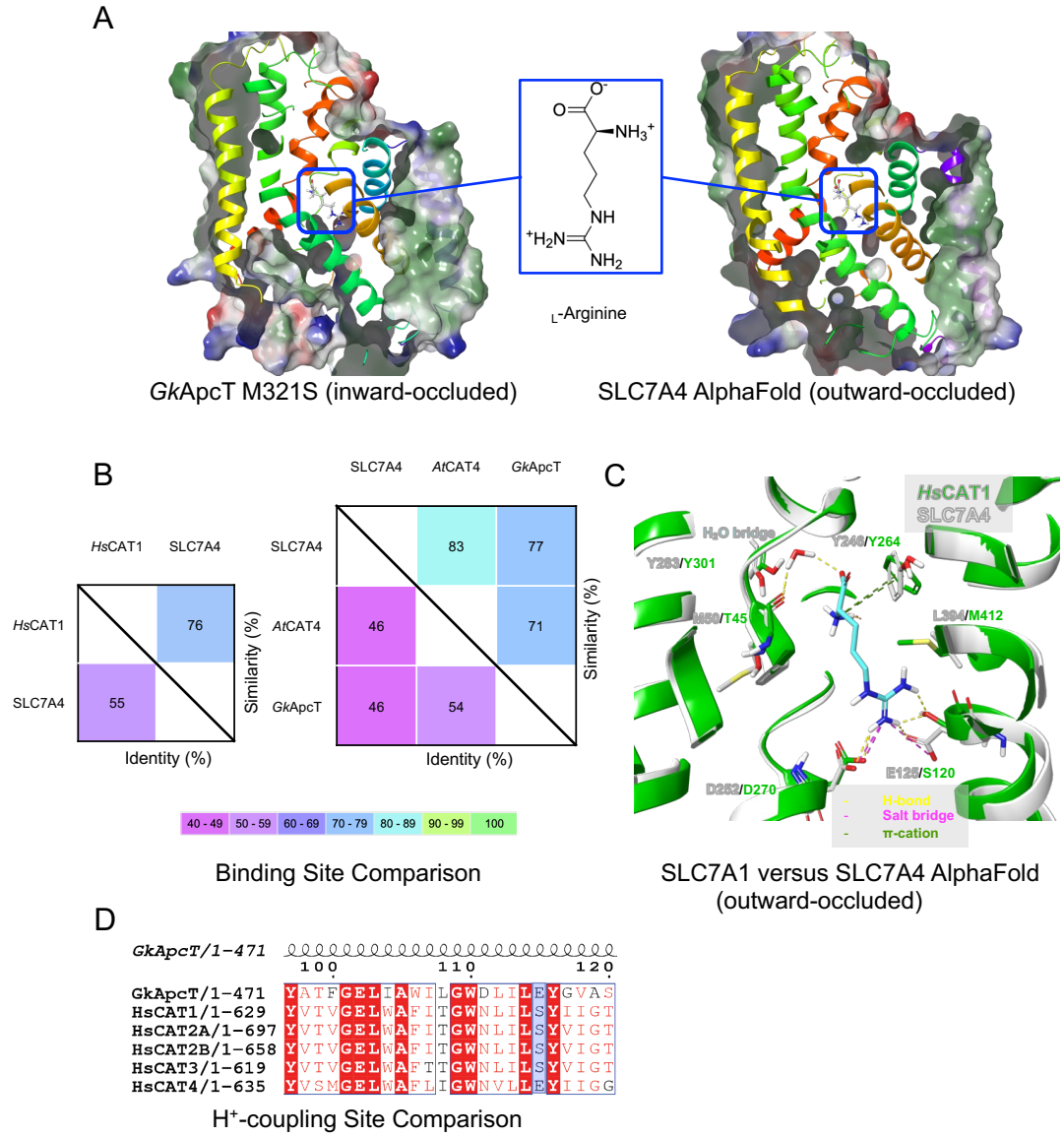


Figure 1.19 Hypothesis for the structure and function of the CAT subfamily member, SLC7A4. (A) Structural comparison of the L-arginine transporter, M321S GkApcT, with the Alphafold2 prediction of SLC7A4 superimposed with L-arginine. (B) Binding site residue (< 6 Å from ligand) sequence identity and similarity between L-arginine transporter *Gk*ApcT, human CAT1 and orphan SLC7A4. (C) Superposition of human CAT1 and SLC7A4 Alphafold2 binding sites shows similar interaction profiles for L-arginine. (D) Multiple sequence alignment of prokaryotic H^+-coupled GkApcT with CAT members shows that SLC7A4 is the only member conserving the E115 H^+-coupling residue.

- How is the function of SLC7A4 related with its localisation in the cell and what are the implications for its role in cellular biochemistry?

This thesis aims to understand whether SLC7A4 might be an amino acid transporter via an integrative structural biology approach coupling cryoEM with functional and biophysical experiments and MD simulation. This included:

1. Identifying an SLC7A4 homologue by a range of experimental procedures including cloning, small- and large-scale expression, purification optimisation (chapter 3).
2. Biophysical characterisation of constructs, ranking according to protein quality and suitability for structural trials (chapter 3).
3. Sybody screening and binding studies to identify a suitable binder for structural studies with nanomolar affinity (chapter 4).
4. Solving the structure of an SLC7A4 homologue and evaluating similarities and differences within the SLC7 family (chapter 5).
5. Performing *in silico* and *in vitro* functional characterisation by binding assays, mutational studies, MD simulation, *X.* oocyte electrophysiological recordings and amino acid transport assays in a human immortalised cell line (HeLa) (chapter 5).

The atomic structure of SLC7A4 would place the answers to the above questions in the context of a molecular, mechanistic basis, and this is therefore the defining question. This thesis presents the SLC7A4 structure and utilises it to test its potential to transport substrates.

Cholesterol interactions with SLC7 transporters are evident both structurally and functionally and therefore promise an interesting means of modulation via well-defined cholesterol binding sites. In the final chapter, the binding of cholesterol to human HAT transporters is investigated by a multiscale MD simulation

pipeline, leading to *in vitro* functional characterisations to understand commonalities and differences between HAT members and their conformations regarding cholesterol binding (**chapter 6**).

Experimental & Computational Methods

CHAPTER 2

The materials and methods chapter of this thesis is organised in two separate parts: (a) the description of experimental methods and (b) the description of computational simulations methods. A section on MD theory and free energy methods used in this thesis is provided in **Appendix B**.

The experiments in this thesis were conducted in the laboratory of Prof. Dr Simon Newstead at the Department of Biochemistry of the University of Oxford. The molecular dynamics (MD) simulations were setup, run and analysed in the research group of Prof. Dr Philip Biggin, using the resources of the Structural Bioinformatics and Computational Biochemistry unit at the Department of Biochemistry of the University of Oxford.

EXPERIMENTAL METHODS

2.1 Cloning & Mutagenesis

2.1.1 Yeast Expression Vectors

GeneArt Strings DNA Fragments of SLC7A4 constructs, codon optimised for expression in *S. cerevisiae*, were ordered with 3' and 5' overhangs, homologous to the insertion site of the SmaI-digested pDDGFP2-Leu2d vector (Joanne L Parker and Newstead, 2014). Vector was linearised by restriction digest with SmaI endonuclease (FastDigest Restriction Enzymes, Thermo Fisher Scientific). If the recombination overhangs were not available in the ordered genestrings, they were introduced by means of Polymerase Chain Reaction (PCR) in a thermocycler (BioRad C1000 Touch Thermal Cycler) using KOD HotStart DNA polymerase (Novagen) and the reaction setup and executed according to the polymerase manufacturer's instructions. The reaction mixture (50 µl final volume) contained 0.4 µM forward primer, 0.4 µM reverse primer, 200 µM of each of the dNTPs, 1.5 mM $MgSO_4$, 1× KOD buffer and 1.0 µl *Thermococcus kodakaraensis* KOD1 DNA polymerase.

Yeast Transformation

S. cerevisiae of the BJ5460 strain were transformed with the AcO^-Li^+ method as follows. 600 µL of an overnight (~ 16-18 h) culture growing in Yeast Extract–Peptone–Dextrose (YPD) was pelleted at 16,000×g for 15 s. The pellet was resuspended in 100 µL 100 mM AcO^-Li^+, 10mM Tris·HCl pH 7.4. Subsequently, it was supplemented with a mixture of 10 µL salmon sperm DNA (ssDNA) and 2 µL of 150 ng μl^{-1} pDDGFP2-Leu2d (SmaI-digested) and 10 µL of 200 ng μl^{-1}

PCR product was added. Subsequently 400 μL of 100 mM lithium acetate, 10 mM Tris·HCl pH 7.4 with 50% (w/w) PEG 3350 was added and the mixture was incubated at room temperature on a rotating wheel for 30 min. 50 μL of DMSO was added and the yeast cells were heat shocked at 42 °C for 15 min. Cells were then pelleted by centrifugation at 500 g for 30 s, resuspended in 75 μL autoclaved milliQ-grade water and spread onto a selective minus Uracil dropout media plate for 3 days at 30 °C, after which they were plated on a minus Leucine dropout media plate for another 2-3 days (until plate 70-80% plate saturation).

Yeast and *E. Coli* Plasmid Minipreps & SLC7A4 Homologue Preparation

8-12 mL of yeast transformed with unsequenced plasmid was grown overnight at 30 °C in minus Leucine dropout media. Cells were harvested by centrifugation at 3,000×g for 8 min and resuspended in 150 μL phosphate buffered saline (PBS) pH 7.4. 200 μL silica beads were added and the mixture was vortexed for 15 min at 4 °C. The resulting lysate was clarified by centrifugation at 500×g for 60 s and used in a standard miniprep kit according to the manufacturer's instructions (GeneJET kit – Thermo Fisher Scientific). The purified plasmid concentration was titled using a NanoDrop ND-1000 spectrometer. If yields were low (<1.5 μg) more yeast culture was miniprepped as above and combined.

The yeast-isolated plasmids contained a combination of empty and insert-containing vectors. This was transformed in *E. coli* (OmniMAX, Invitrogen), plated on Luria Broth (LB) agar (37 °C) grown overnight with ampicillin (100 μg ml^{-1}). Single transformant colonies were upscaled by inoculating to 5-7 ml LB with ampicillin the plasmids miniprepped according to the manufacturer's instructions. Successful gene insertion was verified by diagnostic digest and Sanger

sequencing (Source BioScience). Sequencing results were assessed by careful inspection of resulting chromatograms and pairwise alignment to the expected nucleotide sequence, dictated by the genestring sequence.

*At*CAT4 Mutagenesis

The *At*CAT4/pDDGFP2-GFP plasmid has a molecular weight of 10,489 bp which is too large for routine site-directed mutagenesis. Thus, the mutants were generated by extension of overlapping gene segments by PCR (Heckman and Pease, 2007) which consisted of three-steps shown schematically in Figure 2.1: (a) two PCR reactions yielding, amplicon AB – 5'-overhang with a BamHI restriction site to approximately 10 bp after the mutated nucleobases – and amplicon CD – approximately 10 bp before the mutated nucleobases to a 3'-overhang with a XhoI restriction site (including the C-terminal His_8-$TEV_{cleavage\ site}$-yeGFP nucleobase sequence), such that the 10 bp before and after the mutagenesis site resulted in complementary overhangs, (b) an annealing step of amplicons AB and CD followed by an extension step (without primers) to generate the template *At*$CAT4_{mutated}$-His_8-$TEV_{cleavage\ site}$-yeGFP, (c) an amplification step of the template with 5' and 3' primers to generate amplicon AD. Amplicon AD was gel-extracted (GeneJET kit - Thermo Fisher Scientific) and restriction digested with BamHI/XhoI. The digested amplicon AD was purified from the endonucleases via PCR cleanup and eluted in 20 µl autoclaved milliQ grade water. The product was ligated into a BamHI/XhoI digested and gel-extracted pDDGFP2 vector (Quick Stick Ligase, Bioline; room temperature incubation; 15 min) before transforming the reaction to 50 µl OmniMAX competent cells by standard *E. coli* transformation with 45 min outgrowth in LB. The cells were plated in LB plates

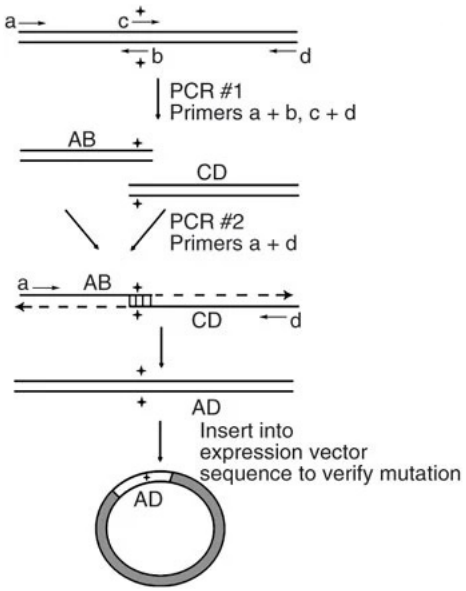


Figure 2.1 Overlap Extension PCR Method. Site-directed mutagenesis is accomplished by using mutagenic primers (b and c) and overhang primers (a and d) to generate intermediate PCR products AB and CD. Products AB and CD have complementary 3'-ends and are denatured when used as template DNA for the second PCR; strands of each product hybridize at their overlapping, complementary regions (containing the desired mutation indicated by the cross). Amplification of product AD in PCR #2 is driven by primers a and d. The final product AD can be inserted into an expression vector, which will be sequenced to ensure mutagenesis success. Figure adapted from Heckman and Pease (2007).

with ampicillin.

2.1.2 Mammalian & Oocyte Expression Vectors

Yeast codon optimised human SLC7A4 and *At*CAT4 were PCR amplified out of the pDDGFP2-Leu2D vector with 5' SacI and 3' FLAG-tag stop-codon XhoI primers and subcloned into pLexM-GFP vector digested with SacI/SalI pair (XhoI and SalI are compatible endonucleases). PCR amplicons were digested (FastDigest Restriction Enzymes, Thermo Fisher Scientific), and ligated (Quick Stick Ligase, Bioline) into pLexM-GFP. The steps were identical for the *X.* oocyte vector, pBF (also known as pFAW), and the amplicons were subcloned in the multiple cloning site with XbaI/XhoI (human SLC7A4) and BamHI/XhoI (*At*CAT4).

Human SLC7A5 (LAT1) was kindly provided by Dr Yongchan Lee in a pEG BacMam vector and was PCR amplified with a C-terminal FLAG tag. The WT LAT1 amplicon was subcloned into a pcDNA3.1(+) vector with HinIII/XhoI endonuclease pairs. Mutants of LAT1 in the pcDNA3.1(+) vector were generated

Protocol Step	T (°C)	Time (s)	cycles
Denaturation	94	120	1
Denaturation	94	15	23
Annealing	52-55	15	
Elongation	72	300	
Elongation	72	420	
End	12	∞	1

Table 2.1 Site-directed mutagenesis PCR amplification protocol steps programmed in the thermocycler.

by site-directed mutagenesis. Complimentary forward and reverse PCR primers (approximately 25-32 bp long) were designed containing the desired mutation in the middle of the primer. PCR was carried out with the KOD HotStart polymerase for 23 cycles (Table 2.1). Amplicons were treated with DpnI restriction enzyme for 3 h at 37 °C to digest methylated template DNA and directly transformed into OneShot® OmniMAX™2 T1R E. coli and plated onto LB agar plates with appropriate ampicillin. Transformants were grown, miniprepped and sequence-verified as described above.

2.1.3 Sybody Vectors

Sybodies selected against *Arabidopsis thaliana* CAT4 by ribosome display and were cloned into the phagemid pDx_init (Addgene #110101) to use during the phage display steps. The binder ORFs were sequentially subcloned into the ELISA vector pSb_init (Addgene #110100) and then the MBP-fusion expression vector pBXNPHM3 (Addgene #110099) and C-terminal His_{10}-tag pBXNPH3

(Addgene #110098). Subcloning steps were performed with the fragment-exchange (FX) cloning method, as outlined in detail in the Sybody publications (Zimmermann *et al.*, 2018, 2020b).

2.2 Protein Expression & Purification

2.2.1 Membrane Protein Expression in *S. cerevisiae*

The following protocol was used for the expression of SLC7A4 homologues for purification in various detergents (chapter 3). AtCAT4 was routinely expressed with this protocol (chapters 3,4,5). *S. cerevisiae* BJ5460 was transformed with sequenced SLC7A4-pDDGFP2-Leu2D plasmid and plated onto minus Uracil drop out agar plates for three days, then moved to minus Leucine drop out agar plates for another two to three days depending on plate saturation with colonies. A 70-80% saturation of plate surface with yeast was allowed.

The yeast was grown in 2× 100 ml overnight cultures in synthetic-complete minus Leucine drop out media (SC-LEU; 1.92 g/l; prepared in-house with L-α-amino acids) supplemented with 2% glucose and yeast nitrogen base (YNB; 6.9 g/l) for 20-22 h or until optical density at 600 nm (OD_{600}), a measure of culture turbidity and growth reached 3.0, after which it was upscaled to a 2×700 ml culture in the same media, grown for another 24 h to an OD_{600} of 5.0. The large overnight culture was then inoculated into 12 flasks containing 800 ml SC-LEU+YNB media with 2% lactate (100 ml culture inoculated in 800 ml lactate media). The cultures were grown in lactate media for 17 h to an OD_{600} of 2.0-2.5. Flasks were induced with 1.5% galactose to express the protein. The cultures

were left to express and harvested 18-20 h after induction. Expression level was quantified by whole-cell GFP fluorescence ($\lambda_{excitation}$= 488 nm, $\lambda_{emission}$= 512 nm), from 8 ml of culture pelleted and resuspended in 150 µl of PBS, measured on a SpectraMax M3 Microplate Reader (Molecular Devices). The OD_{600} before harvesting would routinely range between 4.0-4.5, depending on the preparation. Cells were harvested at 5,000×g for 25 min and the pellet was resuspended in PBS pH 7.4 (5 ml·g^{-1} total volume). Yeast was grown at 30 °C at 240 rpm in flasks. The protocol followed here was adapted from elsewhere (Drew *et al.*, 2008).

2.2.2 Membrane Protein Expression in HEK293F cells

The following protocol was used for the expression of human SLC7A4 for purification in GDN (chapter 5). The sequence-verified plasmids containing SLC7A4 with a C-terminal His_8-TEV-GFP were gigaprepped according to the manufacturer's instructions (PureLink™ HiPure Expi Plasmid Gigaprep Kit, Invitrogen). The DNA was syringe-filtered (0.22 µm) and stored at –20 °C, prior to transfection.

Human embryonic kidney–293F (HEK293F) cells in the FreeStyle 293 Expression Medium (Thermo Fisher Scientific, UK) were used for transient transfection and expression. The cells were cultured in suspension phase at 37 °C and 8% CO_2 and were started from 30 ml (inoculated with 0.3×10^6 viable cells/ml) and sequentially scaled-up by passaging to 100 ml (inoculated with 0.3×10^6 viable cells/ml) to 600 ml (inoculated with 0.3×10^6 viable cells/ml) to 2 l culture (inoculated with 0.7×10^6 viable cells/ml). 24 h later the cells were at a density 1.3-

1.5×10^6 cells/ml at which point they were transiently transfected. For the 2 l culture, 2.2 mg of plasmid DNA was diluted in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific, UK) to a total volume of 30 ml, mixed by gentle inversion and left to stand for 5 min. During incubation, 4.4 ml (from 1 mg/ml of stock) of linear polyethyleneimine (PEI) MAX (MW 40,000; Polysciences Inc., USA) was mixed with DMEM to a total volume of 30 ml. The diluted PEI MAX was added to the diluted DNA and mixed gently. The mixture was incubated for 15 min at room temperature and was added dropwise to the cells with gentle swirling. Syringe-filtered (0.2 μm) sodium butyrate was added at 8 mM final concentration. Cells were returned to the incubator and harvested 48 hours after transfection, washed in PBS pH 7.4, pelleted and stored at -80 °C until required. Routinely, the cell biomass yield was ~ 8 g/l culture.

2.2.3 Membrane Fraction Preparation

S. cerevisiae cells

Yeast cells were lysed by triple passing through a TS Series Continuous cell disruptor (Constant Systems) running at 38 kpsi and equipped with NESLAB ThermoFlex 1400 Chiller (Thermo Fisher Scientific) keeping temperature at 4 °C. Subsequently lysate was clarified by centrifugation at 30,000×g for 30 min. The supernatant was then ultracentrifuged at 235,872 g for 1.5 h to pellet the membranes. The pellet was washed in high-salt buffer consisting of 20 mM HEPES·NaOH pH 7.5, 1 M potassium acetate, and homogenised by Dounce homogenisation. The washed, homogenised membranes were ultracentrifuged at 235,872×g for 1 h and the membrane pellet was resuspended in PBS (5 $ml \cdot g^{-1}$ total volume).

Human HEK239F cells

The HEK 293F cell pellet was thawed and resuspended in cold PBS (3.125 ml·g^{-1} total volume) containing deoxyribonuclease and 2 tablets of protease inhibitors per L worth of cells (SIGMA*FAST*^{TM}; each containing: 2 mM AEBSF, 0.3 μM Aprotinin, 130 μM Bestatin, 1 mM EDTA, 14 μM E-64, 1 μM Leupeptin). Cells were subsequently lysed on ice using a sonicator (Qsonica, USA), with sonication amplitude set to 30% and total sonication time 2 min/L consisting of 20 pulse cycles (2 s on followed by 4 s off). Unbroken cells and cell debris were pelleted at 12,000×g for 10 min at 4 °C, and the supernatant was ultracentrifuged at 235,872×g for 1 hour, washed with 15 mM HEPES·NaOH pH 7.5 and 20 mM KCl and dounce homogenised. The homogenised membranes ultracentrifuged at 235,872×g for 1 hour and the pelleted membranes were resuspended in PBS (10 ml·g^{-1} final volume) and snap-frozen for storage at −80 °C until required.

2.2.4 Fluorescence Size Exclusion Chromatography

200 μl of membrane fractions containing C-terminal GFP-tagged membrane protein were solubilised in 1% (w/v) detergent (Anatrace) containing 10% v/v glycerol for 1 h at 4 °C and then centrifuged at 30,000 x g for 60 min. GFP fluorescence was measured before and after centrifugation to calculate the detergent solubilisation efficiency. For each detergent tested, 100 μl of the supernatant containing detergent-solubilised membrane material was injected on a Superose 6 10/300 GL column (Cytiva) pre-equilibrated in 20 mM Tris·HCl pH 7.5, 150mM NaCl, 0.03% DDM. The column was downstream in-line with a RF-20A Fluorescence Detector (Shimadzu) recording at [$\lambda_{excitation}$= 488 nm, $\lambda_{emission}$= 512 nm]. Samples were run at a flow rate of 0.4 ml·min^{-1}. The protocol followed here was

adapted from elsewhere (Kawate and Gouaux, 2006a; Drew *et al.*, 2008).

2.2.5 Membrane Protein Purification

Membrane fractions were solubilised in 1% (w/v) detergent (DDM, LMNG; Anatrace) or detergent supplemented with cholesteryl hemisuccinate (DDM-CHS, LMNG-CHS; mixed at 5:1 or 10:1 w/w), buffered with PBS pH 7.4 and supplemented with 150 mM NaCl, 10% (v/v) glycerol for 1.5 h. Non-solubilised material was separated by ultracentrifugation at 235,872 x g for 1 h. The solubilised fraction was supplemented with 20 mM imidazole and was incubated to bind HisPur Ni-NTA resin (Thermo Fischer Scientific) for up to 3.5 h or until binding of the GFP-tagged protein to the resin plateaued – monitored by GFP fluorescence. The immobilised species were subjected to immobilised metal affinity chromatography (IMAC). Specifically, resin was washed in a gravity column with 0.1% DDM (± 0.02% CHS) or 0.01% LMNG (± 0.001% CHS), PBS supplemented with 150 mM NaCl, 10% glycerol (buffer A) with 20 mM imidazole, for 20 column volumes (CVs), followed by a second wash with buffer A with 25 mM imidazole (10 CVs). Membrane protein was then eluted with buffer A supplemented with 250 mM imidazole in 4 CVs. The concentration of the eluted GFP fusion protein was estimated as described in (Drew *et al.*, 2008). Equimolar amount of Tobacco Etch Virus (TEV) protease (expressed and purified in-house) was added to the eluant and left to dialyse overnight at 4 °C against 4.0 l gel-filtration buffer in a 3.5 kDa MWCO regenerated cellulose membrane (Spectra-Por tubing). The gel-filtration buffer composition was 20 mM Tris·HCl pH 7.5, 150 mM NaCl and 0.03% DDM (± 0.006% CHS) or 0.003% LMNG (± 0.0003% CHS) - i.e., 3× the Critical Micelle Concentration (CMC) of the detergent in use.

To purify proteins in GDN, the purification was performed exactly as above in DDM+CHS, and the detergent was exchanged to GDN by substituting the 0.1% DDM-CHS in buffer A with 0.1% GDN. From dialysis to all subsequent steps, GDN was kept at a concentration 0.05%.

The next morning, the dialysate was syringe filtered through 0.22 μm, supplemented with 10 mM imidazole, and passed through 5 ml HisTrap column (Cytiva; pre-equilibrated in gel-filtration buffer) to remove the His-tagged TEV protease and liberated GFP-His_8 with Ni-affinity. The flow through was collected and spin-concentrated in a 100,000 MWCO concentrator (Vivaspin 20, Sartorius) to 0.4 mL and injected to a Superdex 200 Increase 10/300 GL SEC column (Cytiva; pre-equilibrated with gel-filtration buffer) and run at flow rate 0.4 $ml \cdot min^{-1}$. Fractions containing protein were analysed with SDS-PAGE for their purity, pooled accordingly and concentrated in a 50,000 MWCO concentrator (Vivaspin 2, Sartorius or Amicon® Ultra) to desired concentration. Protein concentration was determined with a NanoDrop ND-1000 spectrophotometer using extinction coefficients calculated with the Expasy ProtParam server for the purified protein (disulphides considered as reduced). All purification steps were performed at 4 °C.

All protein purification steps were routinely analysed by SDS-PAGE electrophoresis (12% acrylamide) and stained with Coomasie blue dye (InstantBlue, Expedeon) or imaged with in-gel GFP fluorescence by using a Fuji FLA Phosphor imager.

2.2.6 Sybody Expression & Purification

Unique sybodies hits from the ELISA step were miniprepped in the pSb_init (chloramphenicol resistance) vector containing a C-terminal Myc-His_8 tag. *E. coli* MC1061 strain (Lucigen) were transformed with sequence-verified sybodies and plated onto LB agar with 100 $\mu g \cdot ml^{-1}$ chloramphenicol. Colonies were used to inoculate starting cultures 50 ml Terrific Broth (TB) supplemented with chloramphenicol and grown for 11 h at 37 °C and 200 rpm. The starting cultures were then used to inoculate 1-2 l TB-Chloramphenicol at OD_{600} 0.1, which were grown for 85 min at 200 rpm to OD_{600} 0.5, then changed the temperature to 22 °C and continued cultivation for 120 min. Sybody expression was induced by adding 0.02% L-(+)-arabinose (1 ml of 20% stock) and grew overnight at 22 °C and 180 rpm. Next morning and after 13 h of expression time, the cultures were harvested by centrifugation for 25 min at 4,000×g and 4 °C. The cell pellets were immediately used for purification without freezing-thawing.

The cell pellet was resuspended with 50 ml/l culture of periplasmic extraction buffer (20% w/v sucrose, 50 mM Tris·HCl pH 8.0, 0.5 mM EDTA and 0.5 µg/ml lysozyme, prepared fresh and used immediately). The resuspended cells were incubated at 4 °C for 30 min with magnetic stirring. The suspension was diluted with 200 ml/l culture of 20 mM Tris·HCl pH 7.5, 1 mM $MgCl_2$ and incubated on ice for another 30 min, before clarifying the periplasmic extract from cell pellet by centrifugation at 4,000×g for 30 min at 4 °C. The periplasmic extract was supplemented with 15 mM imidazole and additional 150 mM NaCl and incubated with 2 ml of HisPur Ni-NTA resin (Thermo Fisher Scientific) for 1 h. The resin was washed with 20 CVs of 30 mM imidazole, 20 mM Tris·HCl pH 7.5, 300 mM NaCl, before eluting the Myc-His-tagged binder with 300 mM imidazole. The

eluant sybody was dialysed overnight against 4 l 20 mM Tris·HCl pH 7.5, 150 mM NaCl in 3.5 kDa MWCO regenerated cellulose membrane (Spectra-Por tubing). Next morning the dialysate was syringe-filtered (0.22 μm) to remove any visible precipitated Sybody and the filtrate was spin-concentrated to 1.0 ml with 5,000 MWCO concentrators (VivaSpin 20). Any further concentration steps required were done in 3,000 MWCO (Amicon® Ultra concentrator).

2.3 Membrane Protein Characterisation

2.3.1 SEC-MALS

Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) experiments were carried out in the Molecular Biophysics Suite of the Department of Biochemistry, University of Oxford by Dr David Staunton, using an Superdex 200 10/300 GL increase column, in-line with DAWN HELEOSII Scattering and Optilab rEX Refractometry detectors (Wyatt Technology) and analysed with ASTRA software (Wyatt Technology). Protein samples were injected at 1 $mg{\cdot}ml^{-1}$ concentration in 20 mM Tris·HCl pH 7.5, 150 mM NaCl and 3×CMC of the assayed detergent.

2.3.2 Circular Dichroism Spectroscopy (CD)

CD measurements were performed on a Jasco J-815 Spectropolarimeter using 1 mm pathlength quartz cuvettes (Starna Scientific). Concentrated protein samples were diluted in CD buffer, 10 mM K_2HPO_4/KH_2PO_4 pH 7.5, 50mM Na_2SO_4 and 3× CMC of the detergent used for protein purification. Membrane protein

concentrations were 0.2 mg·ml^{-1}. Screening spectra at 20 °C were recorded between 190-260 nm while thermal melt profiles compiled by successive spectra recorded between 195-240 nm wavelength between 20-95 °C at a ramp rate of 1 °C/min. Graphs were plotted in PRISM9 (GraphPad) and analysed using a sigmoidal fitting curve.

2.3.3 Differential Scanning Fluorimetry (DSF)

Differential scanning fluorimetry measurements were performed with a NanoTemper Prometheus NT.48. Wildtype human SLC7A4, *At*CAT4 and mutants F237V, S327L, S327M, and D116S were investigated on their thermal stability in the presence and absence of 10 mM α-amino acids. Protein was added at 0.5-1.0 mg·ml^{-1} and the pH was buffered (a) pH 7.5 with 20 mM Tris·HCl, (b) pH 6.5 with 20 mM MES·NaOH, or (c) pH 5.5 with 20 mM citrate. Other buffer components were 150 mM NaCl and 3× the selected detergent CMC. When lipid was supplemented to the samples prior to measurement, the lipid was firstly disturbed with 15 mM sodium cholate until translucent and incubated with the protein sample at 1:1.2 w/w protein:lipid ratio for 15-30 min before measurements. For concentration-dependent substrate binding analysis, a dilution series of each assayed amino acid was prepared, and the same volume of substrate was pipetted in all samples.

Samples were loaded in standard capillaries (NanoTemper). A 20–95 °C temperature gradient was run at 1 °C·min^{-1}. The resulting melting curves were recorded with the Prometheus software by plotting the F330 nm/F350 nm signal against the temperature gradient. The data were exported and plotted in PRISM9

(GraphPad) using the first derivative of the F330 nm/F350 nm signal against the temperature.

2.3.4 Native-PAGE Analysis

For Native-PAGE analysis we used precast 15-well 4-16% Bis-Tris gels (Invitrogen Novex) according to the manufacturer's instructions. Briefly, 3 µg of protein samples at 0.5 mg·ml^{-1} were loaded per lane with 2.5 µl Native-PAGE 4× sample buffer, 0.15 µl 5% G250 additive and 1.35 µl deionised water. The gel was run at 150 V for 60 min followed by 250 V for 90 min. Gels were then stained with the Fast Coomasie G250 Staining method, i.e., they were fixed for 15 min with 40% (v/v) MeOH, 10% (v/v) acetic acid, destained with 8% (v/v) acetic acid for 15-20 min, before transferring to deionised water for visualisation.

2.3.5 Native Mass Spectrometry (nMS)

For the nMS characterisation of *At*CAT4, training and assistance with experimental setup and analysis was kindly provided by Dr Agni Gavriilidou (OMass Therapeutics). LMNG purified *At*CAT4 was buffer-exchanged into 200 mM ammonium acetate containing 2× CMC LMNG using the 100 kDa MWCO Amicon Ultra 0.5 centrifugal filter device and immediately introduced into a Thermo Scientific Q Exactive UHMR (Ultra-High Mass Range) Hybrid Quadrupole-Orbitrap mass spectrometer. A gentle voltage gradient was applied for the ion transfer: injection flatapole, inter-flatapole, bent-flatapole, and transfer multipole: 8, 7, 4, and 0 V, respectively. To remove the detergent micelle from the protein ions, CID:90V (collision induced dissociation) and further 60V were applied on the

source. Spectra were averaged with a noise level parameter of 3. Backing pressure was maintained at $\sim 2\times10^{-10}$ mbar to allow better transmission of protein ions. Data were analysed with Xcalibur 2.2 SP1.48.

2.3.6 *Xenopus laevis* Oocyte Two Electrode Voltage-Clamp

The ovaries were kindly provided by Isabel Bermudez-Diaz. Ovary digestion was carried out for an hour at 21 °C with collagenase (2 mg/ml) in Barth's solution (88 mM, 1 mM KCl, 1.68 mM $MgSO_4(H_2O)_7$, 10 mM HEPES, 2.4 mM $NaHCO_3$) at pH 7.4. Following the digestion, the ovaries were washed in Barth's solution (supplemented with 0.47 mM $Ca(NO_3)_2(H_2O)_4$ and 0.41 mM $CaCl_2$). The oocytes were sorted and manually defolliculated. They were then stored in ND-96 (96 mM NaCl, 2 mM KCl, 1.8 mM $CaCl_2{\cdot}2H_2O$, 1 mM $MgCl_2$, 5 mM HEPES, pH 7.4). Each oocyte was injected with 50 nl of nuclease-free MilliQ H_2O, CAT-4 cRNA (human: 656 ng/µl and plant: 895 ng/l) in nuclease-free MilliQ. Recordings were made following incubation in ND-96 with gentamycin (50 g/ml) at 17 °C for 3 days unless otherwise stated.

Two electrode voltage-clamp recordings were made using AgCl electrodes placed in micropipettes filled with 3 M KCl (resistance 1.5-5.0 M). The oocytes were placed in ND-96 solution (pH 6, 6.5 or 7) at room temperature for the recordings. A Digidata 1440A (Axon Instruments) converter, an Axopatch (Axon Instruments) amplifier and a HumBug Noise Eliminator (Digitimer) were used with winWCP (Strathclyde Electrophysiology Software) software. The following voltage protocol (voltage-clamping) was used unless otherwise stated: –160 mV to +40 mV with 10 mV steps (800 ms) each followed by 1.5 s at –40 mV (resting

potential). For each oocyte, a recording was made prior to and after the addition of amino acid (all used amino acids are L-amino acids). 30 or 60 µl of 100 mM amino acid (except 60 µl of 50 mM tryptophan and 60 µl of 2.2 mM tyrosine) in ND-96 (pH 6) was added to 570 or 540 µl of ND-96 solution to achieve a 5 mM or 10 mM amino acid concentration, respectively. The measurement made prior to the addition of the amino acid was subtracted from the measurement made after the addition. Baseline adjustments and the data analysis was carried out using Python code based on the pyabf module. Graph generation and statistical analysis were carried out using Origin (OriginLab Corporation). All bar charts are displayed with bars that represent the mean +/− SEM. All electrophysiological experiments were carried out by Sacha Salphati (Newstead group).

2.3.7 Cell-based Radioactive Substrate Transport Assays

HeLa cells were maintained in DMEM/GlutaMAX medium supplemented with 10 % foetal bovine serum under 5% CO2 at 37 °C. For transport assays 2×10^5 cells/ml were seeded into 12 well plates and 24 h later transfected using FuGENE with (a) equal amounts (0.6 µg/well) of SLC7A5/SLC3A2 plasmids for 36-40 h, or (b) with 1.2 µg SLC7A4. Cells were washed twice with 1 ml of PBS pH 7.4 (or other buffers as denoted in chapter 5) before application of 0.3 ml PBS containing 10 µM substrate with trace amounts of radioactive substrate (1:300 dilution, i.e., 1 µl hot-substrate per well). After the desired time (3 or 10 min) the assay buffer was removed, and the cells quickly washed twice with 0.3 ml assay buffer with no substrate. Cells were lysed by solubilisation with buffer [0.05% Triton-X 100 (w/v) in PBS pH 7.4 for 7 min] and placed in a scintillation vial. 4.5 ml scintillation fluid per vial were added. The amount of substrate taken

up by the cells was calculated by scintillation counting in Ultima Gold (Perkin Elmer) with comparison to a standard curve. Assay data was analysed and graphically represented using Prism 9.

2.4 Sybody Screening & Generation

2.4.1 AtCAT4 Biotinylation

For biotinylation an avi-tagged *At*CAT4 construct needed to be purified. Avi-tagged *At*CAT4 was obtained as a synthetic plasmid with a pMK-T vector backbone with Kanamycin resistance codon optimised for *H. sapiens* (Invitrogen). The *At*CAT4 sequence was designed with a direct C-terminal Avi tag, GLNDIFEAQKIEWHE (spaced with a Gly linker). *At*CAT4-Avi was PCR amplified with 5' and 3' homologous recombination primers compatible with SmaI-linearised pDDGFP2-Leu2D (Addgene # 102334). The amplicon was gel-extracted and inserted into pDDGFP2-Leu2D by the yeast homologous recombination method, described §2.1.1. In the resulting plasmid *At*CAT4-Avi was upstream the TEV cleavage site followed by His_8-GFP, identically to the WT plasmid construct.

The avi-tagged protein was expressed and purified identically to the WT, following biotinylation by BirA biotin-protein ligase (Avidity) overnight (Table 2.2), the protein was subjected to a further size-exclusion step to remove excess biotin and BirA ligase. The biotinylation efficiency was semi-quantified by the means of a streptavidin-shift assay in which theoretically biotinylated protein was incubated with threefold excess of streptavidin (Sigma-Aldrich) for 5 min to allow complex formation and the molecular weight shift caused by streptavidin

Final concentration	Stock concentration	For 500 µl mix
*40 µM *At*CAT4-avi	115 µM (7.3 mg/ml)	175 µl
1x biomix A	10×	50 µl
1x biomix B	10×	50 µl
12.5 µg/10 nmol GST-BirA		9.7 µl
**H_2O		top up to 500 µl

Table 2.2 Compositional assembly of the BirA biotinylation reaction. * Higher protein concentrations bring together higher NaCl concentrations which act inhibit BirA reactivity.

binding analysed by SDS-PAGE electrophoresis. The electrophoresis apparatus was immersed in an ice bath and run at low voltage, 120 V, to avoid excessive heat generation and complex dissociation.

2.4.2 Sybody Screening

Sybodies were selected against lipidated *At*CAT4-avi purified in LMNG. Lipidation was afforded by 30 min incubation of *At*CAT4 with Asolectin supplemented with cholesterol (AsoChol; 4:1 w/w) at a ratio 1:1.2 w/w. The *in vitro* selection pipeline is described step-by-step in the detailed Sybody pipeline protocol (Zimmermann *et al.*, 2020b). Key steps will be described below briefly.

2.4.2.1. Ribosome Display

10 µL of PURExpress 2.1 translation mix was prepared, 0.7 µL of each mRNA library (1.6×10^{12} mRNA strands) was added and the reaction incubated at 37 °C for 30 min. The ribosomal complexes were diluted with 100 µL buffer (20 mM Tris·Ac pH 7.5, 150 mM NaCl, 50 mM $MgAc_2$, 0.1% LMNG, 0.5% (w/v) BSA, 5 $mg \cdot ml^{-1}$ Heparin and 40 U of RNasIn). The reaction was then centrifuged at 20,000×g for 5 min at 4 °C. 50 nM of lipidated biotinylated target protein was

introduced and mixed gently by pipetting. The panning mix was incubated for 20 min on ice. 12 µL Dynabeads MyOne Streptavidin T1 were washed twice with 500 µL buffer A (20 mM Tris·Ac pH 7.5, 150 mM NaCl) with 0.5% BSA, 0.1% LMNG and then blocked in the same solution for 30 min. The panning mix was added to the beads and incubated for 5 min. Beads were then washed three times with 500 µL buffer A with 0.003% LMNG. During the last washing step, the beads were placed into a fresh tube. RNA was eluted from the beads with 100 µL 50 mM Tris·Ac, 150 mM NaCl, 50 mM EDTA and 100 µg/ml yeast RNA in nuclease-free water and incubated at room temperature for 10 min. The eluted RNA was purified using the RNeasy micro kit (Qiagen) using the manufacturer's instructions and eluted into 15 µL RNAse free water.

The eluted RNA was used as reverse transcription template by adding 2 µL of 100 µM RT_Primer and 4 µL of 10 mM dNTPs and incubating at 65 °C for 5 min and then on ice. 40 µL of the reverse transcription reaction mix was added according to the manufacturer's instructions (AffinityScript kit, Agilent) and incubated for 1 h at 37 °C followed by 5 min at 95 °C. The produced cDNA was purified using the Macherey-Nagel NucleoSpin Gel and PCR Clean-up kit and eluted in 30 µL of the kit's elution buffer. The cDNA was amplified with non-proofreading polymerase, GoTaq G2, with the following primer pairs: (a) Med_FX_for/ Med_FX_rev for concave/loop sybodies or (b) Long_FX_for/ Long_FX_rev for convex sybodies. The reaction serves to add FX-cloning overhangs to the DNA. The reaction was run in a 1.5% (w/v) agarose gel the results of which are shown in **chapter 4**. The amplified bands of each library were gel-extracted with the Macherey-Nagel kit, according to the manufacturer's instructions. The amplicons and the pDX_init vector were restriction digested with

BspQI endonuclease, gel-extracted and 500 ng digested cDNA insert and 1 μg digested pDX_init backbone were ligated in 50 μl using 12.5 U T4 DNA ligase.

2.4.3 Phage Display

Sybodies cloned in the pDX_init vector were transformed in SS320 electrocompetent *E. coli* by electroporation with a BTX™ ECM™ 399 Exponential Decay Wave Electroporator. Electroporated cells were immediately resuspended in 25 ml SOC media and recovered at 37 °C for 30 min at 160 rpm. Cells were diluted into 225 ml 2YT-Amp media supplemented with 2% glucose and shaken overnight at 37 °C. The next morning, 1ml of culture was backdiluted to 50m 2YT-Amp with 2% glucose and grown to an OD_{600} of 0.6. 10 ml of culture was infected with 2.7×10^{11} plaque forming units of M13KO7 helper phage at 37 °C mixed and incubated without shaking for 30 min. Infected cells were centrifuged and resuspended in 50 ml 2YT, with dual selection of 100 $\mu g \cdot ml^{-1}$ ampicillin, 25 $\mu g \cdot ml^{-1}$ kanamycin and shaken at 160 rpm overnight at 37 °C. Next morning the cells were pelleted (5,000×g, 30 min, 4 °C) and 40 ml of the culture supernatant (containing phages) was mixed well with 10 ml 20% PEG6000 (v/v), 2.5 M NaCl and incubated on ice for >30 min. Precipitated phages were pelleted by centrifugation (5,000×g, 30 min, 4 °C), resuspended in 1 ml PBS and centrifuged again at full speed on a table top centrifuge (20,000×g, 5 min). This phage clarification step was repeated. Final phage titre was determined by UV-Vis spectroscopy at 269nm and 320nm by the equation,

$$\text{Phages}/_{\text{ml}} = (A_{269}\text{-}A_{320}) \times 6 \times 10^{16} /_{4900}$$

For the phage display, a 96-well Nunc Maxisorp plate was prepared such that each well was coated overnight with 100 μl per well of 67 nM neutravidin. The

plate was then washed three times with buffer (Tris·HCl pH 7.4, 150 mM NaCl) and subsequently blocked with buffer containing 0.5% BSA for 1 h. In the meantime, phage stocks of each library were diluted to 1×10^{12} phages·ml^{-1} in buffer with 0.5% BSA, 0.1% LMNG and incubated with 50 nM biotinylated target protein for 20 min (panning step). The plate was washed three times with buffer containing 0.5% BSA, 0.1% LMNG and 100 µL of phage-target protein mixture was added to each well and allowed to incubate for 10 min at 4 °C. The plate was then washed three times with buffer with 0.003% LMNG. The phages were eluted with 100 µl of buffer per well supplemented with 0.25 mg·ml^{-1} trypsin for 10 min at room temperature, and subsequently well eluants pooled together. 0.125 mg·ml^{-1} trypsin inhibitor (AEBSF, 4-(2-Aminoethyl)benzenesolfonyl fluoride) was added to stop the reaction. The 4.7 ml of eluted phages were added to 45 ml of SS320 cells grown in 2YT-tetracycline (OD_{600} 0.6) and incubated at 37 °C without shaking for 30 min. Infected cells were inoculated to 250 ml 2YT-Amp supplemented with 2% glucose and grown overnight at 37 °C resulting in a preculture to produce more phage for the second round. The second round of phage display was performed in the same manner as above, except that magnetic beads (Dynabeads MyOne Streptavidin C1) replaced the Maxisorp immunoplate. The phages resulting from the second round of display were used to infect SS320 *E. coli* cells. Their phagemids were purified by the standard miniprep methods (Qiagen). The purified phagemids contained the selected sybody clones. Sybody clones were subcloned into pSb_init (chloramphenicol resistance) using FX cloning with BspQI (as before) and transformed into *E. coli* MC1061 strain.

95 single clones were picked per library and transferred to 96 deep well with

2 ml well capacity plate where they were precultured in 1.2 ml TB with chloramphenicol. Cultures were inoculated and at OD_{600} 0.4-0.8 were induced with 0.02% L-(+)-arabinose to start sybody expression overnight at 22 °C, 300 rpm.

2.4.4 ELISA Screening

The next morning, the cells were pelleted, resuspended in 100 µL periplasmic extraction buffer (20% w/v sucrose, 50 mM Tris·HCl pH 8.0, 0.5 mM EDTA, 0.5 $\mu g \cdot ml^{-1}$ lysozyme). For the ELISA, two 96-well Nunc Maxisorp plates were used per library (one column for target protein screening – *At*CAT4, one column for negative control protein - $PepT_{So}$). The plates were already overnight coated with 100 µL of 5 $\mu g \cdot ml^{-1}$ Protein A in PBS per well and washed three times with 250 µL TBS and blocked with 250 µL TBS with 0.5% BSA for 30 min. Plates were then washed three times with TBS buffer and incubated with monoclonal Anti-c-Myc antibody (1:2000, Sigma 9E10) diluted in TBS with 0.5% BSA, 0.1% LMNG for 20 min. Plates were finally washed three times with TBS buffer with 0.003% LMNG.

The supernatant periplasmic extracts were seeded in the ELISA plates, adding the same supernatant in adjacent column pairs to compare binding with the negative control. After 20 min, wells were three times washed with TBS buffer with 0.003% LMNG. 50 nM of biotinylated lipidated target protein or negative control protein were added to wells in TBS buffer with 0.5% BSA, 0.1% LMNG, and incubated for 20 min. Plates were then washed three times in TBS buffer with 0.003% LMNG and incubated with 1:5000 diluted Streptavidin-HRP in TBS buffer with 0.5% BSA, 0.1% LMNG for 20 min. After three final washes TBS

with 0.003% LMNG the ELISA was developed by adding 100 µL developing buffer (0.1 $mg \cdot ml^{-1}$ TMB, 50 mM Na_2HPO_4, 25 mM citric acid, 0.006% H_2O_2). Results were measured at SpectraMax M3 Microplate Reader (Molecular Devices) at an absorbance wavelength of 650 nm.

The ELISA sybody hits selected were miniprepped in pSb_init from the frozen preculture pellets in the 96-deep well culture plate according to the manufacturer's instructions (GeneJET kit - Thermo Fisher Scientific). The plasmids were sequence verified, clustered to unique hits (as there is degeneracy between hits). The sybodies in pSb_init were expressed and purified as described in §2.2.6.

2.4.5 Biolayer Interferometry (BLI)

BLI was measured on an Octet Red 384 instrument (Sartorius) with streptavidin biosensor tips pre-equilibrated in buffer and loaded with 50 µl of 25 nM *At*CAT4 in purification buffer (20mM Tris·HCl pH 7.5, 150mM NaCl, 0.003% LMNG) and allowed to equilibrate with a loading step until baseline equilibration was achieved (300 s), followed by a short 60 s baseline measurement. For sybody screening/ranking, 50 µl of 500 nM of Myc-His_6-tagged sybodies were allowed to associate for 5 min followed by a 5 min dissociation step. Binding kinetics of SybB5 were measured with a serial dilution of the sybody, with one reference well without sybody. The association step was performed for 15 min followed by dissociation for 15 min in buffer only wells. Sensogram data was analysed with the Octet v9.0 software package (ForteBio). All raw data was baseline and reference subtracted, in-step corrected, y-axis aligned and filtered with a Savitzky-Golay filter. The processed data was plotted in PRISM9 (GraphPad).

2.5 CryoEM Structure Determination

2.5.1 Sample Preparation & Data Acquisition

Grid preparation and screening were done in collaboration with Dr Kato Takafumi (Newstead group). CAT4-Syb complexes were formed on the day of grid preparation by incubation of 0.3 $mg \cdot ml^{-1}$ purified CAT4/LMNG (from a 32.4 $mg \cdot ml^{-1}$ stock) with molar excess of the Myc-His_6-tagged sybody of choice (1.5-2.0 excess) for 60 min on ice in 400 µl reaction volume buffered with TBS pH 7.5 supplemented with 0.003% LMNG. The formed 1:1 complex was separated from excess sybody by means of gel filtration by injecting the reaction into a Sepax SRT-3C SEC300 (Cytiva) running at 0.4 $ml \cdot min^{-1}$. The pooled fractions (A11-A12 for the sample that led to structure) were analysed with SDS-PAGE for complex component purity and stochiometric composition prior to spin-concentration with 50 MWCO concentrators (Amicon® Ultra) at 4,000×g and 4 °C until the desired complex concentration was achieved, assessed spectrophotometrically by NanoDrop ND-1000 spectrophotometer. The complex extinction coefficient was estimated from the sequence of *At*CAT4 (cleaved at the TEV_{site}) and the corresponding Myc-His_6-tagged sybody with the Expasy ProtParam server.

The grid preparation took place in a specially equipped facility providing 100% humidity at room temperature. A Vitrobot Mark IV (Thermo Fisher Scientific) was used to prepare grid sample. In all cases, 2.5-3.5 µl of complex was dispensed onto glow-discharged holey carbon-coated grid (Quantifoil 300 mesh, Cu or Au R1.2/1.3, Agar Scientific). After allowing 0 or 30 s for sample adsorption, grids were blotted for 1.5 or 3.0 s (vitrobot parameters: blot force –5, 100% humidity, 4 °C) and then plunge-frozen into liquid ethane.

Grid samples (as detailed in chapter 5) were screened, without dataset collection, on a Talos Arctica 200 kV Cryo-TEM (Thermo Fisher Scientific) equipped with a Falcon 3 Direct Electron Detector (Thermo Fisher Scientific). For data collection, micrographs were collected in counted super-resolution mode on a Titan Krios G3 (FEI) operating at 300 kV with a BioQuantum imaging filter (Gatan) and K3 direct detection camera (Gatan) at 105,000× magnification, physical pixel size of 0.832 Å spanning a defocus range –2.0 to –1.0 μm. The total dose rate 19.3 $e^-/Å^2$, exposure time 3.00 s, corresponding to a total dose 58 $e^-/Å^2$ spread over 40 frames. The collection yielded a total of 17,271 movies including 11,166,793 particles extracted at 256×256 pixel boxes.

2.5.2 Data Processing & Model Building

Datasets were processed on-the-fly in SIMPLE 3.0 (Caesar *et al.*, 2020) for patched (15×10) motion-correction, dose-weighting, patched contrast transfer function (CTF) estimation and particle picking. All subsequent processing was performed in either cryoSPARC (Punjani *et al.*, 2017) or RELION 3.1 (Zivanov *et al.*, 2018) interconverting format using the csparc2star.py script within UCSF pyem (Asarnow, D., Palovcak, E., Cheng, 2019). Resolution estimates were derived from gold-standard Fourier shell correlations (FSCs) using the 0.143 criterion as calculated within cryoSPARC. Local resolution estimations were calculated within cryoSPARC.

Comprehensive information on CAT4-SybB5 complex single-particle analysis and processing are presented in chapter 5.

The AlphaFold2 model of *At*CAT4 (ID: AF-Q8W4K3-v3) was manually fitted to the map on ChimeraX to generate the initial model. An alphafold model of SybB5 was generated by ColabFold (Mirdita *et al.*, 2022) with default parameters and with Amber force field relaxation settings, and the SybB5 model was also fitted to the map. After model fitting, the models were combined into one and manually readjusted using Coot v0.9.8 (Emsley *et al.*, 2010), and then refined using PHENIX real-space refinement (Liebschner *et al.*, 2019). A steroid density was modelled as cholesterol with stereochemical parameters generated with AceDRG (Long *et al.*, 2017). The model was protonated and globally minimised with the OPLS force field (heavy atom ΔRMSD < 0.3 Å cutoff) to resolve any remaining steric clashes and geometry outliers. The minimised structure was stripped of hydrogens and manually readjusted in Coot in a second round followed by global refinement in PHENIX to yield the final model. The figures depicting the molecular structures were prepared using ChimeraX (Pettersen *et al.*, 2021) and PyMOL. Secondary structure elements in the figures were calculated with DSSP plugins (Kabsch and Sander, 1983).

COMPUTATIONAL METHODS

2.6 Molecular Dynamics for Substrate Interactions with SLC7s

This section describes the setup, parameterisation, and execution of atomistic simulations of SLC7 transporters (xCT and CAT4) with the selected substrates. The simulation methods are common for both transporters with minor adaptations, where noted. A portion of the reported methods have been published as

part of the publication "Molecular basis for redox control by the human cystine/glutamate antiporter system xc^{-}" (Parker, Deme, Kolokouris, *et al.*, 2021), from where they have been taken and adapted for this thesis.

2.6.1 System Preparation & Amino Acid-Bound MD Simulations

The full heterodimer of system xc^{-} (xCT-SLC3A2) was embedded into a pure 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) membrane via CHARMM-GUI (Jo *et al.*, 2008). The heterodimer contained 500 POPC molecules on both leaflets, with 258,000 atoms in a 12.5×12.5×16.0 nm^3 simulation box. In the cystine-bound simulations, the transporting SLC7A11 domain was simulated with 500 POPC molecules, totalling 227,000 atoms in a 13.5×13.5×12.2 nm^3 simulation box. For substrate-bound CAT4 the system was setup in a 12.0×12.0×11.2 nm^3 box containing approximately 320 POPC molecules and 80 cholesterols. CAT4 does not have a soluble domain and required less waters to avoid minimum image convention violations. The CAT4 systems consisted of approximately 150,000 atoms.

System xc^{-} specifics

The coordinates of b$^{0,+}$AT-rBAT+Arg complex (PDB ID: 6LI9) were obtained, the rBAT domain was deleted and the transporting b$^{0,+}$AT domain superimposed to the apo xCT-SLC3A2 cluster average structure [obtained from 200 ns simulation with *gmx_ cluster*]. Hydrogen atoms were added to cystine so that cystine had –1 net charge, according to a previous publication and similar to the glutamate substrate (Bannai and Kitamura, 1981). The SLC7A11-Arg complex was saved and was subsequently used as a template to guide docking of the

cystine ligand to SLC7A11. We used pre-existing structural data of amino acid binding interactions to other SLC7 members to guide the docking and increase confidence of the docked substrate structural models. Specifically, the $–COO^-$, $–NH_3^+$ groups belonging to half of cystine were restrained to 2 Å of the arginine $–COO^-$, $–NH_3^+$ template coordinates. Additionally, H-bonds with I57, G61, Y244 and the crystallographic water molecule from the *Gk*ApcT structure were set as docking constraints. The docking grid was generated with Glide assigning the OPLS3e force field to protein and cystine (Friesner *et al.*, 2006). Only the anionic cystine was considered as this is the only form of cystine present in physiological conditions (Bannai and Kitamura, 1981). The resulting poses were degenerate and converged within 1 Å. The top-ranked pose according to the scoring function Glide XP (Friesner *et al.*, 2006) was saved and used as starting structure for MD simulations.

*At*CAT4 specifics

For CAT4 simulations, our apo CAT4 structural model was prepared for docking and MD simulation by incorporating the following modifications: (a) the F237 dihedrals were set to $\chi_1 = +178.3°$ (*trans*) and $\chi_2 = 74.1°$, (b) SybB5 was deleted (c) apo CAT4 model was superimposed with the Alphafold2 model (ID: AF-Q8W4K3-v3) and the coordinates for the following residues were copied to our structural model so that missing loops within the sequence were filled: the extracellular loop residues 135-156 (TM3-TM4; EL2), 294-305 (TM7-TM8; EL4) and the long intracellular stretch 409-457 (TM10-TM11; IL5). Initial simulation with substrates showed steric clashes with TM1 which resulted in substrate exiting the binding site. To resolve this, the coordinates of superimposed TM1 from the Alphafold2 model were used since it adopts a substrate-bound conformation.

The resulting structure was minimised and superimposed to $b^{0,+}$AT +Arg and the L-arginine substrate copied to the CAT4 binding site. The L-ornithine and L-lysine ligands were mutated from L-arginine. The substrate-bound complexes were minimised with the OPLS3e force field (heavy atom ΔRMSD < 0.3 Å cutoff).

Modelling & Simulation parameters

For MD simulation xCT was represented with the Amber99SB-ILDN forcefield (Lindorff-Larsen *et al.*, 2010), CAT4 with the Amber14SB (Maier *et al.*, 2015) and the lipids were represented with the Slipid forcefield (Grote and Lyubartsev, 2020). Substrates were represented with the General Amber Force Field (GAFF) (Wang *et al.*, 2004) and the point charges were derived using the semiempirical AM1-BCC method bundled with AmberTools18. The systems were then solvated with TIP3P water model (Jorgensen et al., 1983) and neutralised to an ionic concentration of 150 mM with NaCl.

MD simulations in explicit solvent used Gromacs 2020.3 (Hess *et al.*, 2008). The electrostatic interactions were computed using particle mesh Ewald (Essmann *et al.*, 1995a) with a short-range cut-off of 1.2 nm, which was also used for the Lennard–Jones interactions cut-off. Bonds involving hydrogen atoms were constrained by the LINCS algorithm (Hess *et al.*, 1997), and a time step of 2 fs was used for the integration of the equations of motion. Snapshots recorded every 10 ps during the production MD simulations were considered for analysis steps. The system was minimised with steepest descents and a conjugate-gradient step every 10 steps until convergence to $F_{max} < 500$ kJ/mol. After energy minimisation, the systems were equilibrated with 1,000 kJ mol^{-1} nm^{-2} positional restraints on the heavy atoms for 5 ns with the v-rescale thermostat (Bussi,

Donadio and Parrinello, 2007) at 310 K in the NVT ensemble. The systems were further equilibrated with the same positional restraint in the NPT ensemble with a semi-isotropic Berendsen barostat (Berendsen *et al.*, 1984) at 1 atm and a Nose–Hoover thermostat (Evans and Holian, 1985) at 310 K for 10 ns. The resulting frames were used as the input frames for the production runs. Initial velocities were sampled from the Maxwell-Boltzmann distribution for each repeat. Production runs were performed with the Parrinello-Rahman barostat (Parrinello and Rahman, 1981) and the Nose-Hoover thermostat at 310 K. For CAT4 relaxation and production runs, the v-rescale thermostat was used in all steps, while all other parameters remained the same.

Analysis of ligand interactions with protein from trajectories were done by in-house python scripts using MDAnalysis (Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016) and gromacs-tools.

2.6.2 End-Point MMPBSA Calculations

The section describes the endpoint free energy calculations done in the xCT-cystine interaction investigation. For MMPBSA-TΔS calculations, molecular mechanics energies and the non-polar contribution to the solvation free energy were computed with the mmpbsa.py module of AmberTools20 (Miller *et al.*, 2012). gmx-MMPBSA (v1.4.3) (Valdés-Tresanco *et al.*, 2021) was used to convert the Gromacs trajectories into AMBER format with parmed, and subsequently used mmpbsa.py. Alanine scanning mutations of R135 and R396 were done to quantitatively characterise important residue contributions, as shown previously with

MMPBSA (Liu *et al.*, 2018). Data post-processing of absolute binding free energies (ABFE) from MMPBSA was done with the gmx_MMPBSA_ana analysis module. Even though MMPBSA-TΔS is known to be less accurate than some of the more computationally expensive methods, such as free energy perturbation (FEP) and thermodynamic integration (TI), the qualitative agreement is accurate for the ranking of protein-ligand interactions (by including solvation and entropic terms, as opposed to a simple molecular mechanics energy calculation). Importantly alanine scanning MMPBSA has been used with good success in identifying key residues for binding (Lagarias *et al.*, 2019), justifying the selection of MMPBSA to decompose the relative contributions of R135 and R396 to the binding free energy. The method was specifically chosen in this study due to the lack of experimentally determined structures of L-cystine-bound xCT, which would perhaps raise convergence issues with FE methods.

The relative $\Delta\Delta G$ values reported in **chapter 5** required statistical processing of the ΔG values. To calculate mean of means the formula used was:

$$\bar{x}=\sum_{k=1}^{m}\frac{n_k}{n}\cdot\bar{x_k}\ ,\quad \text{where}\quad n=\sum_{i=1}^{m}n_k \qquad (4)$$

, and $n_1= \ldots =n_k$, the sample sizes from production runs of independent repeats of WT/mutant systems were the same in all cases.

For the calculation of the pooled standard deviation of the difference of the means of absolute binding free energies between WT and mutant, $1\times M$ vectors consisting of ΔG values from selected frames, M, were subtracted and the pooled standard deviation was calculated from the components of the resultant difference vector. The formula used for calculations of the standard deviation of pooled

means (mean of the means of each run) was:

$$\widehat{\sigma}_{\bar{x}} = \sqrt{\frac{(N_1\text{-}1)SD_1{}^2 + (N_2\text{-}1)SD_2{}^2 + \ldots + (N_k\text{-}1)SD_k{}^2}{(N_1\text{-}1)+(N_2\text{-}1)+\ldots+(N_k\text{-}1)}} \qquad (5)$$

, where N_k indexes the statistical degrees of freedom of each sample (i.e., the number of analysed frames), and with the assumption that the independent MD runs are statistically independent. This assumption is justified due to seeding production simulations replicates with velocities randomly selected from the Maxwell-Boltzmann distribution. Therefore, the covariance of the samples is taken as 0 in all cases, and equation (5) holds. Since the number of samples (i.e., MD frames selected for MMPBSA calculations) were the same in all cases, $N_{run1} = N_{run2} = N_{run3}$, equation (5) simplifies to,

$$\widehat{\sigma}_{\bar{x}} = \sqrt{\frac{SD_1{}^2 + SD_2{}^2 + SD_3{}^2}{3}} \qquad (6)$$

2.7 Multiscale Molecular Dynamics Simulations of Lipid Interactions

2.7.1 CryoEM Structural Models & Coarse-graining Models

The atomistic models of the inward-open $b^{0,+}AT$, xCT, LAT1 and outward-occluded LAT1 determined with cryoEM SPA were downloaded from the PDB (PDB ID: 6LI9, 7P9V, 6IRT and 7DSK, respectively). The LAT1-inward loop consisting of residues 223-239 had missing heavy atoms which were modelled with PyMOL (*https://www.pymol.org/*). Side chain protonation states were modelled using pdb2gmx (histidine residues protonated on N_ε) (Søndergaard *et al.*, 2011). The N and C-termini were treated as neutral. N and C-termini absent in the electron density were not modelled as they are commonly missing in atomistic

models of SLC6 and SLC7 transporters, due to their high mobility. Models of the LAT1 mutants were also obtained using the PyMOL mutagenesis tool. Non-protein features of the models, including ligand, lipid, N-acetylglucosamine post-translational modification atoms and modelled waters were removed. The protein structures were energy minimised at atomistic resolution (AT), curated in the PPM2.0 server to assign membrane-placement dummy atoms (Lomize *et al.*, 2012) and mapped to coarse-grained MARTINI v2.2 models (De Jong *et al.*, 2013) with the ElNeDyn22 elastic network to retain secondary structure description during simulation (Periole *et al.*, 2009). Crucially, the SLC7 and SLC3 domains were coarse-grained as one protein particle to permit the formation of an interprotein elastic network. This was important to retain the SLC3-SLC7 interfaces close to the structural model.

2.7.2 Coarse-grained Simulations

Coarse-grained (CG) protein models were placed in a cubic box with dimensions $15\times15\times15$ nm^3, subsequently lipidated with the specified lipid compositions using Insane (*https://github.com/Tsjerk/Insane*) at the PPM-specified coordinates for the membrane and solvated with standard MARTINI water beads with 150 mM NaCl concentration. Charge neutrality within the box was ensured. The system was minimised and subjected to two rounds of MD relaxation in the NPT ensemble; the first with protein position restraining to let the membrane assemble around the transporter for 100 ns; the second unrestrained for 100 ns to allow protein residues to accommodate the membrane lipids. For repeat simulations, special care was taken to setup the systems from the start, i.e., before lipidation and system setup, to ensure that the repeats were free of system configuration

bias. Lipid parameters were obtained from *http://cgmartini.nl/index.php/force-field-parameters/lipids* (Wassenaar *et al.*, 2015), including virtual site Martini cholesterol parameters (Melo, Ingólfsson and Marrink, 2015). The CG systems contained ~ 3.0×10^4 particles. For 10 μs repeats, the initial 2 μs were discarded from analysis as equilibration time and to allow randomisation of lipids around the protein. For the 100 μs simulations the initial 10 μs were discarded.

System relaxation MD were performed at constant temperature 310 K and constant pressure 1 bar regulated by the velocity rescale thermostat (τ^T=1.0 ps) and the Berendsen barostat (τ^T=4.0 ps), respectively, and the equations of motion integrated every 10 fs. Production runs were in the canonical ensemble, NPT, with the Parrinello-Rahman barostat (τ^P=12.0 ps) and the equations of motion were integrated every 20 fs. Explicit electrostatics were modelled for pairs with a potential shift at 1.1 nm, after which point a reaction field (Tironi, Sperb, Smith and van Gunsteren, 1995) was applied (ε_r=15). All CG MD simulation steps were done in GROMACS v2018 (Hess *et al.*, 2008; Pronk *et al.*, 2013)

2.7.3 Potential of Mean Force (PMF) Calculations

The umbrella sampling (US) method (Valleau and J.M., 1977) to construct the PMF landscape of cholesterol binding by the weighted-histogram-analysis-method (WHAM) (Kumar *et al.*, 1992; Hub, de Groot and van der Spoel, 2010). For each site in each transporter assayed, a representative frame from the unbiased CG trajectory, where the site was visually-verified as cholesterol-occupied, was extracted. The membrane composition and configuration was retained, and the system replica was energy minimised. A simple, distance reaction coordinate connecting the cholesterol with the site centre-of-mass (COM; defined as

residues within 0.6 nm of cholesterol) was generated from the bound state to the bulk (~ 1.5-2.0 nm) by steered MD simulation (F_c=1000 kJ/mol/nm^2, 0.0001 nm/ns). Weak translational restraints were applied to four transporter backbone beads in the *xy* plane (F_c=100 kJ/mol/nm^2), to prevent protein rotation and translation during the pulling of a cholesterol. US simulations were performed every 0.05 nm along the reaction coordinate, with each window sampled for 5 μs with an umbrella biasing potential F_c=1000 kJ/mol/nm^2 applied between the site COM and the COM of cholesterol. 34-58 windows were run depending on the site to achieve evenly distributed sampling across windows, verified by inspection of histogram overlapping. PMF profiles were reconstructed with GROMACS WHAM (Hub, de Groot and van der Spoel, 2010).

2.7.4 Atomistic Simulations

Representative CG system frames with all cholesterol binding sites occupied per transporter (verified visually) were converted to AT resolution with the CG2AT2 resolution conversion method (Vickery and Stansfeld, 2021), providing the atomistic cryoEM structure of the apo states as input (H added and minimised at the gas phase). The Charmm36 force field was assigned to transporters and lipids. Atomistic TIP3P/Charmm water model was selected to model water atoms. The atomistic fragments were fitted to the CG structure and minimised, followed by a short 5 ps NVT simulation and a morph using steered MD to drive the backmapped structure to the input atomistic structure. The system was then minimised twice with steepest descents (maximum force tolerance 1000 and 500 kJ mol^{-1} nm^{-1}, respectively), followed by 1 ns NVT relaxation with position re-

straints on the protein C_α atoms and cholesterols (F_c=1000 kJ mol^{-1} nm^{-1}), followed by 1 ns NPT relaxation with the same position restraints, followed by 10 ns unrestrained NPT relaxation. Production simulation repeats with randomised velocities were performed in NPT at 310 K and 1 bar with velocity-rescale thermostat (τ^T=0.1 ps) (Bussi, Donadio and Parrinello, 2007) and Parrinello-Rahman barostat (τ^P=1.0 ps) (Parrinello and Rahman, 1981), 1.2 nm cutoffs for non-bonded interactions with vdW force-switching at 1.0 nm. Long-range electrostatics were modelled with the particle mesh Ewald method (Essmann *et al.*, 1995b) The LINCS algorithm was used to constrain covalent H-bond lengths (Hess *et al.*, 1997). All AT MD simulation steps were done in GROMACS v2020 (Hess *et al.*, 2008; Pronk *et al.*, 2013).

2.7.5 Binding Site Detection and Analysis

Cholesterol time-averaged density heatmaps were generated with VMD 1.9.3 (Humphrey, Dalke and Schulten, 1996). For the binding site classification, trajectories from five independent repeats of 10 μs CG MD of each transporter were analysed. Sites reported in the results were cross-verified to be present in the 100 μs simulation of each transporter. These are presented in **chapter 6**. The binding classification was carried out with the PyLipID python library (Song *et al.*, 2022), described in detail in the respective publication. Briefly, PyLipID creates a 3D array. Each row records the distance of a lipid molecule (e.g., a membrane cholesterol) from a specific transporter residue as a function of time (the array columns are time steps). A graph network is then constructed in which the nodes are protein residues and the edges are the time-dependent Pearson correlation

coefficients of two row vectors (see Figure S10 in Song *et al.*, 2022). Lipid binding sites are defined as clusters of nodes – clusters calculated optimised with the Louvain algorithm (Blondel *et al.*, 2008). With the binding sites defined, features such as site occupancy and lipid residence time are calculated, the latter as the inverse of a lipid k_{off} rate for a site.

SLC7A4 Homologue Screening for Structural Studies

CHAPTER 3

In order to maximise the chances of a successful crystallographic project, it is wise to employ as many means as reasonably possible by filtering a range of protein homologues of interest to identify those which are natively appropriate for structural studies and biophysical analyses. This means that construct expression is sufficient and amenable to purification in high yield and purity, and thermostable. This is because structural studies, especially crystallisation, require high sample purity and concentration while the ease in expression, both quality and yield, facilitates high throughput screening of homologues. A more thermostable homologue reflects a well-folded more rigid molecule that can form and maintain crystal contacts more readily than an unstable one. Thus, it is necessary to perform experimental screening to determine these qualities.

3.1 Homologue Selection & Screening Pipeline

Initially, a set of homologous proteins must be selected before a systematic screening can identify the most suitable construct. The selected set should have satisfactory sequence diversity but enough sequence similarity to ensure the structural elucidation of any member is informative.

In the case of eukaryotic SLC7A4, there is a variety of plant, fungal and mammalian homologues available. Several targets were chosen (Table 3.1) to explore a proton-coupled CAT structure to evolve our understanding from ApcT from *Geobacillus kaustophilus* (*Gk*ApcT). For this purpose, we selected (a) higher mammalian homologs, e.g., *Rattus novergicus* (*Rn*), *Mus musculus* (*Mm*) and *Homo sapiens* (*Hs*), being more informative for hCAT, although possibly less thermostable, and (b) less evolved, e.g., *Danio rerio* (*Dr*) and *Arabidopsis thaliana* (*At*), likely to provide higher expression, stability, and yield. Identification of a well expressing construct is a priority since subsequent thermostabilising mutations can increase the yield of well-behaved monodisperse protein, increase stability in short chain detergents and can stabilise conformational states that are difficult to crystallise (Tate, 2012).

Organism	Uniprot	Gene Name	Length	Mass (Da)
H. sapiens	O43246	*slc7a4*	635	68,268
M. musculus	Q8BLQ7	*slc7a4*	635	68,350
R. norvegicus	B5DFJ0	*slc7a4*	635	68,379
D. rerio	A4IG22	*slc7a4*	677	72,904
A. thaliana	Q8W4K3	*cat4*	600	63,639

Table 3.1 CAT4 homologues screened to identify targets amenable to structural work.

Solving the structure of an SLC7A4 homologue should provide important information about other homologues, as the amino acid sequence is generally conserved across species. Figure 3.1 shows the high degree of similarity, with residue similarities ~ 50% for even the most distant homologues. This primary sequence conservation of SLC7A4 members (Figure 3.2) indicates that although SLC7A4 is an orphan transporter, this transporter's function likely plays an important role

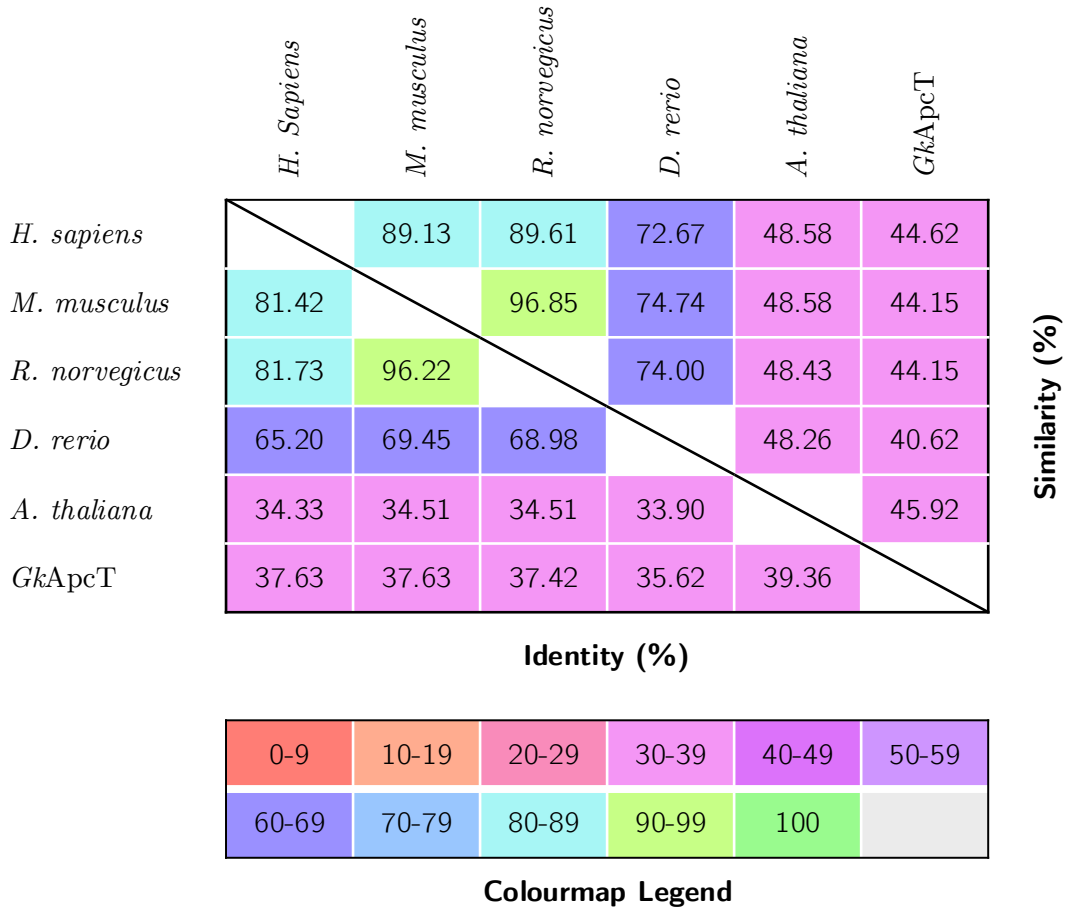


Figure 3.1 Amino acid sequence identities (below diagonal) and similarities (above diagonal) of SLC7A4 constructs from the selected species. The high degree of similarity between the tested human (*Hs*), mouse (*Mm*), rat (*Rn*), zebrafish (*Dr*) and arabidopsis (*At*) constructs, with similarities over 45%. Similarities calculated in *www.bioinformatics.org/sms2* with similar groups defined as: GAVLI, FYW, CM, ST, KRH, DENQ, P.

in eukaryotic cells. If the SLC7A4 transcript was not a functional protein, then it would be a pseudoprotein. However, Wolf *et al.* counter this argument since, (a) SLC7A4 is expressed in several human tissues and (b) an unfunctional *slc7a4* gene would have assimilated ORF-perturbing mutations leading to segregated transcripts (Wolf et al., 2002). Despite of this observation, homology (related to function) and sequence similarity (related to structure) are not "injective functions", i.e., proteins may share structural similarity but diverge functionally.

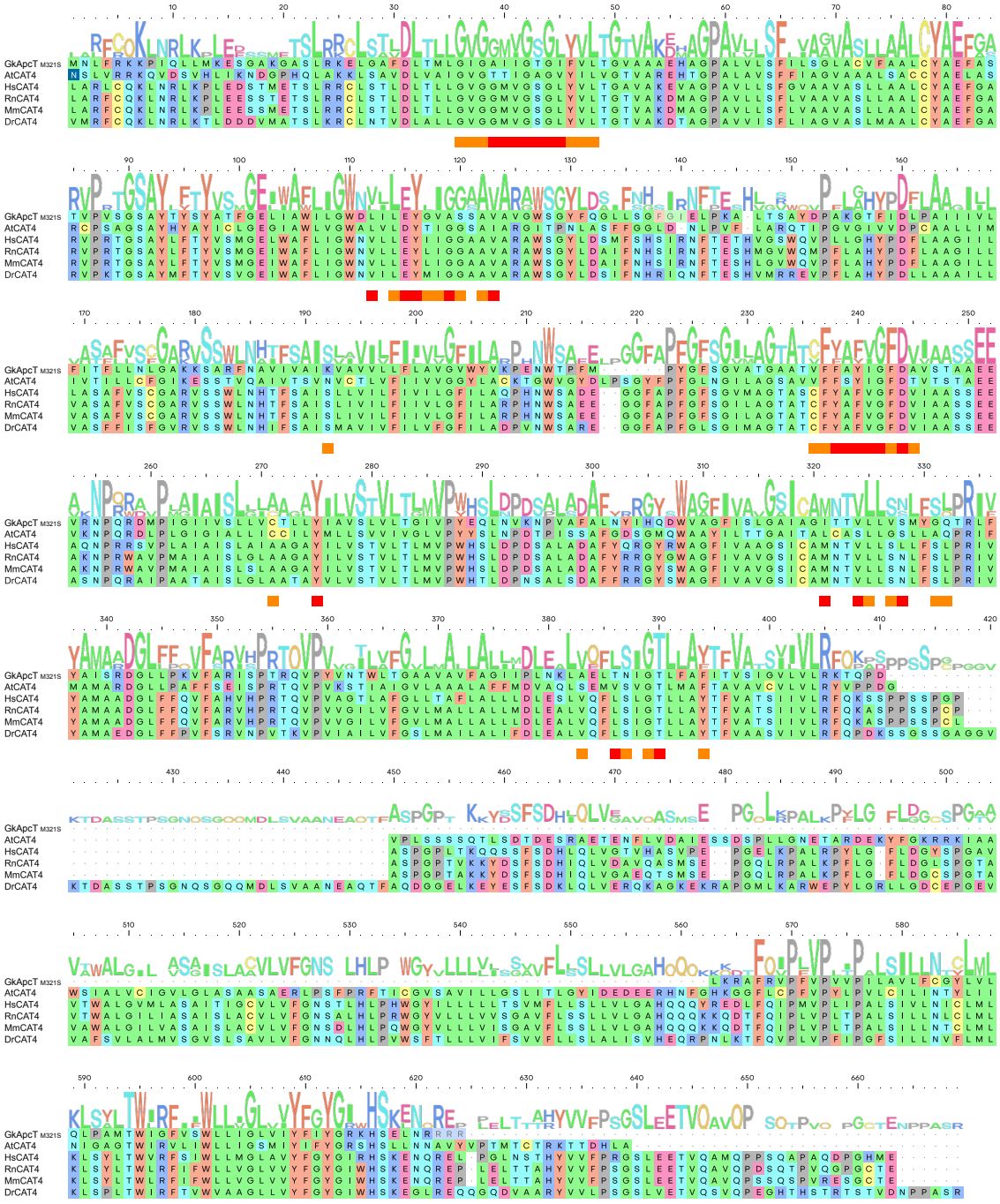


Figure 3.2 Multiple sequence alignment of SLC7A4 homologues used in this project. Residues are coloured based on side-chain chemistry. The consensus sequence is shown in logo format. Residues forming the L-arginine binding site in *Gk*ApcT M321S are indexed by red (within 5 Å) and orange (within Å) boxes below the alignment.

In an attempt to cautiously approach this problem before devoting an expensive structural biochemistry toolkit for study of the aforementioned "homologues",

a refined bioinformatics approach was employed. There is increasing understanding of the required substrate binding site elements that constitute a transporter selective for its cargo (chapter 1). This understanding of determinants has started to develop for the SLC7 family of amino acid transporters, e.g. structural basis for L-cystine recognition by the cystine/glutamate exchanger SLC7A11 (Parker, Deme, Kolokouris, *et al.*, 2021). Another example of binding site responsiveness to small perturbations in the SLC7 family is the transformation of *Gk*ApcT from an alanine to an arginine transporter by the substitution M321S. As the leading hypothesis for SLC7A4 is for it to be an L-arginine transporter similarly to GkApcT-M321S and the human CAT members, the L-arginine binding site similarity was investigated. Multiple sequence alignment revealed that the SLC7A4 homologues highly conserve the binding site characteristics of the L-arginine transporter GkApcT (Figure 3.2). The similarities range from 73% between *A. thaliana* and *Gk*ApcT and reach 100% between higher mammalian constructs and *D. rerio* (Figure 3.3). This was considered a satisfactory conservation of local chemical environment so that all constructs would likely function similarly.

The positively charged guanidino group, $[H_2N–C(=NH_2)–NH–]^+$, is the distinct molecular feature of the L-arginine substrate, whereas other interactions with the substrate backbone, $H_3N^+–(CH)R–CO_2^-$, are largely conserved for α-amino acids transporters in TM1 and 6 (chapter 1). The homologues carry highly similar sequences with GkApcT in TM1 and TM6, i.e., positions 36-58 and 241-250 (Figure 3.2), suggesting their function as α-amino acid transporters. Importantly the H^+-coupling *Gk*ApcT residue E115, located in TM3 (positions 112-123), is also conserved between constructs. E115 also formed a salt-bridge with $[H_2N–C(=NH_2)–NH–]^+$ in the GkApcT structure. Collectively, this suggests SLC7A4 homologues

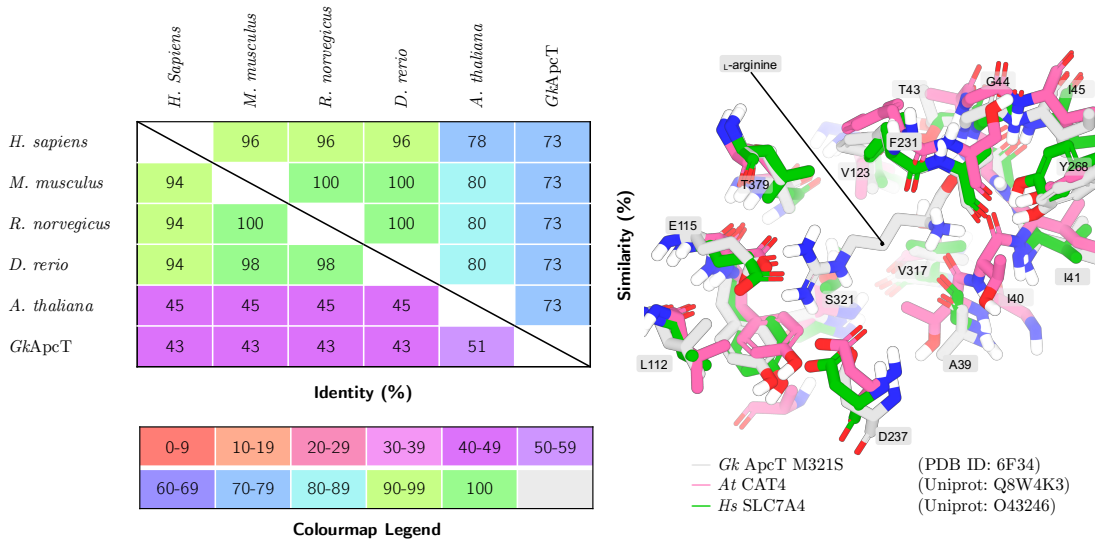


Figure 3.3 Binding site amino acid sequence identities (below diagonal) and similarities (above diagonal) of SLC7A4 constructs from the selected species. Binding site residues are selected as those with at least one heavy atom within 6 Å of L-arginine (left). Structures of the selected binding site residues (in *Gk*ApcT numbering) displayed from the crystal structure of M321S *Gk*ApcT and Alphafold2 predictions of human and plant SLC7A4 (right).

to transporter L-arginine with a H^+-coupling mechanism. This observation further proposes that SLC7A4 may primarily function as an arginine transporter like *Gk*ApcT M321S. Secondly, the second acidic *Gk*ApcT binding site residue, D237, is fully conserved. Lastly, the eukaryotic homologues also retain residues with reduced surface area in position 328, in accordance with the M321S mutation (i.e., plant – S327, human – L336, mouse & rat – N336, zebrafish – N336). Overall, our selection of eukaryotic homologues retains the key characteristics involved in the recognition of L-arginine and the H^+-coupling mechanism.

3.1.1 Expression Tests

The five homologues (Table 3.1) of SLC7A4 were cloned into a pDDGFP2-LEU2D vector for expression in *S. cerevisiae* (Joanne L Parker and Newstead,

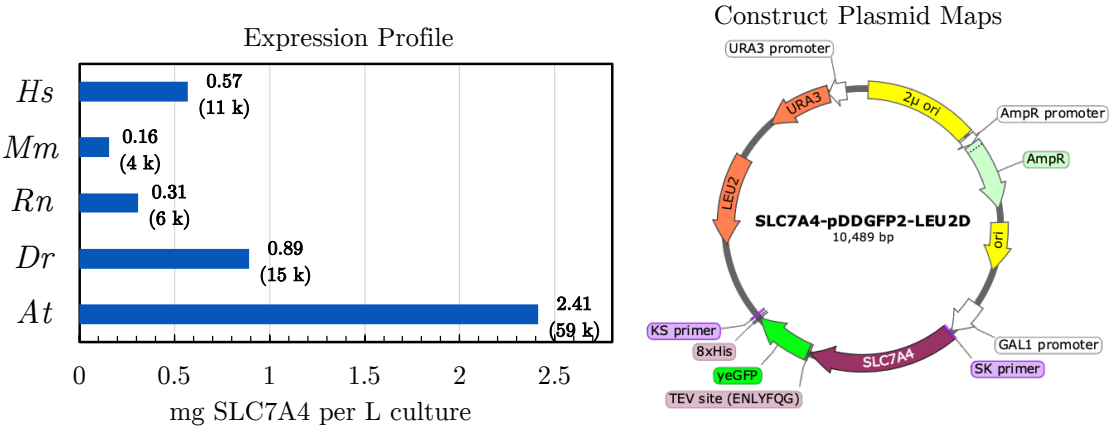


Figure 3.4 Quantification of SLC7A4 construct expression profiles as C-terminal yeGFP fusions in *S. cerevisiae* heterologous overexpression. Primary units displayed are in mg SLC7A4 per L culture, calculated from crude isolated membranes from 40 ml culture and scaled to 1 L. In parentheses are shown the whole-cell GFP fluorescence counts from 8 ml saturated culture reported as thousands (k) relative fluorescence units (left). The SLC7A4 plasmid maps in the yeast overexpression vector, pDDGFP2-LEU2D (right).

2014). These constructs are all of the sequence format, 'N-SLC7A4-TEV$_{cleavage}$-yeGFP-His$_8$-C'. All constructs successfully expressed as TEV$_{cleavage}$-yeGFP-His$_8$ fusions in small-scale screening cultures (50 ml) in synthetic complete media, with whole-cell GFP counts showing a vast range from 4,000 up to 60,000 relative fluorescence units (RFU) per 8 ml of saturated culture. To transform the measured GFP fluorescence counts to generalisable, transferrable units, the isolated membrane counts of each construct were converted to mg SLC7A4 per L *S. cerevisiae* culture according to an established method for yeast overexpression of GFP-tagged membrane proteins (Drew *et al.*, 2008). The construct plasmid and expression results are presented in **Figure 3.4**.

The construct with the highest expression level was the plant SLC7A4 homologue, displaying $>$ 4.0$\times$ higher expression in mg/L culture to human SLC7A4 whereas the zebrafish construct expressed 1.6$\times$ more than human. Surprisingly the other mammalian constructs, rat, and mouse SLC7A4, showed very low levels

of protein expression even though their sequence identity to the human construct is 81-82%. Previous group experience dictates that counts above the 10,000 RFU/ 8 ml threshold (equivalent to ~ 0.4-0.5 mg/L for a transporter the size of SLC7A4) are representative of a construct that overexpresses without anomalies. Consequently, all constructs except *Mm* and *Rn* showed appreciable levels of expression. In general, SLC7A4 overexpression in *S. cerevisiae* is reduced the more evolutionary advanced the originating organism is. This is not surprising if one considers that fungi have tighter evolutionary relationships with plants, having evolved from early protists, and thus share a comparable cell structure and an evolutionary linked protein folding machinery. Structural studies of plant membrane transporters often report *S. cerevisiae* as the heterologous expression host, as reviewed elsewhere (Locascio *et al.*, 2019). All expressing constructs were advanced to the next stage of screening.

3.1.2 Detergent Solubilisation & FSEC Monodispersity

An important parameter in successfully purifying a membrane protein is detergent choice (Bill *et al.*, 2011). An ideal detergent should solubilise the target protein from the membrane to a reasonable degree, stabilise it in aqueous solution and retain its tertiary structure once solubilised. Even closely related proteins are not stable in similar detergents (Meury *et al.*, 2014). The higher the detergent head group charge density is, the harsher it is (Privé, 2007), while, the shorter the alkyl tail length the harsher it will tend to be, due to hydrophobic mismatch between the protein hydrophobic belt and the lipid/detergent acyl tails (Lee, 2005). As a general rule, mild detergents are preferred for sensitive, flexible membrane proteins to avoid denaturation (Privé, 2007).

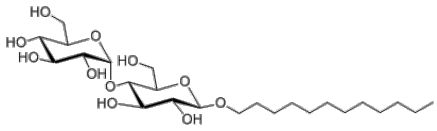
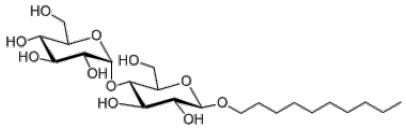
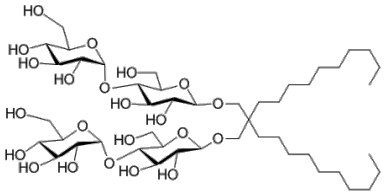
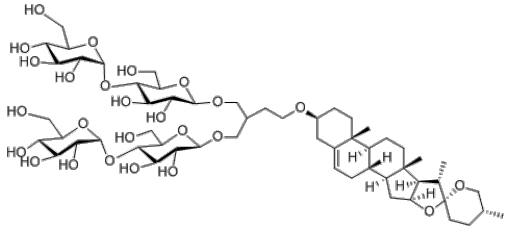
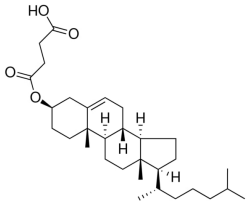
Name	Structure	Class	CMC	Micelle Size
DDM		Non-ionic Maltoside	0.17 mM	72 kDa
DM		Non-ionic Maltoside	1.80 mM	40 kDa
LMNG		Non-ionic Maltose-Neopentyl Glycol	0.01 mM	*91 kDa
GDN		Non-ionic glycodiosgenin	0.018 mM	N.D.
CHS		Cholesterol derivative	N.A.	N.A.

Table 3.2 Seminal information on detergents used for SLC7A4 homologue screening and stability. Information extracted from (Stetsenko and Guskov, 2017; H. J. Lee *et al.*, 2022) and anatrace website (*https://www.anatrace.com/*).

The most commonly used detergents in alpha helical membrane protein crystal structures are maltosides (50%, of alpha helical structures in the PDB), glucosides (23%) and amine oxides (7%) (Parker and Newstead, 2012). Most structural studies of amino acid solute carriers, either with X-ray crystallography or

cryoEM single particle analysis, report the use of DDM and a new class of NG-based detergents. These "base" detergents are commonly mixed with low percentages of cholesterol esters, such as cholesteryl hemisuccinate (CHS) (Coleman, Green and Gouaux, 2016; Bräuer *et al.*, 2019; Parker, Deme, Kolokouris, *et al.*, 2021; Y. Lee *et al.*, 2022), or other native lipids during purification (Coleman, Green and Gouaux, 2016) to form a bicelle, i.e. bilayer micelles consisting of a disc-like lipid membrane whose edges are stabilised by detergent (Sanders and Prestegard, 1990; Poulos *et al.*, 2015). Cholesteryl hemisuccinate stabilises some membrane proteins, even independently of any physiological relevance (Thompson *et al.*, 2011; K. E. J. Jungnickel, Parker and Newstead, 2018), thus base-detergent mixtures with CHS were justified and were also tested (Table 3.2).

To rapidly and quantitatively assess the solubilisation efficiency of different detergents for the SLC7A4 homologue set, an established GFP fusion strategy was employed, as developed elsewhere (Newstead, Kim, Von Heijne, *et al.*, 2007; Drew *et al.*, 2008) As shown in Figure 3.5, all wild-type SLC7A4 constructs are solubilised more efficiently in DDM and DDM-CHS micelles ($\alpha>0.77$) compared to the LMNG and LMNG-CHS ($0.64<\alpha<0.8$). The presented values are estimates as large-scale purifications may not scale linearly. DDM-CHS and LMNG-CHS micelle mixtures are 10:1 w/w, as this ratio is standard in assessing the supplementation of CHS in micelles. Solubilisation efficiency is calculated by measuring GFP fluorescence of solubilisation reaction before and after separating the insoluble fraction using ultracentrifugation. Although very low solubilisation efficiencies ($\alpha<0.5$) could be indicative of unfolded protein species that do not partition in the detergent micelles, the reverse is not necessarily true. That is,

high solubilisation efficiency is not an indicator of construct stability. Thus, we assessed uniform distribution of solubilised protein and aggregation susceptibility.

A downstream application is Fluorescence Size Exclusion Chromatography (FSEC), a more sophisticated approach that reports in protein-in-detergent monodispersity and provides a fast and high throughput method for membrane protein screenings (Kawate and Gouaux, 2006b; Newstead, Kim, Von Heijne, *et al.*, 2007). The solubilised membranes contain an array of solubilised proteins. Among them, the targets of interest are tagged with a C-terminal GFP. Thus, by injecting the solubilised membrane fraction and monitoring GFP fluorescence ($\lambda_{Ex.}$ 488 nm, $\lambda_{Em.}$ 512 nm) the protein of interest is distinguished and the monodispersity of the chromatogram peak can be used to reflect protein stability. In **Figure 3.5** are shown the FSEC profiles for the wild-type constructs; in micelle complexes with a high molecular weight (LMNG, LMNG-CHS and DDM-CHS) SLC7A4 produces large, monodisperse peaks, likely attributed to the detergent's overstabilising properties. In contrast, DDM solubilised SLC7A4 shows variability between constructs and is therefore a suitable discriminator of homologue stability. The plant protein is the only homologue that shows a sharp peak shape in DDM which is indistinguishable from the others (DDM-CHS, LMNG, LMNG-CHS), suggesting this protein is intrinsically more stable. In DM [shorter acyl tail length than DDM (10 versus 12 carbons)], plant produced a broader peak showing construct-responsive destabilisation. GDN produced a monodisperse peak for the plant protein, which overlapped with the remaining detergents, indicating that the construct has reached the maximum stabilisation detectable by FSEC.

Based on the FSEC peak monodispersity in DDM and responsiveness to other

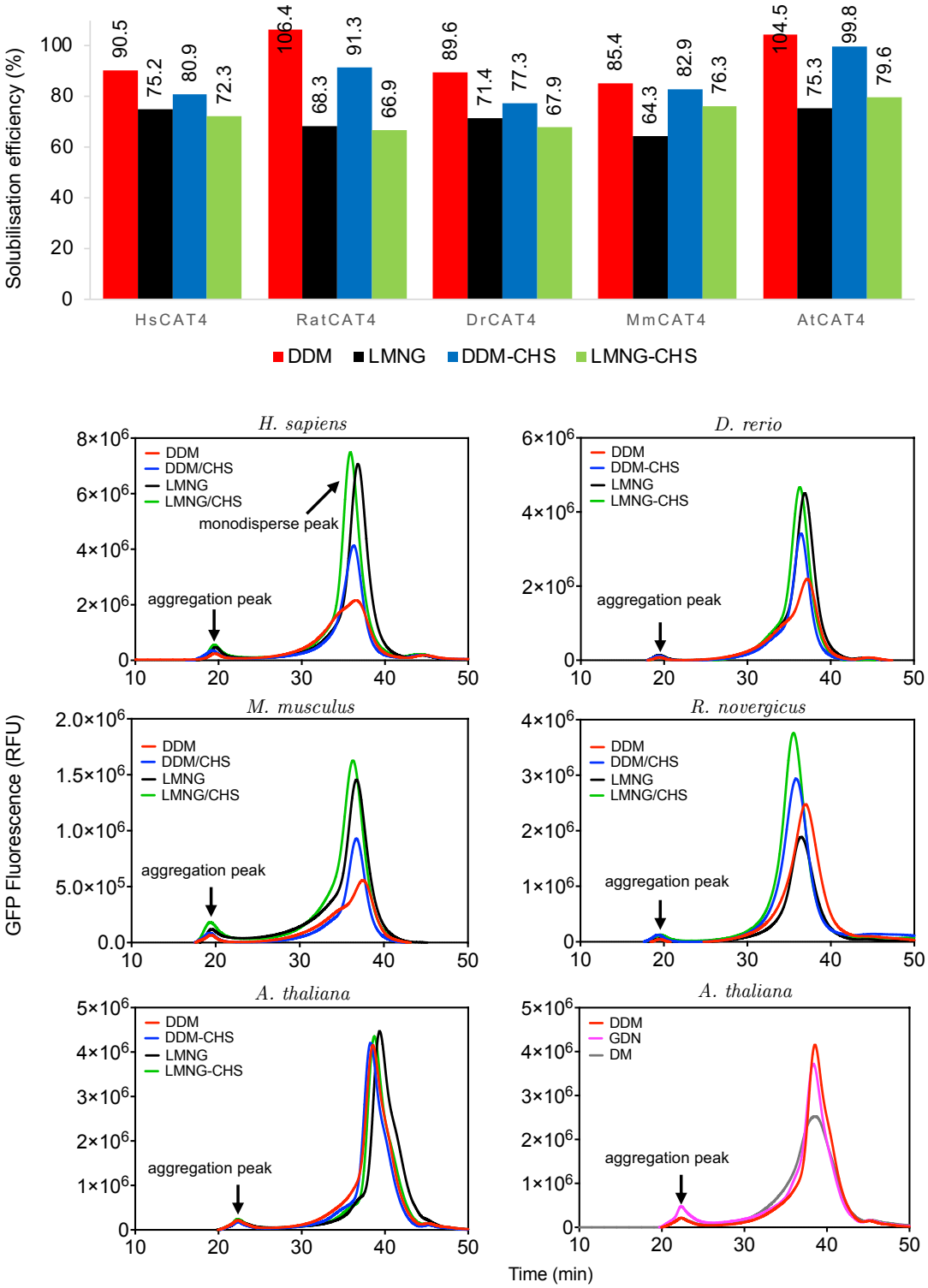


Figure 3.5 Detergent solubilisation efficiencies of various detergents with tested homologues. FSEC traces of homologues tested in this project with detergents. FSECs were run on a Superose 6 10/300 GL gel filtration column. X-axis in min and Y-axis is recorded via a fluorescence detector set to emit 488 nm - absorb 512 nm. No molecular weight standards were used.

detergents homologues were ranked in order of decreasing stability as: $At > Rn$

> *Dr* > *Mm* > *Hs*. Taken together with the expression profiles in *S. cerevisiae*, the highly expressing *Dr* (second only to the plant) and *Hs* (high level of expression) constructs were prioritised for small scale test purifications. The lower expression levels of the *Rn* and *Mm* (see §3.1.1) render these constructs unsuitable for large scale expression in *S. cerevisiae*, even though the *Rn* FSECs were high in ranking. It was decided to proceed with test purifications of *At*, *Dr* and *Hs* SLC7A4, with the plant homologue possibly showing the best thermostability. Plant homologues have shown amenability to crystallisation (Joanne L. Parker and Newstead, 2014; Tao *et al.*, 2015; Guo *et al.*, 2016; Wang, Chen and Zhou, 2021; Löbel *et al.*, 2022).

3.1.3 Trial Purifications & Construct Characterisation

The trial purifications were tested at "small-scale", referring to 1.5-3.0 g of membranes from 2-3 l of *S. cerevisiae* culture. Briefly, the purification involved the yeast membrane solubilisation in detergent, removal of non-solubilised proteins via ultra-centrifugation, purification over a Ni-NTA column, removal of the yeGFP-His_8-tag via overnight TEV-cleavage (detailed in chapter 2). Specifically, after the first round of Ni-affinity purification SLC7A4 should empirically be >80% pure from major contaminants. After the elution fraction is treated with protease, the major components are SLC7A4, cleaved yeGFP-His_8, TEV protease and to a smaller extent, other species which were retained by the Ni resin and co-eluted with the target. The eluant is dialysed overnight and the next day, after the second Ni affinity column, the sample is injected through a gel-filtration column (size exclusion chromatography) as a polishing step. This method, with modifications as to the affinity tag chosen (e.g., FLAG-tag, TwinStrepTactin,

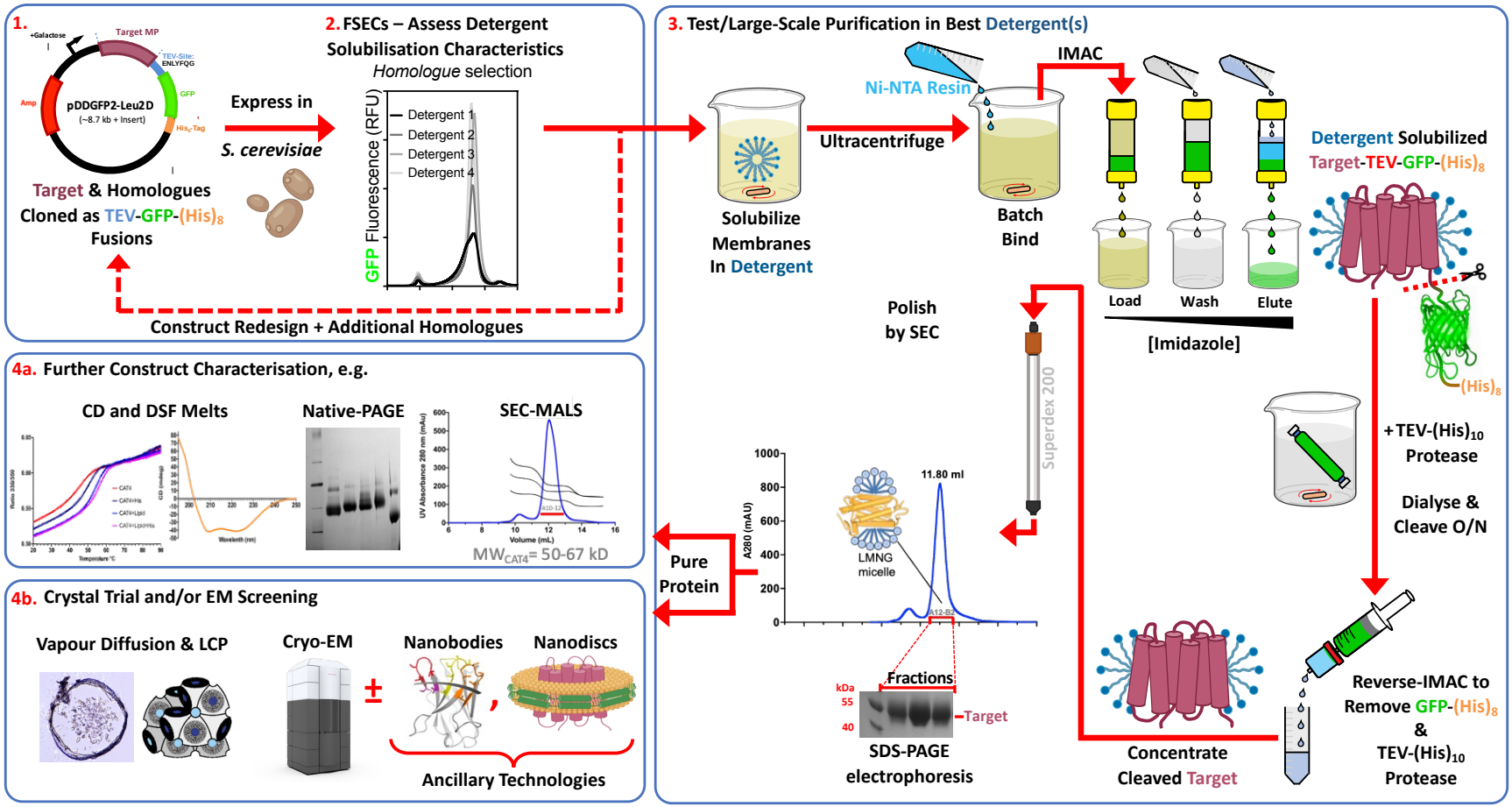


Figure 3.6 Expression, purification, biophysical characterisation, and structural determination pipeline steps employed in state-of-the-art eukaryotic membrane transporter structural biochemistry projects. Data presented in inlets is derived from the plant SLC7A4 homologue. Figure adapted from (Drew et al., 2008) and Dr Oliver Adams' doctoral thesis (available in the Oxford Thesis Repository)

etc.), provides a useful general method for purifying SLC transporters (Drew *et al.*, 2008). An outline of the experimental purification steps undertaken for SLC7A4 homologues is shown (Figure 3.6).

3.1.3.1. *Human* & *Danio rerio* SLC7A4 Purification Trials

The zebrafish construct (*Dr*) being the better expressing homologue among the higher eukaryotes and with monodisperse FSEC peaks continued into purification in all four detergents (DDM, LMNG and CHS supplemented counterparts) (Table 3.3). Affinity immobilisation of tagged proteins is an indicator of protein-detergent quality since it must be well-folded, stable, and exposing the terminal His-tag to bind the resin with fast saturable kinetics. Additionally, the protein must quantitatively elute the resin upon addition of higher imidazole concentrations (250 mM). Zebrafish SLC7A4 was readily soluble in all detergents and showed $>$ 60% His-tag immobilisation to Ni resin in less than 3 h. Furthermore, all protein-detergent complexes eluted the resin without remnants on the resin, as greenish colouring of the resin was not observed after elution. The opposite would indicate unfolded protein non-specifically binding the resin in an uncontrollable manner. This was not observed with *Dr* SLC7A4, suggesting promising biochemical characteristics.

Quality Metrics	DDM	DDM-CHS	LMNG	LMNG-CHS
Solubilisation *eff.* (%)	93%	>100%	91%	>100%
Binding *eff.* (%)	76%	63%	84%	87%
IMAC elution counts/g membrane ($\times 10^3$ GFU)	4,149	6,147	10,226	10,683

Quality Metrics	DDM	DDM-CHS	LMNG	LMNG-CHS
Precipitation	Yes	No	barely	No
SLC7A4/g membrane (mg)	0.144	0.324	0.159	0.277
SEC trace	OK	shoulder	shoulder	monodisperse
Purity (%)	>90%	90%	>90%	>90%
DSF melt profile	linear	linear	linear	linear

Table 3.3 Purification quality control metrics tabulated for *Dr* SLC7A4 overexpressed in *S. cerevisiae*. Solubilisation efficiency is calculated by measuring the GFP fluorescence of solubilisation mixture before ultracentrifugation and the supernatant after ultracentrifugation; >100% values are possible considering pre-ultracentrifugation samples contain impurities that may quench GFP fluorescence.

The set of detergents showed different purification profiles which could not be captured by FSEC analysis. DDM solubilised the protein > 90%, but the amount of eluted protein was lower compared to the other detergents, i.e., < 0.15 mg per g membrane. Precipitation during TEV cleavage demonstrated instability of SLC7A4 in DDM once the C-terminal GFP protein was removed. In the more lipid-like LMNG micelle, protein stability during overnight TEV cleavage was retained, likely due to the more compact micelle. Additionally, the amount of eluted fluorescent fusion protein was over two-fold compared to DDM (calculated from IMAC elution counts/g membrane). Purification size-exclusion chromatograms were comparable for DDM and LMNG and although the protein could be separated from impurities (**Figure 3.7**), it remained unresponsive/linear to a temperature gradient monitored with differential scanning fluorimetry [DSF (§3.1.3.2)] (**Table 3.3**), an omen to partially unfolded, unstable protein.

The effect of CHS in SLC7A4 purification

DLS and SAXS measurements in protein-detergent complexes demonstrate

that when CHS is added to DDM micelles, the formation of a bicelle-like micelle architecture is induced, with a less curved centre region. An increase in the thermal stability of solubilised proteins was also observed (Thompson *et al.*, 2011). It is well known that eukaryotic membrane proteins are on average less stable than their bacterial homologues in detergent micelles; an effect in which delipidation plays a central role (Kotov *et al.*, 2019). However, in the case of difficult eukaryotic targets CHS has a clear beneficial effect in stability, making challenging flexible targets purifiable (Thompson *et al.*, 2011; Meury *et al.*, 2014; Green, Coleman and Gouaux, 2015; Kotov *et al.*, 2019). For example, in the case of eukaryotic KDEL receptors; CHS is supplemented in KDELR-DDM complexes (Bräuer *et al.*, 2019) to avoid precipitation during overnight dialysis, which was confirmed for SLC7A4 (Table 3.3). Although *Gk*ApcT was purifiable in DDM without supplements, it did co-purify with a cholesterol analogue and additional TM-helix chaperone (K. E. J. Jungnickel, Parker and Newstead, 2018). CHS's stabilisation mechanism is either by binding to a cholesterol site on a lipid-facing cleft or by creating a larger-sized more protective micelle (Bräuer, 2019) which introduces lateral compression that retains folding.

Small-scale purifications showed that including CHS reduces or eliminates SLC7A4 precipitation during dialysis/cleavage (Table 3.3), suggesting that zebrafish SLC7A4 unfolds in a detergent environment that is not protective enough or that the cholesteryl moiety binds the protein and stabilises its fold. Although protein precipitation during overnight cleavage was reduced in DDM-CHS (visual inspection), concentration-dependent aggregation before gel filtration suggests that the complex is unstable. Although high levels of overconcentration can be avoided, the observation suggests that the condition is not suitable

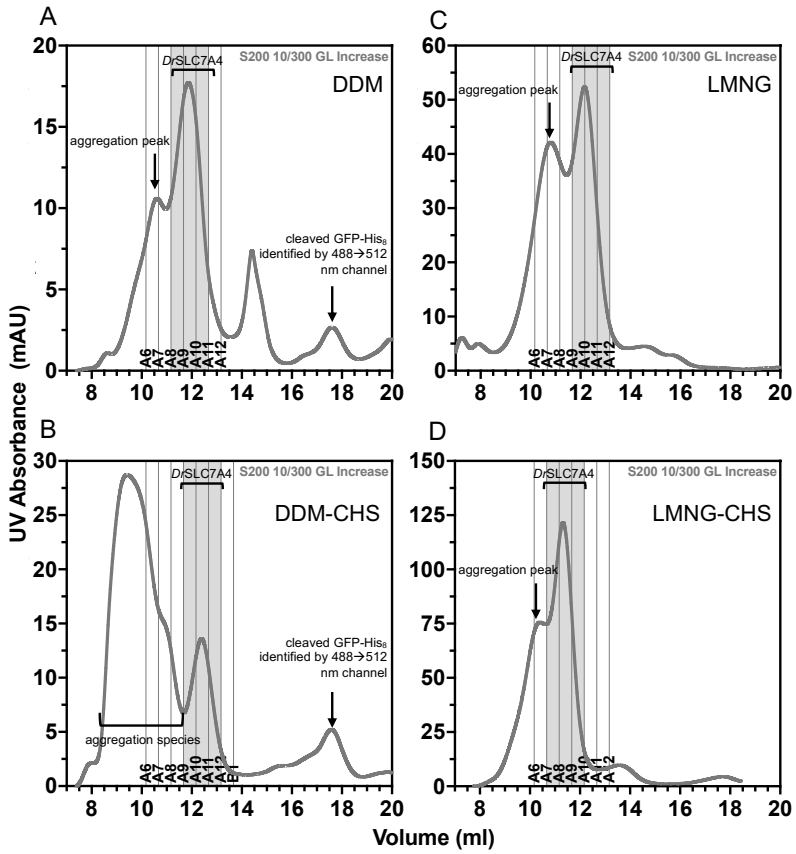


Figure 3.7 SEC traces of small-scale *Dr* SLC7A4 purifications run on a S200 Increase column with different detergents. Columns were not calibrated with molecular weight standards.

for the high protein concentrations (>10 mg/ml) required for LCP crystallisation. Protein aggregation in DDM-CHS is evidenced by the species eluting at the void elution (**Figure 3.7-B**). When LMNG-CHS is used purification steps proceed without precipitation formation and with the protein being cleaved as expected, and the concentrated sample lane primarily containing species at 65 kDa (**Figure 3.8-A**). The LMNG SEC trace shows improved monodispersity (**Figure 3.7-C**), suggesting that LMNG is a more stabilising detergent for SLC7A4 and when combined with CHS assists purification of homogeneous SLC7A4 (**Figure 3.7-D**). The main LMNG-CHS SEC peak (fraction A8-A10) contains a single protein species which runs at approximately 60-65 kDa (**Figure 3.8-B**). The theoretical molar mass

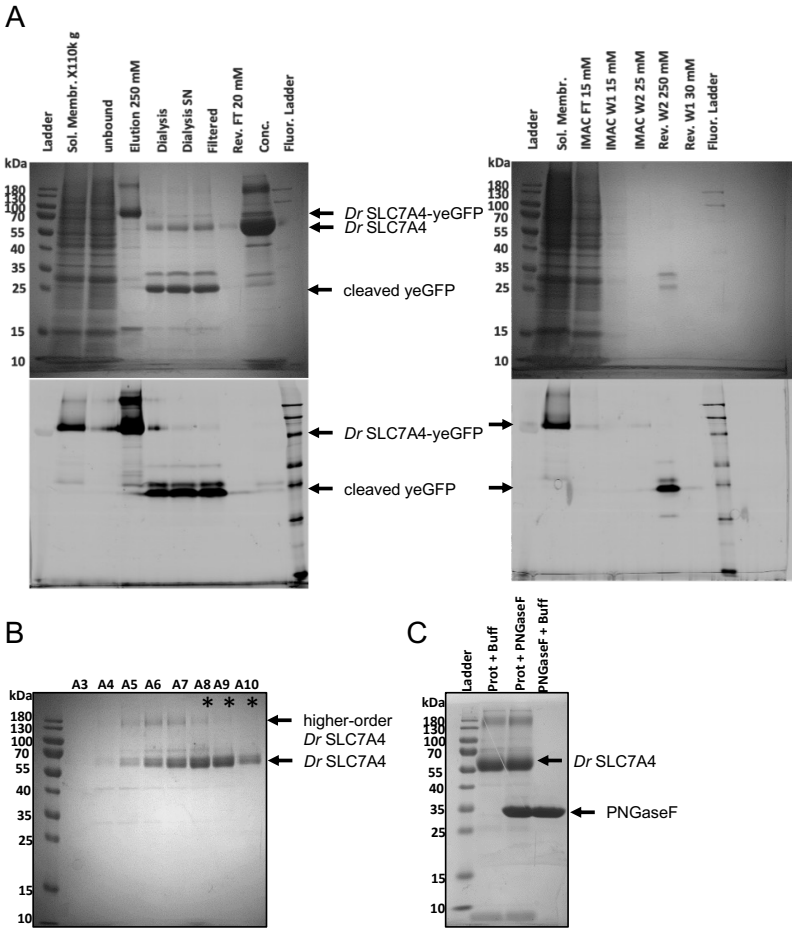


Figure 3.8 SDS-PAGE analysis of zebrafish SLC7A4 (A) purification steps in LMNG-CHS, (B) SEC fractions and (C) PNGaseF deglycosylation assay. Stained with Coomassie brilliant blue. In-gel GFP-fluorescence imaging was also done for the purification SDS-PAGE gels to monitor cleavage of the fusion protein. SEC fractions with a red star were pooled to further characterise with CD and DSF.

of zebrafish SLC7A4 is 72 kDa but detergent-solubilised membrane proteins disrupted by SDS routinely migrate faster through the polyacrylamide gel during electrophoresis due to the encapsulation of the species in higher negative charge. SDS-PAGE analysis of the LMNG-CHS purification showed $>$ 90% SLC7A4 purity (**Figure 3.8-B**).

Processing of the pooled fractions (A8-A10) with Peptide:N-glycosidase F (PNGaseF), a deglycosylating enzyme, for 1.5 h at 4 °C does not affect the band

width of zebrafish SLC7A4 in SDS-PAGE (Figure 3.8-C). However, this does not necessarily preclude the double band observed being due to glycosylation as *S. cerevisiae* proteoglycan branches may lead to "inaccessibility" phenomena for PNGase's. PNGaseF cleaves all asparagine-linked complexes, hybrid, or high mannose oligosaccharides unless the core GlcNAc contains an alpha 1,3-fucose.

3.1.3.2. Biophysical Characterisation of *Dr* SLC7A4 in LMNG-CHS

Analysing Sample Oligomerisation

In Figure 3.7-D two distinct peaks appear in the zebrafish SLC7A4 LMNG-CHS SEC trace, perhaps a higher order oligomer or self-assembled SLC7A4 and a monomeric peak. To some degree, shoulder abundance and resolution would correlate with protein concentration at injection, more suggestive of undesirable self-association. Upon reinjecting fraction "A6" to size exclusion a two-peak profile emerges with higher-order species being prevalent. Material within the shoulder (i.e., fractions "A3-A7") displayed two additional higher order bands by SDS-PAGE (Figure 3.8-B). Protein in the dominant SEC peak (i.e., fractions "A8-A10") retained its elution volume when rerun down the S200 increase column, arguing against a rapid equilibration with the shoulder species. However, separation of the two species by multiple gel filtration rounds of the "monomeric fractions" (i.e., fractions A8-A10) resulted in significant yield losses. This was considered unsustainable for downstream crystallisation which requires considerable protein amounts. As such, subsequent work with LMNG-CHS purified SLC7A4 only used fractions "A8-A10" or later. Accounting for the large size of the LMNG-CHS micelle it is possible that SLC7A4 dimers are stabilised in the earlier fractions. This hypothesis is tempting considering the dimerisation-inducing effects CHS exerts on LeuT (Gupta *et al.*, 2017).

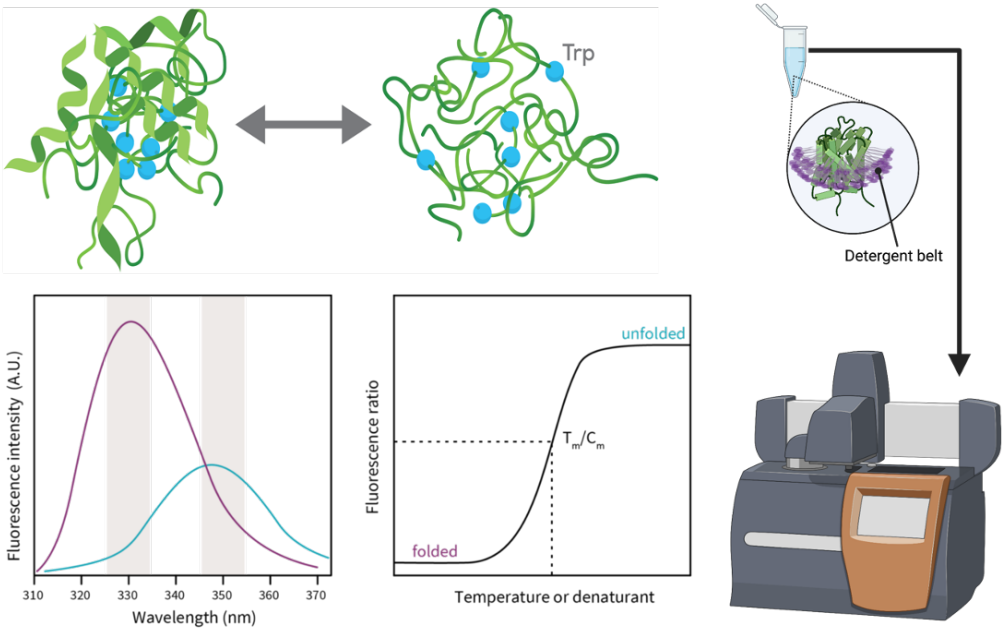


Figure 3.9 Physical principle and experimental setup of differential scanning calorimetry on a Prometheus NT.48 nanoDSF instrument (NanoTemper).The DSF method relies on temperature-resolved fluorescence monitoring. The Prometheus tracks tryptophan fluorescence (absorption 330 nm and excitation 350 nm). As a thermal gradient is applied to the sample, the protein gradually unfolds, and hydrophobic tryptophan residues become solvent exposed and absorb/emit fluorescence. The F350/F330 ratio is measured at each temperature point resulting in a sigmoidal curve representing the transition of the protein from its folded to an unfolded state with the midpoint equivalent to an apparent melting temperature (T_m). In this project, purified monodisperse protein in detergent belt was applied on the Prometheus (shown schematically). Chemical unfolding with a denaturant can also be used to capture the denaturant C_m, relying on the same principle.

Monomeric Sample Responsiveness to Thermal Gradients

At 20 °C the typical α-helical CD profile expected for SLC7A4 is observed (Figure 3.10-A). However, at increasing temperatures the signal is increasingly noisy, an observation attributed to an equilibrium established between well-folded and partially unfolded species. Thus, the melting transition of T_m=80.1 °C calculated by CD spectroscopy (Figure 3.10-B) is unlikely due to unusually high thermostability. A labile solute carrier is not expected to be more stable than an

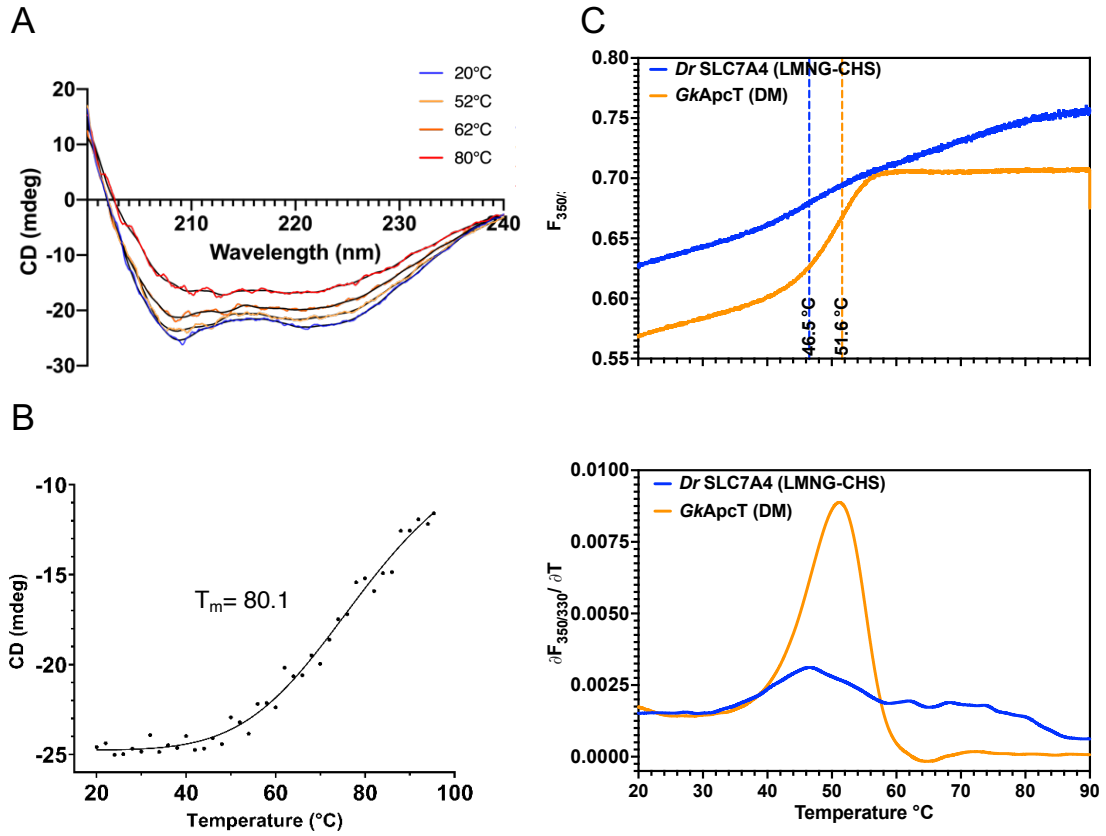


Figure 3.10 (A) Circular dichroism scan of the protein in LMNG-CHS over the temperature range 20-95 °C. CD scans at different temperatures (20, 52, 62, 80 °C). (B) CD signal at 208 nm from 20-95 °C – the temperature-gradient data was interpolated with a sigmoidal to extract the apparent melting temperature (T_m). (C) Differential Scanning Fluorimetry (DSF) signal as $F_{350/330}$ and $\frac{\partial F_{350/330}}{\partial T}$ of zebrafish SLC7A4 (red) and *Gk*ApcT (green, positive control).

evolutionary related bacterial homologue, for example $T_{m,GkApcT}$=72°C (K. E. J. Jungnickel, Parker and Newstead, 2018), or other eukaryotic membrane proteins (Nji *et al.*, 2018). To verify this, differential scanning fluorimetry (DSF) was employed which is independent of secondary structure (Figure 3.9). Zebrafish SLC7A4 has widely distributed tryptophan residues (2 in solvent accessible sites and 10 at membrane-facing sites according to Alphafold2) making zebrafish suitable for DSF. Indeed, cross-validation of zebrafish SLC7A4 and *Gk*ApcT shows that the melting profile for SLC7A4 produced a weak and broad transition at T=48 °C, at lower T than *Gk*ApcT, as expected (Figure 3.10-C).

$F_{350/330}$ at 20 °C is larger for SLC7A4 than *Gk*ApcT and the $\frac{\partial F_{350/330}}{\partial T}$ trace is not Gaussian-like. For a well-behaving protein, a Gaussian heat capacity trace should have narrow width, reflecting a high degree of phase transition cooperativity of particles in *non-degenerate* energy states (Crippen, 2001). To understand this, let us assume an ideal Gaussian (μ,σ) thermal transition of a protein. As $\sigma \rightarrow 0$ (i.e., the narrower the trace width is) the lower the density of microstates, $\rho(E)$, that have energy E_0 at T_0. Macroscopically we would "monitor" this behaviour as more uniform population dynamics, i.e., more cooperative. This can be understood instinctively if we consider that a Gaussian distribution contains 95% of population behaviour within 2σ, therefore the sharper a transition the more uniform the particle distribution. Based on this interpretation, the broad thermal melt profile for purified SLC7A4 suggests low cooperativity in melting transition; a higher density of *non-degenerate* microstates (incl. folded and partially unfolded states) at the transition offset. Thus, SLC7A4 displays intrinsic folding-unfolding dynamics in the LMNG-CHS micelle.

The discrepancy in T_m with CD spectroscopy is attributed to membrane protein α-helices being highly stable. At elevated temperatures the protein potentially is a bundle of loosely associated helices and therefore, a clear transition point cannot be observed.

Outlook on Human and Zebrafish Homologues

Comparison of human SLC7A4 purification and characterisation to zebrafish, along with rationalisation as to the biochemical quality of these constructs for structural studies is presented in **Appendix A**. Overall, the total yields of human

and zebrafish SLC7A4 purified from *S. cerevisiae* were only 0.03 mg/l and 0.08 mg/l of culture, respectively. A typical yield (75% of cases) of higher eukaryotic membrane proteins from *S. cerevisiae* is considered to be ~ 0.1 mg/l of culture (Newstead, Kim, von Heijne, *et al.*, 2007), in order to be cost-effective considering typical large-scale expression range from 10-20 l culture. Although zebrafish SLC7A4 is close to an acceptable yield per l from *S. cerevisiae*, the purified protein is not well-folded and shows aberrant behaviour. Thus, the *A. thaliana* plant SLC7A4 transporter homologue was deemed the most reasonable next step since:

- expression is >3.9× zebrafish (Figure 3.4) – assuming expression scales linearly with purification yield, at least a 0.31 mg/l protein yield can be expected which is above the 0.1 mg/l criterium for *S. cerevisiae.*
- plant membrane proteins are expected to crystallise either in detergent or lipidic cubic phase more readily than higher homologues.

3.2 The *Arabidopsis thaliana* SLC7A4 (AtCAT4)

Plant species have been used to understand human transporters (Tao *et al.*, 2015). The plant SLC7A4 is predicted to have 14 TM helices like its eukaryotic counterparts through high sequence homology (Figure 3.1 & Figure 3.2), a feature that the best "model" for cationic amino acid transport, *Gk*ApcT, lacks. Therefore, plant SLC7A4 was considered of high biological interest, superseding *Gk*ApcT as a eukaryotic protein, and of high similarity to mammalian counterparts. This homologue is known as *At*CAT4 – likely due to phylogenetic clustering with hCAT1 and hCAT2 – and is localised in the vacuole (Yang *et al.*, 2014). A genome-wide-association study reported *At*CAT4 to affect the levels of

L-histidine (Angelovici *et al.*, 2017). Additionally, *At*CAT2 – 64% sequence identity and 80% similarity to *At*CAT4 – reconstituted in proteoliposomes was shown to transport L-arginine in a pH-dependent manner (Cosco *et al.*, 2019b) and *At*CAT5 – 31% global identity and 43% binding-site identity – transported L-arginine in a pH-dependent manner (Y.-H. Su, Frommer and Ludewig, 2004). These studies suggest *At*CAT4 to be an amino acid transporter homologue.

3.2.1 Characterisation & Quality Control of AtCAT4

In *S. cerevisiae*, *At*CAT4 expressed 2.3 mg per g membrane (equiv. 2.0 mg/l culture), compared to 0.277 mg/g membrane for zebrafish (Table 3.4). In §3.1.2, detergent solubilities and FSECs were presented, and good profiles were retrieved in all detergents, except DM (Figure 3.5). Loading the same amount of membrane

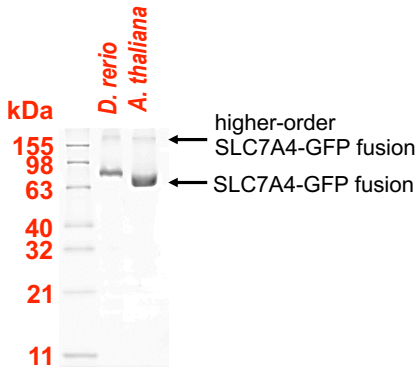


Figure 3.11 In-gel GFP fluorescence imaging of fusion zebrafish SLC7A4 and arabidopsis CAT4 from isolated *S. cerevisiae* membranes.

in SDS-PAGE gives a stronger in-gel fluorescence signal for *At*CAT4 (Figure 3.11).

Metrics	*D. rerio*	*A. thaliana*
Whole cell GFP Counts (488→512nm, GFU)	15,972	57,769-78,000
Optical Density ($OD_{600\ nm}$)	5.3	3.2-4.5
GFP Counts/ $OD_{600\ nm}$	3,013	18,053
GFP counts in membrane fraction (enriched fraction, 488→512nm, GFU)	17-20 k	55-75 k
Solubilisation *eff.* (%)	>100%	>100%
Binding *eff.* (%)	87%	70-85%
IMAC elution counts/g membrane ($\times 10^3$ GFU)	10,683	36,355
Precipitation	No	No
SLC7A4/g membrane (mg)	0.277	2.344
SEC trace quality	Monodisperse *with* shoulder	Monodisperse *without* shoulder

Table 3.4 Comparison of *D. rerio* and *A. thaliana* SLC7A4-yeGFP-His_8 expression profiles in *S. cerevisiae* and purification in LMNG-CHS. The metrics show superiority of plant SLC7A4 (CAT4) for the purposes of the project; purification of large quantities of likely, stable, protein. Solubilisation *eff.* >100% values are possible considering pre-ultracentrifugation samples contain impurities that may dampen GFP fluorescence.

3.2.1.1. Purification Optimisation

Purification in large scale *S. cerevisiae* cultures showed that, *At*CAT4 is more stable compared to the other constructs. Overnight aggregation during yeGFP-His_8-tag cleavage was avoided, although DDM was still deemed unsuitable due to protein precipitation in the final overconcentration step, evidenced by the large aggregation peak during gel-filtration (**Figure 3.12-A**). In complexes with DDM-CHS and LMNG-CHS (10:1 w/w) *At*CAT4 produced monodisperse peaks at elution volumes consistent with monomeric protein (V_R=[11.0-11.8] ml), without visual overconcentration-related self-aggregation, shown by the two gel-filtration

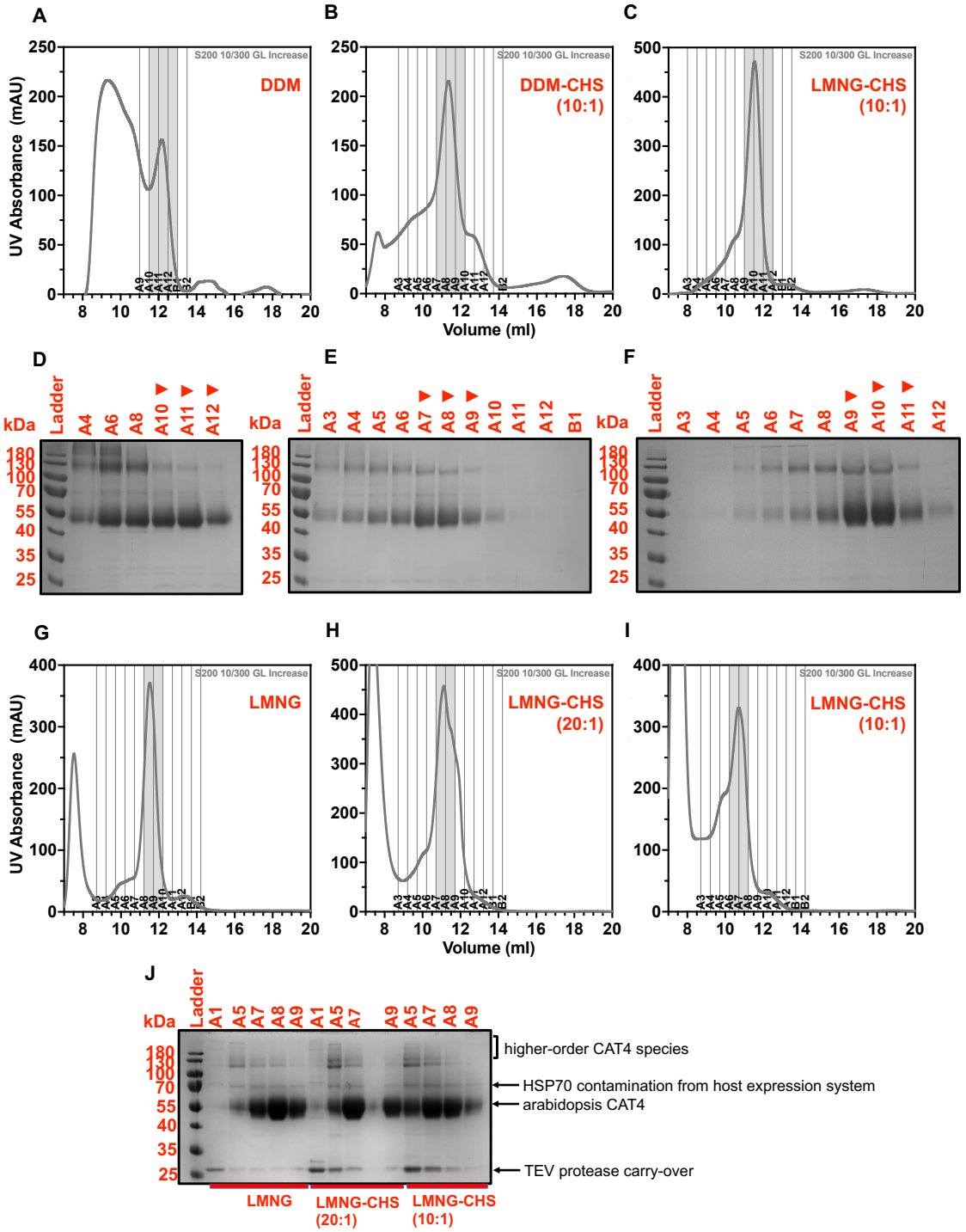


Figure 3.12 Purification screening of arabidopsis CAT4. SEC traces in (A) DDM, (B) DDM-CHS and (C) LMNG-CHS followed by (D,E,F) SDS-PAGE of the SEC fractions, respectively. Optimisation of CHS content in LMNG micelle purifications ranging from (G) none, (H) 20:1 LMNG:CHS, (I) 10:1 LMNG:CHS, followed by (J) SDS-PAGE of the SEC fractions.

traces (**Figure 3.12-B,C**). This is in line with previous observations that eukaryotic

SLC7A4s require a more stabilising environment to avoid self-aggregation effects. The height of the left-shoulder is similar in both selections (approx. 100 mAu) and material within the shoulder (i.e., fractions A3-A6 DDM-CHS; A4-A8 LMNG-CHS) displayed higher order bands by SDS-PAGE (Figure 3.12-E,F). The species isolated from the main peak retained its elution volume when rerun down the S200 column. Nevertheless, the LMNG-CHS purification shows considerable improvement in trace shape. The protein eluted in a narrower range of fractions separating it from other species. Homogeneous protein fractions from both detergents were pooled on the basis of SDS-PAGE, marked by ▼ in Figure 3.12, although the visible higher order species would engage us in evaluating the oligomeric state of the purified protein (§3.2.1.2).

*At*CAT4 was purifiable in LMNG with no addition of CHS (Figure 3.12-G,H,I), producing a narrow monodisperse main peak. The shoulder species in LMNG-only micelles contain less contaminants than LMNG-CHS micelles (Figure 3.12-J) making the former advantageous. Thus, purer CAT4 and higher yield is achieved in LMNG-only. Ni resin immobilisation was increased without CHS, i.e., from 70.2 % LMNG-CHS 10:1, to 89.4% LMNG-CHS 20:1, to 91.1 % LMNG-only (over 3 h saturated binding).

3.2.1.2. Investigating the Oligomeric State of AtCAT4

SDS is known to cause artifacts in electrophoresis samples, for example laddering. The higher-order bands from the SEC main peak initiated careful inspection of the purified species to ascertain that monomeric CAT4 is being purified. To investigate this, we resorted to couple SEC with multiple angle light scattering

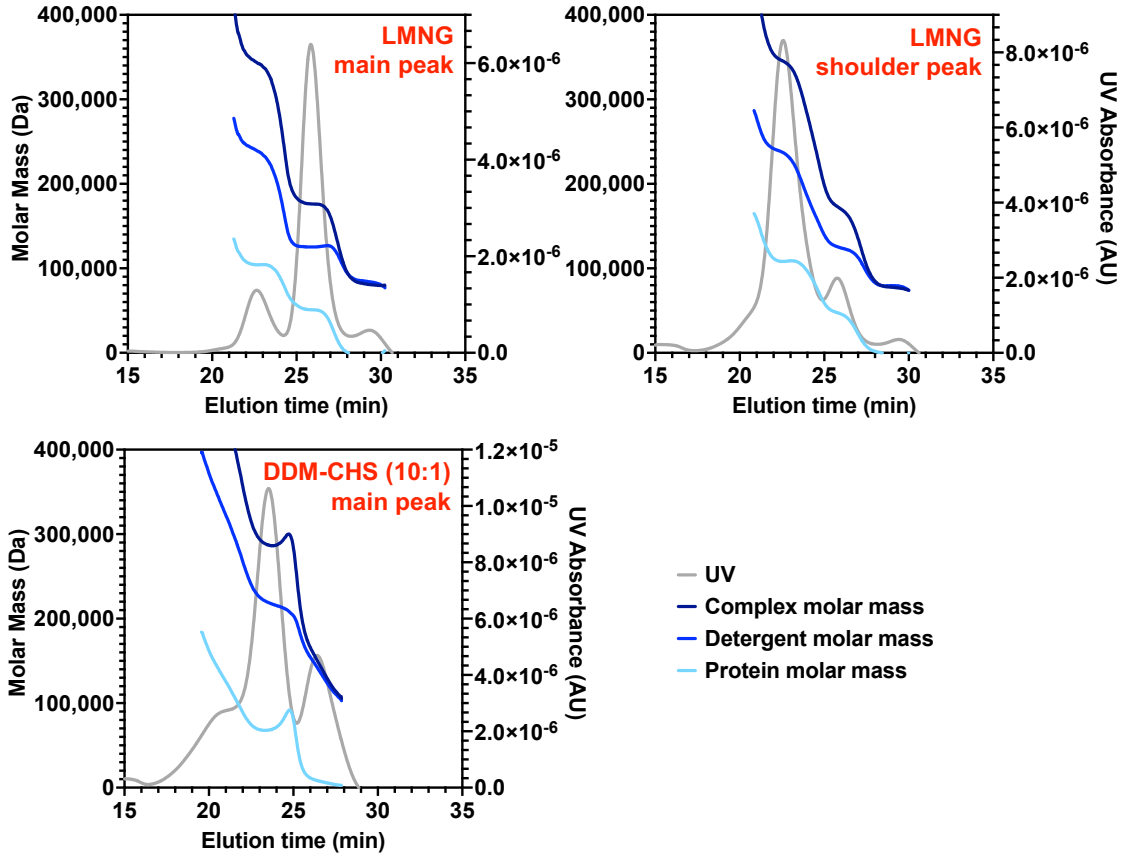


Figure 3.13 Evaluation of arabidopsis CAT4 oligomeric state in DDM-CHS and LMNG at 1 mg·ml concentration using multiangle light scattering (SEC-MALS). Pooled fractions (marked by ▼ in previous section) were injected to the SEC-MALS system. UV absorbance of the chromatogram and the deduced complex, detergent and protein-only molar masses are shown.

(SEC-MALS). The method couples fractionation by SEC (in-line with UV absorbance) with light-scattering intensity measurements (affected by concentration and molar mass), and refractive index measurements, to calculate species molar mass as they exit the column. The light scattering system indicates where peaks are composed of different oligomeric mixtures (mass varies across peak). The proportion of protein to conjugate (detergent, carbohydrate, etc.) in a peak can be determined by using the UV absorption, refractive index and light scatter signals (Dobrin and Randall, 2006). For a detergent-engulfed membrane protein three traces are produced: (a) complex molar mass, (b) detergent-only molar mass

and (c) protein-only molar mass taken as the difference of and a and b.

In LMNG, the peak material (i.e., shoulder-free) showed a near-constant total mass across its SEC-MALS elution profile, corresponding reasonably to the expected size of a *At*CAT4 monomer (~50 kDa) in an LMNG micelle. In DDM-CHS the main peak is ~68 kDa, close to the theoretical molar mass for CAT4, but the mass measurement is not constant across the peak pointing to an inseparable mixture of states. Corroborating this, the DDM-CHS shoulder species had a highly variable mass, more consistent with aggregates than true oligomers. The LMNG shoulder peak results in species of defined molar mass (100-108 kDa), roughly double the monomer size, suggesting that CAT4 dimer is isolated (**Figure 3.13**).

3.2.1.3. Investigating the Helicity & Thermal Stability of AtCAT4

The above results demonstrate that *At*CAT4 is purifiable in DDM-CHS and LMNG±CHS. The protein is generally monomeric, suggesting that CAT4 exists predominantly as a monomer with differential tendency to self-associate depending on detergent selection; LMNG stabilising a monomer better than DDM-CHS.

*At*CAT4 thermal stability and helicity were tested by differential scanning fluorimetry (DSF) and circular dichroism (CD), as described above. CD measurements (**Figure 3.14**) confirmed the α-helical secondary structure expected for *At*CAT4 pooled fractions, with significant negative bands near 222 nm and 208 nm and a positive band in 193 nm (Greenfield, 2007), in all tested detergents.

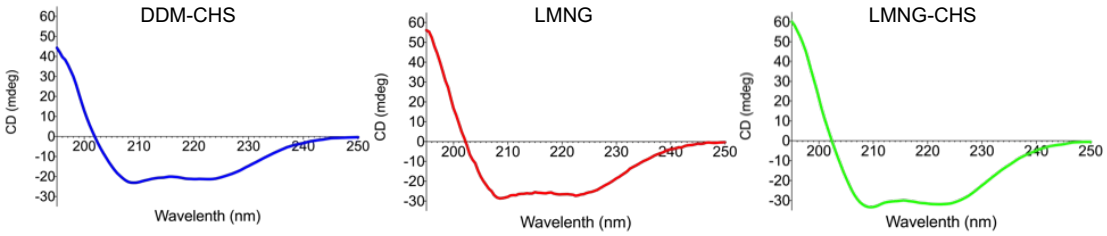


Figure 3.14 *At*CAT4 biophysical characterisation with CD. CD scan profiles of in DDM-CHS (blue), LMNG (red) and LMNG-CHS (green) showed a characteristic α-helical profile.

	Inflection Point (°C)
Sample	$T_{m,app}$
DDM-CHS (10:1)	44.7 °C
LMNG-CHS (10:1)	53.9 °C
+ L-Ala	54.1 °C
+ L-Lys	54.7 °C
+ L-Arg	55.5 °C
LMNG CAT4	45.8 °C
+ L-His	50.3 °C
+ Lipid	54.6 °C
+ Lipid + L-His	56.5 °C

Table 3.5 Apparent melting temperatures ($T_{m,app}$) extracted from melting trace inflection points, determined by DSF. All measurements were in 20 mM Tris, 150mM NaCl.

Previous analysis of zebrafish SLC7A4 (§3.1.3.2) showed that CD spectroscopy is not the method of choice for analysing thermal stability, while DSF showed reasonable results for zebrafish SLC7A4. Arabidopsis CAT4 is amenable to condition screening with DSF, having seven tryptophan residues (uniport ID: Q8W4K3). A strong signal was recorded when protein was screened at 1.0 mg·ml (15.2 μM). DSF analysis showed *At*CAT4 is better thermally stabilised in LMNG-CHS compared to DDM-CHS by ~ 9.2 °C (Table 3.5). However, adding CHS to LMNG was found to cause phase shifts in LCP crystallisation and was deemed unsuitable for crystallography (§3.3.1).

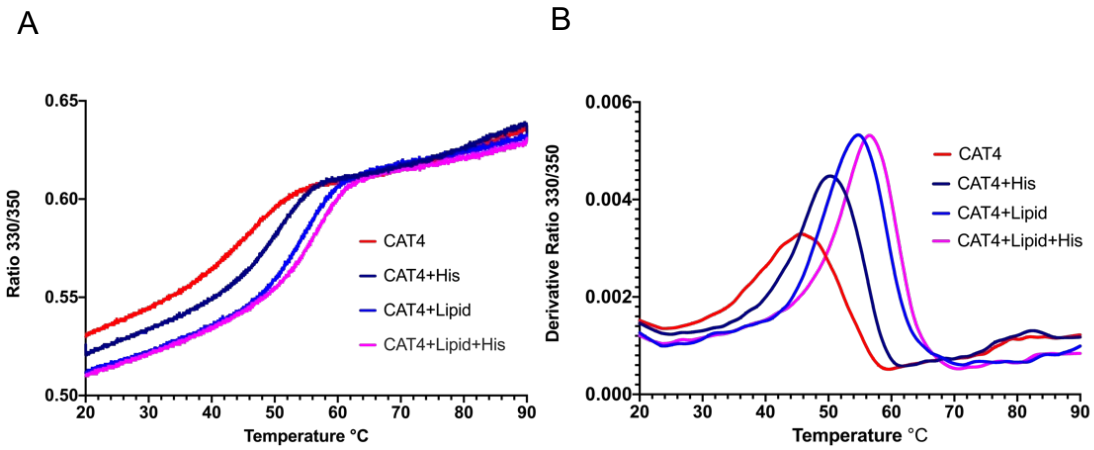


Figure 3.15 *At*CAT4 in LMNG thermostabilisation by substrate and lipid supplementation at pH 7.5. (A) $F_{350/330}$ and (B) $\frac{\partial F_{350/330}}{\partial T}$ signals were recorded, and the apparent T_m extracted from the Prometheus software, reported in **Table 3.5**.

The addition of other stabilising components to CAT4 in LMNG-only was investigated. Thermal stabilisation was conferred with the addition of a putative ligand, L-histidine, by 5 °C, and by 10 °C with the addition of soy phospholipid total extract (asolectin). Asolectin comprises roughly of equal proportions of lecithin, cephalin and phosphatidylinositol. L-histidine is a substrate for human SLC7A1 (hCAT1) (Kim *et al.*, 1991). The maximum stabilising effect was recorded with the simultaneous incubation of CAT4 with substrate and lipids (**Table 3.5**). In this case the melting curves give a noticeable shift compared to apo, with improved melting cooperativity (**Figure 3.15**), characteristic of stable protein. Addition of L-arginine did not result in significant thermostabilisation of lipidated CAT4 (< 3 °C). Collectively, this data demonstrates the purified transporter is responsive to substrate, and that relipidation increases its stability. Relipidation brings CAT4/LMNG T_m to similar value as having CHS in the purification (**chapter 5**).

Detergent	Conc.	Lipid	Substrate	Temperature	Outcome
DDM-CHS	20 mg/ml	MO	+R/K/H	4 & 18 °C	GB
LMNG-CHS	20 mg/ml	MO	+R/K/H	4 & 18 °C	GB
LMNG	30 mg/ml	MO	Apo+R/K/H	18 °C	GB
LMNG	30 mg/ml	MO+CHL	+R	18 °C	GB
LMNG	50 mg/ml	MO+CHL	+R	4 & 18 °C	microcrystals

Table 3.6 Setup and result of *At*CAT4 LCP crystal screens. The purification detergent and the concentration of the protein (measured with $A^{1\%}_{280nm}$=11.28 on nanodrop) prior to mixing with lipid are shown. Lipids tested were MO (monoolein, 9.9MAG) and MO+CHL (10:1 ratio). Amino acid single letter codes are given in the substrates column. Crystallisation pH was buffered with 20 mM Tris, 150 mM NaCl pH 7.5. In the outcome column, "GB" stands for "glowing bolus".

3.3 Crystallisation Trials

3.3.1 Lipidic Cubic Phase (LCP) Crystallisation Trials

The Lipidic Cubic Phase (LCP) is a liquid crystal that self assembles by creating a homogeneous mixture of lipids with an aqueous phase (*in meso*). This results in a quasi-solid bilayer environment. The protein, detergent-solubilised in the aqueous phase, is reconstituted in the bilayer, stabilisation is conferred, and crystallisation is enabled. The reconstituted protein lies on a three-dimensional curved bilayer, formed with a grid of aqueous channels. In this, proteins are free to diffuse laterally to create crystal contacts (Caffrey, 2015). One must thoroughly mix the phases, most commonly via a dual-syringe coupler, in the correct ratio, which is lipid-dependent such that the system is in the cubic portion of its phase diagram. For monoolein (9.9 monoacylglycerol - 9.9 MAG), a traditionally successful lipid used in LCP, this is achieved with a ratio of ~ 2:3 protein:lipid (w/w) at STP conditions (Kulkarni *et al.*, 2011; Caffrey, 2015).

Satisfied with a set of stabilising conditions for *At*CAT4, LCP crystal trials begun due to the method's distinct applications for membrane proteins (Parker and Newstead, 2012). *At*CAT4 purifies well in lipid-like detergents (LMNG) and is thermostabilised with excess lipid (§3.2.1.3) suggesting *in meso* crystallisation to be beneficial. LCP crystallised the CAT homologue, *Gk*ApcT, in cholesterol-doped monoolein and led to 2.86 Å diffraction (K. E. J. Jungnickel, Parker and Newstead, 2018).

LCP crystallisation conditions were setup with a Gryphon LCP robot (Art Robbins) and screened at 4 °C and 20 °C, in monoolein (MO), using custom crystallisation screens, covering a range of polyethylene glycols (PEGs), pH, buffer and salt conditions. These were either the commercial "MemMeso" and "MemMesoGold" or customised screens. Crystal formation was checked with the use of a light microscope equipped with a cross-polariser and a UV fluorescence microscope. Incubation – up to 14 days – of *At*CAT4 in monoolein or cholesterol-doped monoolein did not result in crystal formation at high protein concentrations (20-30 $mg{\cdot}ml^{-1}$).

The conditions tested for crystallisation are shown in Table 3.6. Visual inspection of all LCP batches displayed some degree of residual "haziness" (glowing bolus); indicating that formation of the desired mesophase was suboptimal, hindering crystallogenesis. This was most evident for samples containing CHS. A higher concentration without CHS could be beneficial to mediate crystal contacts. *At*CAT4 was setup at 50 $mg{\cdot}ml^{-1}$ in the presence of 10 mM L-histidine with monoolein supplemented with cholesterol. After 4-5 days needle-like microcrystals were observed, ~ 30 μm length (Figure 3.16). No considerable growth was recorded

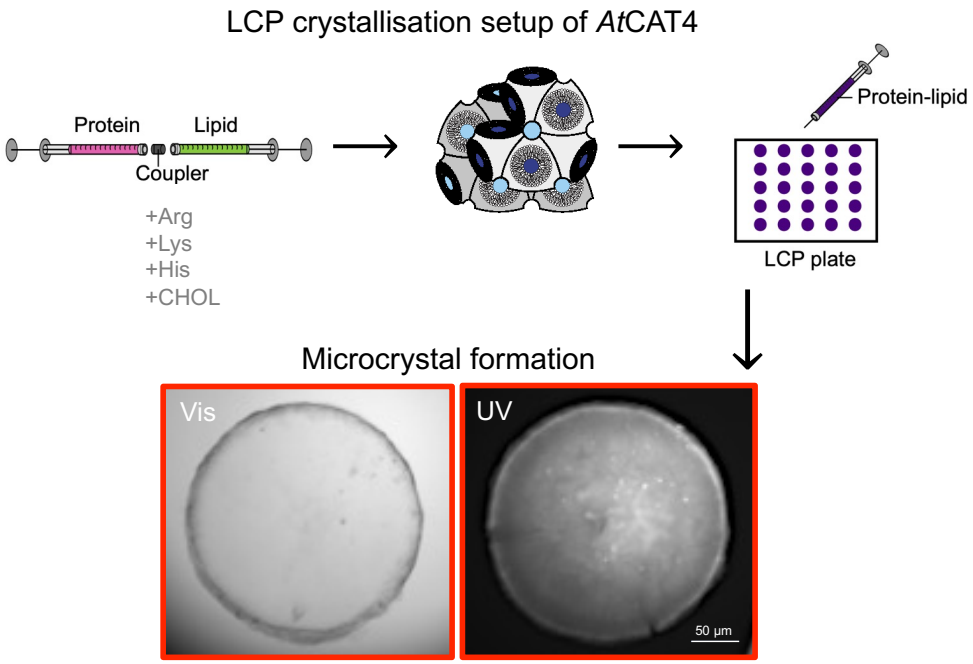


Figure 3.16 LCP method outline for *At*CAT4 and microcrystal formation in select few wells without the glowing bolus effect (Tokyo Set 2 screen). Example proteinaceous species in crystallisation wells in LCP trials. Crystallisation wells are viewed under cross-polarised visible light or UV fluorescence. Under UV, protein crystal particles appear white.

after 3 weeks at 18 °C. Unfortunately, these microcrystals were not suitable for diffraction experiments, as they would likely not survive X-ray radiation.

3.3.2 Vapour Diffusion (VD) Crystallisation Trials

With the concomitant discovery of the 10 °C thermal shift asolectin provides CAT4, we incubated overnight CAT4 in asolectin at a 3.2:1 lipid:protein molar ratio followed by traditional vapour diffusion crystallisation. These trials showed good precipitation of protein-like species and perhaps microcrystal showers but did not result in well-formed crystals (**Figure 3.17**). Other SLC6 and 7 transporters have been crystallised using high lipid-detergent (HiLiDe) vapour diffusion crystallisation previously (Yamashita *et al.*, 2005c; Gao *et al.*, 2010; Penmatsa,

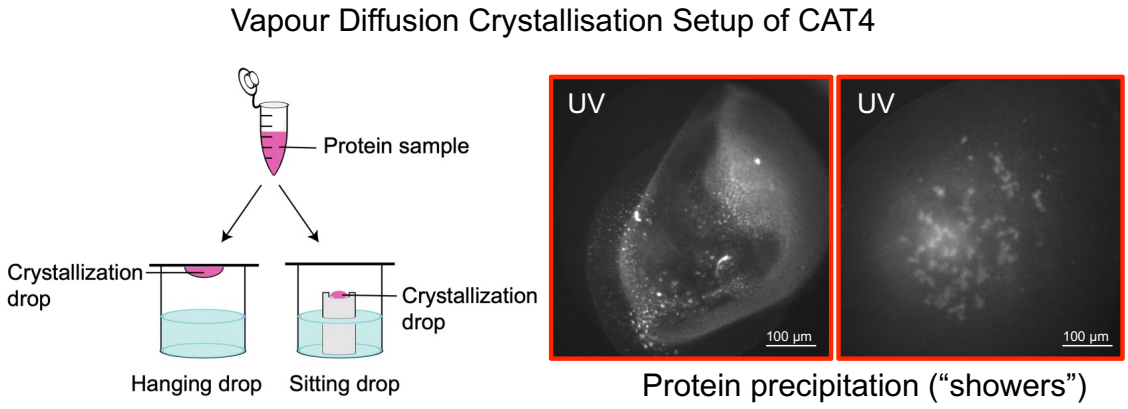


Figure 3.17 Traditional crystallisation methods using vapour diffusion and setup of *At*CAT4 with the hanging drop method resulting in protein species bulk precipitating (MemGold2 screen).

Wang and Gouaux, 2013; Wang, Penmatsa and Gouaux, 2015; Coleman, Green and Gouaux, 2016; Errasti-Murugarren *et al.*, 2019b). However, this crystallisation method requires significant amounts of protein. This limitation convinced us to initiate nanobody binder generation for *At*CAT4 which would perhaps induce stability and mediate crystal contacts and open the avenue for cryoEM single particle analysis (SPA).

3.4 Chapter Conclusions

After an exhaustive screening analysis trying a range of SLC7A4 homologues, detergent combinations, purification optimisation and an array of biophysical characterisation techniques to assess the well-foldedness of purified protein, a viable construct was selected. *At*CAT4 expressed well in *S. cerevisiae*, purified to homogeneity in very high yields reaching 2.0 mg/l culture, and was identified as stable, folded, and responsive to substrates by DSF measurements. Full-length *At*CAT4 did not produce crystals of acceptable size and thus, the option of selective molecular chaperones to aid either crystallisation attempts or render the

target amenable to cryoEM SPA was decided. A synthetic nanobody as a molecular chaperone would provide a soluble domain as an imaging fiducial to render *At*CAT4 suitable for cryoEM analysis. The generation of Syb binders against detergent-purified, lipid-stabilised CAT4 was undertaken (**chapter 4**).

Synthetic Nanobody Generation for AtCAT4

CHAPTER 4

In the previous chapter, a suitable plant SLC7A4 homologue for structural studies was identified, *At*CAT4, which however failed to crystallise. The use of a structural chaperone such as a nanobody would likely prove advantageous by ordering possible flexible regions and mediating crystal contacts, or alternatively, rendering *At*CAT4, a small membrane protein, amenable to single particle cryoEM analysis as a >75 kDa complex with a prominent soluble domain. This chapter reports the *in vitro* generation of synthetic nanobodies for *At*CAT4 and their biophysical characterisation.

4.1 Sybodies: Synthetic Nanobodies

When designing a nanobody binder generation strategy for a target, it might be advantageous to utilise more than one means of selection. Experience in the Newstead Group has shown that combinations of different Nbs/Sybs (binding discrete epitopes) with the target protein may result in different resolution apo and substrate-bound structures (Löbel *et al.*, 2022). However, for CAT4, only sybodies as structural chaperones were realised since attempts to reconstitute CAT4 into proteoliposomes were unsuccessful. LMNG has a CMC in the nanomolar range [CMC_{LMNG} ~ 11.3 nM, CMC_{DDM} ~ 130 μM, (Chae *et al.*, 2010)],

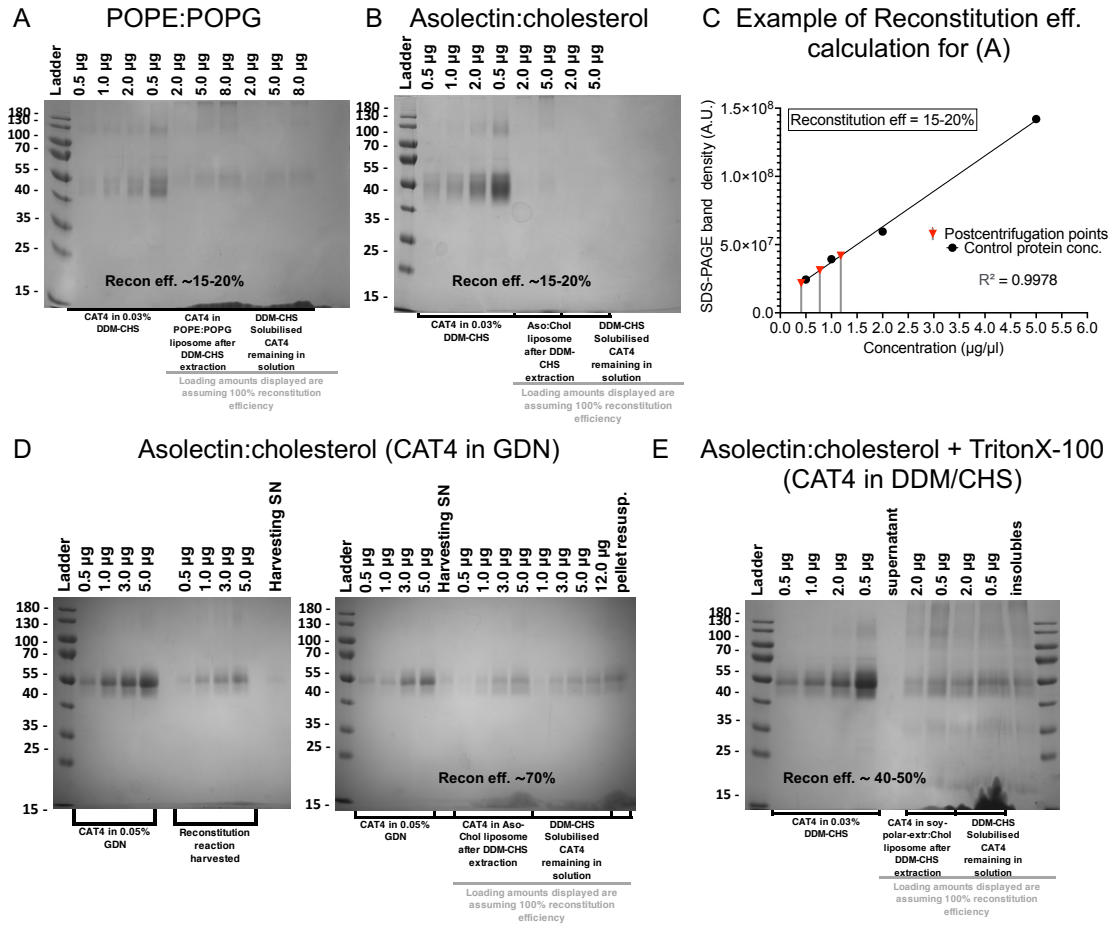


Figure 4.1 Calculation of reconstitution efficiencies of plant CAT4 in proteoliposomes from SDS-PAGE band density quantification. Each gel shows the "standard" band densities from a given amount of CAT4 in detergent (e.g., 2 µg) followed by densities produced from the same amount of CAT4 in proteoliposomes, assuming 100% efficiency. The standards are used for linear calibration.
Reconstitution of protein purified in DDM:CHS 10:1 in (A) POPE:POPG 3:1 w/w and (B) Asolectin:Cholesterol 9:1 w/w. (C) Example of standard curve for efficiency calculation. (D) Test reconstitution of CAT4 in GDN in Asolectin:Cholesterol 9:1 w/w proteoliposomes. (E) Test reconstitution of CAT4 in DDM:CHS in TritonX-100-disturbed Asolectin:Cholesterol 9:1 w/w proteoliposomes. Reconstitution efficiencies – calculated as in (C) – are shown in each gel.

exhibiting an off-rate that is 4 orders of magnitude lower than that of DDM (Chung *et al.*, 2012). Thus, for reconstitution CAT4 was required to be in a DDM-CHS micelle, a non-optimal detergent choice (chapter 3). The tested reconstitution lipids were: (a) POPE:POPG (3:1 w/w) shown to form H^+-tight AdiC proteoliposomes (Tsai and Miller, 2013), and (b) soy lipid extract, as a native-

like mimic for plant CAT4. However, the reconstitution efficiency was below 20% in both lipid selections, indicating either unstable protein in DDM-CHS or resistance to reconstitution (Figure 4.1-A,B,C). Much later in the duration of the project and after binders were selected for, it was discovered that either, GDN purified CAT4, or preparatory disruption of the lipid mix with Triton-X 100, would increase the reconstitution efficiency to $> 50\%$ (Figure 4.1-D,E). However, at the time of strategy design, this information was unavailable and CAT4 in LMNG was the sample of choice. Camelid injection of LMNG-solubilised β2-adrenergic receptor-Gs complex successfully resulted in good quality Nbs (Rasmussen, Choi, *et al.*, 2011), likely owing to the slow dissociation of this detergent (Chae *et al.*, 2010). However, it was decided that detergent solubilised CAT4 may not withstand the harsh environment of the camelid immune system.

A comprehensive outline of the Syb pipeline for CAT4 is presented in Figure 4.2. Each of the three libraries are separately passed through: ribosome display, and two rounds of phage display. Finally, an ELISA assay is performed followed by single Syb purifications which can be then separately expressed/purified and characterised for their binding to the target with biophysical methods, such as surface plasmon resonance. The use of different display systems removes binder selection bias arising from the display format. The display processes are performed with alternating target immobilisation (streptavidin T1 and C1 Dynabeads, and neutravidin coated plates) thus presenting different surface chemistries for every round to avoid selection bias and background binders. In addition to this, solution panning was carried out, so display binders were incubated with the target in solution, followed by subsequent target-pull down (Zimmermann *et al.*, 2018).

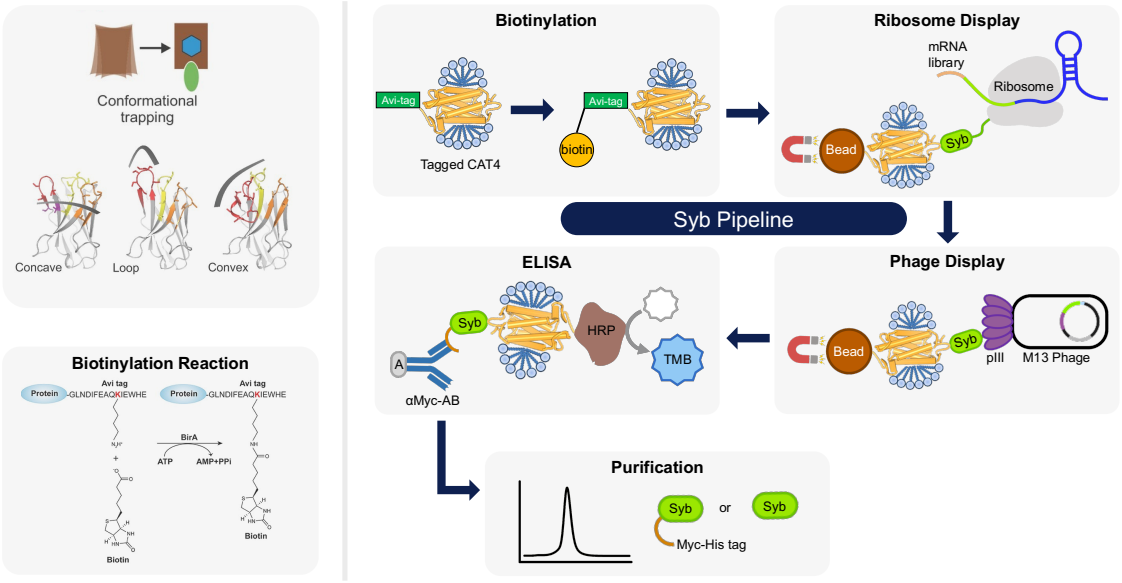


Figure 4.2 Schematic outline of Syb screening pipeline with comprehensive description. The three Syb libraries are shown and labelled with the paratope geometries which are selectively used to trap the target protein in a specific conformation. The protein must be tagged with an Avi-tag to be enzymatically biotinylated (E. coli biotin ligase BirA). Ribosome display: The tagged protein is non-reversibly immobilised on streptavidin-coupled magnetic Dynabeads (T1 variety) and the ternary complexes of the library mRNA are *on-the-fly* translated to Syb (via a stalled ribosome) and presented to the target. Non-binding Syb are washed away, and the remaining binders are eluted. The enriched binders' mRNA is reverse transcribed to a cDNA library and cloned in a phagemid vector (pDx_init) by fragment exchange (FX) cloning.
Phage display: The phagemids are cloned into M13 helper phages (propagated in *E. coli*). Each phage presents a specific Syb which is fused to pIII coat protein. The Syb-phage complexes are presented to immobilised CAT4 for 2 or more consecutive rounds of phage display, until sufficient library enrichment is reached. First via immobilisation on a neutravidin-coated MaxiSorp plate, and secondly in solution; again, to be adsorbed onto magnetised streptavidin Dynabeads (C1 variety). Non-binding Syb-phages are washed away, and binding Syb-phages are lysed and their phagemids purified. The Syb open reading frames from the purified phagemids are subcloned into a vector with an N-terminal PelB periplasmic targeting sequence and a C-terminal Myc-His_6-tag (pSb_init).
ELISA: The selected Syb are expressed, periplasmically purified and immobilised on a MaxiSorp plate coated with α-Myc antibody. Each well has one unique Syb. The biotinylated membrane protein target is presented to immobilised Syb, and upon incubation, the binding is tracked via streptavidin-conjugated HRP-chemiluminescence on TMB (3,3',5,5'-tetramethylbenzidine). ELISA wells which illuminate are identified, and the Syb expressing clone is isolated and plasmid purified. The isolated Syb-expressing plasmids are then ready for medium and large scale Syb expression and purification for downstream applications. Figure partially adapted from Zimmermann *et al.*, 2018.

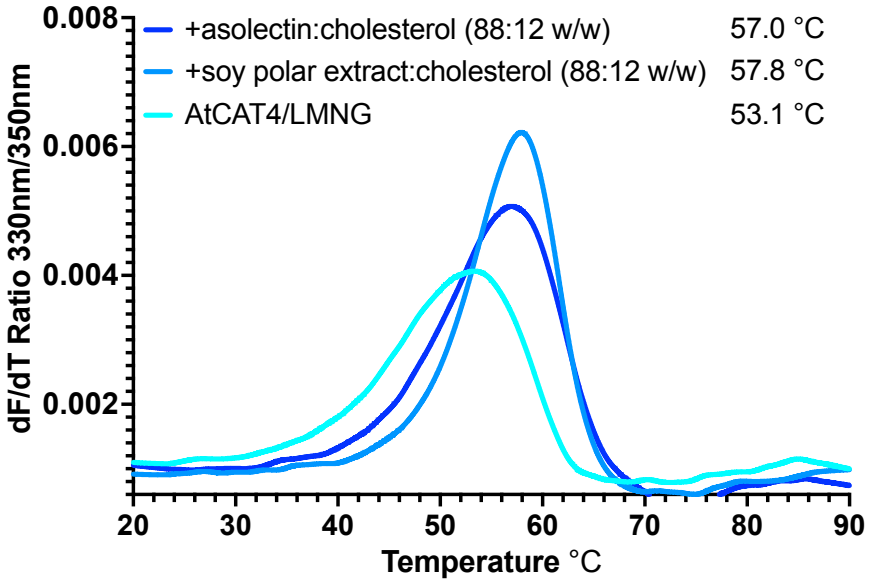


Figure 4.3 Thermostabilisation response of plant CAT4 to soy-derived lipid extracts. Asolectin (total extract) and soy polar extract. Both mixes were supplemented with 11.27 % w/w cholesterol, corresponding to 4:1 mol Asolectin:Cholesterol, taking assumed $MW_{asolectin}$ = 760 Da (O'Brien, Costa and Ott, 1977; Niu, Kim and Khorana, 2002). CAT4:lipids mixed at 1:1.2 w/w, corresponding to CAT4:asolectin:cholesterol 1:80:20.

4.2 Selection Against Lipidated CAT4

4.2.1 Sample Preparation for Sybody Selection

Sybody generation for *lipidated* CAT4/LMNG at neutral pH 7.5 was decided, with the lipids offering ΔT_m +4.0-5.0 °C during selection (**Figure 4.3**) and increasing melting cooperativity (**chapter 3**). Asolectin:cholesterol and soy polar extract:cholesterol offered comparable thermostabilisation, and since the former is cost-effective it was used during selection. CAT4 was preincubated with LMNG-solubilised asolectin-cholesterol (4:1 mol) mixture prior to all selection steps. Previous experience in the group demonstrates that a less stabilising detergent, e.g., DDM, would expose slightly unfolded regions to the Syb during panning, which subsequently contaminate and dominate the epitope selection process.

4.2.1.1. Generation of CAT4-Avi for Enzymatic Biotinylation

Avi-tag Cloning

To conduct selection with the different display methods, CAT4 must be covalently biotinylated, so that it can be immobilised for panning via biotin-streptavidin interactions. This is possible by enzymatical biotinylation of an accessible Lys contained in an Avi-tag (N-Gly-Leu-Asn-Asp-Ile-Phe-Glu-Ala-Gln-**Lys**-Ile-Glu-Trp-His-Glu-C). In the Newstead Group, C-terminal cloning of the Avi-tag to the *cat4* DNA sequence is produced by PCR with 5' and 3' homologous recombination primers compatible with SmaI-linearised pDDGFP2-Leu2D vector. The process begun with the successful production of a single PCR amplicon containing a portion of the Avi-tag, templated from WT CAT4/pLeu2D. The product was gel extracted and used as a template for a second PCR with HR primers containing the remaining Avi-tag sequence. A single amplicon product was detected by agarose DNA electrophoresis (Figure 4.4-A). Following, the amplicon was mixed with linearised pLeu2D processed according to the schematic in Figure 4.4-B. Before sequencing the cloned plasmids, a restriction digestion with BamHI/XhoI were run to excise the expected CAT4-Avi-GFP dsDNA (2.6 kb) and visualise the band with DNA electrophoresis. The resulting digests did not produce the desires 2.6 kb band, rather most single colonies contained an unknown 0.8 kb insert while one contained a 1.0 kb insert (Figure 4.4-C). This is in contrast to the control digestion with WT-CAT4/pLeu2D which excises the CAT4-GFP insert at 2.6 kb correctly (Figure 4.4-C). Repetition of the whole HR process using the same PCR amplicons resulted in the same digestion products (Figure 4.4-D), suggesting that cloning failure is reproducible.

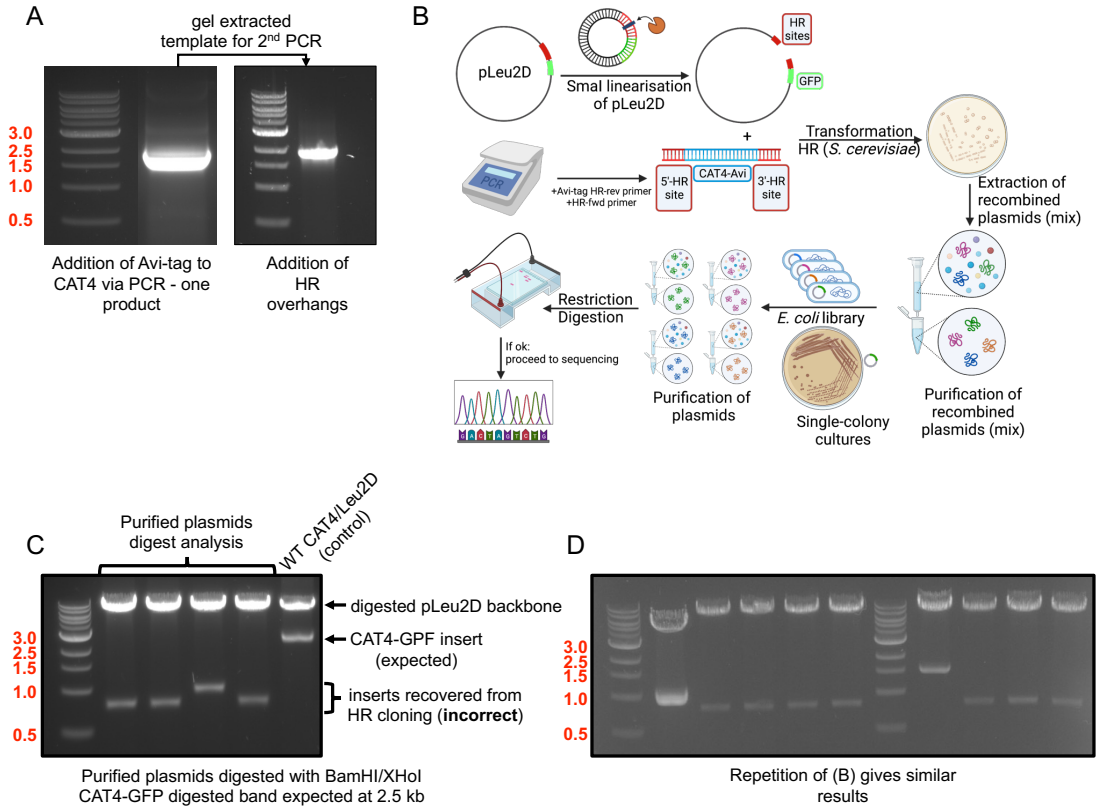


Figure 4.4 Overview and troubleshooting of HR cloning CAT4-Avi into pLeu2D. (A) Amplification of WT-CAT4 template with HR primers containing the Avi-tag sequence (B) Comprehensive diagram of HR cloning steps from PCR amplification to sequencing. (C) After processing outlined in (B), the resultant restriction digests are run on 1% agarose gel and imaged with SYBR Safe dye. Insert bands should be at 2.6 kb (CAT4-Avi-GFP) but are $<$ 1.0 kb from multiple colony digestions. (D) Restriction digests from the repetition of the whole HR-cloning process shown in (B). Failure of cloning is reproducible.

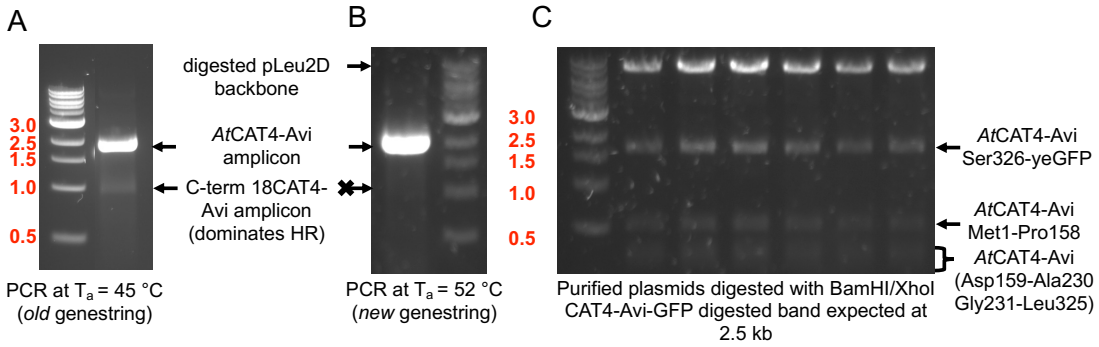


Figure 4.5 Comparison of PCR amplicons from *old* and *new* *At*CAT4 genestrings used for this project. (A) PCR amplification at low annealing temperature T_a = 45 °C uncovers the non-specific 0.8 kb product of the last eighteen avi-tagged CAT4 residues. (B) PCR amplification with HR primers of *new* genestring generates a single product that homologously recombined with SmaI-linearised pLeu2D. (C) Analytical BamHI/XhoI restriction digests of the purified plasmids from the HR pipeline report the expected bands for the *new* genestring. Cloning success of plasmids was determined by sanger sequencing.

An extensive troubleshooting of the *At*CAT4-avi assembly with different cloning strategies was performed, presented in the appendices (**Appendix A – AtCAT4-Avi Cloning Troubleshooting**). The investigation uncovered that the genestring used to express WT *At*CAT4 had a sequence section with >85% sequence identity with a PCR primer resulting in production and persistent carry-over of a non-specific PCR by-product. Due to its smaller size, the by-product cloned more efficiently than the full-length *At*CAT4-Avi product, hindering downstream assembly with pLeu2D and explaining the 0.8-1.0 kb digested bands in **Figure 4.4-C,D**. We subsequently designed and received a yeast-codon-optimised *At*CAT4 genestring lacking sequences homologous to the PCR primers. The desirable PCR amplicon without contaminating by-products was generated (**Figure 4.5-B**) and the HR processing pipeline resulted in the desirable restriction digests (**Figure 4.5-C**) which were sequence verified.

Expression and Purification of CAT4-Avi

The avi-tagged construct was expressed according to the WT protocol. This resulted in 0.62 mg CAT4-avi per l culture upon purification, lower than WT (2.0 mg/l) due to the presence of the Avi-tag. The purification gel reported protein purity ~ 82.2% prior to the polishing gel-filtration step, while the gel filtration trace showed a monodisperse main peak (**Figure 4.6**). The main peak (fractions A10-A12) was pooled whose purity was sufficient for biotinylation (**Figure 4.6-C**).

4.2.1.2. Biotinylation and Streptavidin-Shift Assay

The purified protein was overnight incubated with excess biotin and recombinant BirA; a biotin ligase catalysing biotinylation of a singular lysine residue

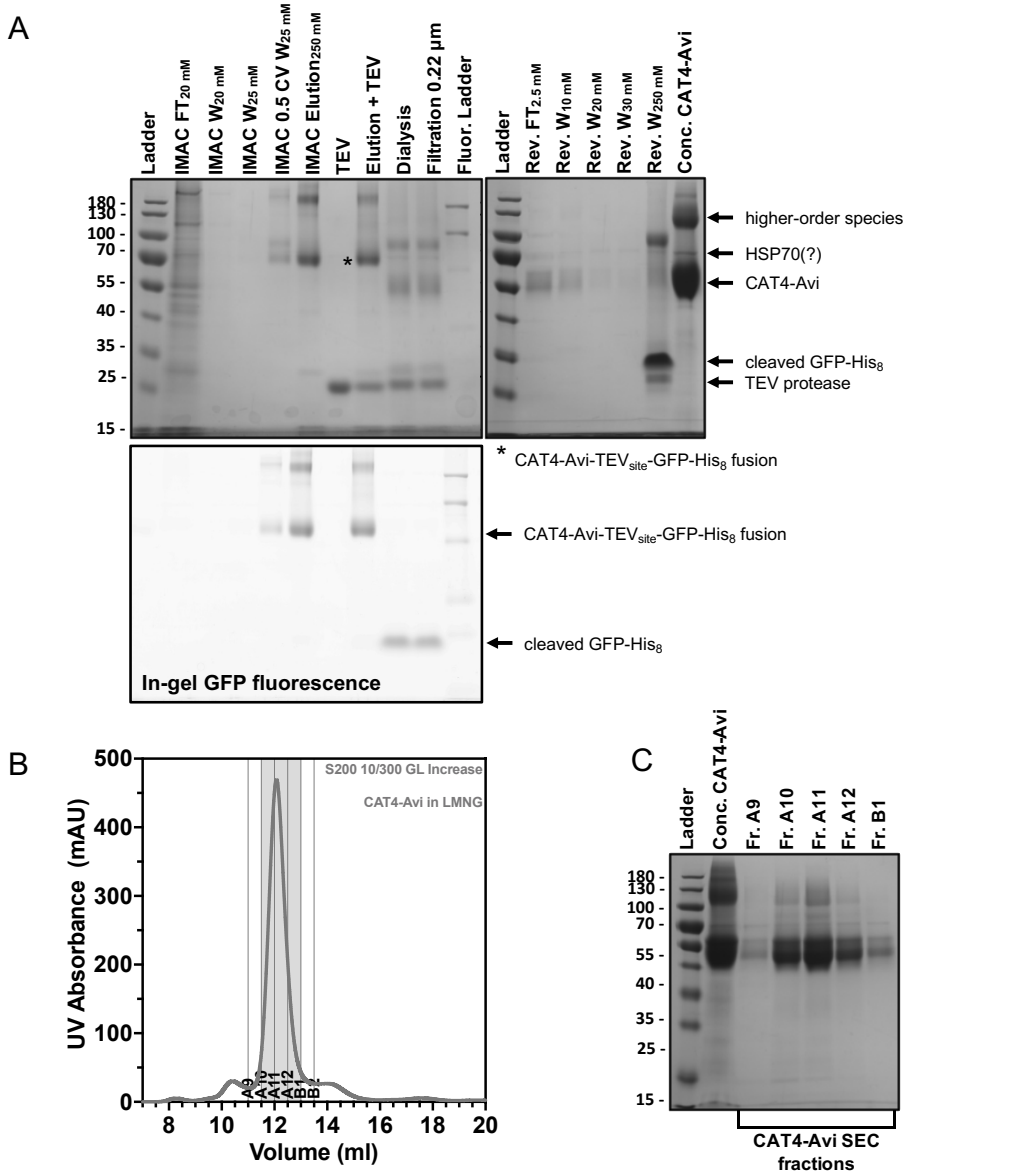


Figure 4.6 Purification of CAT4-Avi in identical manner to WT CAT4. (A) 12% acrylamide SDS-PAGE analysis of purification steps from IMAC up to gel-filtration injection. Both coomasie-blue and in-gel GFP fluorescence imaging is shown. (B) Representative gel-filtration trace of CAT4-Avi in 0.003% LMNG (1 of 4 runs). Fractions pooled are highlighted in grey.

introduced in the avi-tag in an ATP-dependent manner. The reaction is setup with excess biotin and in low ionic strength (70 mM NaCl), as higher salt concentrations inhibit the progress. CAT4 species (biotinylated and non-biotinylated) were separated from excess biotin and BirA via gel-filtration (**Figure 4.7**)

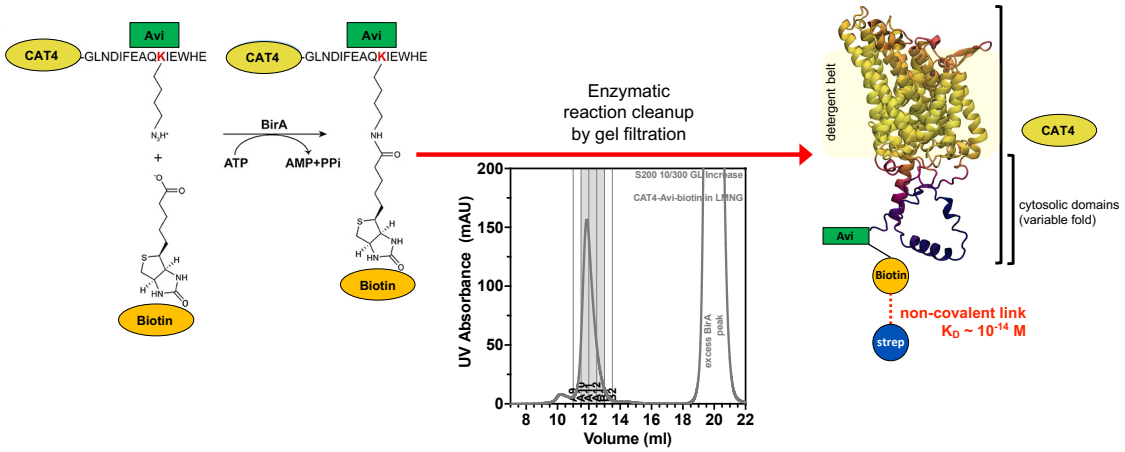


Figure 4.7 Enzymatic biotinylation reaction and cleanup of avi-tagged CAT4. A singular Lys present in the avi-tag sequence (highlighed in red) is selectively derivatised by BirA with concomitant consumption of ATP. The reaction mixture is subsequently separated into its components: (a) CAT4 species (both biotinylated and non-biotinylated CAT4-Avi, grey-highlighted fractions) (b) BirA and (c) biotin, by size exclusion chromatography. The resulting CAT4-Avi-biotin can be linked to streptavidin-coated surfaces.

and the main peak species were analysed to check biotinylation success.

The reaction product was assessed by a streptavidin-shift assay. In this assay the reaction product, consisting of an unknown composition of biotinylated and non-biotinylated protein, is saturated with molar excess streptavidin before SDS-PAGE analysis. The biotinylated species will non-covalently link to the supplemented streptavidin by exploiting the femtomolar affinity [$K_d \sim 4\times10^{-14}$ M, (Michael Green, 1990)]. In this case the avidin will retard CAT4-Avi electrophoretic migration compared to non-biotinylated species, evident by an apparent "shift" in SDS-PAGE (**Figure 4.8**). Semiquantitation of the shift assay bands showed that $>$ 90% of CAT4 species are biotinylated. This population is required for Syb selection (Zimmermann *et al.*, 2020a). In line with the thermal shifts of WT CAT4 (**Figure 4.3**), both CAT4-avi and CAT4-avi-biotinyl show a ΔT_m of +7.8 °C upon relipidation demonstrating that the protein responds to exogenous lipid, similar to WT. No additional increase in ΔT_m with L-histidine was recorded.

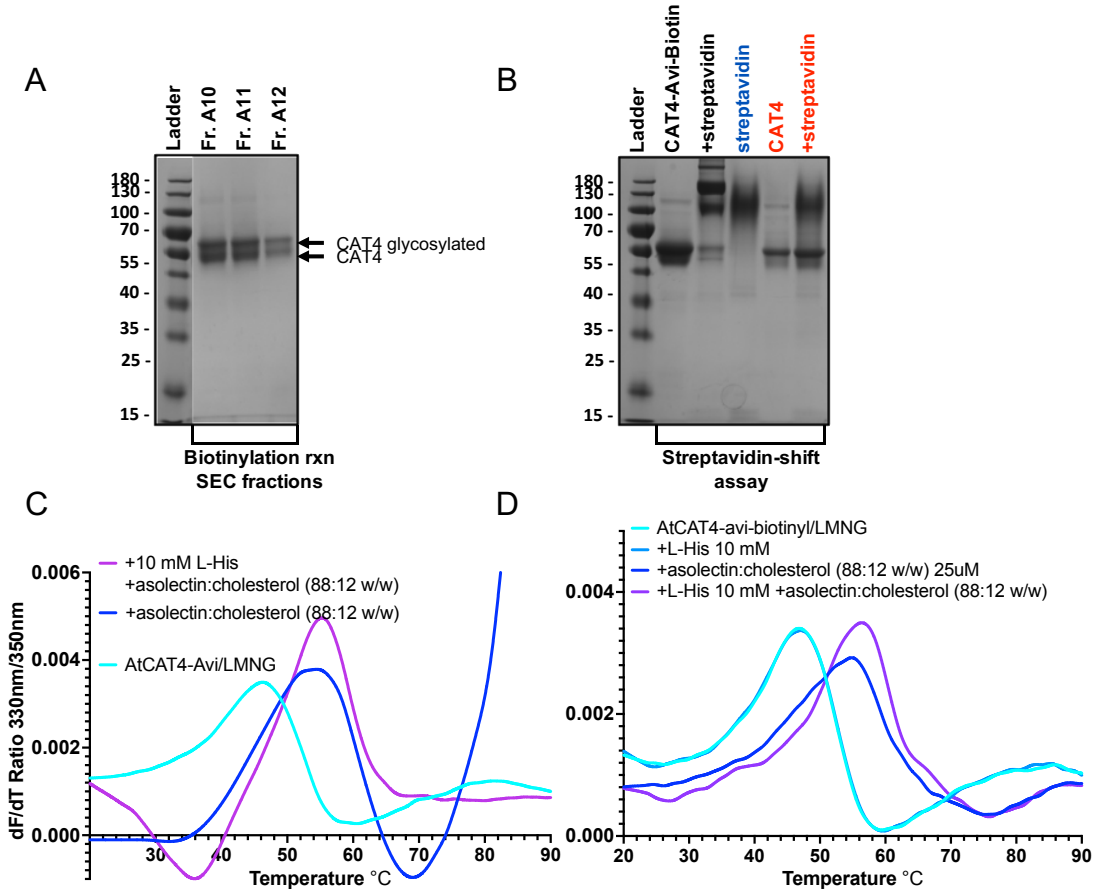


Figure 4.8 Analysis of biotinylated CAT4. (A) Gel-filtration fractions from biotinylation reaction cleanup. (B) Streptavidin-shift assay showing shifting of >90% of reaction product after addition of streptavidin whereas no shift occurs when WT CAT4 is supplemented with excess streptavidin. DSF thermal melt scans (C) of CAT4-avi and (D) of CAT4-avi-biotinyl.

Collectively, no unexpected deviations of biotinylated CAT4 biophysical behaviour from the WT protein were recorded.

4.2.2 Ribosome Display

The three sybody libraries are formulated as mRNA with 5' and 3' stem loops for stabilisation. The libraries were *in vitro* translated by reconstituted ribosomes and the translation was stalled after the sybody CDS by a TolA-based spacer, an unstructured domain used in ribosome display which allows the nascent Syb to

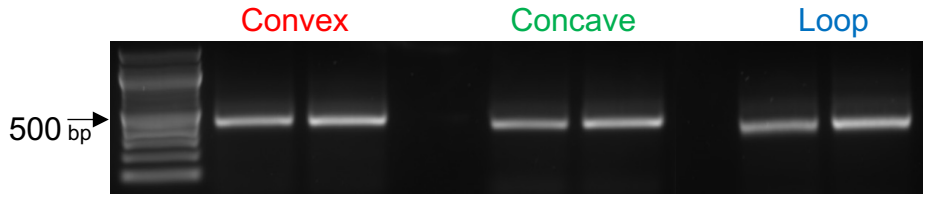


Figure 4.9 cDNA libraries (convex/red, concave/green, loop/blue) produced by reverse-transcription PCR run on 1% agarose gel prior to gel extraction and PCR cleanup.

project from the exit tunnel and fold outside of the ribosome (Dreier and Plückthun, 2011). After forming a ternary Syb-ribosome-mRNA complex, the complexes, braced by added magnesium, were panned against lipidated CAT4-biotin before magnetically pulling-down on streptavidin-coated beads. After washing out weakly bound mRNAs, the remaining mRNAs were eluted with EDTA and reverse transcribed into a cDNA library by RT-PCR. The cDNA enrichment levels for each library were comparable by visual inspection of agarose gel showing enrichment for all Syb libraries (**Figure 4.9**). The amplified cDNA library was cloned into the pDX_init vector for phage display.

4.2.3 Phage Display

Electrocompetent *E. coli* cells were transformed, phages produced overnight. and precipitated from the culture supernatant. Purified phages were incubated with lipidated CAT4 and the complexes immobilised on a neutravidin-coated maxi-sorp plate. After washing out weakly bound phages, the remaining phages were eluted and used to infect a second batch of cells for the second round of phage display, identical to the first in all aspects except that streptavidin-coated magnetic beads were used instead. The second round was performed in the presence of 100-fold molar excess of non-biotinylated lipidated CAT4 aimed to bias recovery towards slower off-rate and select for the tightest binders. During each

	Phage Display Library Enrichment	
Syb Library	First Round.	Second Round
Concave	4×	20×
Convex	1×	1×
Loop	4×	1×

Table 4.1 Convex/concave/loop library enrichments for CAT4 over off-target PepTSo (control) after each round of phage display. The enrichment was calculated as the ratio of phage eluted by CAT4 panning to the phage eluted by $PepT_{So}$ panning; measured by qPCR quantification of pDx_init phagemids. Enrichment factors are rounded to the nearest integer.

phage display round, each library was also screened against biotinylated negative control protein, $PepT_{So}$, a bacterial peptide transporter. This allowed the monitoring of the excess phage numbers eluted in a round of display, quantified by qPCR, and comparing the phage enrichment factors for the target versus the control. The enrichment of the convex and loop libraries was poor after the second round of phage display (Table 4.1). Only the concave library showed significant CAT4-specific enrichment by qPCR, and thus was forwarded to the ELISA step. As general guidance from Zimmermann *et al.*, enrichment factors higher than five are advised to progress to the next screening stage, a cutoff surpassed only by the concave library.

4.2.4 ELISA Screening

The concave library phagemid output was subcloned into an ELISA vector, pSb_init, i.e., a vector with an N-terminal periplasm-targeting PelB sequence and a C-terminal Myc-His_6 tag. The Myc-tag is used for ELISA immobilisation of the binder, in contrast to streptavidin-based target-protein immobilisation done in the previous panning steps. 95 single clones were selected, propagated, and expressed for expression and then subsequently purified via osmolytic

periplasmic extraction (at high $MgCl_2$ and high ionic strength). Syb-laden periplasmic extracts were individually immobilised, each to a well, by Myc-tag affinity against ELISA-plate wells coated with α-Myc antibody. Subsequently, Syb-laden extracts were panned against lipidated CAT4 and negative control $PepT_{So}$.

After three rounds of plate washing the low affinity binders are expected to be washed away while the strong interactors remain bound to target CAT4. With the addition of the chromogenic 3,3',5,5'-tetramethylbenzidine (known as, TMB), the wells retaining immobilised CAT4-Syb complexes chemilluminate in a CAT4-concentration-dependent manner. Direct chromatometric comparisons were made on the same plate with the $PepT_{So}$ wells to identify Syb specific for CAT4 and not $PepT_{So}$ (**Figure 4.10-A,B,D**).

The illuminescence signal strength is not necessarily a good predictor of Syb affinity to target and therefore any well displaying $>1.5\times$ more signal than the equivalent control $PepT_{So}$ well is identified as a "hit" (Zimmermann *et al.*, 2020a). To test these cutoffs distribution analysis of the readouts was performed (**Figure 4.10-C**). The $PepT_{So}$ signal (σ) distribution analysis shows that $1.5\times\sigma_{PepTSo,mean} \sim 0.09$, while the maximum value was ~ 0.08. On the other hand, more than 50% of the binders had a signal >0.10 in the CAT4 wells, which is over the 0.09 cutoff, justifying the $1.5\times$ specificity criterium. Additionally, the probability is bimodal for CAT4 wells with a second peak at $\sigma_{CAT4} \sim 0.20$. These binders represent the top 25% of the total binders with CAT4-specificity. To

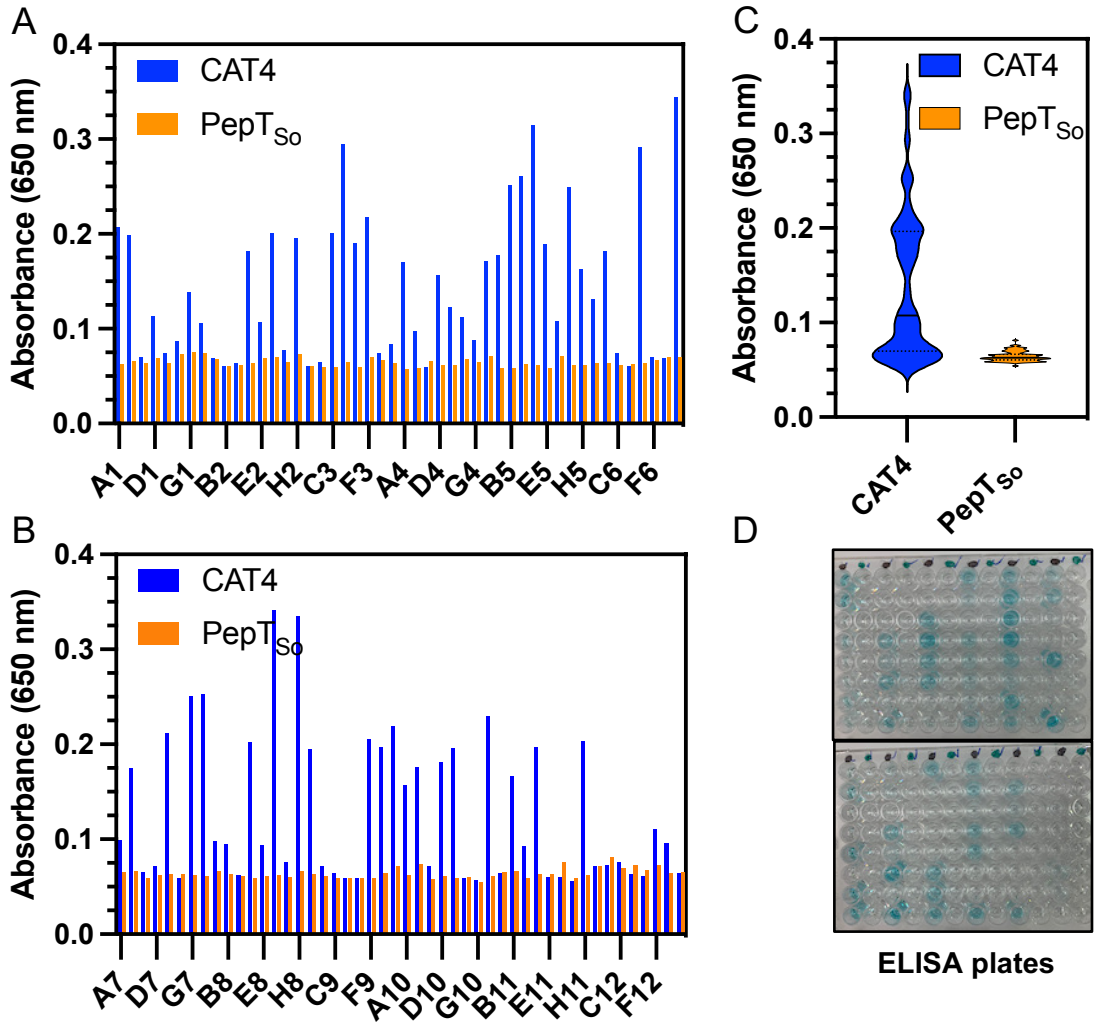


Figure 4.10 (A,B) ELISA of the concave library Syb against CAT4 (blue) and negative control PepTSo (orange). Absolute chromometric readouts of the ELISA wells (A1-G12, 96 wells) at A_{650} from the two ELISA plates. (C) ELISA readout distribution analysis in a violin-style plot showing the three percentiles of the distribution as dashed lines. (D) Photograph of the ELISA plates after 30 min incubation with TMB.

capture these binders for downstream analysis an enrichment factor >3.0 is required. Therefore, while the >1.5× criterium recommended by Zimmermann et al., (2018) works for CAT4, it is likely that the specific binders will be those with >3.0× CAT4/$PepT_{So}$ signal. Overall, the concave library produced 44 Syb hits in ELISA out of the 95 theoretically possible (~ 45% success rate). It is common that "different" hits are degenerate binders. Thus, these hits were sequence-verified to establish uniqueness.

4.2.5 Sybody Sequence Analysis

Sequencing of 40 ELISA hits in pSb_init was undertaken and 37/40 were successfully sequenced for their full ORF. Surprisingly the verification resulted in just 5 of the 37 binders were unique. The unique hits were chosen to be named according to ELISA well coordinates – i.e., Syb H7, D8, E7, A6 and B5. The degeneracy degree for SybE7 was the highest with 16 sybodies carrying the same sequence, whereas the most unique sybody analysed was Syb H7 with a degeneracy degree of 2 (Figure 4.11-A).

The degeneracy of the isolated sybodies is remarkably high considering that the concave library theoretically is reported to have a diversity set of 8.3×10^{17} binders (Zimmermann *et al.*, 2018). The sequence stringency of the concave binders to CAT4 suggests that there is an immunodominant CAT4 epitope that the concave library has an unequivocal preference for. Considering the topology map of CATs (chapter 1) this is not very surprising given that 49.6% of CAT4 residues are within the 14 TM helices sequence. The remaining 50.4%, including termini, was not expected to show globular domain features which would span more than a single polypeptide stretch to render epitope combinations possible. Short extracellular loops are predicted, EL1, 2 and 3, with EL3 being equivalent to EL3 in mouse SLC7A1 (*Mm*CAT1; residues 213-239), the established epitope for murine ecotropic leukaemia viruses (Verrey, Ellen I Closs, *et al.*, 2004) with two glycosylation sites. Heavy glycosylation may hinder epitope presentability during selection. A perceivable exception could be the intracellular loop IL5 (404-462).

The 5 unique binders possess variability in all CDR regions suggesting that a

systematic experimental bias during selection is unlikely to have occurred (Figure 4.11-B). Aligned CDR3 sequences showed the largest agreement, whereas CDR1 showed the lowest. In CDR1, SybB5 is 50%/75% identical/similar to SybH7 but only 25% identical to SybD8. SybD8 is only 25% identical to SybH7. SybE7 and SybA6 are also 25% identical in their CDR1s. Overall CDR1 sequence diversity is large. In CDR2 all Syb are only 50% pairwise identical whereas similarity varies from 50% in most cases. Only the SybA6-SybD8 pair was 75% similar in CDR2. SybA6-SybD8 are evolutionary related in the dendrogram (Figure 4.11-A) and may behave bind CAT4 with similar affinity. In CDR3, sequence identities are high given the diversity expected in this region. All Syb pairs (excluding SybB5) are 75-88% identical, with SybA6-SybE7 being 100% similar. SybB5 is the most diverse – it is 62% identical to D8 and E7, while only 50% to the remaining. Among binders, SybB5 is the least sequence-related to others within the lot. Overall, binder diversity is clustered within the CDRs (Figure 4.11-A).

A typical Nb will use all three CDRs, however, structural metanalyses show CDR3's involvement in affinity regulation of a binder (Mitchell and Colwell, 2018). With the exception of SybB5, Sybs selected against CAT4 seem to conserve traits in their CDR3, showing specific requirements to confer CAT4 binding whereas other CDRs are free to explore more sequence space. The residues in CDR1 have a hydrophilic tendency (Figure 4.11-B), as do the ones in CDR2, albeit less pronounced. V98 and V100 in CDR3 are hydrophobic residues present in all the binders as a result of them being constant residues of the concave template (Zimmermann *et al.*, 2018). The remaining CDR3 residues are not hydrophobic.

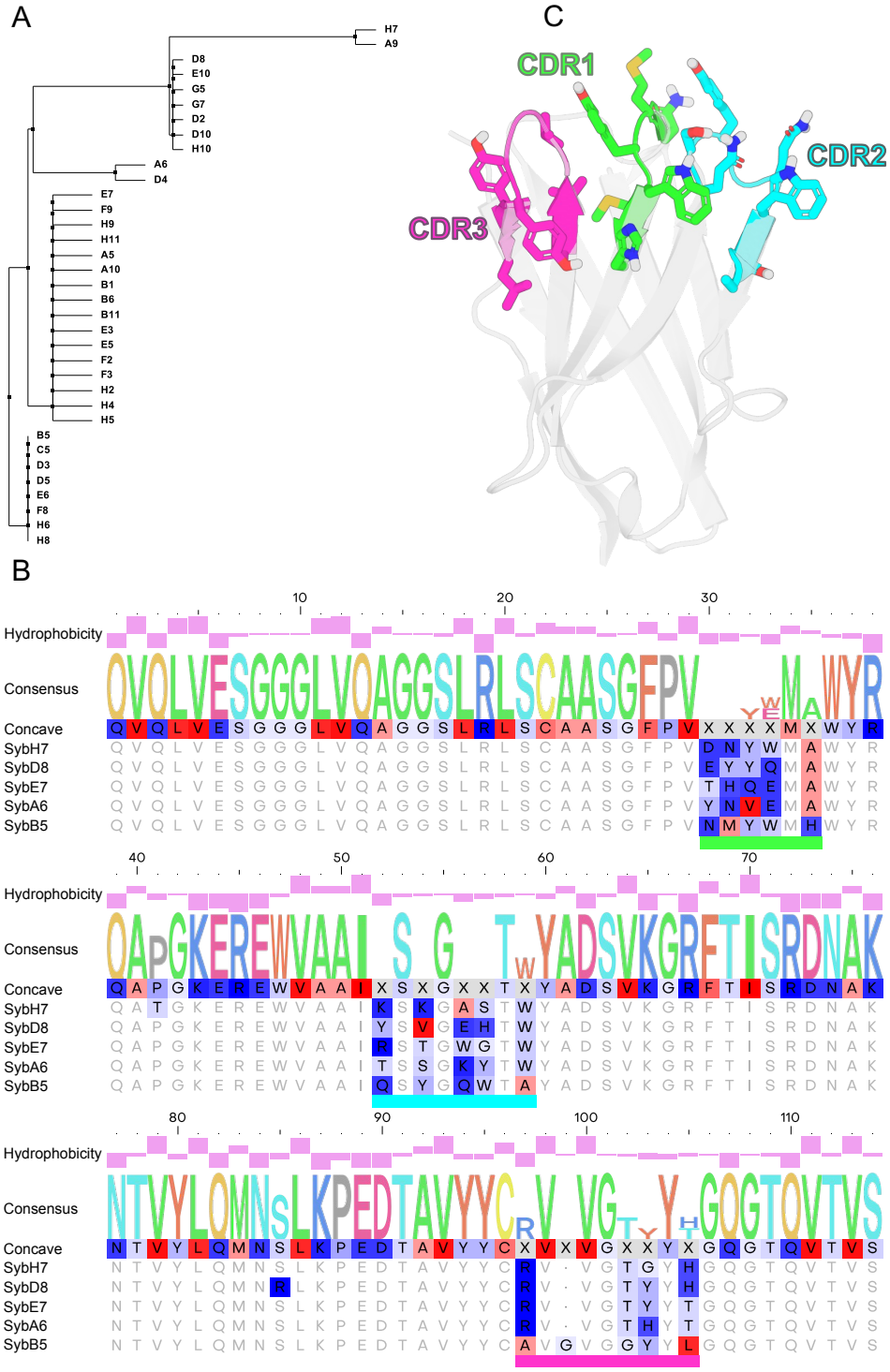


Figure 4.11 Concave Syb unique hits sequences and degeneracy degree. (A) Clustering of sequence-degenerate ELISA hits in the phylogenetic scheme. (B) Multiple sequence alignment of Syb sequences from each cluster (concave template as reference). Reference residues are coloured with the Kite-Doolittle hydrophobicity scale. Residues deviating from the consensus are also coloured. CDR1 – green; CDR2 – cyan, CDR3 – magenta. (C) Concave Syb model (SybB5; Alphafold2) with CDR residues in sticks.

Overall this observation is a good predictor against hydrophobicity, a feature

desirable as sybodies may be hydrophobically "sticky" (Zimmermann *et al.*, 2020a).

4.2.6 Screening Sybodies for Purification and Biochemical Stability

With the unique number of binders being limited to five, all binders were carried forward for affinity characterisation. The 5 unique binders (Syb B5,D8,E7,H7 and A6) were subcloned into the tag-free expression vector, pBXNPHM3, which is an arabinose-inducible vector with a leading PelB sequence similar to pSb_init but exchanges the C-terminal Myc-His_6 tag for an N-terminal His_{10}-MBP fusion that is 3C-protease cleavable. This vector provides the advantage that it is amenable to crystallisation since it lacks flexible tags (i.e., Myc-tag). The MBP-tag assists in stable and comparable expression of binders.

During the purification process all 5 binders proved problematic when expressed in pBXNPHM3. Specifically, cleavage of the His_{10}-MBP-tag was incomplete at 1:1 Syb:3C molar ratio (Figure 4.12-A; example SybD8). Even the fraction of SybD8 that cleaved would stick to Ni resin in the reverse Ni-affinity step. Alas, an additional reverse amylose affinity step was included after subjecting the SybD8 to threefold excess of 3C during overnight dialysis and before the reverse Ni-affinity step (Figure 4.12-B,C). 3C still did not cleave to completion and SybD8 displayed non-specific interactions with the amylose resin, similar to Ni. A fraction of binder managed to elute the amylose resin, escorted by contaminants (Figure 4.12-C). To remove persistent contaminants, either size-exclusion chromatography or an additional reverse Ni IMAC step with high ionic strength washes (up to 1 M NaCl; reduces stickiness and aggregation) were carried out. The latter

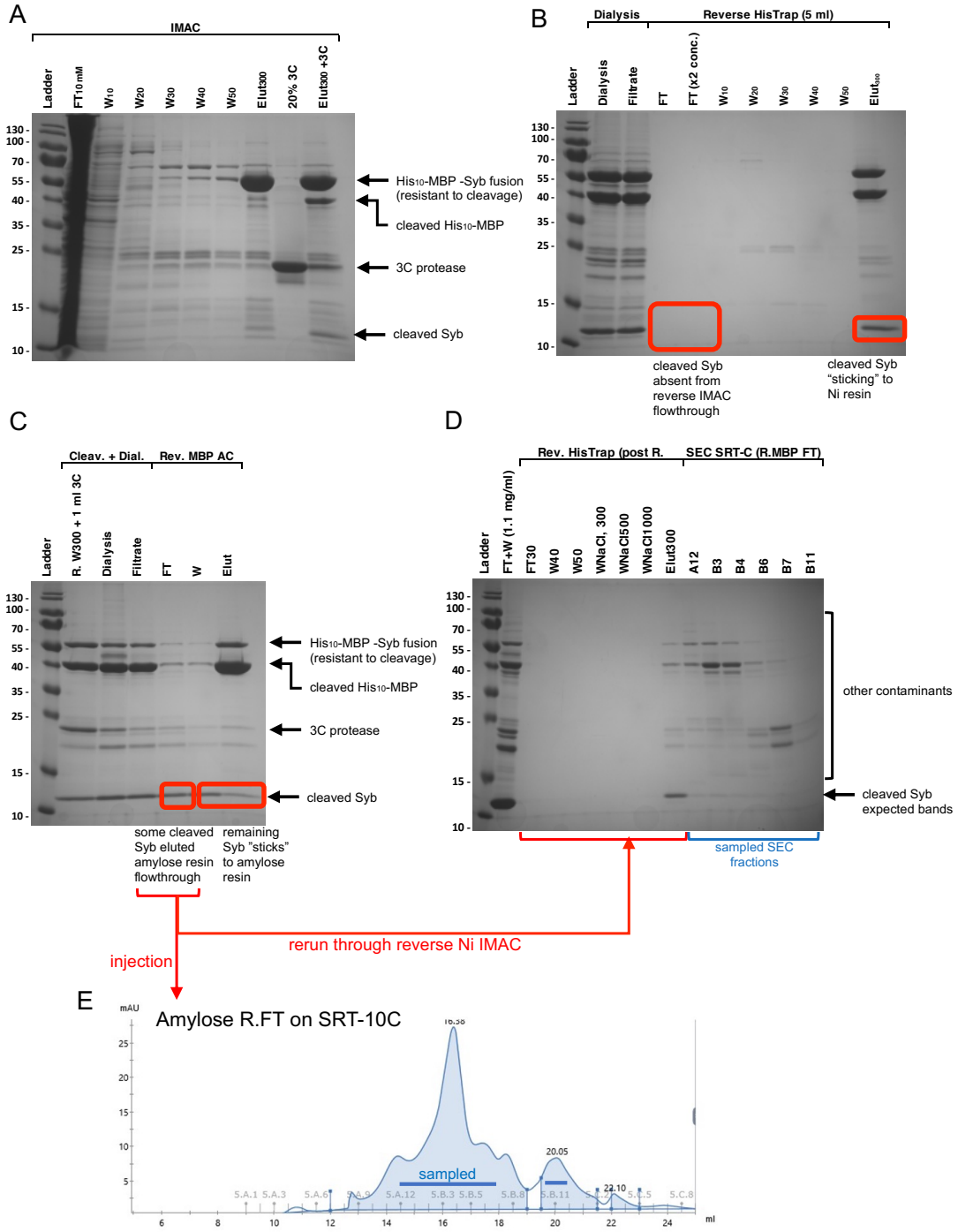


Figure 4.12 Exemplary purification attempt of a SybD8 as His_{10}-MBP fusion.

showed that SybD8 displayed aberrant stickiness to Ni resin even at high ionic strengths, while in size-exclusion, the Sybs spread over multiple fractions escorting persistent contaminants (**Figure 4.12-D,E**).

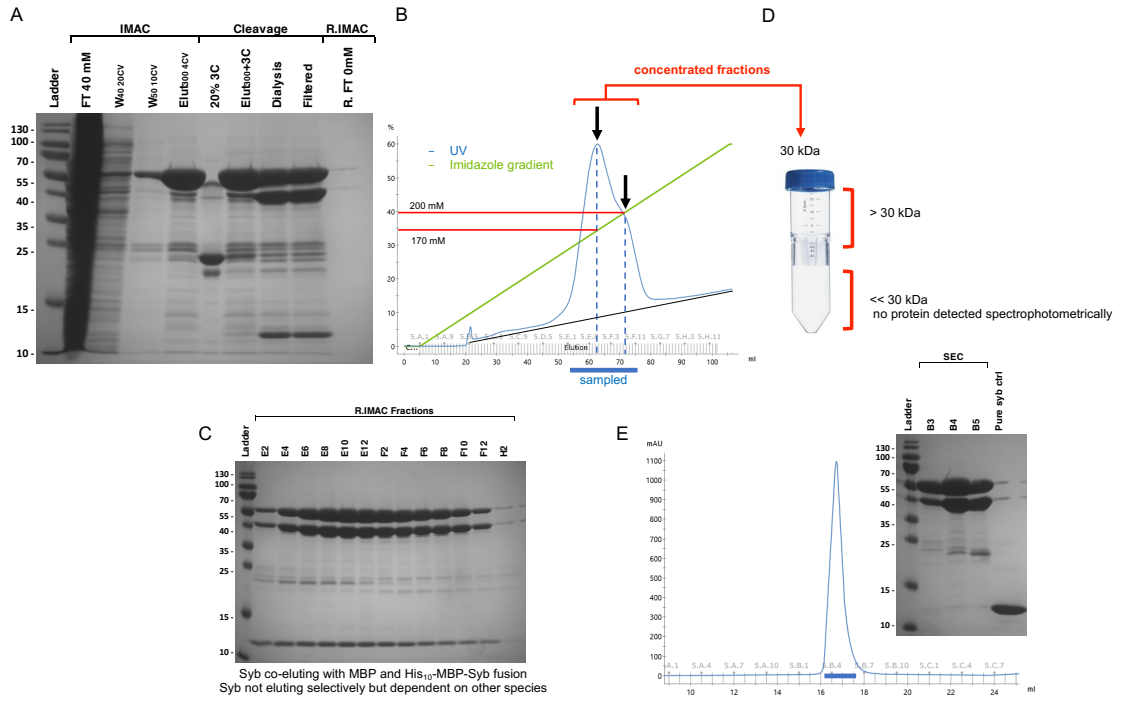


Figure 4.13 Further troubleshooting with SybB5. (A) Ni-base IMAC, 3C cleavage/dialysis and reverse Ni IMAC, not showing eluant Syb in reverse flowthrough (R. FT). (B) UV chromatogram after subjection of sample-loaded Ni column to a linear imidazole gradient. (C) SDS-PAGE analysis of chromatogram fractions showing cleaved Syb co-eluting with cleaved MBP and protease-resistance His_{10}-MBP-Syb fusion. (D) Concentrator-based separation with a 30 MWCO concentrator failed to separate cleaved Syb based on molecular weight. (E) UV chromatogram after injection of concentrated mixed species (Cleaved Syb + cleaved MBP + protease-resistant His_{10}-MBP-Syb fusion) to an SRT-10C SEC300 column showed a single peak containing cleaved MBP and protease-resistance His_{10}-MBP-Syb fusion. Cleaved Syb did not elute the column. A Syb available in the lab was run as positive control in SDS-PAGE.

SybB5 did not elute the reverse Ni-affinity step after protease cleavage (even at 0 mM imidazole) (**Figure 4.13-A**), similar to SybD8. We wondered whether the cleaved SybB5 would separate from cleaved His_{10}-MBP and protease-resistant uncleaved His_{10}-MBP-SybB5 fusion, at different imidazole concentrations. A continuous linear imidazole gradient was used. However cleaved SybB5 coeluted along with cleaved His_{10}-MBP and His_{10}-MBP-SybB5 fusion in all fractions (**Figure 4.13-B,C**) as did SybD8 in the step-wise separation protocol (**Figure 4.12-B**). Separation of the mixture with a hard molecular weight cutoff (**Figure 4.13-D**) or

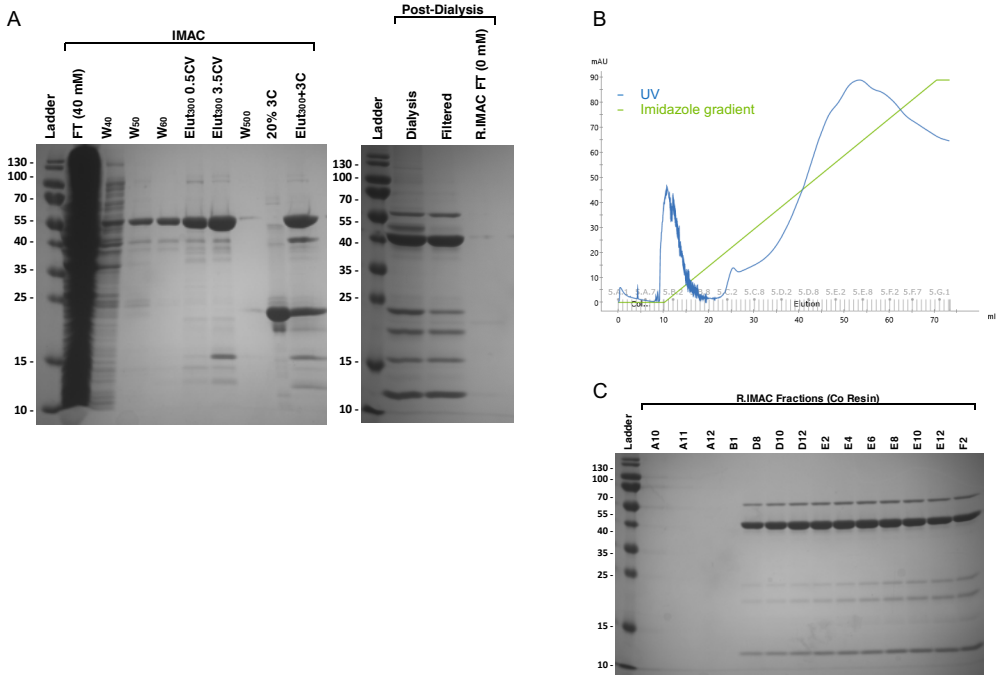


Figure 4.14 Further troubleshooting with SybH7. (A,B) Ni-base IMAC, 3C cleavage/dialysis and reverse Ni IMAC, not showing eluant Syb in reverse flowthrough (R. FT). (C) UV chromatogram after subjection of sample-loaded Co column to a linear imidazole gradient. (D) SDS-PAGE analysis of chromatogram fractions showing cleaved Syb co-eluting with cleaved MBP and protease-resistance His_{10}-MBP-Syb fusion.

gel-filtration were unsuccessful (**Figure 4.13-E**). The latter did not elute cleaved SybB5 from the column matrix; suggesting a case of a sticky Syb binder, known to be problematic (Zimmermann *et al.*, 2020a). Co-affinity purification with a linear gradient was also attempted with SybH7 but resulted in identical co-elution profiles of cleaved SybH7 with cleaved His_{10}-MBP and His_{10}-MBP-SybH7 fusion (**Figure 4.14**).

The multistep troubleshooting of different Syb binders was unsuccessful to provide pure binders. The above observations hint towards problematic and misfolded binder expression in the host *E.Coli* (MC1061) with the binders remaining in solution owing to His_{10}-MBP's solubility-enhancing properties (Sun, Tropea

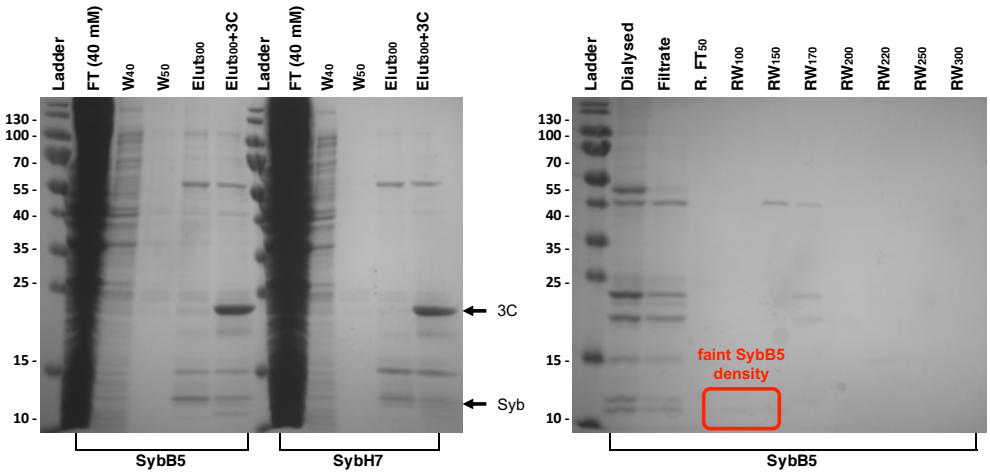


Figure 4.15 Troubleshooting purification of SybB5 and SybH7 without the MBP, as cleavable Syb-His10 fusions. Compared to the strong band densities shown previously for His$_{10}$-MBP-Syb fusions, the His$_{10}$-Syb fusions at 300 mM imidazole elution have weak densities, comparable to the contaminants (left). Upon overnight 3C processing the solution showed white precipitation. This was attributed to precipitating Syb, as shown for SybB5 (right gel) if one compares band density at dialysis step and after precipitant particle filtration. The resulting cleaved Syb remaining in solution eluted the reverse Ni IMAC column at high imidazole concentrations (100 mM) as shown by the faint density which was however of very low yield.

and Waugh, 2011; Salema and Fernández, 2013), explaining the problems arising upon cleavage. To troubleshoot this SybB5 and H7 were cloned into an identical vector lacking the MBP-tag, pBXPC3H. A marked reduction in SDS band density was seen in the Ni IMAC elutions and after overnight protease processing the binders precipitated, most likely due from solubility decrease after MBP removal (**Figure 4.15**). Therefore, the Sybs selected for CAT4 require a solubility-enhancing tag to express well and upon cleavage they present features of misfolded proteins. Misfolded Nbs are common and their purification is challenging with the binders trapped in purification-resistant inclusion-bodies (Xu, Song and Jia, 2017; Ban, Sharma and Shetty, 2020). A strategy that circumvents this problem is periplasmic extraction which takes advantage of the selective periplasm trafficking of only the well-folded binder populations, which can subsequently be hyperosmotically extracted without cell lysis, avoiding liberation of

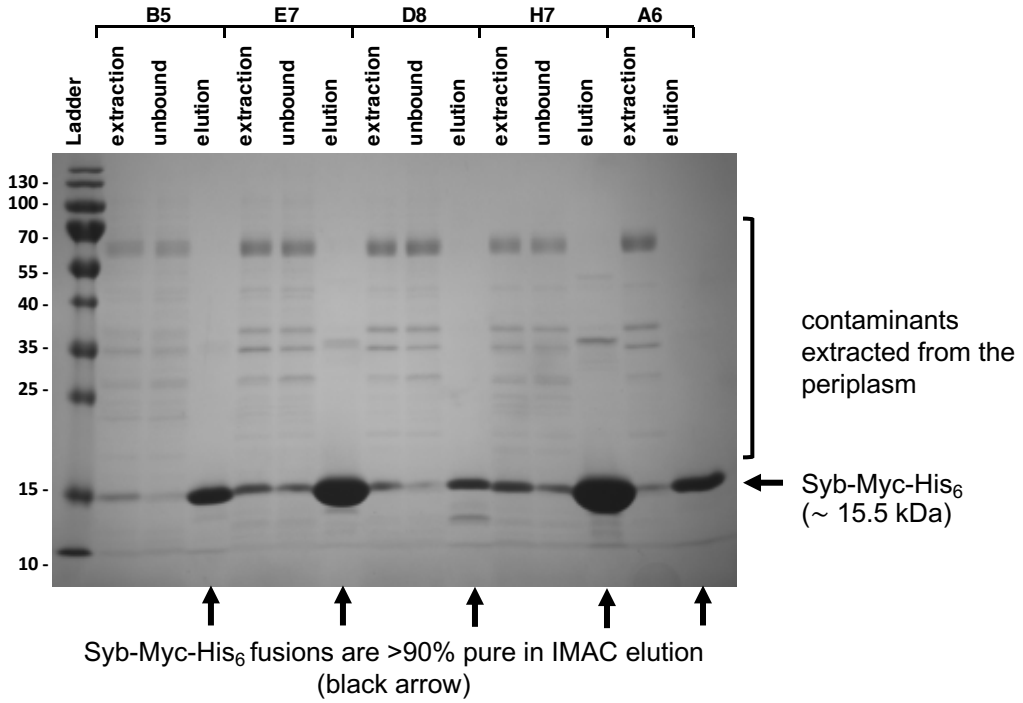


Figure 4.16 SDS-PAGE analysis from periplasmic extraction of binders in pSb_init. Samples from periplasmic extract, unbound flow-through from Ni affinity chromatography and elution are shown. All Myc-His$_6$ tagged binders are expected to be ~15 kDa.

misfolded binders from inclusion bodies and the cytosol.

4.2.6.1. Purification by Periplasmic Extraction

Traditional Nb purification pipelines are designed with periplasmic extraction and are reported to improve recalcitrant binder yields compared to cell lysis (Pardon *et al.*, 2014). Additionally, trafficking to the periplasm is a guarantor of binder quality and well-foldedness (Kariuki and Magez, 2021). For periplasmic extraction, we utilised the pSb_init plasmid for expression and proceeded with periplasmic extraction of the C-terminal Myc-His$_6$-tagged binders . However, with this tag the binders would be unsuitable for crystallisation purposes and cryoEM would be the structural method of choice. After expression, periplasmic extraction, and Ni-affinity purification, the binders purified to purity (>90% pure) (Figure 4.16). Based on the SDS-PAGE analysis, the binders can be ranked in order

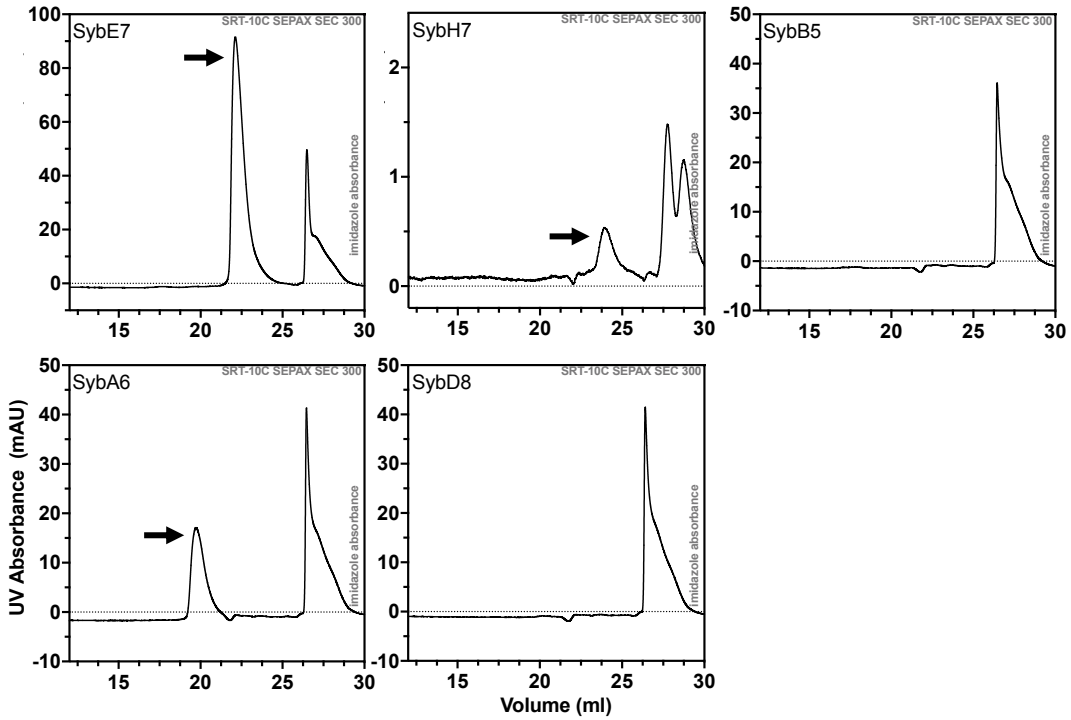


Figure 4.17 Gel-filtration traces of injected IMAC elutions (300 mM imidazole) containing Syb as Myc-His_6 fusions. SybE7 H7 and A6 eluted the column at V_R indicated by the black arrows.

of highest to lowest expression/purification as Syb H7>E7>B5>A6>D8. Satisfied with this result inspection of binder monodispersity by SEC was carried on (**Figure 4.17**). Only 3/5 binders eluted gel-filtration column (SybE7, SybA6 and SybH7) whereas the remaining displayed aberrant interactions with the column matrix, hindering elution. Lack of SybB5-Myc-His_6 fusion elution is in line with lack of cleaved SybB5 eluting the gel-filtration column in the previous attempt (**Figure 4.13-E**). This suggests that even with a hydrophilic Myc-His_6-tag, aberrant column-matrix interactions are persistent and are an intrinsic trait of SybB5. In general, gel-filtration elution at $V_R < 22$ ml is a testament against undesirable Syb hydrophobicity. However, given the restrictive number of available binders for CAT4 and since periplasmic extraction/IMAC provided sufficient purity without gel-filtration, the latter was deemed unnecessary, the binders were scored based on their interactions with CAT4.

4.3 Analysis of Syb Binding to AtCAT4

Binding affinity towards CAT4 was assessed with, (a) affinity pull-downs followed by SDS-PAGE electrophoresis, (b) biolayer interferometry, and (c) co-elutions of CAT4-Syb complexes by SEC.

4.3.1 Syb Affinity Pull-downs

CAT4 was incubated with 1.5× molar excess binder for 1 h at 4 °C followed by immobilisation of Syb-Myc-His_6 on Ni resin. It was expected that a binder with good affinity against CAT4 would retain the protein on the Ni resin via strong paratope-epitope interactions despite three rounds of washes with high ionic strength buffer. These wash steps electrostatically screen Syb interactions with membrane protein (either CAT4 or the negative control $PepT_{So}$) to dissociate weak binders and increase selection stringency. Sybs are ranked with SDS-PAGE based on the amount of CAT4 retained after the washes compared to a sample taken before the wash steps.

The Ni affinity pulldowns showed no observable CAT4 or $PepT_{So}$ densities when imaged with simple coomasie dye, suggesting that the retained amount of membrane protein samples was below the detectable limit for this dye (Figure 4.18). Silver-staining however increased the sensitivity of the imaging and allowed the observation of membrane proteins. With silver-staining, the Sybs showed variable ability to retain CAT4, whereas they had no specificity towards $PepT_{So}$. Other, low-abundant species were detected in the CAT4 gel, attributed to CAT4 purification, since they were not detected in the Syb-$PepT_{So}$ gel. In the silver-

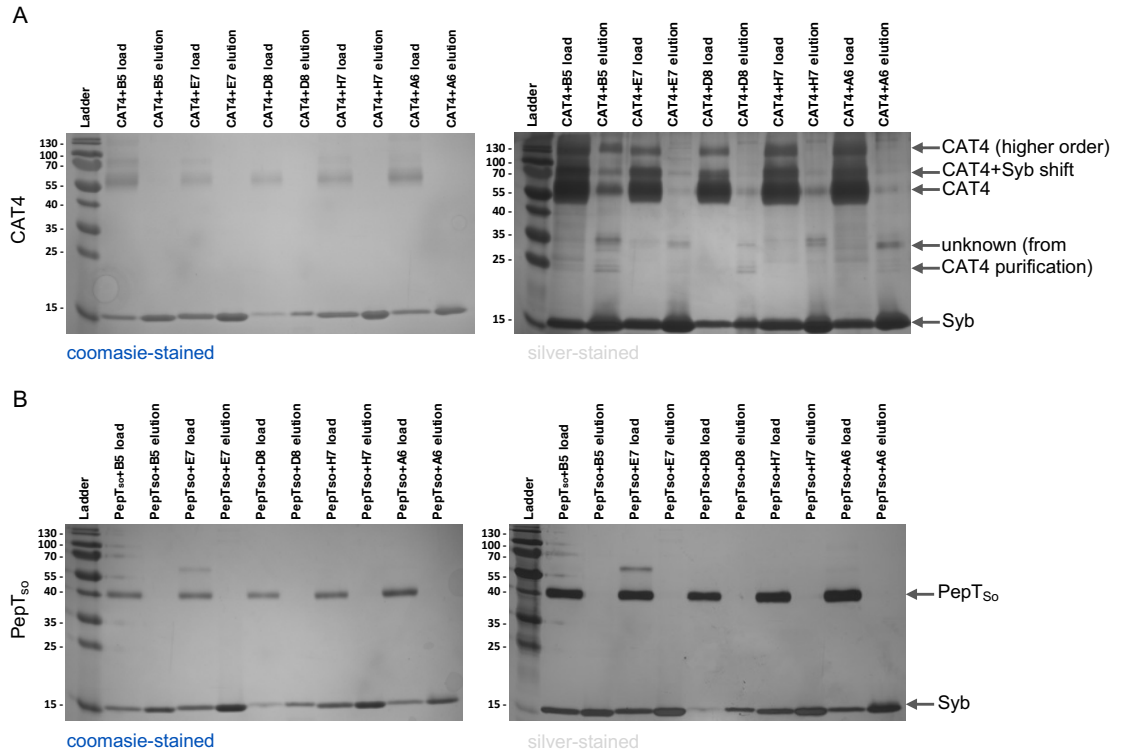


Figure 4.18 SDS-PAGE Coomasie (left) and silver-stained (right) gels reporting Ni resin pulldowns by the 5 unique Myc-His6-tagged sybodies. Adjacent lanes show sample composition of the load prior to wash steps and after washes and imidazole elution. Pulldowns of (A) CAT4 and (B) negative control $PepT_{So}$. Membrane proteins and Syb are indicated with black arrows.

stain, SDS artefactual higher-order laddering is seen for CAT4, previously discussed in chapter 3. With regards to Syb affinity, based on the CAT4 band intensity after elution, the apparent ranking of Syb is the following: B5 >> H7 > D8 ~ A6 > E7. The first-ranking Syb was found to be SybB5 with Ni-pulldowns.

4.3.2 Biolayer Interferometry Analysis of Binding

Biolayer interferometry (BLI) is a biophysical method using an optical biosensor to distinguish biomolecular interactions on the basis of wave interference. Specifically, BLI uses evanescent waves; an optical disturbance generated when a light ray has an incident angle larger than the critical angle of two interface

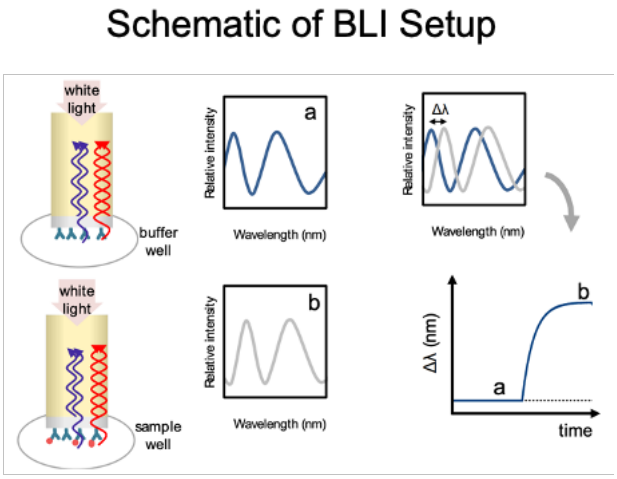


Figure 4.19 Schematic overview of the experimental setup of the method and the recorded sensogram outputs (adapted from online resource *https://cmi.hms.harvard.edu/biolayer-interferometry*).

media, causing total internal reflection of the refracted ray. Although no net energy flows between the media, an evanescent wave emerges. Intuitively, this happens because electric and magnetic fields cannot be discontinuous at a boundary. The evanescent wave's electric field amplitude (E_0) decays exponentially with distance from the interface instead of varying sinusoidally with distance in all directions (like the incident wave). Since the amplitude of the evanescent wave decays exponentially with distance, its intensity ($I \sim E_0^2$) drops by a factor of $e^{-2}\times$ faster. Thus, only media very near the interface is illuminated by the evanescent wave. In BLI the intensity of the electric field of the evanescent wave, is measured. When a particle with a refractive index different from the surrounding media is immobilised near the refracting interface on the side the evanescent wave is generated, scattering occurs which is very sensitive to the particle's height dimensions, h ($I_{\text{scatter}}(h) = I_0\, e^{-\beta h}$, where I_0 is the intensity at zero particle-interface separation and β is the evanescent wave penetration depth). Therefore, if a membrane protein is immobilised at the BLI interface, it scatters based on its height dimensions causing an interference pattern different from the baseline,

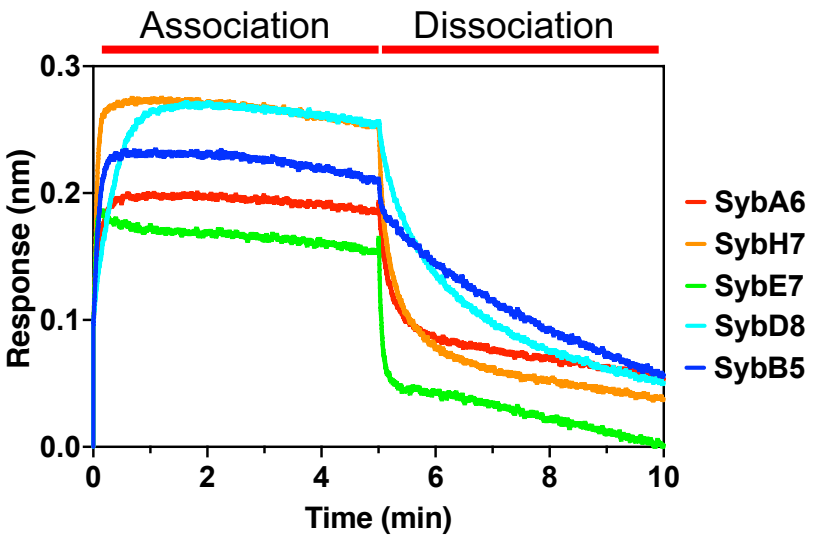


Figure 4.20 Association/dissociation sensograms of 500 nM Sybs to 100 nM CAT4 over 10 min time-course. The response of the interference shift caused by Syb association and dissociation was measured in nm.

recorded with a sensogram. The addition of a Syb will change the height of the particle, changing the interference, resulting in detection of Syb binding (**Figure 4.19**).

Biotinylated CAT4 was selectively immobilised to streptavidin-coated fibre-optic biosensors. A CAT4 loading baseline was achieved. Sequentially, the CAT4-loaded biosensors were presented to Syb binders (association phase) followed by fresh LMNG assay buffer (dissociation phase). The changes in immobilised protein shape during association/dissociation yield an interference shift $\Delta\lambda$ (nm) and are reported as biphasic sensograms over time. Sybs B5, H7, E7 and A6 all display comparable association rates while SybD8 is slower to associate (**Figure 4.20**). The Sybs have different response heights ranging from +0.18-0.26 nm. This property is attributed to differences in binding conformation between the binders and is not connected with affinity, and thus is considered secondary in selecting a binder. Due to their quick dissociation, Sybs H7, A6 and E7 are ranked less

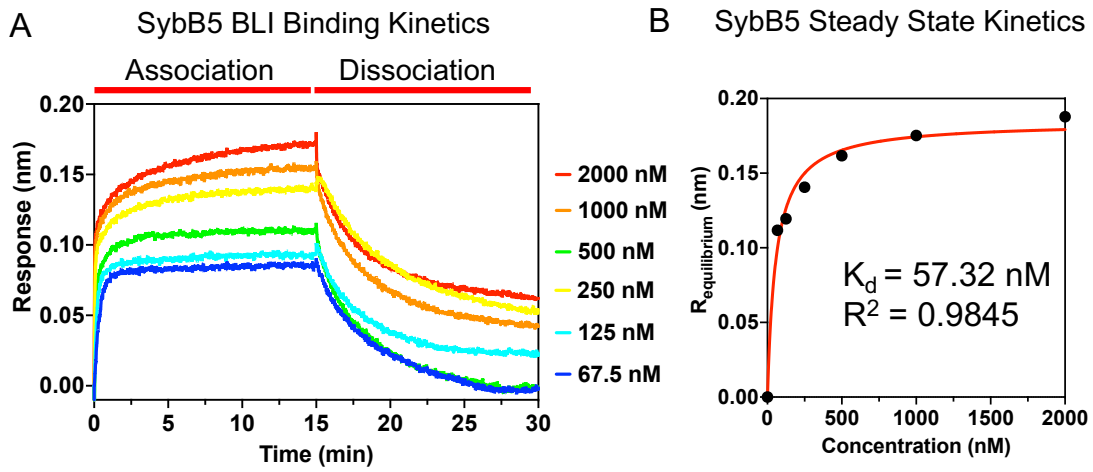


Figure 4.22 BLI binding kinetics of the CAT4-SybB5 complex. (A) Concentration gradient ranging from 0-2000 nM SybB5 to 50 nM CAT4 over 30 min time-course. (B) $R_{equilibrium}$ steady-state kinetics of sensograms shown in A, calculated with non-linear regression analysis of unimolecular reaction (1:1) association and dissociation responses (available within the OCTET Red384 analysis software). The apparent K_d and R^2 quality of fit are presented.

favourably whereas SybD8 is ranked second best. However, SybB5 dissociation proceeds markedly slower than the others without plateauing. A low dissociation rate is desirable in both crystallography and cryoEM single particle analysis (SPA). Moreover, the dissociation rate is an intrinsic property of a complex, i.e., it cannot be regulated by increasing binder concentrations, like the association rate. In agreement with the Ni affinity pulldowns, the sensograms report SybB5 as the tightest binder for CAT4, evidenced by the steep gradient in the initial association phase and the slowest and incomplete dissociation over the recorded 10 min.

Binding kinetics by BLI were performed for the highest ranking SybB5. The kinetics presented a $K_{d,app}$ of 57.32 nM for the CAT4-SybB5 complex (**Figure 4.22**), an affinity comparable with successful Syb and Nb binders of a small mem

brane protein (Löbel *et al.*, 2022). Additionally, the binding showed concentration-dependence in the sensograms providing greater confidence in its selection as the main CAT4 binder.

4.3.3 Gel-filtration Co-elution Analysis

In Ni pulldowns and BLI, CAT4 and Sybs were incubated, and the binding characterised with different surface chemistries resulting in a homonymous decision towards SybB5. This does not however provide information on the monodispersity of the resulting complex. A Syb suitable for structural applications must be able to co-purify with CAT4 through a gel-filtration column to ensure stochiometric complexation (1:1) in the complex peak fractions. Of seminal importance for a binder-target complex is it surviving large volume dilution whilst traversing the 30 cm SEC column (dilution factor $> 60\times$). This dilution lowers the concentration of the free, unbound Syb – disfavouring the equilibrium kinetics of complex formation – and assesses whether the CAT4-bound Syb dissociates. This stringent test qualitatively reports on the Syb's off-rate which ideally does not dissociate.

We proceeded to test co-elution with SybE7 – being the binder that gave the most prominent gel-filtration peak after its periplasmic extraction/IMAC purification – and SybB5 – previously not eluting the column – (Figure 4.17). Analytical-scale (< 40 µg CAT4) co-elutions were carried out with both binders by incubating them at $1.5\times$ molar excess for 30-60 min on ice before injecting on a SEPAX 10/300 SRT-10C column pre-equilibrated with LMNG 20 mM Tris/150

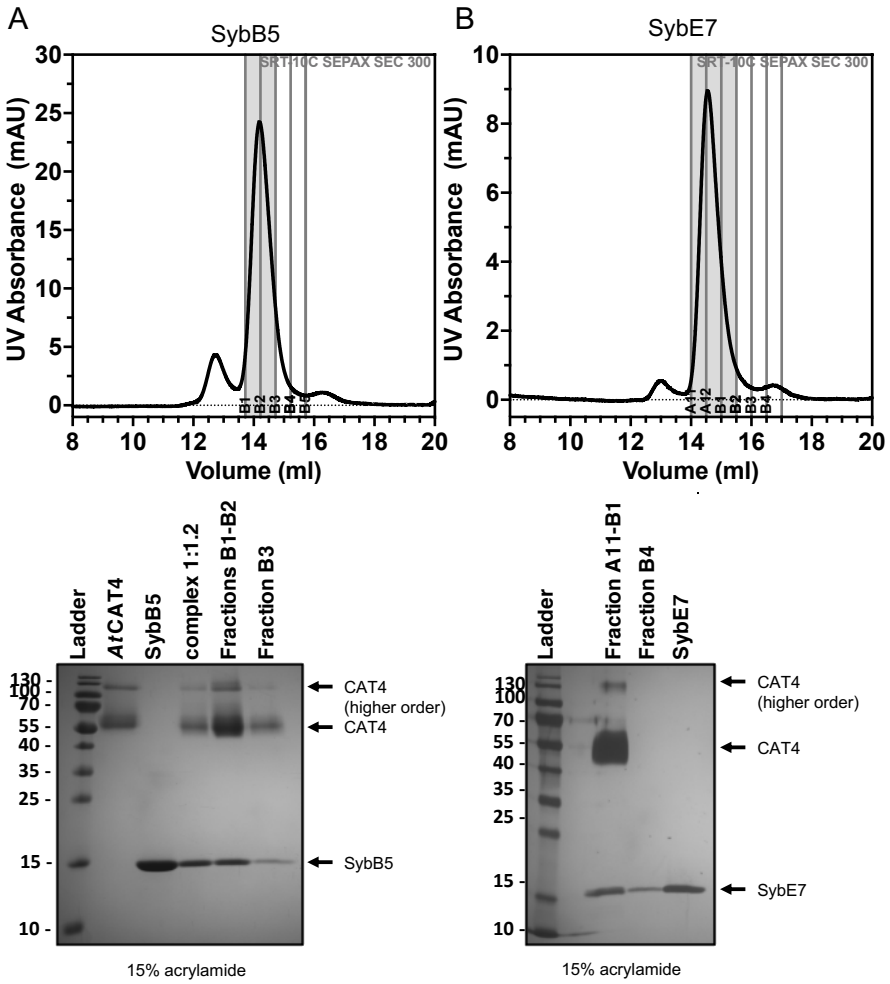


Figure 4.23 Analytical-scale co-elution analysis of purified CAT4 with purified (A) SybB5 and (B) SybE7. Gel-filtration fractions from the main chromatogram peak were analysed for co-existence of CAT4 and Syb bands, in which case CAT4 coelutes with a Syb.

mM NaCl buffer. Both Sybs E7 and B5 elute gel-filtration within the same fractions as CAT4 by forming a complex (**Figure 4.23**). This suggests that irrespectively of whether a binder elutes the column on its own, in the presence of CAT4 they can elute as complexes, which is the desirable trait.

For structural applications larger than analytical scale quantities of the complex need to be isolatable and retain monodispersity. Based on the BLI binding screening (**Figure 4.20**), the Ni pulldown (**Figure 4.18**) and analytical co-elution

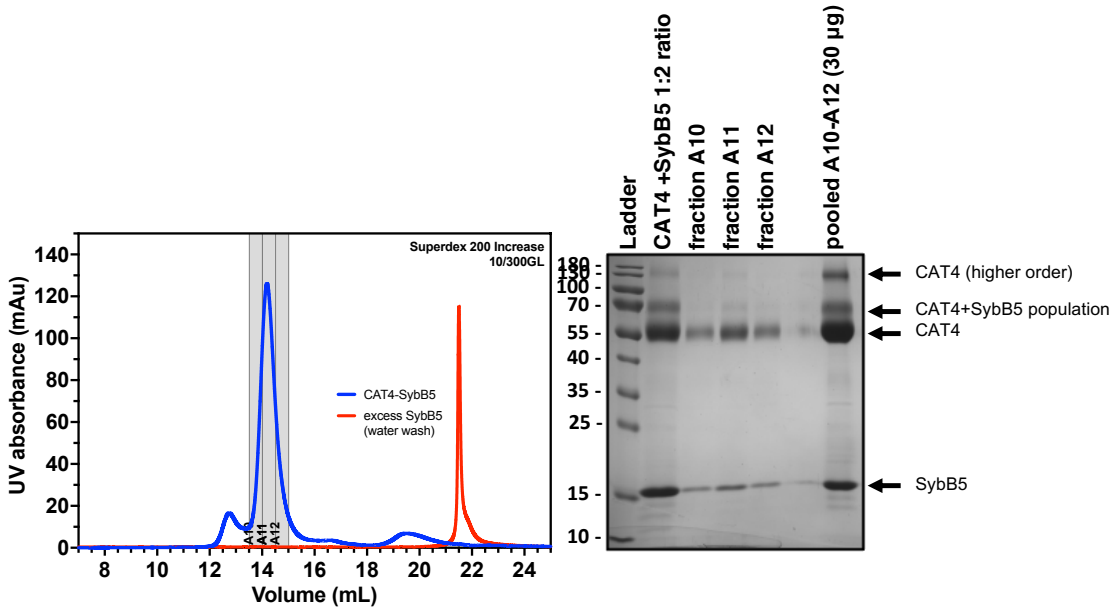


Figure 4.24 Preparative-scale co-elution of CAT4-SybB5 complex (incubated at 1:2 CAT4:SybB5 mol). The gel-filtration chromatogram of the complex is shown in blue. Excess SybB5 was eluted from the column in the water wash (red). SDS-PAGE analysis of the complex formation reaction, gel-filtration fractions. A sample of the polled fractions after overconcentration is shown.

(**Figure 4.25**), SybB5 was selected for preparation-scale co-elutions. This is presented in **Figure 4.24**, where, upon a 1:2 mol CAT4:SybB5 complex formation, CAT4 elutes as a complex with SybB5 within a monodisperse peak (1.5 ml volume; fractions A10-A12) on a Superdex 200 size-exclusion column. As this Syb had shown to not elute without CAT4, the 1× molar excess was found to elute in a water wash with a peak that is comparable in height to that of the full complex. The SDS-PAGE analysis of the overconcentration pooled fractions showed $> 90\%$ purity and the sample was deemed appropriate for single particle analysis and crystallisation trials.

4.4 Chapter Conclusions

In vitro screening for CAT4 binders was presented to augment CAT4 crystallisation or render the protein amenable to single particle analysis. Initial challenges in cloning an avi-tagged version of CAT4 were realised to be due to non-specific PCR primer annealing and were surpassed by generating a new gene-string. The expression, purification and thermostabilisation with lipids of avi-tagged CAT4 was comparable to WT and the construct was entered in the Syb pipeline. The Syb screening procedure provided 5 unique Syb binders from the concave library as candidate molecular chaperones. The purification of the binders initially proved challenging, owing to their aberrant biochemical quality and "stickiness" when expressed as N-terminal MBP fusions and purified after cell lysis. Periplasmic extraction was found to be the method of choice for CAT4 binders. This preference was subsequently tested by other group members in cases of purification-resistant binders for other projects and was verified to produce pure Sybs.

Screening the binding affinity towards CAT4, SybB5 was found to be the top-ranked by Ni affinity pulldowns and biolayer interferometry. The latter reported SybB5 to have a K_d in the low nanomolar range which was deemed satisfactory for structural applications. Previous experience shows that Syb in the 150-300 nM range were able to produce protein crystals in LCP (Löbel *et al.*, 2022). Satisfied with this result, complexes with CAT4 were prepared and crystallisation trials as well as cryoEM grid screening was attempted, reported in **chapter 5**.

Structure, Dynamics & Characterisation of AtCAT4

CHAPTER 5

The cryoEM structure of CAT4 was determined at 3.3 Å resolution in its apo-state. Firstly, grid optimisation and single particle analysis are presented. Two datasets were collected, and two maps were generated of CAT4; one of apo- CAT4 and one where CAT4 was supplemented with L-ornithine, which unfortunately failed to reveal resolvable density for the substrate. A new conformational state for eukaryotic SLC7s and structural comparisons with other members are presented. Analysis revealed binding-site features suggestive of an amino acid transporter and new insights into thick gate electrostatic rearrangements upon substrate coordination of TM1α.

A binding assay is developed and applied to WT-*At*CAT4 and mutants to study α-amino acid binding. MD simulations of CAT4-substrate interactions are also presented and compared to L-cystine recognition by 4F2hc/xCT. Transport testing of amino acids was tested for human SLC7A4 and *At*CAT4. Expression of *At*CAT4 in mammalian cells was not achieved. TEVC recordings of influx currents with α-amino acids are presented. Cell-based uptake assays and binding experiments showed that human SLC7A4 has a distinct amino acid profile. The chapter reports structural and MD data of *At*CAT4 and the progress towards functional deorphanisation of SLC7A4.

This chapter contains simulation data also published in the manuscript:
"Molecular basis for redox control by the human cystine/glutamate antiporter system xc⁻" Parker, J.L.; Deme, J.C.; **Kolokouris, D.**; Kuteyi, G.; Biggin, P.C.; Lea, S.M.; Newstead, S. Nat Commun 12, 7147 (2021)., doi: 10.1038/s41467-021-27414-1

5.1 CryoEM Pipeline for CAT4

With the discovery and generation of a CAT4-specific Syb displaying the desired traits (SybB5), progression into structural elucidation of the complex was warranted, to first, solve the putative transporter in an apo- state before attempting substrate-bound conformations.

The purified complex (reported in §5.1.1) was setup for LCP crystallisation initially, at complex concentrations 30 mg·ml^{-1} prior to lipid reconstitution. Crystallogenesis was not immediately achieved in the presence of the binder and "haziness" or "glowing bolus" similar to the crystallisation attempts of the apo-protein were seen (chapter 3). The soluble and highly flexible Myc-His$_6$-tag on SybB5 was realised as a likely source of crystallisation inhibition, albeit of low biological significance. However, such flexible regions pose no problem for cryoEM single particle analysis (SPA), as unstructured regions are usually not observed in electron density maps. Therefore, it was decided to proceed with CAT4-SybB5 SPA.

Being only 63-64 kDa with unknown glycosylation afforded by *S. cerevisiae*, arabidopsis CAT4 was a relatively unattractive target for cryoEM. Additionally, as proven by SEC-MALS, CAT4 in LMNG is a strict monomer, which precludes

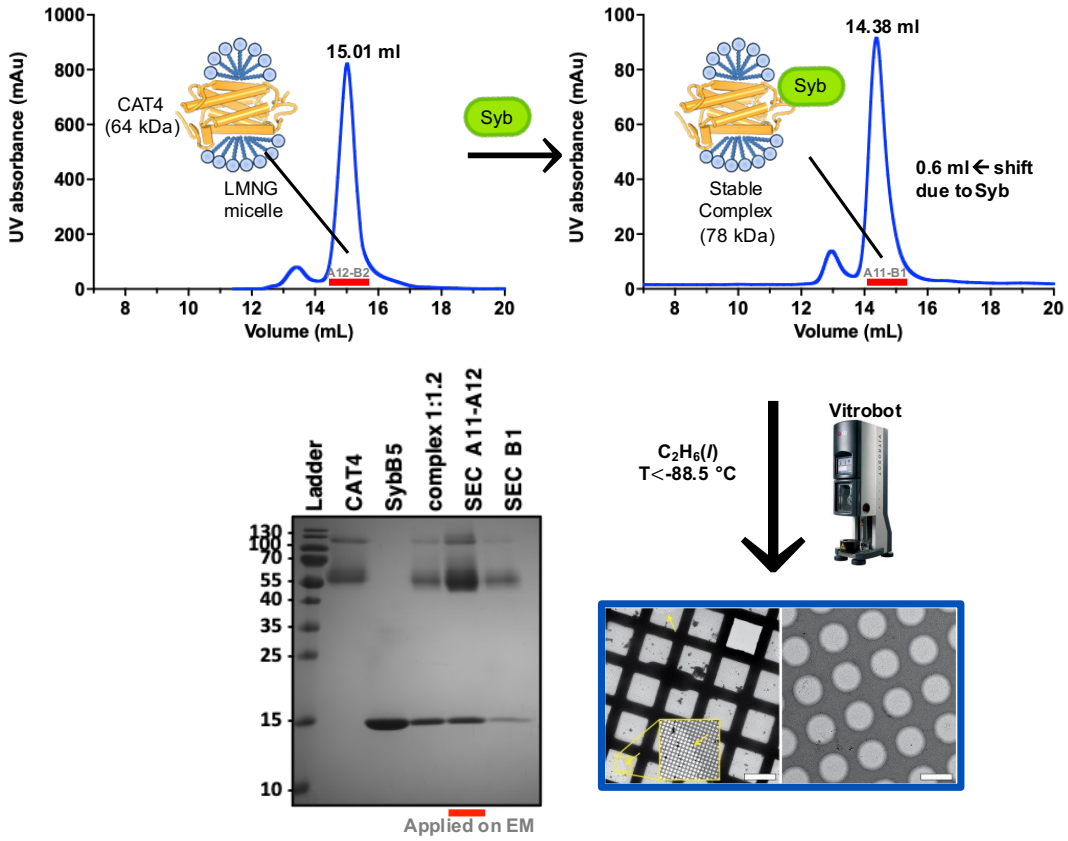


Figure 5.1 Pipeline showing the preparation of CAT4-Syb sample for cryoEM analysis. Purified CAT4 in LMNG, fractions "A12-B2" are complexed with excess Syb and purified via SEC. Monomeric, shifted fractions "A11-B1" were pooled for vitrobot screening of grids. Sample purity was assessed with SDS-PAGE electrophoresis; higher-order species are artefactual and attributed to SDS. Vitrobot prepared grids were vitrified in ethane liquidified by liquid nitrogen to ~ –88.5 °C to instantaneously vitrify the protein sample.

the use of symmetrical operators in SPA. Given that SLC7 transporter complexes have been reported with global FSC resolution < 4.0 Å, e.g. 4F2hc/xCT heterodimer (Parker, Deme, Kolokouris, *et al.*, 2021), CAT4-Syb complexes would likely succeed in providing structural information. Grid optimisation and the development of a particle analysis pipeline for CAT4, involving a non-proteinaceous density cleanup strategy, led to a high-resolution structure of CAT4.

5.1.1 Sample Preparation & Grid Optimisation for SPA

As CHS hindered crystallisation (chapter 3), sybodies were screened for

LMNG-only CAT4 (chapter 4). Thus, CAT4 in LMNG was selected for SPA. The lack of CHS supplement likely favoured cryoEM reconstructions since CHS induces large micelles of variable size. In general, small, tight micellar densities are preferred to encapsulate protein species, impeding relative movements of protein domains, and "forcing" overlapping particle conformations, allowing enrichment of the particle pool, and boosting signal/noise. Purified CAT4 in LMNG was incubated with 1.2-1.5× molar excess SybB5 (as Myc-His_6 tag) on ice for 15-30 minutes before size-exclusion chromatography to yield a monodisperse and homogeneous sample consisting of the pure complex (Figure 5.1). The complex was then concentrated to 3.0-8.0 $mg{\cdot}ml^{-1}$ to screen for optimal concentrations. In addition, loading volumes – 2.5 and 3.5 μl – and grid surface chemistry – either Au or Cu – was tested to optimise particle density in parallel (Table 5.1).

A qualitative ranking system was adopted to classify grid conditions for the final data collection session. A condition is considered "GOOD" when visualisation under a 200 kV FEG electron microscope shows:

- homogeneous distribution of observed, discoidal particles,
- particles similar in size and shape,
- high density of particles, and
- particles without significant overlap.

Similarly, the rank "OK" is given when at least one of these conditions is not met, while "Thick ice" pertains to a microscopy image that may or may not show particles and is characterised by a reduced signal contrast to background due to the dense, vitrified sample. The rank "POOR" means that no particles were observed irrespective of thin/thick ice squares; observed in instances of high surface tension which forces the particles to the grid edge. Examples of "OK", "GOOD"

Conc.	Grid	Sample vol.	Wait time	Rank
3.4 mg·ml^{-1}	Au	2.5 µl	0 s	OK
3.4 mg·ml^{-1}	Cu	2.5 µl	0 s	GOOD
3.4 mg·ml^{-1}	Cu	double load 2.5 µl	0 s	Thick ice
3.4 mg·ml^{-1}	Au	3.5 µl	0 s	Thick ice
3.4 mg·ml^{-1}	Cu	3.5 µl	0 s	GOOD
5.0 mg·ml^{-1}	Au	2.5 µl	0 s	OK
5.0 mg·ml^{-1}	Cu	2.5 µl	0 s	OK
5.0 mg·ml^{-1}	Au	3.5 µl	0 s	OK
5.0 mg·ml^{-1}	Cu	3.5 µl	0 s	OK
5.0 mg·ml^{-1}	Au	3.5 µl	30 s	Thick ice
8.2 mg·ml^{-1}	Au	3.5 µl	0 s	POOR

Table 5.1 Grid optimisation screen of CAT4-SybB5 complexes. Screened parameters are complex concentration, Quantifoil R1.2/1.3 grid chemistry (Au or Cu), volume of applied sample and sample incubation time before plunge freezing. In one instance the "double loading" sample application technique was tested, wherein sample is applied on both sides of the grid hole to increase particle density. The imaged results are ranked according to an order of increasing desirability: "POOR", "Thick ice", "OK", "GOOD".

and "POOR" grids are shown in (**Figure 5.2**).

In Cu and at lower concentrations (3.4-3.5 mg·ml^{-1}), CAT4 complexes showed high particle numbers in both the centre and ring-edge regions of the grid holes (**Figure 5.2**). The particle borders could be easily discerned, and the distribution was homogeneous. In Au, CAT4 was observed to compartmentalise in high-particle density regions whilst migrating away from other regions, creating a two-part "patch-like" distribution. This is undesirable since particle migration to highly dense regions are expected to overlap resulting in particle-picking difficulties downstream (**Figure 5.2**).

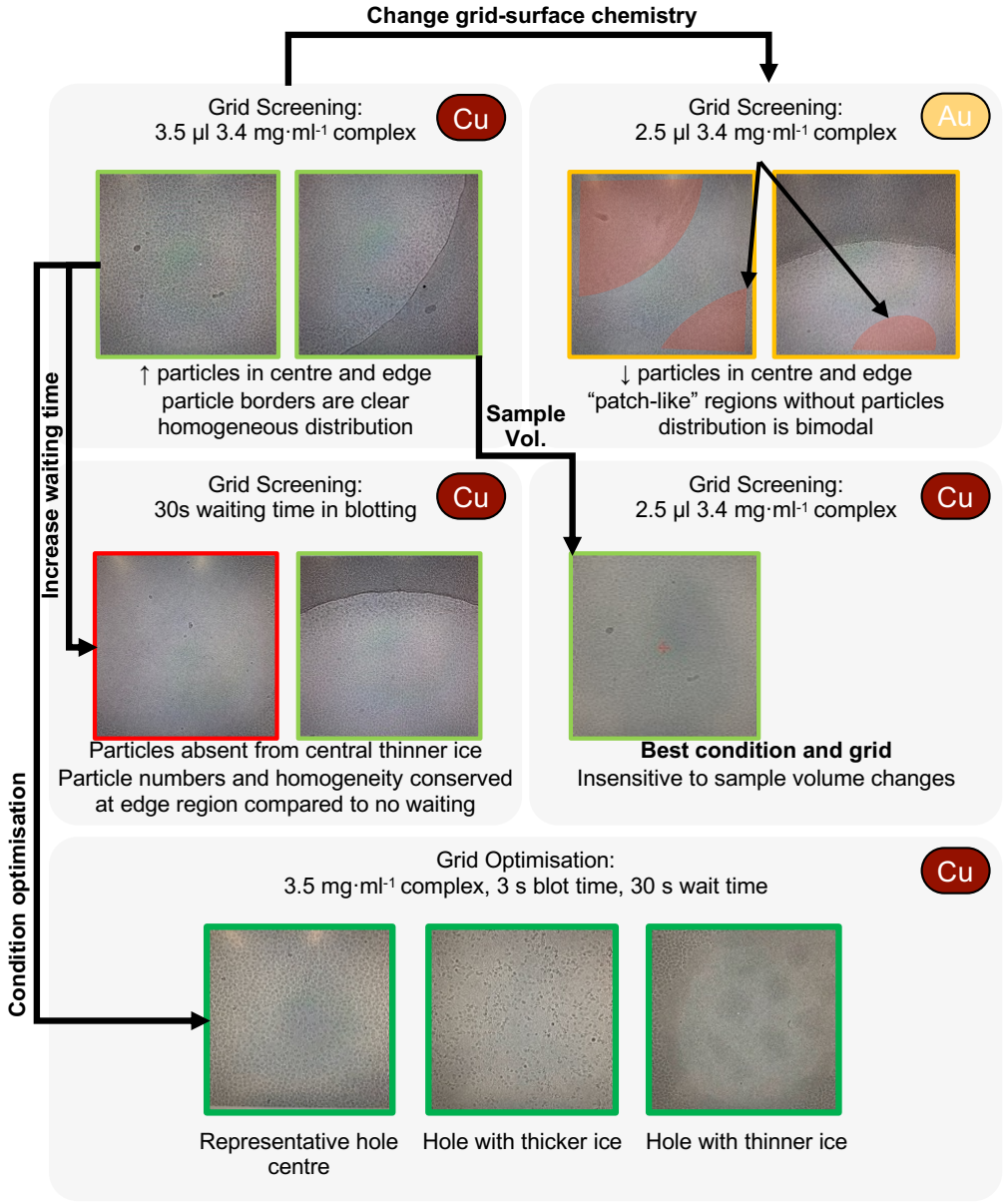


Figure 5.2 Representative micrographs from grid screening and optimisation acquired with Talos Arctica cryo-TEM equipped with Falcon 4 direct electron 200 kV detector. Grid conditions were classified, as in **Table 5.1**, based on the micrographs. Examples of "GOOD" micrographs are highlighted in green, "OK" in orange and "POOR" in red.

Higher concentrations of complex ($>$ 4.0 mg·ml^{-1}) did not increase particle numbers and were less optimal due to patch-like distribution formation; an effect likely owing to either surface tension effects from detergent overconcentration, and/or sample-specific self-attraction of unknown identity. Since commonly these

observations are sample-specific, there is generally little to gain from further investigation; especially when suitable conditions for data collection have already been identified. Therefore, further optimisation of the latter conditions was prioritised. Optimisation of Cu grids at concentrations 3.5 $mg{\cdot}ml^{-1}$ yielded largely similar micrographs showing that sample concentration is the major deciding factor on micrograph particle quality; largely insensitive to sample volume, waiting time before blotting, and blotting time (duration) (Figure 5.2).

5.1.2 Single Particle Analysis of CAT4

Being satisfied with the behaviour of the CAT4-SybB5 complex in LMNG detergent in vitreous conditions, micrograph data collection on a production-level FEI TITAN KRIOS equipped with a Gatan K3 detector and beam-image shift acquisition. A high frame-rate of 300 movies/h of data collection was achieved, permitting the collection of >12,000 movies. This setup has had widespread success in membrane protein structural elucidation and is considered, at the time of writing, the "gold-standard" for single particle analysis. The acquired CAT4-SybB5 movies were patch motion corrected and had their CTF estimated (SPA pipeline; Figure 5.3). 11,166,793 particles were picked "on-the-fly" with SIMPLE from collected micrographs. Two rounds of 2D classification took place. In the 1st round (350 classes) all proteinaceous densities were selected (7,053,972 particles, 63.2% remaining). These were passed to a 2nd round with stricter selection criteria, i.e., CAT4-SybB5 imitating densities displaying characteristic Syb protrusions and helical densities (2,459,086 particles, 22.0% remaining). For CAT4, most classes showed a top/bottom view of the complex (~ 70%) whereas the remaining were side views.

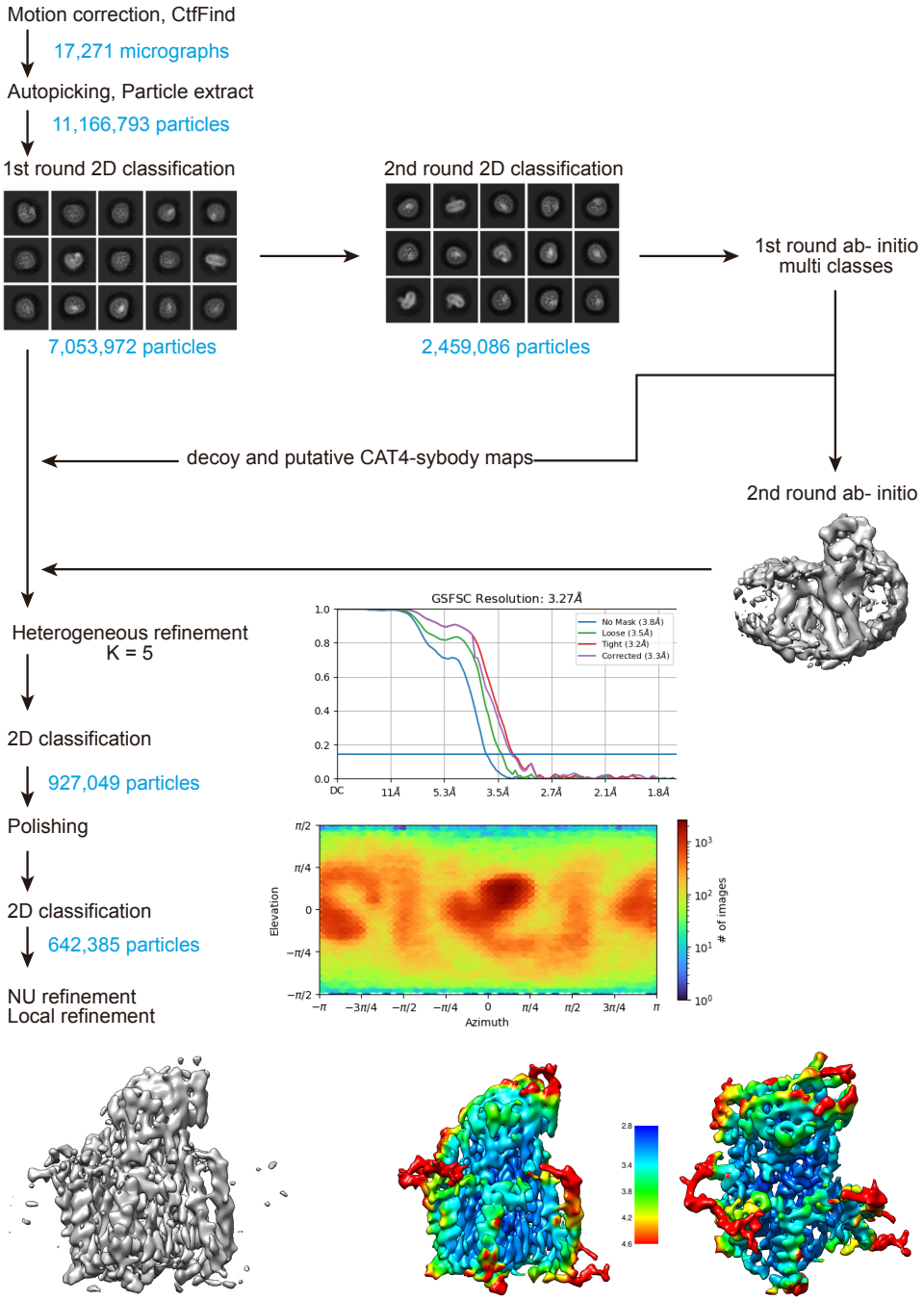


Figure 5.3 CAT4 CryoEM SPA workflow and map resolution. Details of the individual pipeline steps are given in the main text. The final 3D reconstruction coloured according to real space resolution is shown.

The 2D class averages (2D-CAVG), calculated in cryoSPARC (Punjani *et al.*, 2017), provided discernible structural elements with the most profound being the protruding density of the putative Syb binder and the α-helical-like signals in

some 2D-CAVGs. Multi-class (N=6) *ab-initio* models were generated with particles form the 2nd 2D-CAVG round, one of which gave an apparent Syb-bound transporter-like architecture with discontinuous helical densities (509,029 particles, 4.6% remaining). The class particles were submitted to a further *ab-initio* round to regenerate the model.

Initially, the 3D maps retained disconnected densities of seemingly random orientations even after non-uniform refinement. These were attributed to stubborn, contaminant "junk" particles causing compositional heterogeneity, and which could not be discarded from the dataset with conventional 2D and 3D classification methods. Recently, a clever hack was reported which leverages existing 2D classification algorithms to *in silico* purify particles with decoy "junk" maps (Straub *et al.*, 2021). The classes corresponding to "junk" particles in the dataset (N=5, from the 1st *ab-initio*) were used as reference guide along with the protein-target maps (from the 2nd *ab-initio*) to dictate separation in distinct classes by the cryoSPARC heterogeneous refinement functionality (K=5) (**Figure 5.3**).

The heterogeneous refinement, albeit to a low resolution (low-pass-filtered at 10 Å), provided a clear CAT4-SybB5 class, without broken densities and with CAT4 14TM α-helices intact. Subsequently, 2D-CAVG (927,049 particles, 8.3% remaining) were CTF-refined and followed by non-uniform refinement and finally masked-local refinement to yield a reconstruction at global FSC resolution 3.6 Å. The particles were Bayesian-polished and classified in 2D to generate a subset 642,385 cleaned and polished particles followed by homogeneous refinement, non-uniform refinement and finally, masked-local refinement yielding a final reconstruction at 3.3 Å, with the local resolution varying between 2.8-4.4 Å (**Figure 5.3**,

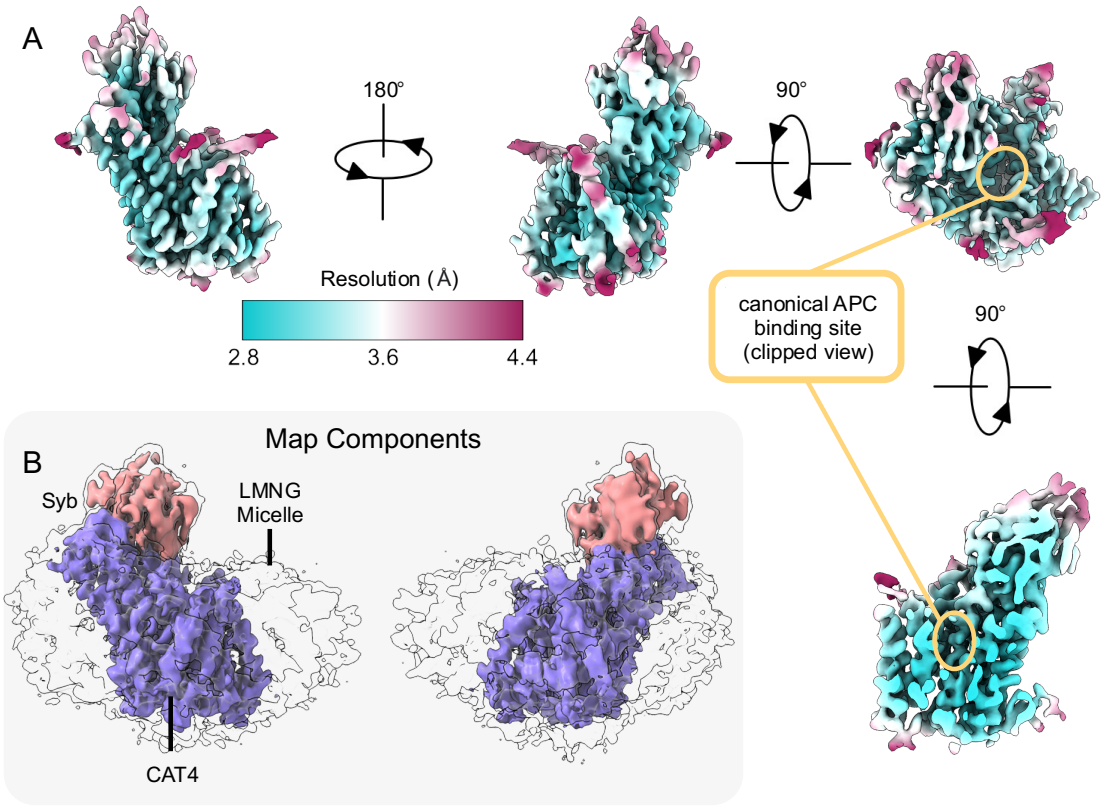


Figure 5.4 High-resolution cryoEM electron density maps of the CAT4-SybB5 complex. (A) Local resolution values ($FSC_{0.143}$) projected on the masked CAT4 density map and coloured from high (cyan) to low (maroon) resolution. The canonical APC binding site is highlighted. (B) Whole-map components; CAT4 (violet), SybB5 (salmon), LMNG detergent belt (transparent, outlined).

Figure 5.4-A). Both Bayesian polishing (per-particle, reference-based beam-induced motion correction) and CTF refinement (per-particle defocus, magnification, and higher-order aberration estimation) improved CAT4 map resolution. The final reconstruction showed a "tight" LMNG micelle encapsulating the CAT4-SybB5 complex (Figure 5.4-B), likely rigidifying it into the presented conformation.

5.2 CAT4 Structural Analysis

CAT4 has 14 transmembrane helices (Figure 5.5-A,B) with the first 10 being

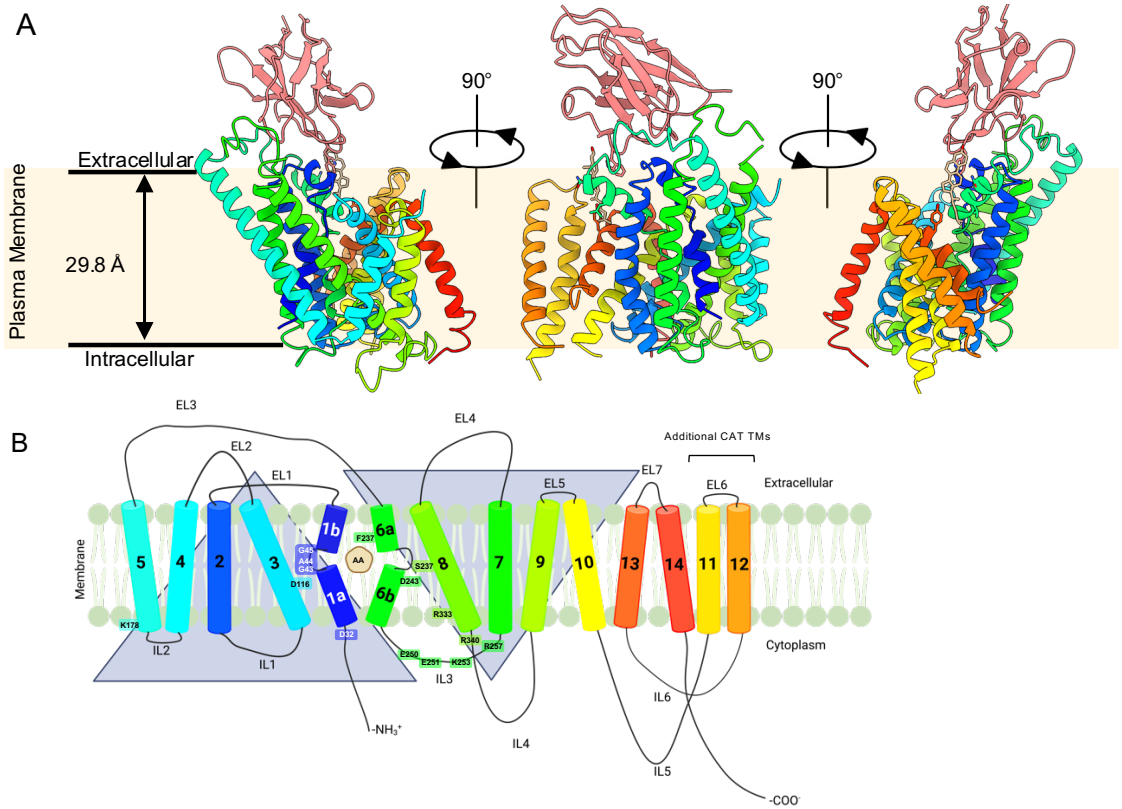


Figure 5.5 Atomic model of the CAT4-Syb complex. (A) Cartoon representation of CAT4-Syb complex from three 90° directions. CAT4 is coloured in chainbow from blue to red. SybB5 is in salmon. The cholesterol moiety is shown in stick (tan). A bilayer is shown as a continuum model (cantaloupe), based on CAT4 hydrophobic thickness; calculated on PPM3 (Lomize, Todd and Pogozheva, 2022). Structures and volumes prepared with UCSF ChimeraX v.1.2.5 (Pettersen et al., 2021). (B) Topology diagram of CAT4.

"canonical" TM helices of the APC fold. These are arranged in a pseudosymmetric internal repeat of TM1-5 and TM6-10, in accordance with the APC fold. The helices are tightly packed and form an outer *scaffold domain* (TM3-5, TM8-10) and an inner, central core (TM1-2, TM6-7). The outer scaffold contains the *hash domain* of the APC fold (TM 3,4,8,9) (Kowalczyk *et al.*, 2011b) – taking its name from the "#" shape formed by the helices. The central core, named, *bundle domain*, is traditionally thought mobile, pivoting on the stable scaffold between conformations (rocking bundle mechanism – chapter 1). From the remaining 4 TM helices, the final 2 helices of CAT4, have an inverted V-shape and superimpose with other SLC7 members, e.g., *Gk*ApcT, AdiC, BasC, xCT, LAT1 and

$b^{0,+}$AT [PDB IDs: 5OQT,3LRB,6F2G,7P9V,6IRT,6LI9, respectively]. The prediction that CATs have 14 TM helices (Verrey, Ellen I Closs, *et al.*, 2004), with the additional 2 TM helices expected to wedge between "TM10" and "TM11" of other APCs, is verified by the structure. In the region of the canonical APC binding site, CAT4 has a local resolution ~ 2.8 Å, higher than most other regions (Figure 5.4-A). Comprehensive analysis of the binding site is reported in §5.3.

5.2.1 CAT4-SybB5 Interface

The extracellular surface of CAT4 presents a polar, negatively charged epitope formed by close packing of 3 extracellular loops; EL1 (between TM1β-2; T52-P60), EL3 (between TM5-6α; G213-F223), EL4 (between TM7-8; G283-Y288) (Figure 5.5-B). The SybB5 paratope closely engages with the negative CAT4 epitope by electrostatic and surface complementarity (Figure 5.6-A). The SybB5 paratope surface has a net positive charge partially explaining the affinity via coulombic attraction of the two interfaces.

The contacting residues of the CAT4 epitope and SybB5 paratope have extended surface areas, 2173.5 $Å^3$ and 2176.6 $Å^2$, respectively. Surface complementarity permits the close proximity of the binder, likely modifying the loop regions (EL1,3,4), and increasing affinity by the hydrophobic effect. To quantify this interaction, the surface complementarity of an interfacial SybB5 residue to surrounding CAT4 residues (defined as < 4 Å to the SybB5 residue) is calculated as the scalar product of the normal to one surface at each surface point and the normal to the other surface at the point nearest to the first surface, with a

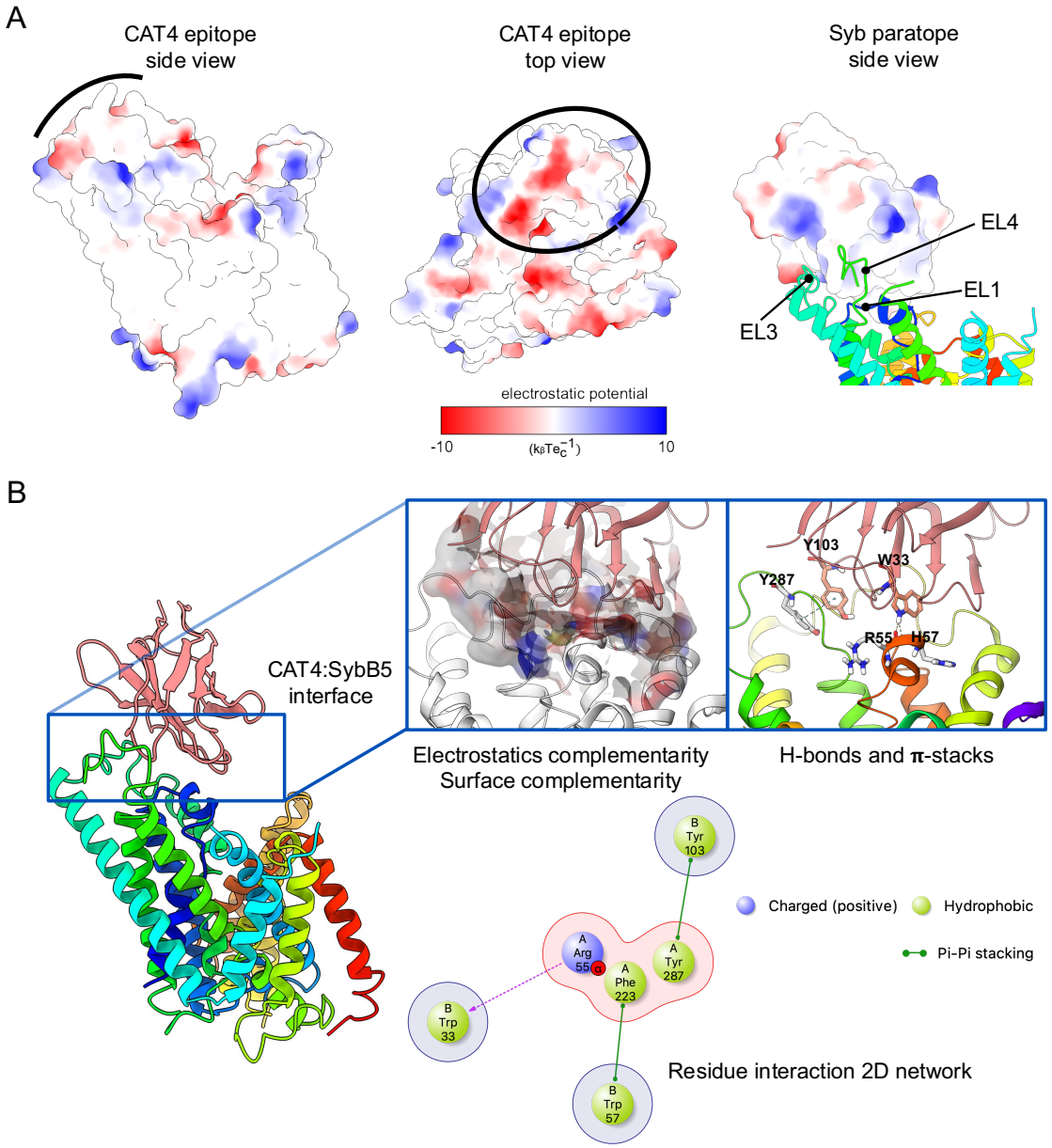


Figure 5.6 CAT4-SybB5 interactions. (A) Coulombic electrostatic surface (ESP) analysis of the complex. From left to right, CAT4 ESP side view and top view followed by a side view of the SybB5 ESP surface. Negative (CAT4) to positive (SybB5) complementarity at the interface is seen. (B) Close-up view at the CAT4-SybB5 interface. Mixing of the individual surface ESP charges "neutralises" the interface. Within the interface specific per-residue interactions appear (H-bonds & π-stacks). A 2D network representation of these is shown.

weighting factor for the distance between the points. The surface complementarity score (α) is the normalised sum over surface points, such that perfect complementarity (parallel normals) has a value of 1, and no complementarity has a

value of 0 (Norel *et al.*, 1994). It is evaluated over the buried solvent-accessible surface area (SASA). SASA – a degree of Syb "buriedness" – was calculated as the % of a residue's SASA that is not solvent exposed but surrounded CAT4 residues. The high α and buried SASA values reinforce the surface complementarity at the interface and suggest a novel feature for binder selection other than CDR residue atomic interactions (Table 5.2).

Atomic-level interactions are also present within the interface that augment the interaction rising from coulombic and surface matching. Specifically, R55 (EL1), F223 (EL3) and Y287 (EL4) are involved in a tri-part interaction with SybB5 residues, M31 and W33 (CDR1), W57 (CDR2) and Y103 (CDR3). The atomic interactions are mediated by hydrogen bonding and π-stacking. A cartoon representation and a 2D residue network model of the main interactions are shown below (Figure 5.6-B). Therefore, each SybB5 CDR provides a residue to interact with one of three CAT4 extracellular loops. Other auxiliary interactions in the form of vdW forces are also present. These are located at or around the CDRs, extending past the strict CDR definition (Table 5.2).

Overall, this analysis shows that the extended CAT4 epitope spans three extracellular loops, and that each Syb paratope residue is in at least one of three interaction categories: (a) surface complementarity and interface proximity packing, (b) classic coulombic electrostatic patch interactions that serve the initial attraction of the components, and (c) the specific atomistic interactions that link and stabilise the final complex with the target. *In silico* mutagenesis and MMP-BSA calculations to rank specific interactions along with SPA attempts with SybD8 are reported in **Appendix A – Other Sybodies as CryoEM SPA Candidates**.

Syb	CAT4	Distance	Interaction Type	Surface Complementarity (α)	Buried SASA
M31	**R55**	3.2 A	1× hb to R55	0.88	40.0%
Y32	P286 R55	3.4 A 3.5 A		0.78	78.5%
W33	**R55** **H57** E56 T58	2.9 A 3.2 A 3.3 A 3.5 A	1× hb to R55 1× vdW to H57	0.84	95.6%
H35				0.27	96.4%
Y37	**P217**	3.1 A	1× vdW to P217	0.72	75.7%
W47	**S218** **P217**	3.3 A 3.4 A		0.69	80.2%
Q52	**E56** **F223** H57	2.7 A 3.1 A 3.8 A	1× vdW to E56 3× vdW to F223	0.79	96.7%
S53	**E56** R55	3.2 A 3.7 A	1× vdW to E56	0.56	100.0%
Y54	T52	3.1 A		0.45	24.1%
Q56	E56	3.3 A	1× hb to E56	0.89	23.0%
W57	E56 **F223** A230 H57	3.1 A 3.5 A 3.5 A 3.6 A	1× pi stack to F223	0.27	52.3%
T58	**F223**	3.6 A	1× vdW to F223	0.66	8.0%
A59	**F223** P222 S218	3.4 A 3.9 A 4.0 A	1× vdW to F223	0.74	92.9%
G99	V285	3.9 A		0.74	100.0%
V100	T58 **L216** V285	3.3 A 3.6 A 4.0 A	1× vdW to L216	0.77	92.5%
G101	L216	3.7 A		0.79	99.9%
Y103	G213 Y288 V285 **Y287** Y214	2.9 A 3.4 A 3.5 A 3.7 A 3.8 A	1× pi stack to Y287	0.76	92.8%
Y104	Y288	3.1 A		0.45	62.3%

Table 5.2 Interaction analysis of CAT4-SybB5 interface. SybB5 residues are grouped to the closest CDR (CDR1 – magenta, CDR2 – cyan, CDR3 - green). Residues strictly in a CDR definition are shown in bold. All CAT4 residues within 4 Å of a Syb residue are listed. CAT4 residues in specific atomistic interactions are shown in bold. The interaction analysis was done in Schrodinger 2020-2 Maestro v12.8.117

5.2.2 Conformational State & Comparison to Other APC Transporters

The structure shows CAT4 in an outward-open conformation. One indication of this is that the thin hydrophobic gate in TM6, F237, is swung open, exposing the canonical binding site from the extracellular side (§5.2.2.1). Superimposition and comparison of TM1 helical kinks categorises the present structure with the AdiC outward-open state (PDB ID: 3LRB), with TM1 angle > 150° (Table 5.3).

SLC7	Conformation	TM1 angle	(°)	TM6 angle	(°)	RMSD (Å)
CAT4	Outward open (cryoEM – apo)	D32-G45-T52	152.1	L225-F237-T246	125.8	0.0
	Outward occluded (Alphafold2)	D32-G43-T52	150.7	L225-F237-T246	127.7	1.954
BasC	Inward-open (6F2G – AIB)	T11-G19-G34	128.6	L187-F199-G208	115.2	3.573
*Gk*ApcT	Inward-occluded (5OQT – L-Ala)	D31-G42-G58	132.6	F219-F231-S240	135.2	2.949
$b^{0,+}$AT	Inward-open (6LI9 – L-Arg)	L32-G45-K53	132.4	A220-W230-N239	135.8	3.759
LAT1	Inward-open (6IRT – BCH)	L52-G65-T73	132.1	N242-F252-F262	143.2	3.857
	Outward-occluded (7DSQ – Diiodo-Tyr)	L52-G65-T73	140.5	N242-F252-F262	151.5	2.561
	Outward-occluded (7DSK – JX-075)	L52-G65-T73	141.4	N242-F252-F262	142.8	2.778
AdiC	Outward-open (3LRB – apo)	L12-G25-S37	153.1	A192-W202-S211	137.5	2.529
	Outward-semioccluded (3OB6 – L-Arg)	L12-G25-S37	137.7	A192-W202-S211	145.9	2.716
	Outward-occluded (3L1L – L-Arg)	L12-G25-S37	135.1	A192-W202-S211	165.2	3.298

Table 5.3 Comparison of TM1 and TM6 helical kinks between SLC7 structures in different states, superimposed on TM3-5 & TM8-12 (CAT4 reference residues 87-224, 282-579; scaffold domain). All backbone atoms were superimposed for RMSD calculations.

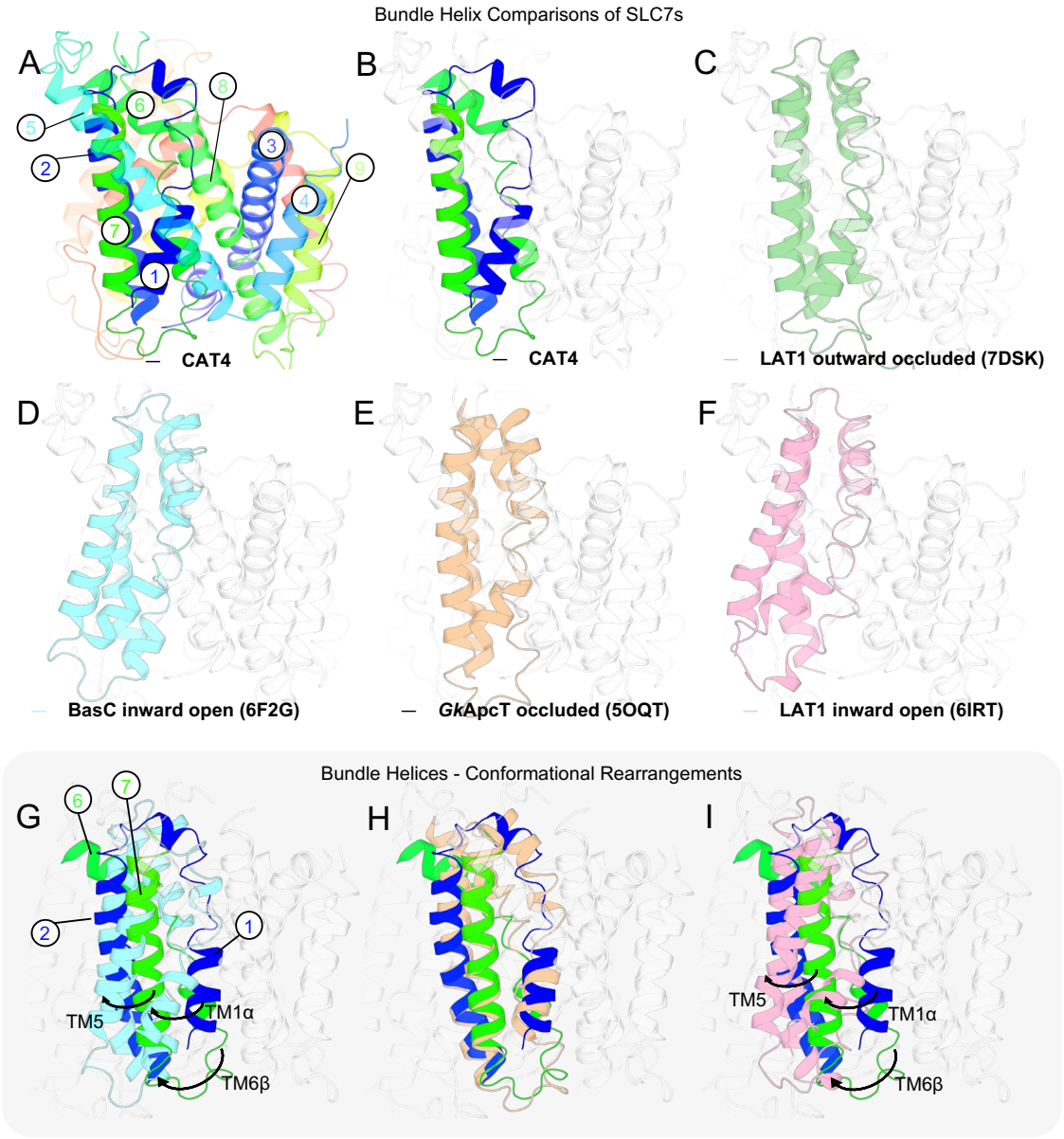


Figure 5.7 Comparison of SLC7 homologue bundle helices (TM1-2, TM6-7) and conformational rearrangements. (A) CAT4 outward-open structure coloured according to residue position (N terminus – blue, C terminus – red). CAT4 TM3-5 and TM8-12 are in transparent off-white and TM1-2 & 6-7 are coloured in each case showing: (B) CAT4 outward open, (C) LAT1 outward occluded, (D) BasC inward open, (E) *Gk*ApcT inward occluded, (F) LAT1 inward open. (G,H,I) Comparisons of CAT4 outward open state with BasC (G), *Gk*ApcT (H), and LAT1 inward open (F). CAT4 deviates from inward structures and overlaps with occluded *Gk*ApcT, although deviations in TM1 are seen. The PDB IDs used for structural comparisons of SLC7 homologues are shown.

CAT4 presents the first eukaryotic structure within the SLC7 family to be captured in this conformational state, since human LAT1 is locked by inhibitors

(7DSK,7DSQ) and AdiC (3LRB,3OB6) is a prokaryotic transporter. Inhibitor-trapped conformations may have key residues biased in orientations that would otherwise be energetically unfavourable and non-physiological (Yan *et al.*, 2021). CAT4 was solved in the outward-open state without small molecules. Interestingly, all mammalian SLC7 structures solved to date are in the inward-open conformation highlighting that there is likely an intrinsic energetic preference for this conformation in the absence of conformational selection. Even substrate-bound structures are inward-open. This observation along with the fact that SybB5 binds TM1α and TM6β suggests that the captured conformation is a result of binder conformational selection. Comparison with SLC7 structures in a variety of conformations (**Figure 5.7-A-F**) confirms CAT4 similarity with other outward-facing structures, with AdiC outward-open being the structurally closest to CAT4 (RMSD ~ 2.5 Å) (**Table 5.3**).

Helical Discontinuities Suggest an Induced-Fit Mechanism

Superimposing the structure with the Alphafold2 prediction (over all heavy atoms) produced an RMSD of 1.54 Å (**Figure 5.8-A**), suggesting that there are structural differences between the AF2 outward-occluded-like state and the CAT4 structure solved. These were primarily located in TM1 and TM6 conformations which line the binding site (**Figure 5.5-B**). In CAT4, TMs 1, 6 and 8 are discontinuous in the outward open state. Specifically, TM1 A44-T52, TM6 F236-F242, TM8 S327-Q331 are coiled residues.

TM1 and TM6 are characteristically discontinuous in SLC7 structures and predicted so in the Alphafold2 model. However, the discontinuous range is considerably longer in the outward-open state, spanning residues 44-52, than the

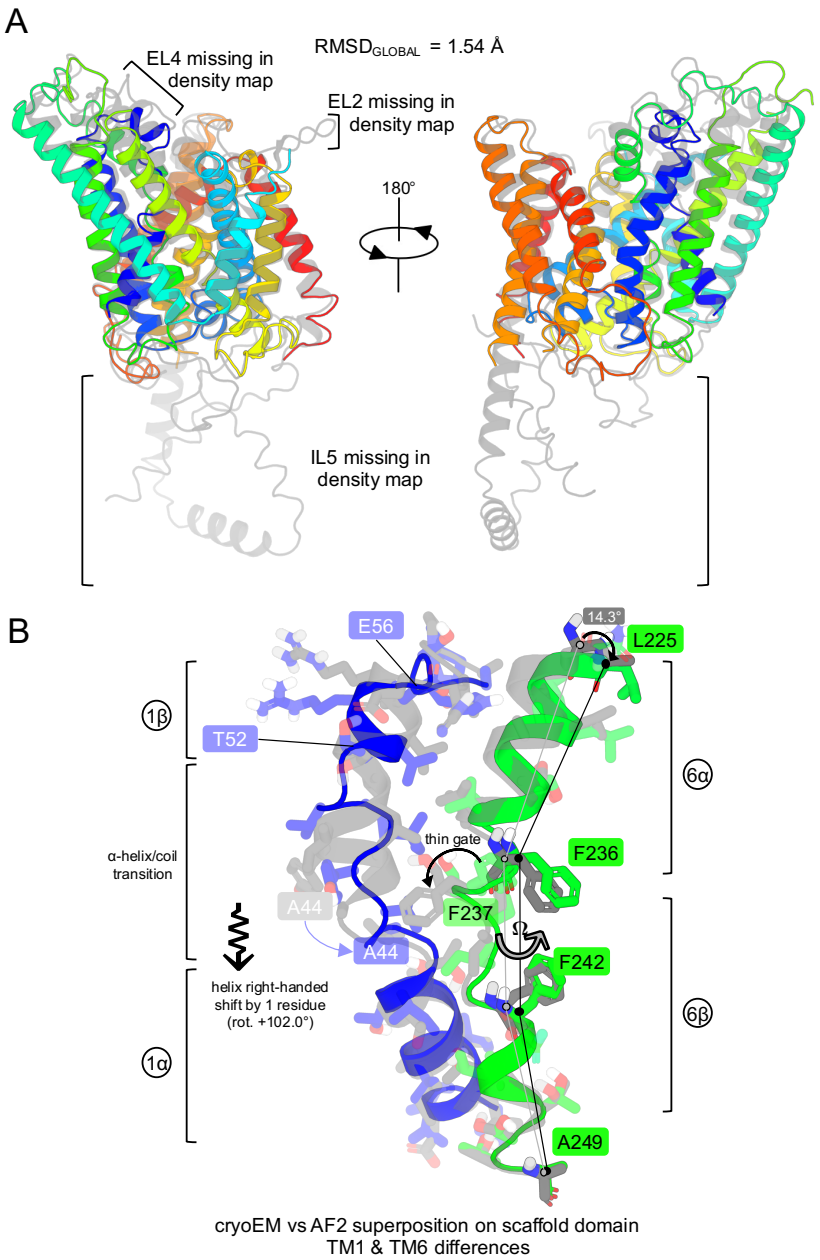


Figure 5.8 Comparison of cryoEM outward-open and Alphafold2 outward-occluded structures, superimposed on their heavy atoms (cryoEM model: rainbow, Alphafold 2 model: grey). (B) Comparison of TM1 and TM6 geometrical traits (cryoEM model: TM1 – blue, TM6 – green, Alphafold 2 model: grey). TM6 helix residues used for the definition of the dihedral, Ω, were L225-F236-F242-A249. All other residues are shown as transparent sidechains. TM helices 1α/β and 6α/6β are annotated. The transition of the thin gate, F237, is shown. The rotational and translational motion of A44 of the TM1 GAG motif is also highlighted. The Alphafold2 predicted helices TM1 and TM6 are $>$ 90% confidence pLDDT score.

AF2 outward-occluded, spanning residues T41-A44 (**Figure 5.8-B**). The coordi-

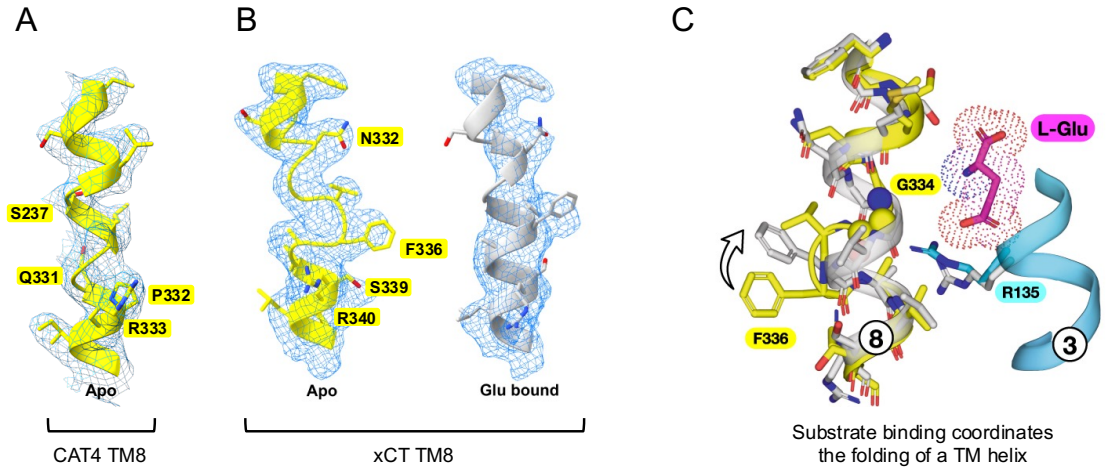


Figure 5.9 TM8 arrangements in CAT4 and xCT. (A) apo CAT4 TM8 discontinuous region S327-Q331. (B) apo xCT TM8 discontinuous region N332-F336 transitions to A-helix upon L-Glu binding. (C) Structural differences on TM8 between bound and unbound forms in relation to the re-positioning of R135 on TM3 and subsequent repositioning of F336. Panels B and C adapted from (Parker, Deme, Kolokouris, *et al.*, 2021).

nates of TM1 of the two structures coincide only towards the end of TM1β, T52-E56, while residues upstream have an offset of 1 residue per turn (**Figure 5.8-B**). Similar loss of TM1 helicity is present in the outward-facing structure of AdiC, suggesting similar outward-open conformations for other SLC7s would be expected.

The cryoEM and Alphafold2 CAT4 structures provide the end states of a hypothetical substrate-coordinated induced-fit mechanism for TM1. It is suggested that, initially, the transporter is idle in the outward-open state and the substrate enters the outward-facing cavity. Before substrate entry, TM1β is in the loosened conformation. As the substrate transiently interacts with the loosened binding site, TM1 residues rearrange and chelate the substrate (T41-G45; canonical discontinuous region in the AdiC and *Gk*ApcT structures). The rearrangement entails the corkscrew transition of TM1α.

A similar substrate induced-fit mechanism, involving a local coil/α-helix transition of TM8, was reported by us for the cystine/glutamate antiporter, xCT (Parker, Deme, Kolokouris, *et al.*, 2021). Upon glutamate binding to the apo state (PDB ID: 7P9V), TM8 coiled residues pack into a canonical α-helix to generate the substrate bound state (PDB ID: 7P9U). A parallelisation between CAT4 (S327-Q331) and xCT (N332-F336) regarding a TM8 induced-fit can be made, although CAT4 TM8 is not as loose as the xCT TM8 (Figure 5.9). Functional analysis of TM8 CAT4 mutants would be necessary to assess whether this statement is true or a structural extrapolation.

5.2.2.1. Structural Features Stabilising the Outward-Open State

The Thick Gate

The transporter is stabilised in the outward-open conformation by the formation of a tripartite extended salt bridge, consisting of K178-E251-K253, which effectively intracellularly bridges the bundle domain helices, TM6β and TM7, to TM5 from the scaffold domain (Figure 5.10-A,B). Augmenting this is another intracellular salt bridge, E250-R340, bridging TM6β with TM8, respectively. The conformation is further reinforced by the close packing of TM1α and TM7 via an additional D32-R257 salt-bridge interaction. The observation that salt bridges are formed between the scaffold domain helices, TMs 3, 5, 8, 10, and bundle domain helices (TM1-2, TM6-7) (refer to topology diagram Figure 5.5-B) forms a structural basis for the closed intracellular thick gate by close packing of transmembrane helices; explaining the outward-facing state captured.

The outward-occluded substrate-bound state – produced after docking and MD simulation of CAT4 with L-Orn – showed rearrangement of the thick gate

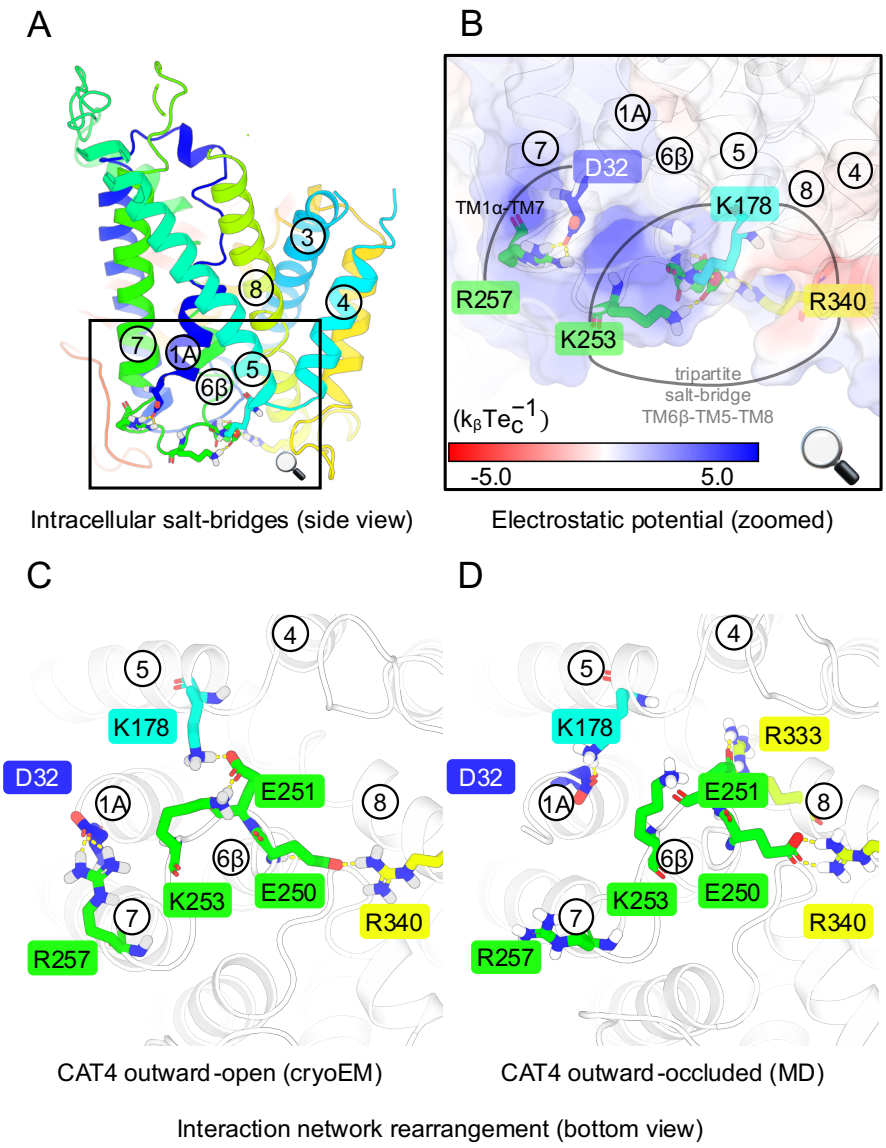


Figure 5.10 Network of intracellular salt-bridges forming the thick gate. (A) Side view of outward-open CAT4 with salt-bridge residues shown in sticks. (B) Zoomed in view of residues shown in (A) with electrostatic potential surface calculated with APBS (Baker et al., 2001). (C,D) Bottom-view of salt-bridge network forming the thick gate with atomistic interactions (yellow dashes) of; the cryoEM outward-open state (C) and; the MD refined L-Orn docked outward-occluded state (D). CAT4 residues shown as sticks and carbons coloured according to TM helix colouring.

interactions (**Figure 5.10-C,D**). The transition from apo- to occluded state requires the right-hand upward "twisting" of TM1 – by 102.0° or approx. one residue– to expose the GAG-motif backbone to L-Orn in order to achieve stable MD simulation of the bound substrate (**Figure 5.8-C**). TM1 rotation upon substrate binding

has been recorded in other SLC7s (Yan *et al.*, 2021). This corkscrew twisting of TM1 repositions D32 (TM1α), breaking the salt bridge with R257 (TM7) and forming one with K178 (TM5). As a result, K178 is unavailable to interact in the tripartite interaction with E251 and K253. Consequently, E251 reorients by forming a new salt-bridge with R333, assisting the E250-R340 interaction in bridging TM6β with TM8. K253 shifts downward and breaks its interactions. Importantly, the described interaction-shift sequence (**Figure 5.10-C,D**) achieves the breaking of the extended network between TM6β and TM7. The deformation of the thick gate is likely an initial event in the opening to the cytosol, suggested to proceed via substrate driven TM1 twisting.

With a transition mechanism suggested for CAT4, we investigated whether other APC transporters show evidence of similar electrostatic networks (**Table 5.4**). Examination of outward-open LeuT revealed a structurally equivalent salt-bridge network. AdiC does not have an identical electrostatics network, but instead has a IL1-TM8 salt bridge (R78-D307) to replace the CAT4 TM6β-TM8 bridge. Although *Gk*ApcT is solved in an inward-facing state, equivalent residues to CAT4 are conserved and could make similar salt-bridges in an outward-facing state. The outward-occluded substrate-bound LAT1 structure has the bridges E265-R355 and E266-R348 (Yan *et al.*, 2021). Collectively, this highlights a conserved role for TM6β-TM8 bridging in gating. Comparison with inward-facing conformations of HATs [LAT1, LAT2 and xCT (PDB ID: 6IRT,7CMI,7P9U)] suggested that TM6β movement results in salt bridges dissociation. In SLC7s, and perhaps other APC members, a well-coordinated intracellular network of coulombic interactions stabilises the tight packing of the helices on the intracellular side, permitting an open arrangement on the extracellular side.

SLC	Structure	Structurally equivalent thick-gate salt bridges		
CAT4	Outward open (cryoEM – apo)	D32-R257 (TM1-TM7)	E250-R340 (TM6β-TM8)	E251-K178 (TM6β-TM5)
	Outward occluded (MD – L-Orn)	D32-K178 (TM1-TM5)	E250-R340 (TM6β-TM8)	E251-R333 (TM6β-TM8)
LeuT	Outward-open (2A65 – L-Leu)	R11-D274 (TM1-TM7)	R5-D369 (N-ter-TM8)	E6-K189 (N-ter-TM5)
LAT1	Outward-occluded (7DSK – Diiodo-Tyr)	–	E265-R355 (TM6β-TM8)	E266-R348 (TM6β-TM8)
*Gk*ApcT	Inward-occluded (5OQT – L-Ala)	D31-R251 (TM1-TM7) D31-R182 (TM1-TM5)	E244-R334 (TM6β-TM8)	E245-R182 (TM6β-TM5)
AdiC	Outward-semioccluded (3OB6)	–	R78-D307 (IL1-TM8)	–

Table 5.4 Salt-bridges structurally equivalent to those identified in CAT4 in other outward-facing APC transporters. The residues shown for *Gk*ApcT are not interacting in the inward-occluded structure but would perhaps form salt-bridges in an outward-facing state.

The Thin Gate

The local resolution at the canonical APC-fold binding site is 2.8-3.0 Å, higher than the overall global resolution of 3.3 Å (**Figure 5.4-A**). The lining of the substrate binding site consists of the TM-helical pairs, $TM1_{\alpha/\beta}/TM6_{\alpha/\beta}$ and TM3/TM8, which in other APC transporters provide the residues implicated in substrate recognition and selectivity.

The thin gate of the transporter, F237, is solved in a swung-open conformation, with $\chi_1 = -53.1°$ (*gauche*$^+$) and $\chi_2 = 130.9°$. In comparison, the Alphafold2 prediction has a different F237 conformation, $\chi_1 = +178.3°$ (*trans*) and

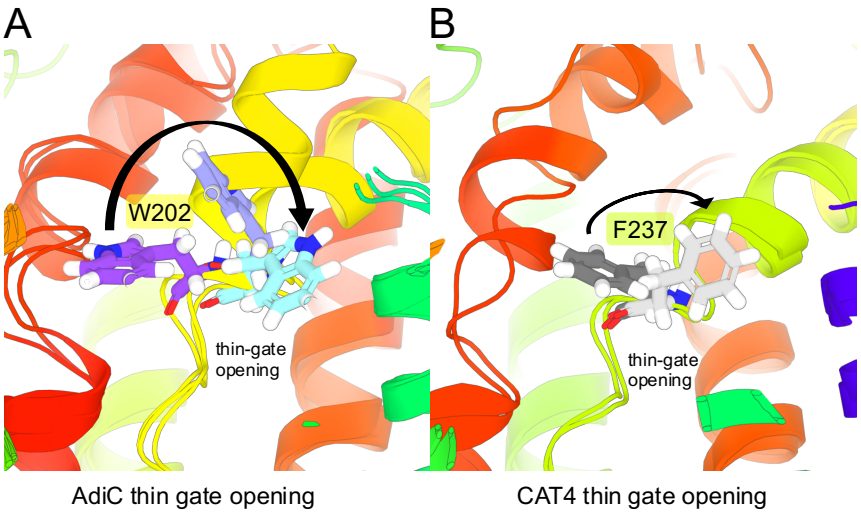


Figure 5.11 Comparison of the thin gate conformations in CAT4 and AdiC. Structures were aligned on their scaffold domain (excludes TM1-2 & TM6-7) to portray differences within the bundle domain. TM1 – red TM6 – yellow.

Open-to-closed transition of: (A) the W202 thin gate in AdiC crystallographic structures and (B) the F237 thin gate in CAT4 cryoEM and Alphafold2-predicted structures.

$\chi_2 = 74.1°$, and a slightly closed TM6 orientation when the scaffold heavy atoms are superimposed ($\Omega_{L225\text{-}F236\text{-}F242\text{-}A249} = 113.6$ vs 127.9) (**Figure 5.8-C**). The movement of the thin gate is characteristic of the binding site state. For reference, the available X-ray structures of AdiC measure W202 dihedral pairs in three states: (a) the outward-facing W202-open state (PDB ID:7OB2) [χ_1, χ_2] ← [–67.2,103.6], (b) in the "intermediate" outward-facing W202-semi-open (PDB ID:3LRB) [χ_1, χ_2] ← [93.0,131.4], and in the outward-facing substrate-occluded W202-closed state (PDB ID: 3L1L) [χ_1, χ_2] ← [164.5,85.6] (**Figure 5.11-A**). The cryoEM CAT4 resembles the AdiC open state and the Alphafold2 CAT4 the AdiC substrate-bound state (**Figure 5.11-B**). The outward-open state captured for CAT4 is the second transporter of the family with the thin gate in an open state, i.e. in a *gauche*$^+$ state, the other being of bacterial AdiC (PDB ID: 3NCY,3LRB,7OB2), which preferentially crystallises as outward-facing in all known structures (Fang *et al.*, 2009; Gao *et al.*, 2009b, 2010; Kowalczyk *et al.*, 2011b; Ilgü *et al.*,

2016, 2021). Overall, SLC7s seem to exchange F237 [χ_1,χ_2] configurations to regulate the binding site accessibility from the extracellular side.

The thin gate in SLC7s is a semiconserved hydrophobic aromatic residue that acts as a gate (Nicolàs-Aragó *et al.*, 2021). However this is not sufficient to qualify as a *hydrophobic gate* by the strict definition; i.e. a mechanism of local "dewetting" of an otherwise hydration-accessible site of a pore, such that dewetting is not superseded by immediate steric occlusion of the pore (Seiferth, Biggin and Tucker, 2022). In the case of APC transporters, the thin gate does not cause dewetting as waters can migrate but acts as a mediator of transporter occupancy state via a van der Waals mechanism.

5.2.3 TM Helices 11&12 and IL5 – Functional Implications

EL2, EL4, IL5 loops could not be modelled

The topology diagram presented in Figure 5.5-B shows the locations of extracellular and intracellular loops for CAT4. Based on the density map, the extracellular loop residues 135-156 (TM3-TM4; EL2), 294-305 (TM7-TM8; EL4) and the long intracellular stretch 409-457 (TM10-TM11; IL5) could not be modelled (Figure 5.8-A). Although for IL5 there was no available density, the two extracellular loops had available densities, albeit diffuse, discontinuous and of variable local quality which prohibited atomistic modelling. The N141 glycosylation site in EL2 and the SybB5 binding EL4 are likely underlying factors affecting the local density quality in these regions. The electron density map corresponding to EL2 and EL4 and the intrinsically disordered propensity of IL5 are shown in Appendix A – AtCAT4 Unmodelled Loops.

IL5 phosphorylation identified by Bottom-Up Mass-Spectrometry

The composition of IL5 is 63.6% polar/charged residues, a positive predictor for disorder allowing unpacking and interaction with water outside of the membrane (Van Der Lee *et al.*, 2014). IL5 has a net charge of –6 (Q_{neg} = -12, Q_{pos} = +6 at pH 7.5, pI = 4.27), and is unlikely to interact with negative lipid headgroups (Tusnády, Dobson and Tompa, 2015), such as phosphoinositides, known for their functional importance (Corradi, Besian I Sejdiu, *et al.*, 2019; Noack and Jaillais, 2020). Expression of IL5 in *E. coli* would be informative as to the isolated behaviour of this domain by crystallography (Kato *et al.*, 2019) or CD. Most likely, IL5 structural propensity could be regulated by unknown binding partners.

To further understand the role of IL5 within CAT4, we proceeded to analyse CAT4 by bottom-up tandem mass spectrometry (LC-MS/MS) to identify possible post-translational modification in the solute-exposed loop. In this experiment purified CAT4 was trypsinised and peptide products were analysed by mass spectrometry. Trypsin cleaves peptide bonds at the C-terminal side of lysine and arginine residues, except for -Arg-Pro- and -Lys-Pro- bonds which are normally resistant to proteolysis (Manea *et al.*, 2007). The term "bottom-up" implies that final information is reconstructed from individually identified peptide fragmentations, typically generated with nanoelectrospray collision-induced dissociation (nESI) (Zhang *et al.*, 2013). CAT4 N-terminal S12 and C- terminal T595, T596 were found to be phosphorylated terminal residues belonging to regions not captured by cryoEM SPA. Interestingly, multiple IL5 residues were found to be phosphorylated. Specifically, two IL5 peptide fragments were recorded in the analysis with the sequences: "N-$_{405}$YVPPDGVPL**S**S**S**SQ**T**L**S**D**T**DESR$_{427}$-C" and

"N-$_{428}$AE**T**ENFLVDAIE**SS**D**S**PLLGNETAR$_{452}$" (residues in bold were interchangeably phosphorylated). In some peptide fragments, double phosphorylation of combinations of the above phosphorylation sites occurred suggesting IL5's potential role as a functional cytosolic domain.

The TM10-IL5-TM11-TM12 Domain's Role in CAT4 Dimerisation

IL5 connects TM10 with TM11, absent in other APC amino acid transporters. A potential clue to understanding IL5 functionality might be linked to the additional TM helices of CATs. *Ab-initio* prediction with Alphafold multimer generated a CAT4 homodimeric model, with the dimerisation interface consisting of the additional TM helices 11 and 12 and stabilised by positively charged residues of one protomer interacting with negatively-charged residues of the other (Figure 5.12). In the cryoEM map, these helices have variable local resolution ranging from 3.5 to 4.5 Å, suggesting high mobility due to insufficient stabilisation; perhaps linked with their putative tendency to heterodimerise. With these observations, a reasonable hypothesis would be that CAT4 has the potential to homodimerise via its additional TM helices; an event perhaps linked to cooperative IL5 phosphorylation and/or folding.

Relipidated CAT4 Gives Rise to Homodimer EM Classes

To further understand the role of TM 11,12 and IL5, a CAT4-SybB5 sample at 3.5 mg/ml, *a posteriori* enriched in soy polar lipid extract was prepared for single particle analysis. Lipids have been shown to drive, increase and modulate the clustering and oligomerisation properties of proteins in the membrane. As CAT4 is a homologue isolated from plant, lipid extract from soy [45.7% PC, 22.1% PE, 18.4% PI, 6.9% PA and 6.9% unattributed (w/w)] was deemed as an

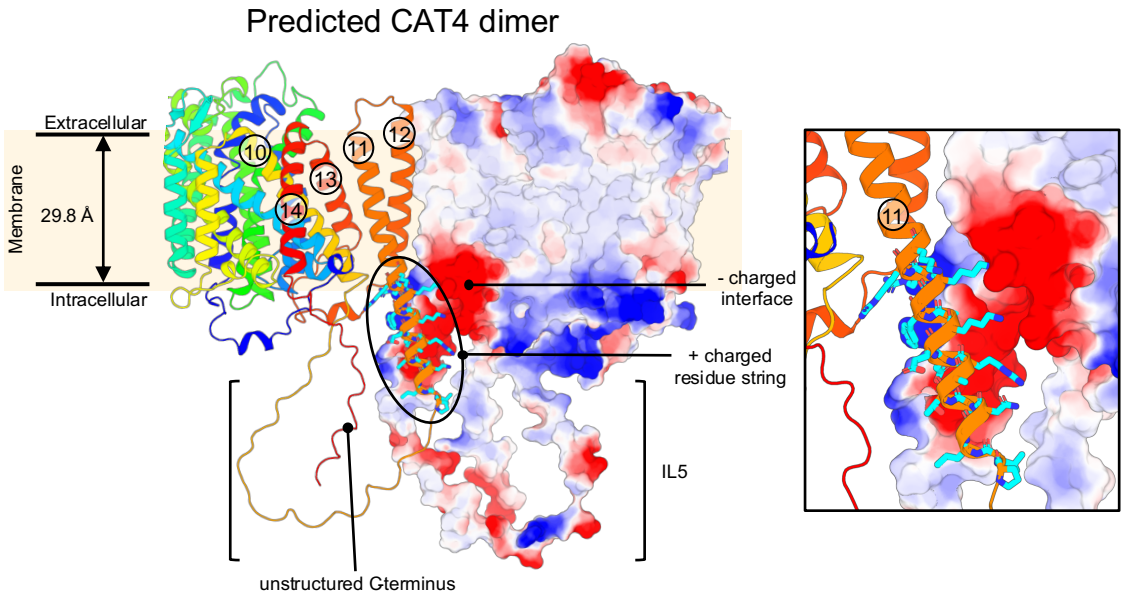


Figure 5.12 Predicted CAT4 homodimer generated with Colabfold2 (Mirdita *et al.*, 2022). The TM11-TM12 interface is shown with positively-charged residues in one protomer contacting a negatively-charged surface in the other protomer (enlarged in inset).

appropriate, native-like supplementation. This was supplemented with 11.27% w/w cholesterol, a steroid lipid recognised for its functional implications across species (Corradi, Besian I Sejdiu, *et al.*, 2019), in order to yield relipidated CAT4 with an array of lipid classes commonly found in the plasma and vacuolar membrane.

The sample was also incubated with 10 mM L-ornithine, a binding amino acid (see §5.3.2.5), to test dimerisation and yield a substrate-bound electron density map. The sample was prepared as the apo- dataset, in Quantifoil R1.2/1.3 Cu grids at 3 and 5 mg/ml, grids ranked with TALOS Arctica Falcon III (Figure 5.13-A), and the best grid subjected to 72 h data collection on Titan Krios equipped with a K3 detector. 2D classification of the extracted particles resulted in largely monomeric CAT4-Syb complexes. Interestingly, with this relipidated sample, previously unobserved dimer classes were observed (Figure 5.13-B). The classes show

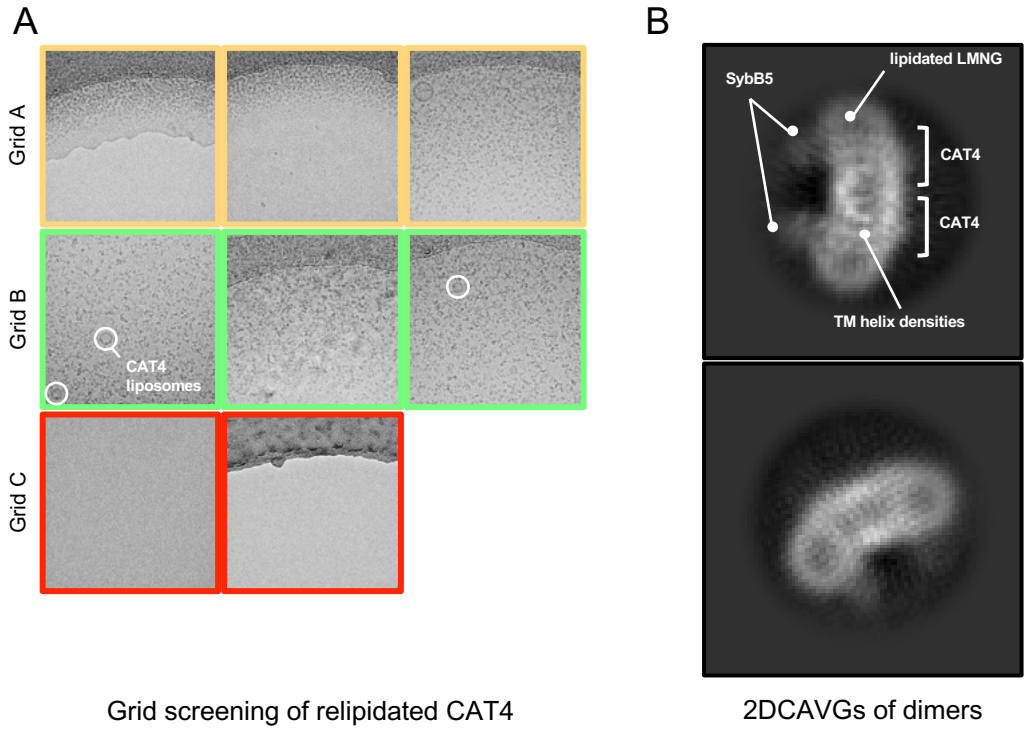


Figure 5.13 CryoEM trials for a CAT4-SybB5 homodimer structure (1:1.2 w/w CAT4:lipid). (A) Exemplary GOOD (green), OK (orange) and POOR (red) micrographs – some self-assembled liposomes showing protein-like densities could be observed. (B) 2DCAVGs of the obtained motion-corrected micrograph containing dimeric species are shown with annotated components.

SybB5 protruding densities facing away from the dimeric interface. In the apo reconstructed map discussed previously, the TM helices furthest away from SybB5 are TMs 11 and 12, suggesting they may be forming the dimer interface. The number of dimeric CAT4 classes was prohibitively low to yield a 3D reconstruction but is suggestive of an *in vitro* formed dimer in a lipidated micellar environment. Increasing lipid:CAT4 ratios did not increase dimer/monomer ratio, studied by Native Blue PAGE (**Figure 5.14**). Therefore, a higher lipid concentration was not investigated. A possible alternative may be disulphide crosslinking of opposing TM11 and TM12 residues based on the ColabFold predicted dimer structure. Collectively, the lower resolution of TM11,12 in the apo- structure, the predicted dimer interface and their absence from HATs suggests these helices are

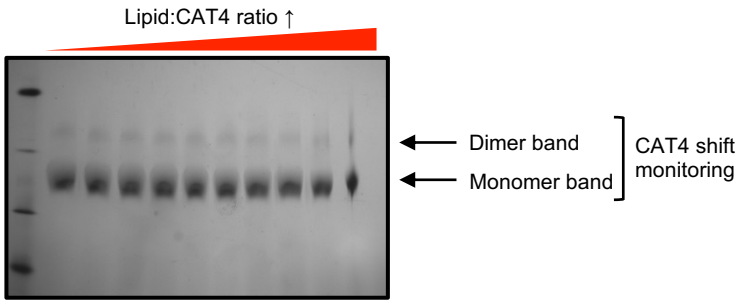


Figure 5.14 NativePAGE analysis of CAT4 dimerisation shift monitoring by increasing lipid:CAT4 ratios – CAT4:lipid w/w ratios from left to right: CAT4-only (delipidated sample), 1:0.01, 1:0.03125, 1:0.0625, 1:0.125, 1:0.25, 1:0.5, 1:1, 1:5, 1:10. No shift is noticeable; monomer/dimer relative populations remain constant.

implicated in the dimerisation process. TM11,12-mediated homodimers stabilised by phosphatidylcholines at the protein interface have been recorded from the SLC12 cation-chloride cotransporter NKCC1 (PDB ID: 6PZT,7ZGO) (Yang, Wang and Cao, 2020; Neumann *et al.*, 2022).

5.3 Substrate Recognition in the Binding Site

The structural data obtained for CAT4 help to form a substrate-binding hypothesis and offers the possibility of investigating the residues involved in a corresponding translocation mechanism. The following section discusses the binding site characteristics, substrate screening and binding assays for CAT4, and the interaction profile of CAT4 via molecular dynamics simulations.

5.3.1 Structural Insights Into Substrate Specificity via the Binding Site

There are 23 proteinogenic amino acids (which differ in size, shape, hydrophobicity, polarity, side-chain charge), non-proteinogenic α-amino acids (e.g., citrulline and ornithine; arginine metabolic products), and unbranched chain amino

acids and analogues where the amino group is in the β or γ position (e.g., taurine, GABA). Carriers vary greatly in substrate specificity with some transporting a single type of amino acid (with extreme selectivity), others accepting almost all amino acids (with varying levels of discrimination), with most falling somewhere in between (Edwards *et al.*, 2018).

Important residues for CAT4 substrate recognition can be classified into two categories. The first comprises of residues chelating the α-amino and carboxy termini of the amino acid substrate. These residues form the helical discontinuities in TM1 and 6 and interact with the substrate via their main chain amido and carbonyl groups. In CAT4 they are T41 G43 A44 G45 (TM1; GAG motif) and I240 S238 F237 (TM6). The second category are residues that are important for substrate side chain specificity. For example, in the L-arginine transporter, *Gk*ApcT (with M321S mutation), these are E115 (TM3) and S321 (TM8). CAT4 conserves these as TM3 D116 and TM8 S327 for cationic substrate selectivity along with the unwound residues of TM1 and 6. Reasonably one can expect that CAT4 is likely to display a substrate preference similar to *Gk*ApcT M321S (**Figure 5.15-A,B**). L-arginine (148 $Å^3$) is the second bulkiest proteinogenic amino acid, second only to L-tryptophan (163 $Å^3$). Since L-arginine fits the M321S *Gk*ApcT binding site and CAT4 has conserved binding site features, L-arginine and smaller cationic amino acids present a reasonable set to test binding to CAT4.

Another key observation in predicting the substrate of orphan CAT4 is that the Poisson-Boltzmann electrostatics modelling of the transporter in a fully solvated membrane – POPC:Cholesterol 4:1 mol and 150 mM NaCl in water – provides an electrical potential gradient for positively charged substrates (**Figure 5.15-**

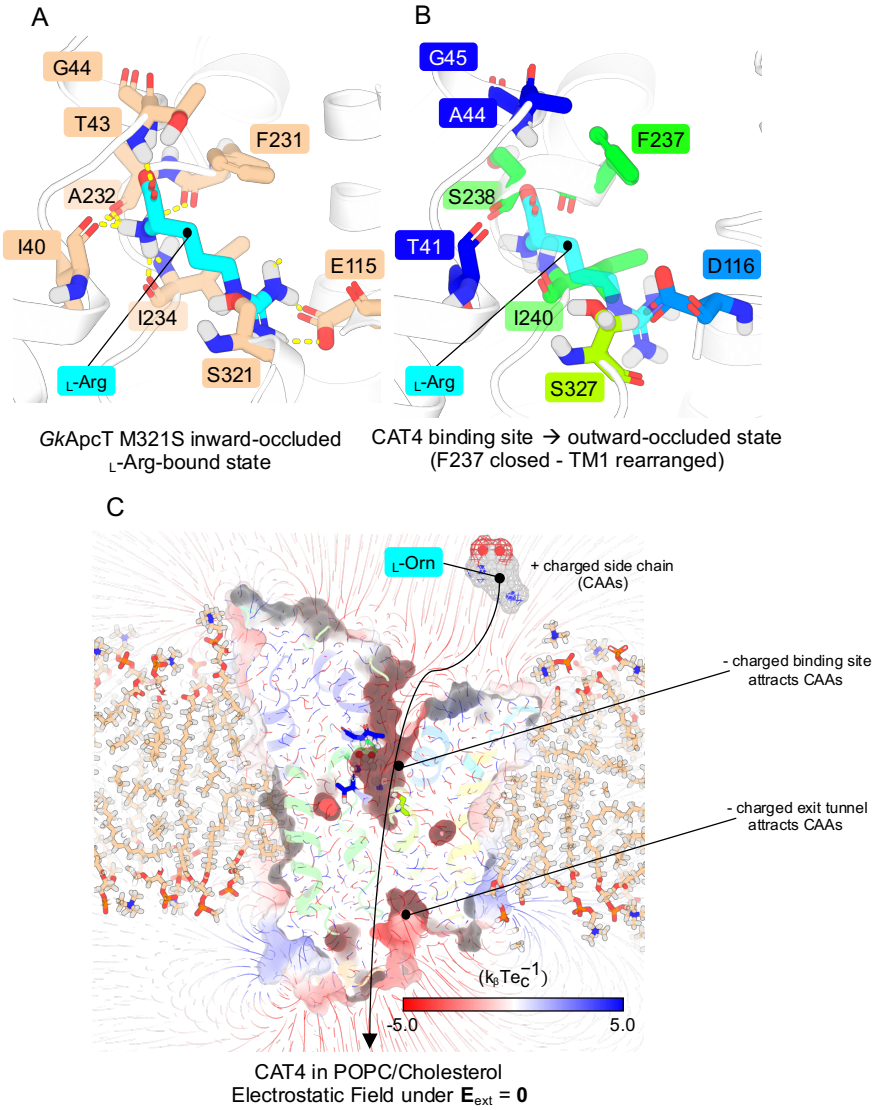


Figure 5.15 Essential binding site ingredients of CATs. (A) The interaction profile of GkApcT M321S with L-arginine, yellow dashes represent H-bonds. (B) Equivalent residues found in the CAT4 binding site with L-arginine shown as semi-transparent to depict that fitting in the CAT4 binding site is possible. The F237 thin gate was closed by setting, $[\chi_1,\chi_2] = [+178.3°,74.1°]$, and TM1 conformation was homology modelled with AF2 prediction as template. The modified CAT4 outward-occluded structure was minimised. (C) Poisson-Boltzmann electrostatic field analysis of a fully lipidated and solvated CAT4 system with L-ornithine docked in the binding site. Field lines calculated as, $\mathbf{E} = -\nabla V_E$. Negative surface charge is calculated in the binding site under conditions of zero external electric field.

C). No external electric field is required for CAT4 to have a negatively potential in the outward-open binding site, increasing the attraction of a positively charged

substrates, as displayed by the electric field lines. Interestingly, the surface charge of the closed inner thick gate is also negative, providing an electrostatic component to the substrate translocation mechanism.

5.3.2 Substrate Binding Assay

5.3.2.1. Assay Development

To begin investigations of the substrate binding properties and preferences of CAT4, a suitable *in vitro* assay needed to be identified. Transport assays requiring radiolabelled amino acid substrates are not cost-effective for the medium throughput required to deorphanise CAT4. A plethora of drug binding assays are available for purified protein studies or even cellular lysates, such as the grating-coupled interferometry, surface plasmon resonance and biolayer interferometry. These techniques are subject to changes in protein refractive index upon drug binding, which would not necessarily be robustly detectable because amino acids substrates:

- are small compared to the transporter to cause significant protein size changes,
- have low affinity to transporters ranging from μM to mM (Errasti-Murugarren *et al.*, 2019a),
- may non-specifically bind which may bring difficulties in gaining reliable data [e.g. KDEL receptor binding non-specifically to small radioactive tripeptides (Bräuer, 2019)].

Differential scanning fluorimetry does not have the above limitations; it is

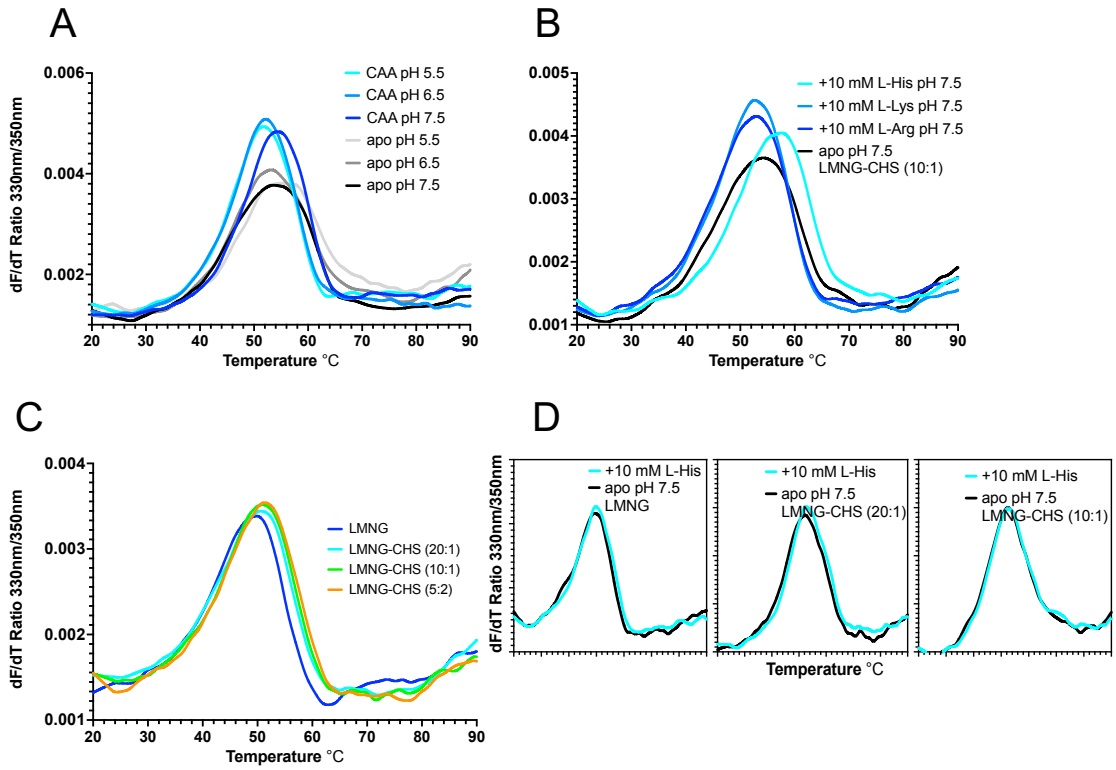


Figure 5.16 Initial pH and detergent optimisation of the differential scanning fluorimetry analysis for substrate binding screening. (A) CAT4 purified in LMNG-CHS (10:1) in pH gradient and in with cationic amino acid (CAA) mix; L-arginine/L-lysine/L-histidine 10 mM each. (B) Screening of individual CAAs at optimal pH 7.5. (C) Detergent belt CHS ratio optimisation at pH 7.5. (D) Repetition of L-histidine shift assay at pH 7.5 with varying CHS concentration; failure of shift reproducibility of L-histidine interaction with CAT4, as originally shown in panel B.

quick, robust, requires limited reagents and is semi-quantitative. An end-point thermostabilisation shift is measured the magnitude of which provides information on the binding of substrates/inhibitors/lipids, demonstrated for other SLC transporters (Crichton *et al.*, 2015; Ruprecht *et al.*, 2019). DSF assays were successful in studying binding of *Gk*ApcT with L-alanine and L-arginine (K. Jungnickel, Parker and Newstead, 2018). Other thermal shift assays (with a thiol reactive probe, CPM) have deorphanised and even thermodynamically characterised solute carriers (Majd *et al.*, 2018; Mavridou *et al.*, 2022).

pH and Detergent Optimisation

At first, the derivative of the tryptophan 330/350 nm readout was used to optimise CAT4 conditions prior to screening amino acids. The derivative trace was preferred for screening and comparisons since, as discussed extensively in §3.1.3.2, its peak height and width provide qualitative metrics of readout quality and protein-substrate thermal cooperativity. CAT4 in LMNG based detergent belt is more thermostable at pH 5.5 by 1.3 °C compared to 7.5, although the peaks have flattened tops making accurate comparison challenging (Figure 5.16-A). The peak quality and the pH preference are reversed upon addition of a cationic amino acid mix – composed of L-arginine/L-lysine/L-histidine at 10 mM each – where CAT4 is thermostabilised at pH 7.5 by 2.4 °C. This shift was the first indication that CAT4 is responsive to external substrate addition and that DSF can be optimised into a binding assay. Seeing the shift with the CAA mix, the next step was to decompose the individual substrate effects. Figure 5.16-B shows that the thermal shift is primarily due to L-histidine. To augment the assay's ability to distinguish binding events the titration and optimisation of CHS in the LMNG detergent belt was done by increasing CHS from 0-20% mol. This resulted in a small thermostabilisation from 49.0 °C in 0% CHS to 52.1 °C in 20% (Figure 5.16-C). However, even with varying CHS concentrations, the CAT4 thermal shift produced by L-histidine could not be reproduced reliably (Figure 5.16-D), suggesting a shift of artefactual nature.

To troubleshoot the L-histidine binding discrepancy, a binding test with all proteinogenic amino acids at 10 mM final concentration was performed, but clear thermal shifts compared to apo CAT4 could not be observed. At this stage it was realised that either the binding assay was not satisfactorily robust or CAT4 may

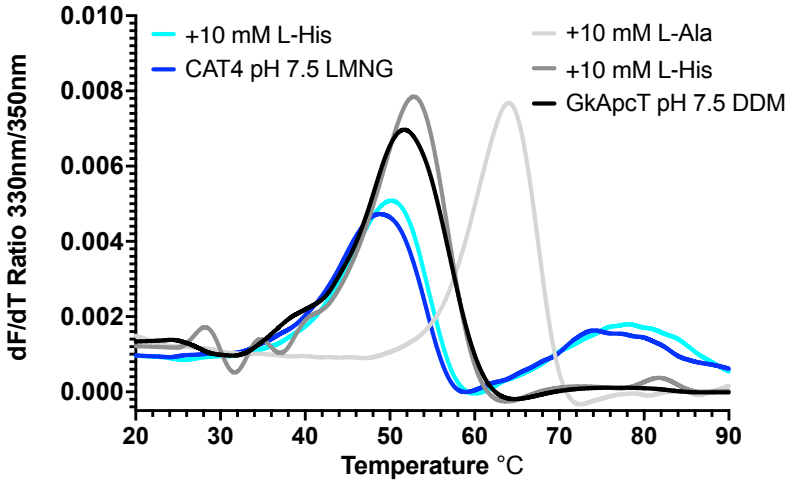


Figure 5.17 Positive control test of the DSF binding assay with the L-alanine *Gk*ApcT transporter and comparison with CAT4 data.

not be responding to substrates as expected. A positive control of the binding assay with purified *Gk*ApcT in DDM produced a significant thermal shift with 10 mM L-alanine, demonstrating that the assay performs as expected for a characterised transporter (**Figure 5.17**). Due to the similarities in the CAT4 and *Gk*ApcT binding sites and overall architecture, it was hypothesised that the assay bottleneck might not be CAT4 itself as it might be the detergent belt, composed of LMNG in CAT4 as opposed to DDM in *Gk*ApcT.

5.3.2.2. nMS Analysis shows LMNG is tightly bound to CAT4

It is a well-known empirical observation that reconstituting membrane proteins from a LMNG detergent belt to a proton-tight proteoliposome is difficult, owing to the detergent's slow off rate, and specialised reconstitution methods are required (Godoy-hernandez *et al.*, 2022). LMNG is known to be a "tight" detergent, accordingly to its development as a highly stabilising detergent for challenging and delicate membrane proteins (White *et al.*, 2012; Kang *et al.*, 2015; Chen *et al.*, 2016).

To investigate whether LMNG may be tightly binding CAT4 – which on one hand may be stabilising but on the other prohibiting responsiveness to substrate – a native mass spectrometry experiment was performed. Native mass spectrometry (nMS) involves the analysis and characterisation of macromolecules, such as proteins and protein complexes, such that the native structural features of the analytes are retained (Keener, Zhang and Marty, 2021; Bennett, Nguyen and Donald, 2022). CAT4 was desalted from Tris buffer and resuspended in a volatile buffer ($AcONH_4$), as required by nMS (Hernández and Robinson, 2007) (**Figure 5.18-A**). The protein provided a relatively clear spectrum showing distinct charge states for CAT4 (**Figure 5.18-B**), with persistent low-level noise in the baseline which did not resolve by modifying recommended instrument parameters (Hernández and Robinson, 2007) (incl. acceleration voltage & pressures, capillary inner diameter, capillary voltage, backing pressure, position of the capillary relative to the cone and the flow of desolvation gas.). To clarify the baseline the level of LMNG was reduced by purifying fresh protein in only 1× CMC_{LMNG}. However rather counterintuitively this attempt caused an increase in baseline noise (**Figure 5.18-C**) and detergent concentration reduction strategies were abandoned. Analysis of the nMS spectrum revealed that LMNG is non-specifically associating to the protein even after applying high desolvating energies to the sample to remove as much adduct detergent as possible without compromising a clear signal for the protein charge states (**Figure 5.18-D**). This demonstrates the high affinity LMNG has for CAT4, potentially causing conformational biasing of the transporter in a semi-locked state.

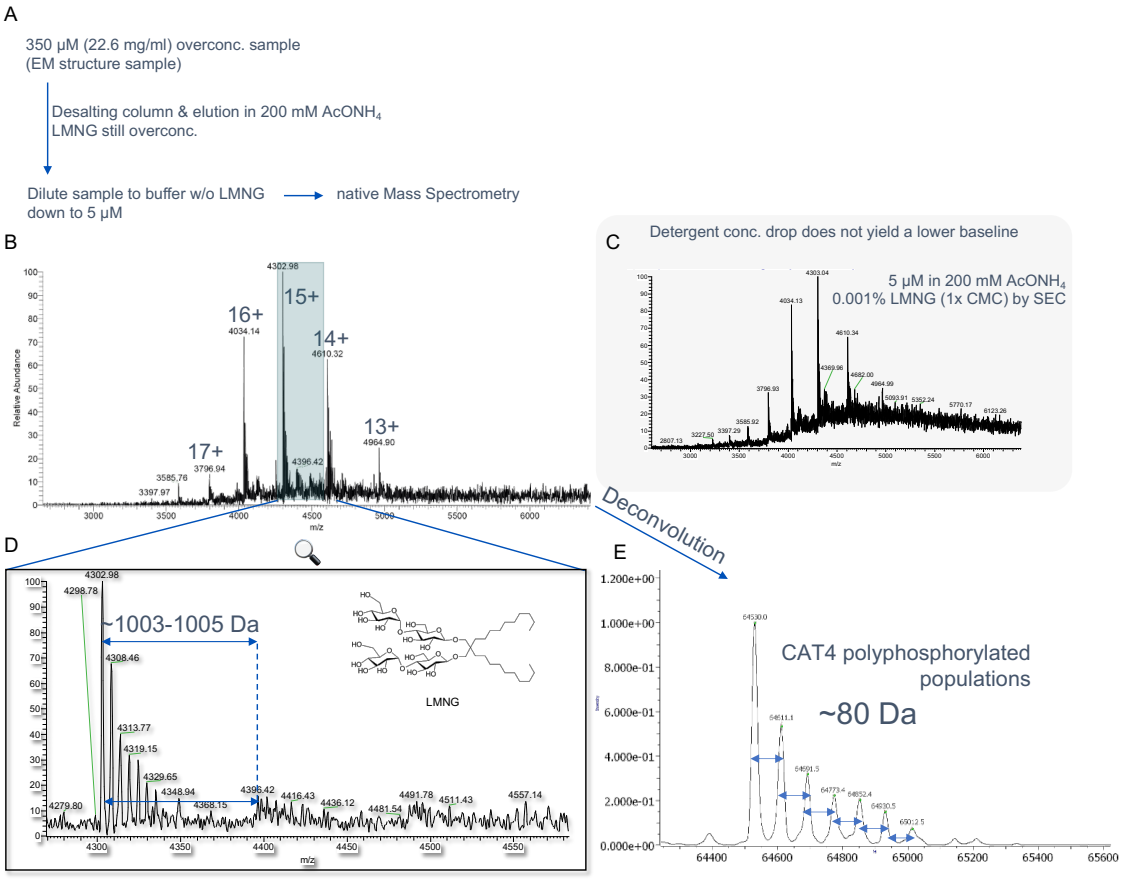


Figure 5.18 nMS analysis of CAT4 in LMNG. (A) CAT4 processing and salt exchange to ammonium acetate pH 6.95, a buffer compatible, due to its votility, with native MS applications. (B) Native MS spectrum acquired for 5 µM CAT4 with reduced LMNG concentration; the charge series detected are highlighted and the most abundant m/z states are shown. (C) Equivalent nMS analysis of CAT4 purified in minimal LMNG concentration, 1× CMC, which retains the LMNG clusters discussed in panel D and shows a more curved, undesirable baseline. (D) Expansion of data in panel B for 15+ charge state, between 4250-4600 highlighting the main population and persistent LMNG clusters. (E) Deconvolution of nMS spectrum from panel B to molecular weight (reported in Da) which shows the CAT4 subpopulations with an approx. 80 Da modification unto the experimental molecular weight MW_{CAT4} = 64530.0 kDa.

5.3.2.3. Relipidation of CAT4 unlocked responsiveness to substrates

The observation that membrane proteins respond to re-lipidation in their micellar form after having been purified has been reported previously. For example, the MscS channel has been solved in multiple conformational states by cryoEM

depending on the lipids retained in the micelle during purification (Flegler, 2021; Flegler *et al.*, 2021). One way the authors regulated the lipids was to change DDM and LMNG concentrations during IMAC, whereas the external addition of soy lipids after purification also proved useful in solving a different conformational state. Another example of relipidation in detergent that resulted in a cryoEM structure is that of the Slo1 K^+ channel (Tao and Mackinnon, 2019).

Gaining inspiration from these studies, CAT4 purified in LMNG was relipidated with soy lipid extract (asolectin), supplemented with 10% w/w cholesterol. The motivation to supplement cholesterol in an SLC7 transporter is clarified in detail in **chapter 6**. Although cryoEM SPA of relipidated CAT4 resulted in the same outward-open apo- structure, repetition of the binding assay at 1:1.2 CAT4:lipids w/w (**Figure 5.19-B**) gave the desirable result. Upon relipidation, CAT4 recovered its binding potency towards cationic amino acids (**Figure 5.19-C**). The highest thermostabilisation of lipidated CAT4 is consistently found with L-ornithine, seconded by L-lysine and thirdly by L-arginine (**Figure 5.19-C,D & Figure 5.20-A,B**). Even delipidated CAT4 binds L-ornithine and shifts by 1.4 °C (compared to delipidated apo CAT4), remarkably more than the other cationic amino acids tested (**Figure 5.20-A**). When relipidated, the thermal shift increases to 4.4 °C (compared to relipidated apo CAT4) followed by L-lysine and L-arginine at 2.6 and 2.4 °C, respectively (**Figure 5.20-A**). These results (a) support the hypothesis formed in **§5.3.2.2** regarding the locking of CAT4 in the tight LMNG micelle, and (b), suggest that DSF is suitable to rank amino acid substrates for CAT4 binding, if the detergent belt is relipidated.

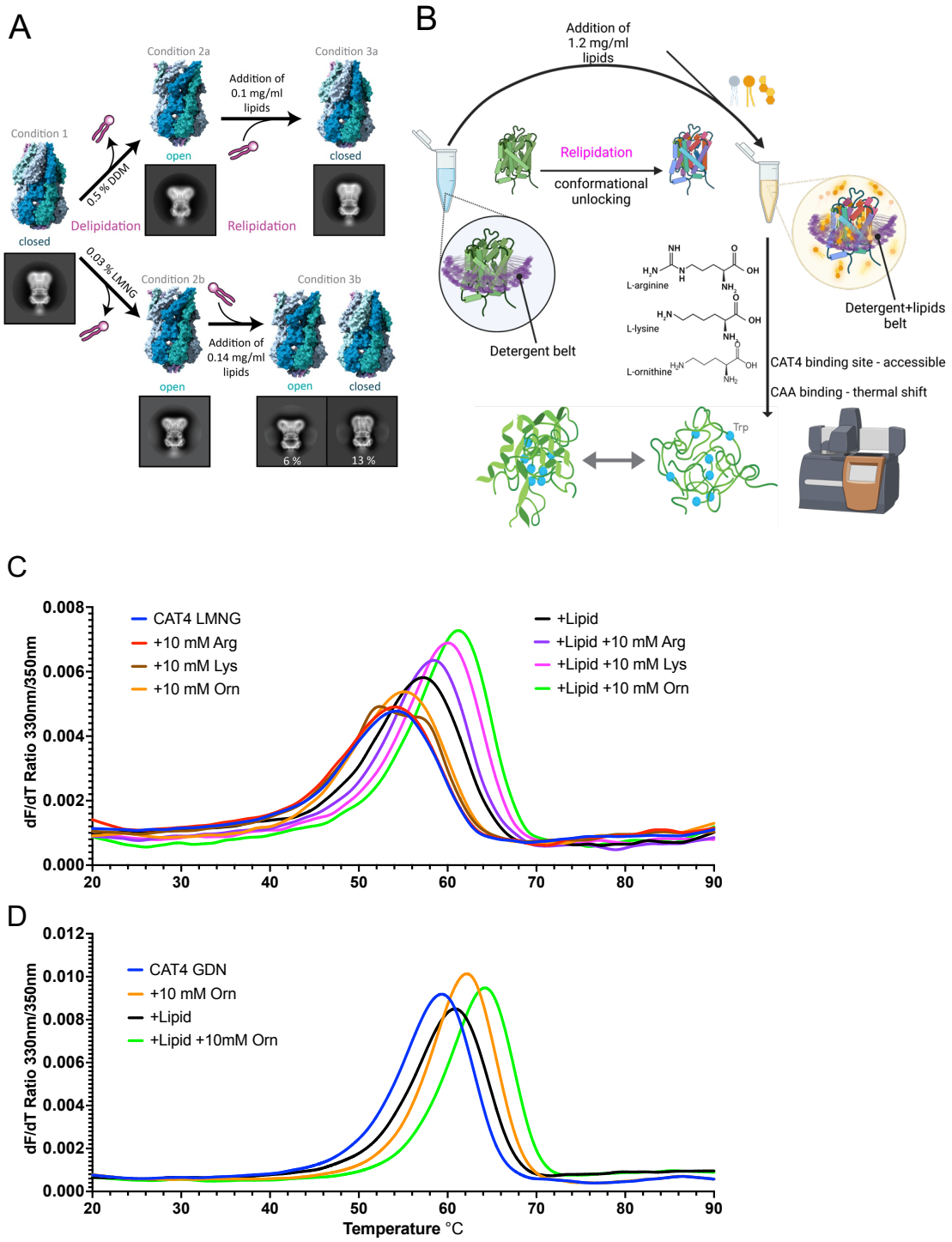


Figure 5.19 Relipidation results for CAT4. (A) Figure from Flegler et al., 2021 showing purified MscS channel conformational state modulation by delipidation and relipidation. (B) Schematic of CAT4 relipidation procedure followed by differential scanning fluorimetry measurements. Prepared with Biorender (except protein unfolding figure from *unchainedlabs.com*) (C) CAT4/LMNG DSF CAA binding assay before and after relipidation with 1.2 mg/ml asolectin-cholesterol lipids. Final CAT4 concentration is 15.7 µM or 1.0 mg/ml. (D) Experiment as in panel C but with the LMNG detergent belt exchanged with the cholesterol-like GDN detergent.

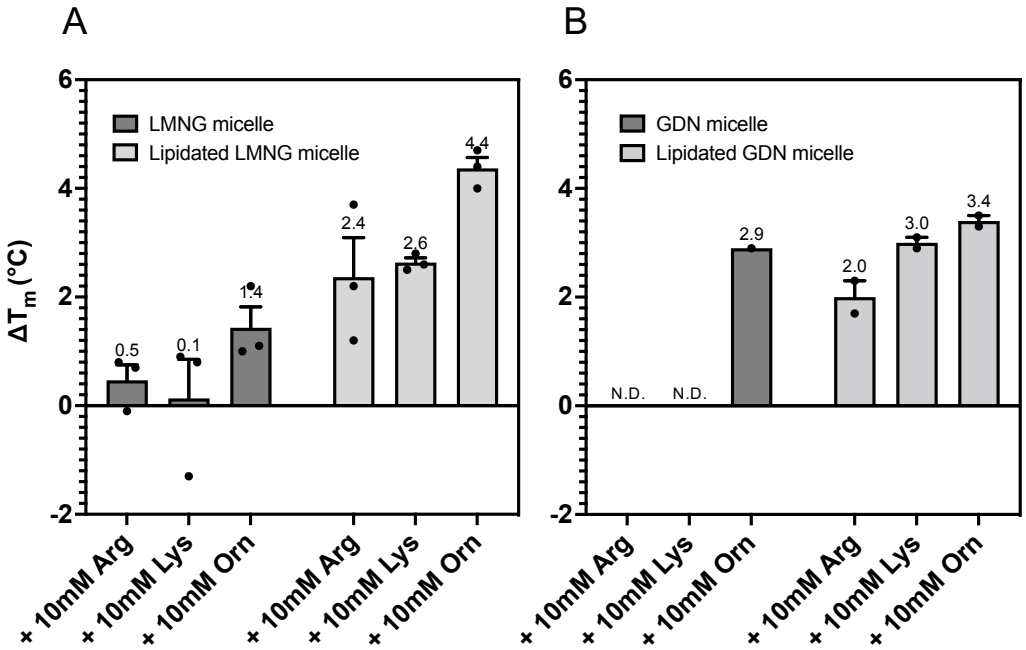


Figure 5.20 ΔT_m thermal shift analysis from DSF of CAT4 with 10 mM cationic amino acids, (A) as purified in LMNG micelle compared to a relipidated LMNG micelle, and (B) as purified in GDN micelle compared to a relipidated GDN micelle. The data presented presents the endpoints of **Figure 5.19-C,D** and technical repeats over different days.

Cross-verification of the above was sought by changing the detergent belt to GDN. As GDN is not known to have very high solubilisation efficiencies, i.e., low extraction yield from host membrane, DDM-CHS (5:2) was used to extract CAT4 and exchanged to GDN during IMAC, overnight dialysis, and gel-filtration to a final 0.05% w/v GDN. CAT4 purified to homogeneity in GDN, with a monodisperse SEC trace and with comparable yields to LMNG (**Appendix A – AtCAT4 Purification in GDN**). Comparison of **Figure 5.19-C** and **D** shows that apo CAT4 is thermostabilised in GDN at pH 7.5 compared to LMNG by 5.1 and 3.7 °C, delipidated and relipidated, respectively. Importantly, responsiveness to amino acid substrate binding is not reserved for the relipidated particles when GDN is used, but instead CAT4 shows a similar thermal shift with the best ranked substrate, L-ornithine (**Figure 5.19-C**), in both delipidated and relipidated GDN, i.e. 2.9 and 3.4 °C (**Figure 5.20-B**). Although the L-ornithine-bound $\Delta\Delta T_{m,relipidation}$ is 0.5 °C

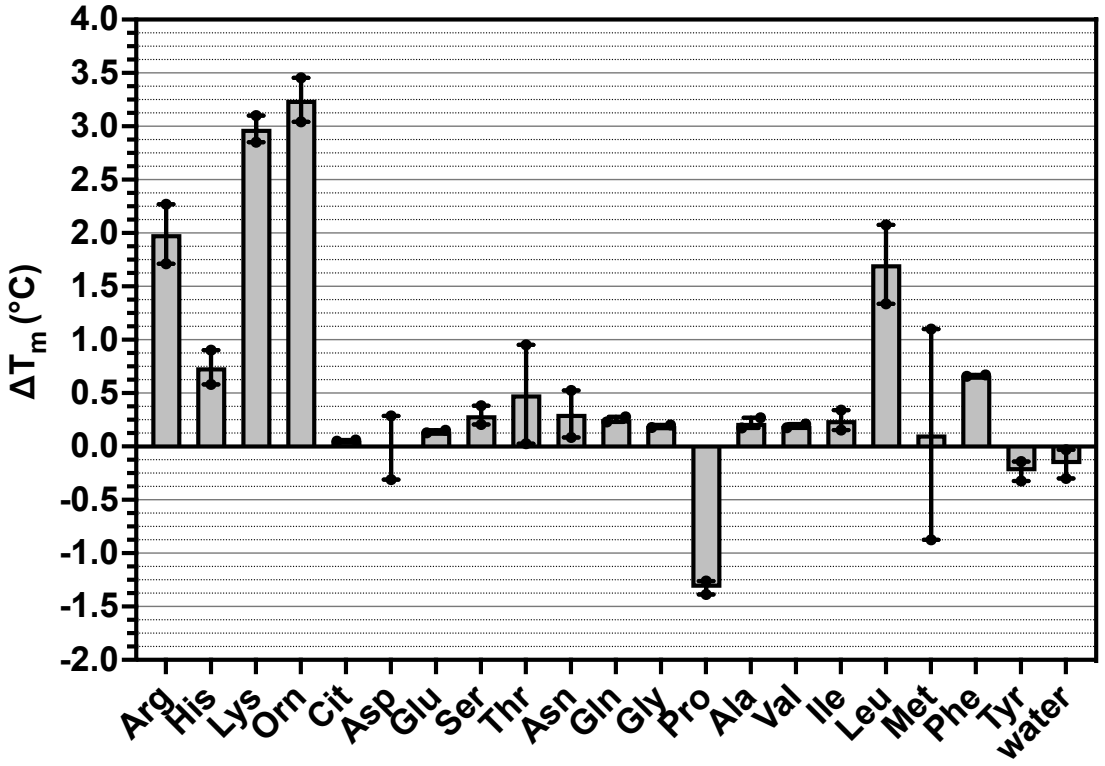


Figure 5.21 CAT4 α-amino acid binding assay – reported as ΔT_m – of all proteinogenic amino acids (excluding L-tryptophan and L-cysteine) and including non-proteinogenic L-ornithine and L-citrulline.

in GDN, it is 3.0 °C in LMNG, highlighting that GDN is flexible enough to allow CAT4 the flexibility to respond to L-ornithine without the relipidation required by LMNG.

5.3.2.4. Amino Acid Binding Analysis to AtCAT4

Having setup the conditions required to measure a detectable thermal shift effect produced by α-amino acid binding to CAT4, an amino acid screen was done at 10 mM – i.e., ~ 640× molar excess ([CAT4] = 15.7 μM). The results of the binding analysis are presented in **Figure 5.21**. The analysis showed that CAT4 binds cationic amino acids in order of apparent melting temperature (here used in equivalence to affinity): ornithine > lysine > arginine > leucine. All the other substrates, including the previously thought substrate L-histidine, produced a

$\Delta T_m < 1.0$ °C. The result that CAT4 preferentially binds cationic amino acids is in line with the literature of human CAT1-3, although the binding with primary amino side chain substrates is pronounced compared to L-arginine; although for mouse CAT1 the three substrates have similar K_M of transport values in oocyte uptake (Wang *et al.*, 1991). Thermal shift with L-leucine is interesting, i.e., there have not been similar observations in the literature.

5.3.2.5. L-Ornithine Binding is Binding Site Specific

Before proceeding with further characterisation of substrate specificity with transport assays, it was necessary to determine whether the observed thermal shifting effects were in fact reporting substrate binding via the APC binding site or whether the effects recorded are non-specific, off-site related. The latter was however not expected due to the high sequence similarity CAT4 shares with other cationic amino acid transporters (chapter 3), exemplified by the structural coincidence shared with *Gk*ApcT M321S L-arginine binding site (§5.3.1).

Binding site mutants of CAT4 were designed and purified, identically to the WT protein. The thin gate mutant F237V was designed as the binding "dead" mutant. The thin gate's role in substrate recognition and transport has been demonstrated in a plethora of SLC7 studies (K. E. J. Jungnickel, Parker and Newstead, 2018; K. Jungnickel, Parker and Newstead, 2018; Errasti-Murugarren *et al.*, 2019a; Lee *et al.*, 2019b), where mutating the aromatic residue results in transport deficiency. The mutation to the similarly hydrophobic Val (Wimley, Creamer and White, 1996), lacking aromaticity, was decided as a semiconservative mutation. Other binding site mutants were D116S, S327M and S327L. The design principles of the mutants are described in Table 5.5.

Mutation	Design justification	Comments
D116S	*Hs Mm Rn Dr* SLC7A1,2,3	Equivalent to H^+-coupling residue in *Gk*ApcT; is a Ser in mammalian CATs
F237V	Binding deficient	Transport "dead" in homologous SLC7 transporters
S327M	WT *Gk*ApcT	Opposite of M321S mutation; shifts *Gk*ApcT to an L-arginine transporter
S327L	*Hs* SLC7A4	Leu in this position is unique in human SLC7A4 among all homologues

Table 5.5 Design principles of binding assay mutants used in DSF assay to investigate the binding site specificity of CAAs to CAT4.

F237V CAT4 did not produce a thermal shift with L-ornithine similar to WT CAT4 (**Figure 5.22-A,D**). However, F237V showed an identical response to relipidation with soy lipid extract as the WT transporter, suggesting that the mutant diverging behaviour is not due to aberrant protein but due to the modulation to the binding site (**Figure 5.22-A**). Similarly, the D116S mutation failed to produce a similar degree of thermal shift as the WT CAT4, even as relipidated to increase responsiveness to substrate (**Figure 5.22-B**). Therefore, the thermal shifting of WT CAT4 with L-ornithine, and by extrapolation likely for other CAAs, is a binding site-specific effect. This highlights that the thermal shift observations are reporters of amino acid binding to the CAT4 binding site. Similarly, the S327 mutations did not produce a thermal shift (**Figure 5.22-C**). A comparative bar plot analysis of the binding site mutants to the WT CAT4 response to L-ornithine is shown in **Figure 5.22-D**.

Both the S327 mutations to Leu and Met abolish their L-ornithine-dependent thermal shift. The S327M mutation behaved according to our expectations, i.e.,

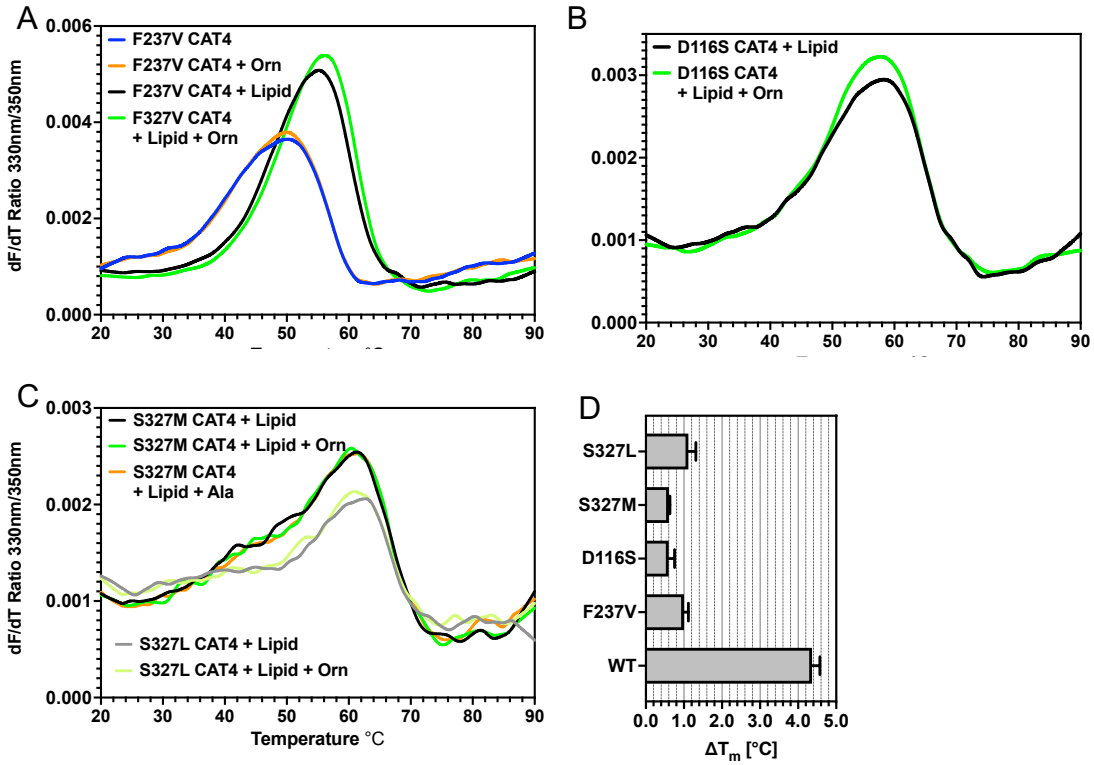


Figure 5.22 Relipidated CAT4 binding site mutant results with the L-ornithine DSF binding assay. Representative binding assay thermal scans of, (A) delipidated and relipidated F237V mutant thermal shift (B) relipidated D116S and (C) relipidated S327M and S327L mutants. (D) Mean value and standard error of the mean of three independent technical repeats of each mutant and comparison to the WT CAT4 binding results ($N = 3$).

to lose CAA recognition due to a bulky Met residue extending in the binding site causing steric hindrance. This agrees with the comparison between WT and M321S *Gk*ApcT crystal structures (K. E. J. Jungnickel, Parker and Newstead, 2018), the former binding L-alanine and the latter L-arginine. However, the S327L mutation lacking L-ornithine thermal shift ability suggests an interesting possible deviation of the human SLC7A4 transporter compared to the plant CAT4 investigated. A hydrophobic Leu residue in this position is exclusive to human SLC7A4, whereas other SLC7A4 homologues, i.e. mouse, rat and zebrafish have a polar Asn [discussed later, in §5.4.2.1 (Figure 5.31)]. A polar residue in this position such as Ser or Asn seems to be the consensus within the CATs. Taken together, a possible different binding specificity for mammalian SLC7A4 might

be expected.

5.3.2.6. Concentration-dependent Binding of Cationic Amino Acids

Cationic amino acid thermal shift of CAT4 is related to binding site residues. The thermal shift is suggested to be largely dependent on specific binding to the binding site. Further characterisation of the binding specificity and the CAT4 binding preferences was carried out by concentration-dependent binding analysis, reported in this section. With the DSF setup, it is expected that thermal shift at each substrate concentration will be positively correlated with the substrate occupancy of the accessible binding sites, available at the selected concentration of CAT4 (15.7 μM). The rate of thermal shift increase with substrate concentration should in theory be larger at lower concentrations and gradually diminish until a plateau, steady state is reached. In this state CAT4 would be fully occupied by the excess substrate. Calculation of the apparent dissociation constant ($K_{d,app}$) with a DSF binding assay has previously been demonstrated (Yoshida *et al.*, 2019), assuming that the unfolding of the protein is reversible. However, this is true only under quasistatic processes (i.e., extremely slow, to maintain equilibrium), thus the term binding thermodynamics is not strictly correct.

The results of the binding analysis are shown in Figure 5.23 and the apparent binding parameters extracted from the fitting in Table 5.6. The results for the substrates show a linear increase of response with concentration, suggesting a non-specific component to thermostabilisation which arises at high amino acid concentrations ($>$ 10 mM). This is more pronounced for L-lysine – by having a steeper slope – whereas it is comparable for L-arginine and L-ornithine – slopes are parallel. The non-specific component justified the kinetics to be modelled with

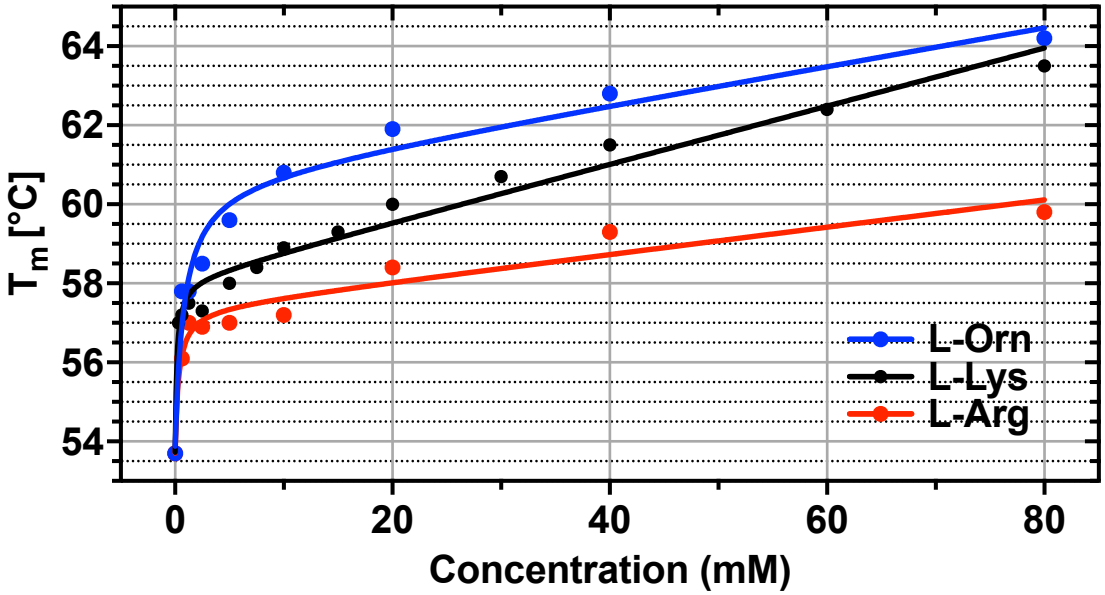


Figure 5.23 Binding analysis of CAAs to CAT4/GDN. A non-specific linear increase in response (presented as T_m) with substrate concentration is seen. The total binding equation, $Y = B_{max} \times \frac{X}{K_d + X} + NS \times X + Background$, including one-site specific and linear non-specific modelling was used. Kinetics modelling was done in Prism 9.

Substrate	B_{max} (°C)	$K_{d,app}$ (μM)	NS (°C/mM)	Background (°C)	R^2
L-arginine	3.683	318.4	0.03425	53.70	0.9622
L-lysine	4.359	153.6	0.07345	53.73	0.9796
L-ornithine	6.804	783.1	0.04809	53.88	0.9755

Table 5.6 Binding parameters extracted from the fitting shown in **Figure 5.23**. The parameters B_{max} and K_d refer to specific binding. Kinetics modelling was done with the One Site – Total binding model of the Saturation Binding analysis module in Prism9.

the equation, $Y = B_{max} \times \frac{X}{K_d + X} + NS \times X + Background$, where B_{max} is the maximum specific binding, K_d is the equilibrium dissociation constant, NS is the slope of nonspecific binding, and background the apparent melting point of lipidated apo CAT4 without substrate.

The binding response, B, in this case the ΔT_m increase with increase of substrate concentration, has a similar initial rate for all three cationic substrates (Figure 5.23). However, B_{max} in decreasing order is L-ornithine > L-lysine > L-arginine, meaning that at the same concentrations L-ornithine occupies more of the available CAT4 binding sites. On the other hand, the extracted apparent K_ds are in decreasing order L-lysine > L-arginine > L-ornithine. The conclusion from this analysis is that L-ornithine is the most thermostabilising substrate for CAT4 while having the lowest apparent affinity for the transporter. From a structural standpoint, L-lysine is longer by one methylene group compared to L-ornithine. This suggests that a longer substrate will bind stronger to CAT4. L-arginine is even longer, but its guanidinium group is likely creating steric hindrance in the binding site. The lack of L-ornithine density in the cryoEM 3D reconstruction is in line with the higher $K_{d,app}$ presented here.

5.3.3 MD Simulations of CAT4-CAA and Comparison to xCT-Cystine

MD simulations of CAT4-cationic amino acid complexes

Molecular Dynamics (MD) simulations were setup to produce the canonical ensemble interaction fingerprint of CAT4 with L-arginine, L-lysine, and L-ornithine (CAAs). As a structure of CAT4 bound with L-ornithine was not afforded with the relipidation treatment of the cryoEM sample (§5.2.3), the cationic amino acids were docked into the outward-open CAT4 structure with F237 thin gate dihedrals adjusted to the closed state. To this end, the coordinates of $b^{0,+}$AT-rBAT+L-Arg complex [PDB: 6li9] were used as a docking reference guide as we have previously described for SLC7A11 L-cystine docking (Parker, Deme, Kolokouris, *et al.*, 2021). We used pre-existing structural data of amino acid

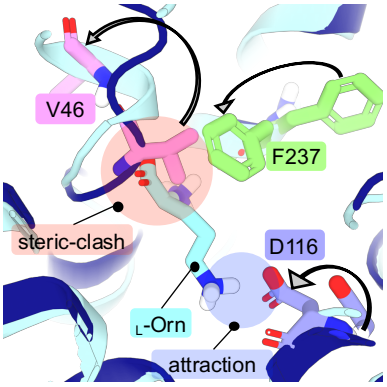


Figure 5.24 CAT4 outward open, modified to accept substrates. Binding site modifications to the outward-open cryoEM structure required to simulate steady L-ornithine binding; including TM1 righthanded upwards twisting to match the Alphafold2 structure (uniprot entry: Q8W4K3), D116 nearing L-ornithine, and F237 thin gate closure.

binding interactions to other SLC7 members to guide the docking grid. The docked substrates were subjected to MD simulation.

Initial MD of L-ornithine led to dissociation of the substrate and exit into the water bulk; a result which is usually attributed to overlooking system setup details rather than rapid unbinding in the short simulation time observed (within 50 ns). Careful inspection of the trajectory revealed that due to the corkscrew TM1 motion previously described in §5.2.2.1, V46 was occupying the ligand cavity and creating steric clashes with the substrate (**Figure 5.24**). This suggests an interesting mechanism by which the apo state is stabilised by V46, a bulky hydrophobic residue, partially occupying the substrate binding site while the F237 gate is swung-open.

By modelling the TM1 helix to match that of the Alphafold2 outward-occluded prediction, stable substrate binding simulation was observed. Interaction fingerprints of the three binding ligands are shown in bar plot format in **Figure 5.25** and in 2D format in **Figure 5.26-B,C,D**. The interactions are similar with

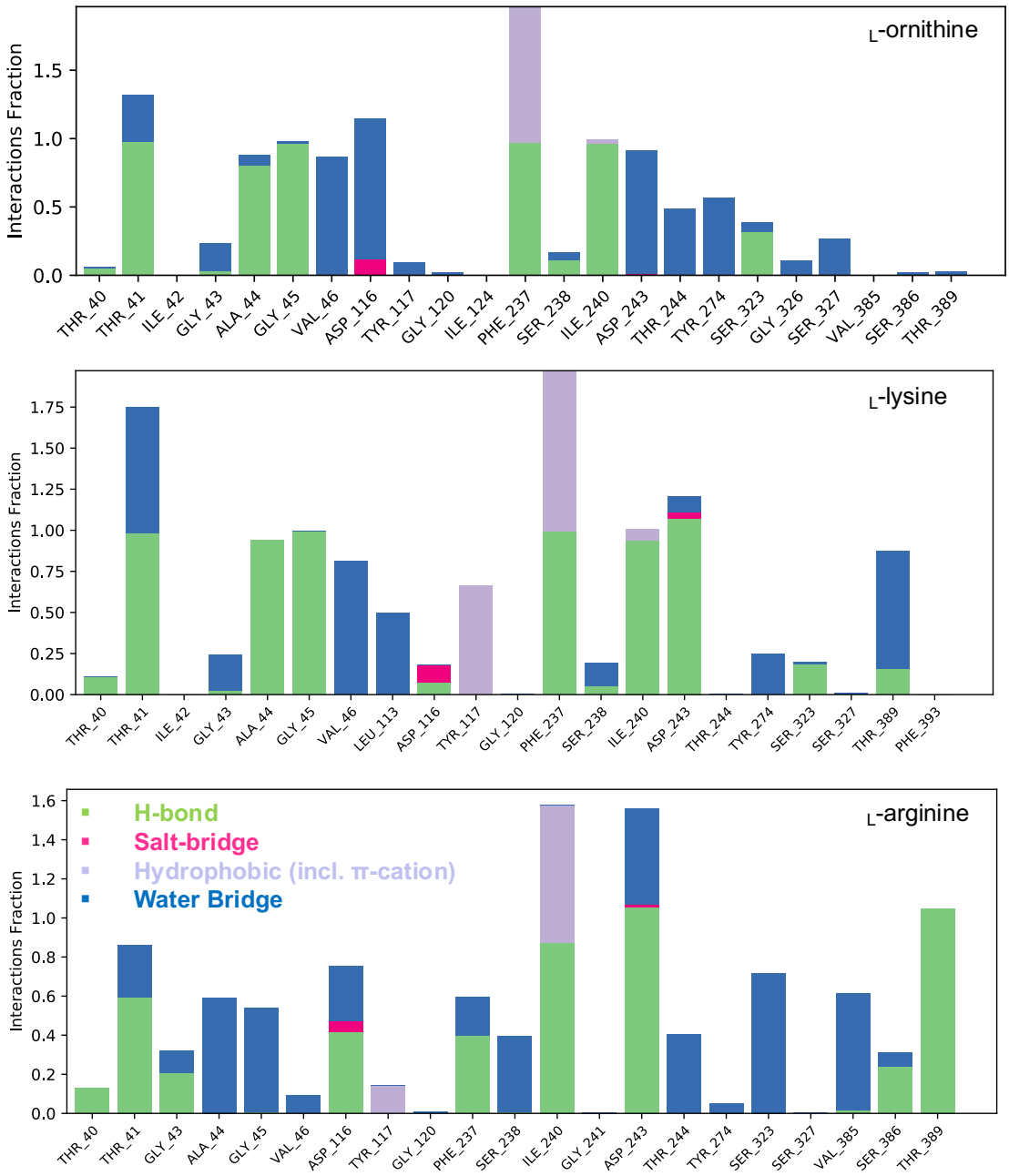


Figure 5.25 MD analysis of L-ornithine L-lysine and L-arginine-docked. Bar plot representation of the time-resolved (500 ns) CAT4 residue interaction fractions with substrates. The interactions are uncapped (can be > 1.0) when more than one interaction between a residue and the substrate can occur in an MD frame. The first 50 ns of production were discarded from the analysis as equilibration time.

regards to their carboxy and α-amino groups, forming degenerate interactions with TM1 and TM6, including TM1α residues T41, G43, A44, G45 and TM6

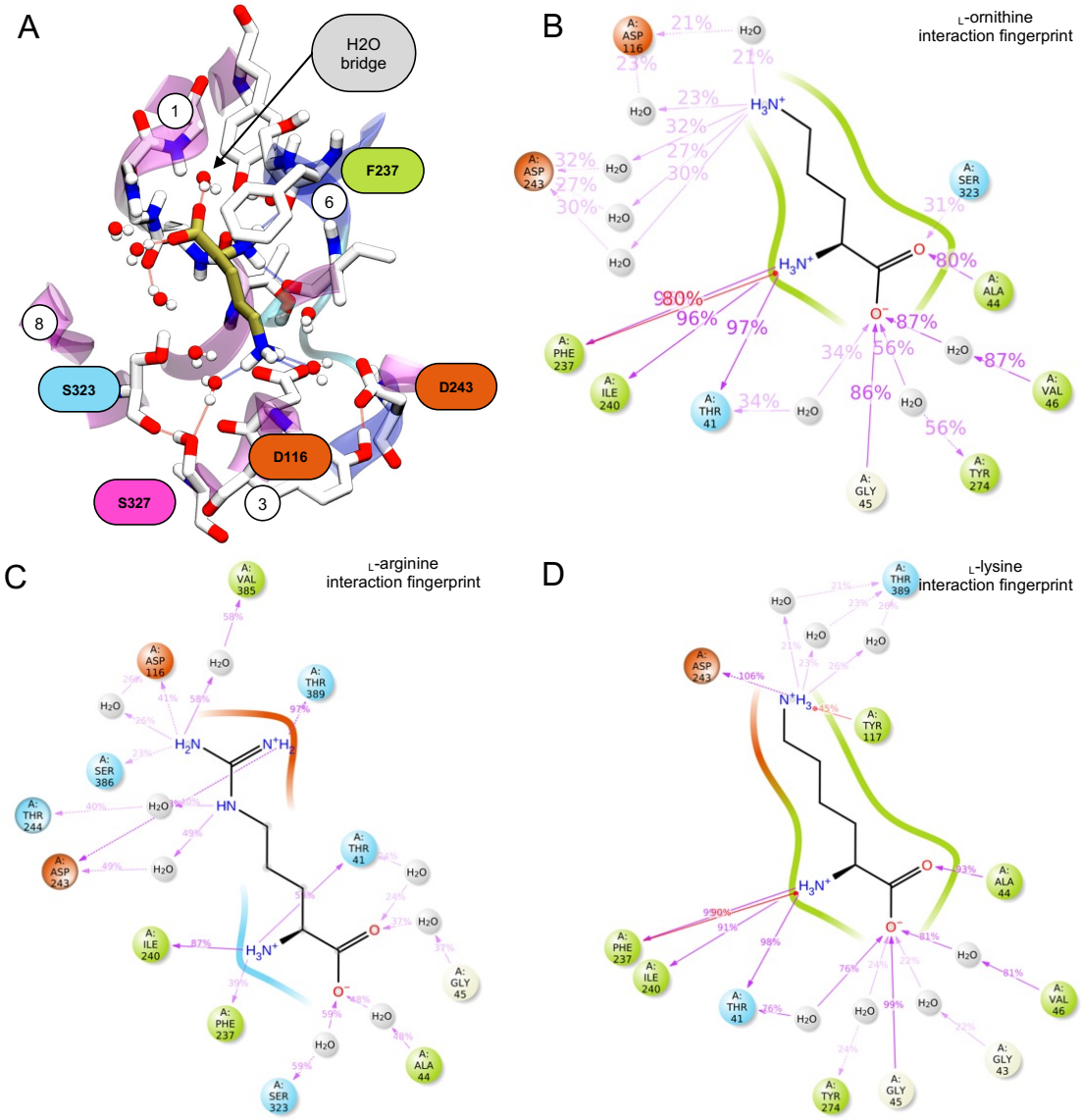


Figure 5.26 Interaction fingerprints of CAAs analysed. (A) Representative pose for L-ornithine binding to CAT4. 2D schematic of substrate interactions with CAT4 residues and water networks created, with freq% of the interactions during 500 ns simulation reported for (B) L-ornithine, (C) L-arginine and (D) L-lysine.

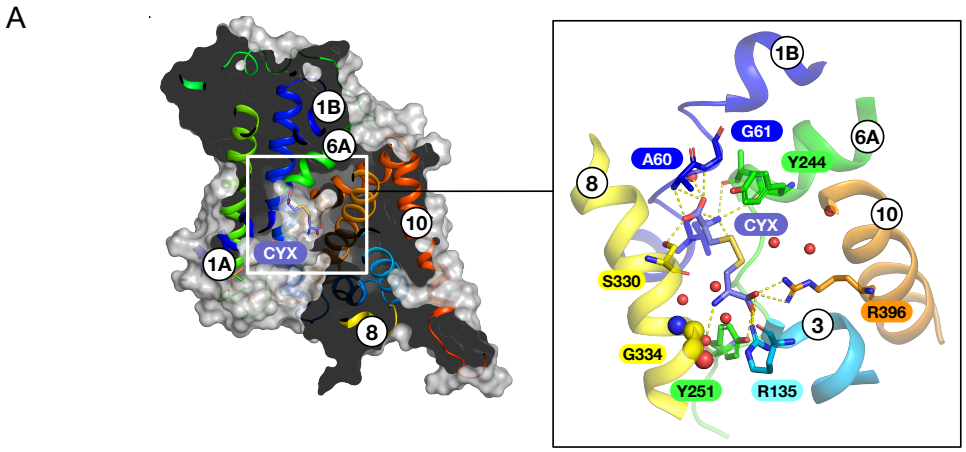
residues S238, F237 and I240. Other common interactions involving the substrates' side chains are with T389 (TM10; only lysine and arginine), D116 (TM3; only ornithine and arginine) and D243 (TM6α; all substrates). Ornithine interacts with D116 (H-bonds, water- bridges, salt-bridge) and D243 via a water bridge, as it is not long enough to reach this remote residue directly (**Figure 5.26-A**). Arginine also interacts with both aspartate residues, but due to its longer side

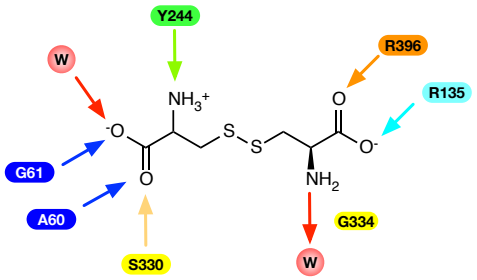
chain can reach deep enough to D243 to make H-bonds. Lysine has a preference for D243 over D116 and is unique in making a π-cation interaction with Y117 (TM3). Water-bridge interactions with TM8 residue S237 is only seen for ornithine, as arginine and lysine orient their side chains to interact with D243 in TM6α instead. The ornithine side chain sits in between TM3 and TM8 making a water bridge with both. Due to their longer side chains, arginine and lysine can also reach TM10 by H-bonding with T389.

From this analysis it is clear that the binding modes of the three substrates have similarities and differences, which likely explain the differences in $K_{d,app}$ and B_{max} in the binding assay. An energetic approach such as absolute binding free energy calculations, which have been used in transporter-substrate studies (Wu *et al.*, 2019; Parker, Deme, Kuteyi, *et al.*, 2021) would enable ranking of substrates and provide insight into selectivity, but has not yet been explored.

MD Simulations of xCT-cystine complexes show similarities with CAT4-CAA complexes

The docking-MD approach, combined with approximate end-point time-resolved MMPBSA calculations, was used for the xCT-CD98hc heterodimer to analyse residues giving rise to the specificity to the anionic form of L-cystine, which predominates at physiological pH (Parker, Deme, Kolokouris, *et al.*, 2021). L-cystine makes several interactions within the binding site that correlate with CAT4 (Figure 5.27-A,B). The α-carboxylate group interacts with the backbone amines of A60 and G61 in the discontinuous region of TM1 (A44, G45 in CAT4). Further interactions were observed to the side chain hydroxyl of S330 on TM8 (S327 in CAT4) and a water molecule, which coordinates interactions to Y281





C

	ΔΔG (WT→R135A) (kcal/mol)	ΔΔG (WT→R396A) (kcal/mol)
Run 1	8.9 ± 3.9	7.8 ± 3.8
Run 2	6.4 ± 2.5	8.4 ± 2.6
Run 3	5.9 ± 3.1	9.0 ± 2.4
Average	**7.1 ± 4.5**	**8.4 ± 4.5**

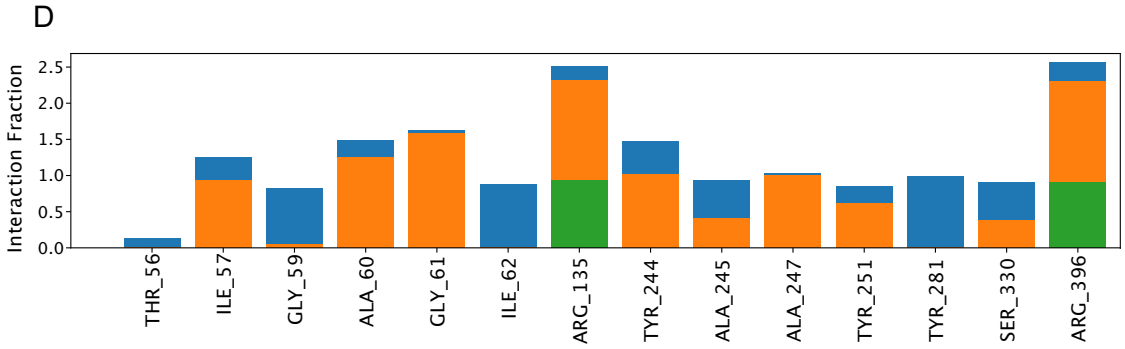


Figure 5.27 MD simulation analysis of L-cystine binding to xCT. (A) Slice through the molecular surface representation of xCT indicating the position of final stable pose obtained for L-cystine (CYX). Inset shows polar and hydrogen bond interactions (dashed lines).(B) Schematic of L-cystine binding interactions, arrows indicate direction of hydrogen bonds. W refers to water molecules, key amino acids are labelled. (C) The calculated binding free energies observed for L-cystine from three independent 200 ns simulations (Runs 1-3) in the WT and mutated models. (D) Interaction fingerprint, calculated as a fraction of the total simulation time from 3 unbiased simulations (3 × 200 = 600 ns) – orange: H-bonds, blue: water bridge, green: salt bridge interactions.

on TM7 and the discontinuous region of TM1 (Y274 in CAT4 seen in L-Orn and L-Lys). At the other end of the binding site, L-cystine interacts with R135 (TM3,

D116 in CAT4) and R396 (TM10; T389 in CAT4), which act as hydrogen bond donors and salt bridge partners. Molecular mechanics interaction frequencies and MMPBSA analysis suggests R135 and R396 contribute equally to the binding free energy within the error of the method (Figure 5.27-C,D), highlighting the importance of the cationic arginine clamp for L-cystine specificity. This clamp is unique to xCT and absent from other SLC7 members. Although a similar energetic analysis is pending for CAT4, we suggest that a similar anionic clamp mechanism involving D116 and D243 may be a selectivity filter for longer CAAs L-arginine and L-lysine. L-ornithine, being shorter in length, could compensate with an extended water-bridge network explaining its lower apparent affinity (§5.3.2.6).

Taken together, the CAT4 binding site residues along with the electrostatic modelling of the outward-open state and the specific binding site interactions provide clear suggestion that CAT4 is a cationic amino acid transporter, similar to hCATs 1-3, which are L-arginine, L-lysine, and L-ornithine transporters. In addition, commonalities in substrate specificity mechanisms are observed between CAT4 and xCT with TM1 and 6 offering recognition of α-amino acids and TM3, 6, 8 and 10 providing residues to engulf the side chain and providing substrate specificity.

5.4 Progress Towards Functional Characterisation of SLC7A4

The definitive experiment to understand the functional behaviour of a membrane transporter and to deorphanise SLC7A4, is a substrate transport assay. This serves to demonstrate not only that SLC7A4 binds CAAs, but also that it

can transport them through the membrane. In the previous section, the results demonstrate that plant CAT4 binds cationic amino acids when unlocked by relipidation and offer specific interactions in the binding site. The hypothesis was then formed that human SLC7A4 would be able to show transport activity of cationic amino acids in an *in vitro* setting. Towards this end, a functional assay for SLC7A4 would need to be developed. This paragraph describes the progress made towards this end, which however proved to have unforeseen obstacles.

5.4.1 Electrophysiological Characterisastion

An electrophysiological approach to functional characterisation was initially deemed appropriate for CAT4 as it was expected to be a proton-coupled cationic amino acid transporter, since:

- the D116 residue of CAT4 is structurally equivalent to E125, the proton-coupling residue in *Gk*ApcT,
- CAT4 conserves D243 which in *Gk*ApcT was shown to be important for the proton translocation pathway, and
- the above arguments hold for the human SLC7A4.

Here, the activity of human SLC7A4 is characterised using Two Electrode Voltage-Clamp (TEVC) in *X. laevis* oocytes. TEVC was chosen for characterisation as: TEVC enables the detection of weak proton currents, CAT activity can be stimulated by cell membrane hyperpolarisation, and human SLC7A4 expresses in oocytes (Kavanaugh, 1993; Nawrath *et al.*, 2000; Ruivo *et al.*, 2012). The method has been used for H^+-coupled transport of glycyl-sarcosine by PepT2 (Beale *et al.*, 2015). TEVC measurements were undertaken by D.Phil student, Sacha Salphati. The interpretation of these data is shared.

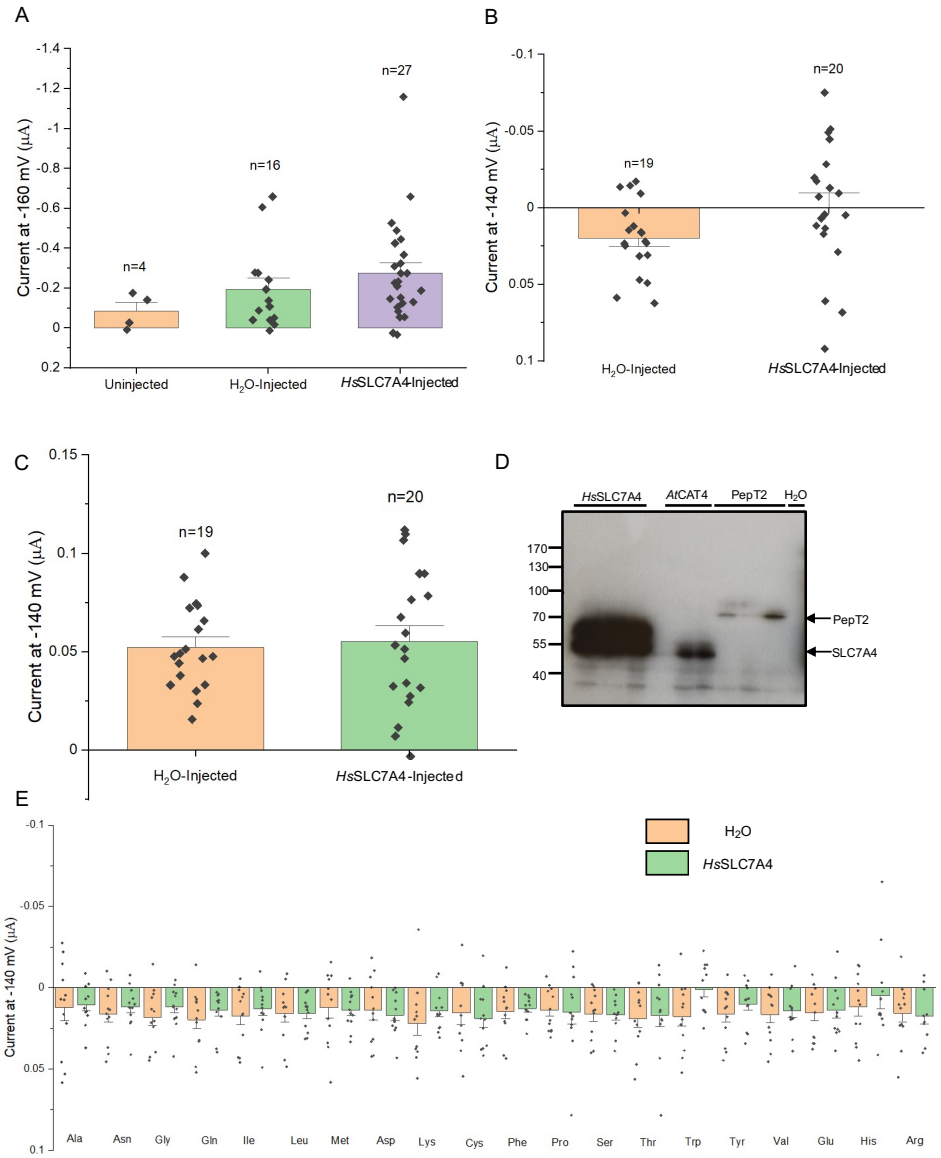


Figure 5.28 Human SLC7A4 TEVC recordings. (A) Currents measured at –160 mV with 10 mM L-lysine pH 6.5 (–180 to +20 mV at 10 mV steps). (B) Currents measured at –140 mV with 10 mM lysine pH 6.5 (–160 to +40 mV at 10 mV steps). (C) Currents measured at –140 mV without substrate injection (–160 to +40 mV at 10 mV steps). (D) Anti-FLAG western blot of water, PepT2, human and plant SLC7A4 injected oocytes. (E) All amino acids currents recording at –140 mV (–160 to +40 mV at 10 mV steps, charged: 5 mM, Cys: 1 mM, remaining: 10 mM). Biological repeats N=2. $^{*}P< 0.05$ with a two-sample t-test assuming normality; tested with Kolmogorov-Smirnov Test. Figure kindly provided by Sacha Salphati.

In preliminary experiments, SLC7A4 failed to produce currents with 10 mM L-lysine at –140 mV, consistently with other oocyte TEVC studies (Sperandeo

et al., 1998; Wolf *et al.*, 2002). A current was generated with the addition of 10 mM L-lysine at –180 mV, although this was identified to be an injection-induced artifact (–0.180 A, SEM ± 0.050 A) (Figure 5.28-A). The protocol was optimised to avoid artefactual recordings by starting from –160 mV and in some cases, by lowering the amino acid concentration to 5 mM; however, no significant currents were generated at –140 mV with L-lysine (Figure 5.28-B). Unlike the proton-coupled PepT2, SLC7A4 was unable to produce currents in the absence of any substrate (Figure 5.28-C). Subsequently, all amino acids were tested (Figure 5.28-E). Only 5 mM tryptophan was found to produce a statistically significant effect (p=0.03671) (0.001 A, SEM ± 0.004 A) relative to the water-injected control (0.018 A, SEM ± 0.006 A). However, the result was not reproducible upon repetition of the experiment. Apart from the small tryptophan current (–17 nA), no other potential ligands could be identified. A whole-cell anti-FLAG western blot showed that the lack of strong currents was not a result of lack of SLC7A4 expression (Figure 5.28-D).

*At*CAT4 did not generate currents greater than –25 nA in the presence of any amino acid at –140 mV (Appendix A – Electrophysiological Characterisation of AtCAT4 with TEVC). *At*CAT4 may generate currents that are beyond the detection threshold of the assay for the conditions used. The lack of observable currents is not the result of weak *At*CAT4 expression as whole-oocyte western blot demonstrates expression of *At*CAT4 (Figure 5.28-D). It is possible that the plant transporter does not reach the plasma membrane.

Radioactive uptake assay characterisation in *X. laevis* of human SLC7A4 has

been reported in the literature previously, wherein no transport activity was recorded with proteinogenic amino acids and polyamines, even though the protein expressed to the oocyte plasma membrane (Wolf *et al.*, 2002). This result is in agreement with the observations made here. Similarly, arabidopsis CAT1 and CAT5 were shown to be L-arginine transporters by transport assays carried out in *S. cerevisiae*, but failed to produce currents in *X. laevis* oocytes (Y.-H. Su, Frommer and Ludewig, 2004). Taken together, our observations and the literature seem to suggest that *X.* oocytes patch clamp TEVC may not be suitable for the functional characterisation of SLC7A4 homologues.

5.4.2 Whole-Cell Radioactive Uptake Assays

L-arginine and L-ornithine uptake assays in HeLa cells expressing SLC7A4 were undertaken. This approach has recently been used to characterise the substrate preferences of SLC7A8, large neutral amino acid transporter-2 (Rodriguez *et al.*, 2021). In these assays, one typically expects 10-20% uptake of the supplied amino acid with an overexpressed transporter which is a good indication of protein-substrate specificity.

Preliminary uptake experiments with L-arginine showed no uptake at pH 5.5-7.5. Uptake via human SLC7A4 was tested with L-ornithine (**Figure 5.29-A**). At pH 7.5 there was no recorded uptake of L-ornithine compared to the controls, including the thin gate mutant Y246V. At acidic pH 5.5, SLC7A4-dependent uptake was recorded, not seen in the transport-dead mutant, albeit, limited to $<$ 1% of available substrate. This very low capacity of transport suggested that human SLC7A4 does not transport L-ornithine. We tested transport with 10 µM

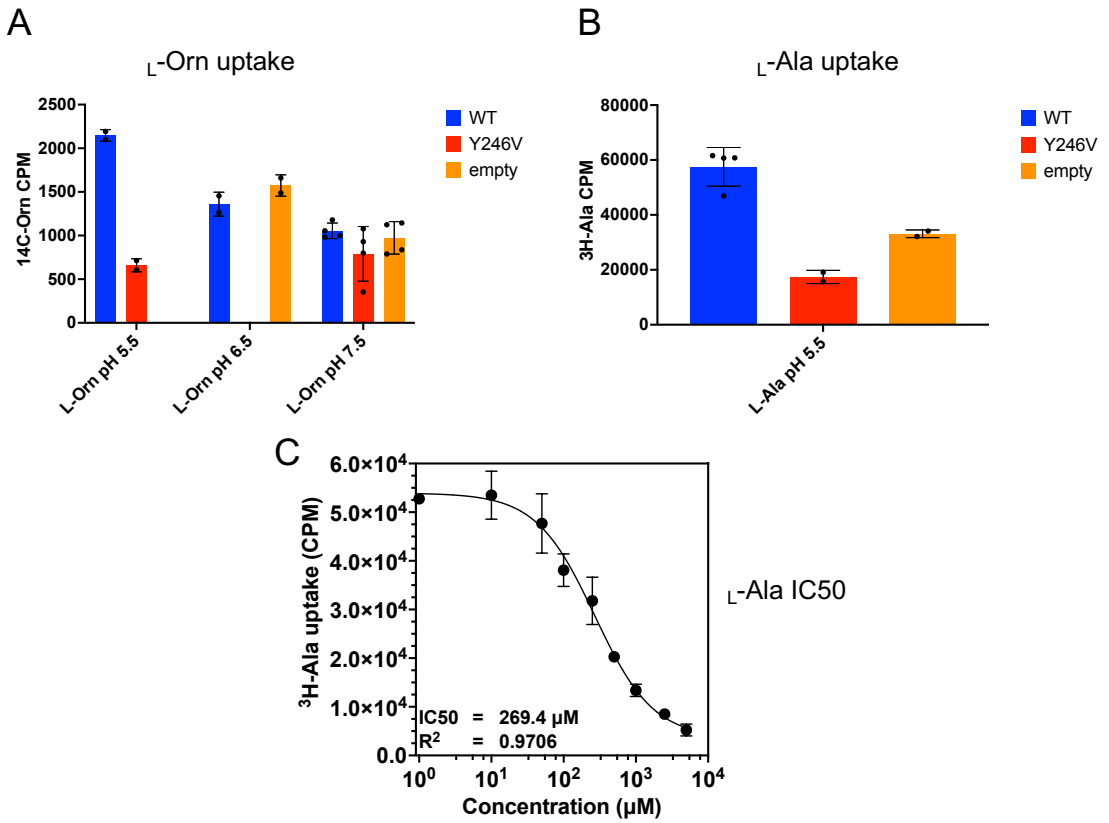


Figure 5.29 Transport of 10-µM radiolabelled (3H) L-alanine, L-isoleucine, and (14C) L-ornithine in HeLa cells expressing WT human SLC7A4 (blue bars), the transport dead Y246V human SLC7A4 (red bars) and negative, vector-only control (orange bars). Data are expressed as mean ± SEM from at least two biological repeats (independent experiments). (A) pH optimisation for SLC7A4 L-ornithine transport, (B) testing optimal pH with L-alanine, (C) transport pH test with an uncharged amino acid L-isoleucine. (D) L-alanine self-competition IC_{50} determination of SLC7A4-expressing HeLa cells.

L-alanine as a positive control, since it has a small uncharged side chain. L-alanine showed a 10% uptake (**Figure 5.29-B**) at pH 5.5 and an IC50 (by self-competition) at ~ 270 µM (**Figure 5.29-C**). L-alanine competition with other amino acids (250 µM) showed decreased L-alanine transport, initially suggesting their transport by SLC7A4 (**Table 5.7**). However, similar reductions in L-alanine uptake were seen in the controls (Y246V and vector-only) which should have remained unaffected if transport was SLC7A4-mediated (**Table 5.7**). This suggested that L-alanine uptake and competition by other amino acids was mediated by other cell-surface transporters, expressed natively in HeLa cells, and not by SLC7A4. The result

Competition	WT (N=2)	Y246V	Vector	$R_{WT} - R_{empty}$	$R_{WT} - R_{Y246V}$
250 μM Asp + 10 μM Ala	27,837	20,574	39,303	-11,466	7,263
250 μM Lys + 10 μM Ala	29,835	29,070	38,591	-8,755	765.5
250 μM Ser + 10 μM Ala	14,030	13,026	15,698	-1,668	1,004
250 μM Leu + 10 μM Ala	14,770	18,498	24,376	-9,606	-3,728
250 μM Arg + 10 μM Ala	39,804	31,617	43,321	-3,516	8,187
10 μM Ala	40,333	37,722	46,949	-6,616	2,611

Table 5.7 Scintillation-counter uptake values of 10 μM [^{3}H]-L-alanine (t=10 min) under competition conditions with 250 μM Asp, Lys, Ser, Leu, Arg and differences, R, between WT, Y246V and vector-control.

was not due to aberrant expression of WT and Y246V, as both constructs showed expression with anti-FLAG western blot (**Figure 5.30-C**), while WT SLC7A4 plasma membrane trafficking was verified (**Figure 5.30-D**). The previous SLC7A4-dependent L-alanine transport (**Figure 5.29-B**) may have been artefactual and reflective of cell health. These observations collectively mean that background L-alanine transport events are present in HeLa cells, reducing signal/noise such that SLC7A4 transport is masked.

An attempt to increase the assay S/N ratio was the utilisation of a sodium-depleted assay buffer as most SLC7 members are sodium-independent transporters. Thus, the transport activity of other sodium-dependent amino acid transporters should be reduced whereas SLC7A4 should remain unaffected. This workaround for noisy cell assays has been utilised in SLC7 studies (Geier *et al.*, 2013; Brown *et al.*, 2014; Chien *et al.*, 2018; Errasti-Murugarren *et al.*, 2019a;

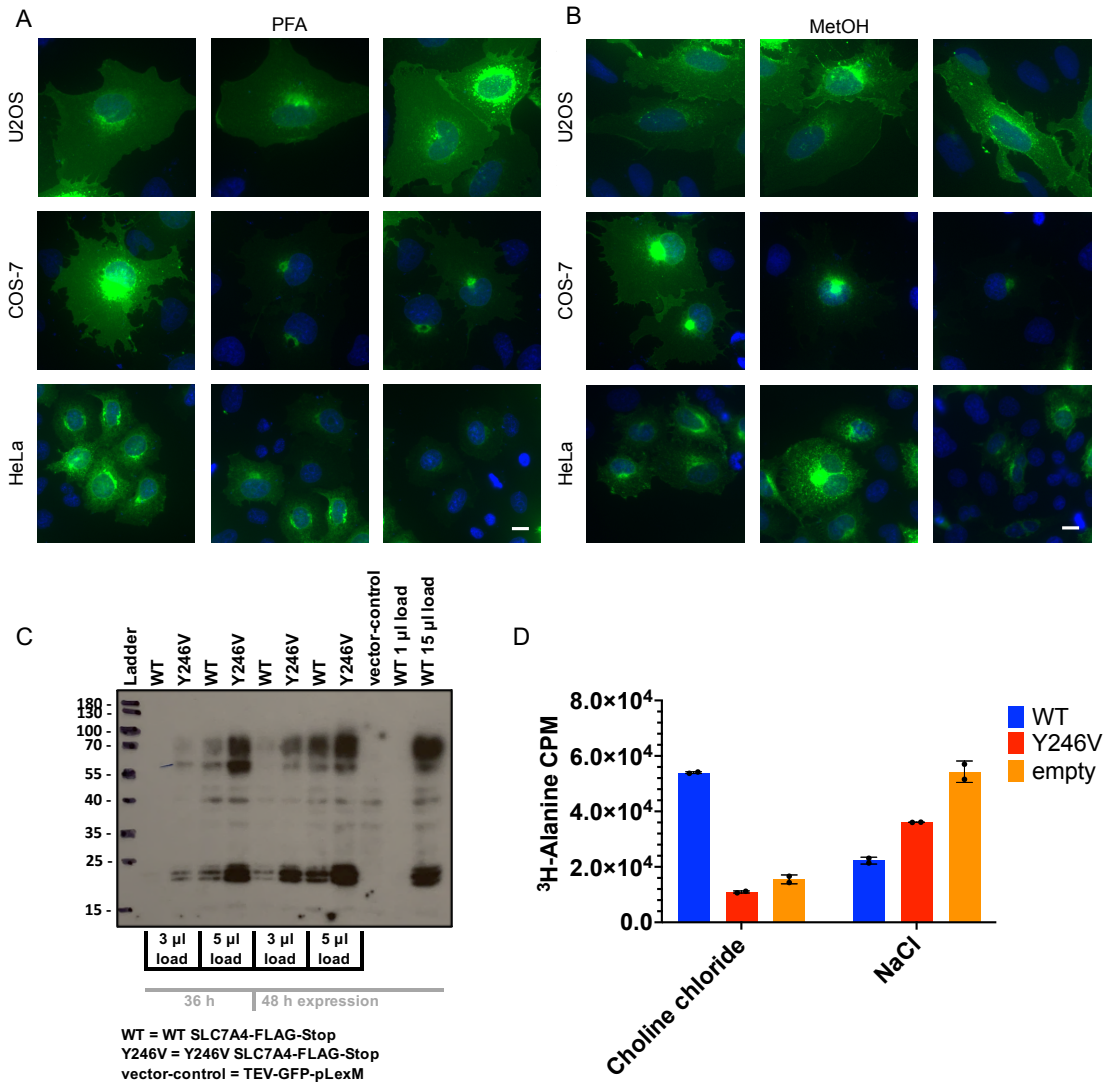


Figure 5.30 Troubleshooting the whole-cell uptake of L-alanine in SLC7A4-transfected HeLa cells. (A) GFP imaging of SLC7A4-GFP transfected U2OS, COS-7 and HeLa cells with the PFA fixing method and (B) with the MeOH fixing method. (C) Anti-FLAG western blot of WT, Y246V and vector control transfected HeLa cells used for uptake assays. Samples taken at 36 h and 48 h post-transfection and 3 µl and 5 µl loading volumes screened. (D) 10 µM L-alanine uptake over 10 minutes (mixed with 0.03 µM hot (^{3}H)- L-alanine) in buffer composed of: 10 mM MES·KOH pH 5.5, 137 mM choline chloride or NaCl, 5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgSO_4$ and 10 mM D-glucose.

Rodriguez *et al.*, 2021). The preliminary results produced before writing this thesis were interesting; SLC7A4-dependent uptake was 3-fold higher than controls (Figure 5.30-D). However, the repeats done were $N = 2$ meaning that more biological repeats are needed to increase confidence. This would be a starting point for

the optimisation of the competition assays.

5.4.2.1. Potential for L-leucine as a substrate of human SLC7A4

Concurrently to the above troubleshooting of radioactive uptake assays in cells, human SLC7A4 was purified from HEK293F cells to homogeneity in GDN – a detergent choice shown to retain substrate sensitivity of plant CAT4 without relipidation (§5.3.2.4). A DSF binding assay screen identical to plant CAT4 was undertaken uncovered a 9.1 °C binding shift induced by L-leucine (Figure 5.31-A,B), while other hydrophobic amino acids produced at most 1.2 °C (Figure 5.31-A). Why does human SLC7A4 shift with different α-amino acids compared to arabidopsis CAT4? Human SLC7A4, in contrast to other homologues including the plant CAT4, has a hydrophobic clamp created by L334 (TM8; S327 in *At*CAT4) and L394 (TM10; T389 in *At*CAT4) which increases hydrophobicity and steric clashes (AF2 predictions; Figure 5.31-C). This would pose a likely explanation for the lack of L-ornithine and L-arginine binding and transport. Other than L-leucine, the shorter (L-valine) or longer (L-isoleucine, L-methionine) hydrophobic substrates did not produce a similar shift (Figure 5.31-A), suggesting that L-leucine may be a specific substrate previously not explored in radioactive uptake assays.

To investigate L-leucine uptake mediated by SLC7A4, radioactive ^{14}C-L-leucine was procured. A time-dependent uptake assay was carried out to observe the uptake trend amongst the constructs (Figure 5.32). Although a high relative background is persistent with L-leucine uptake as was with L-alanine, WT SLC7A4 has a steeper uptake slope and final uptake at $t = 5$ min, suggesting a degree of specificity in transport. However, before decisive conclusions can be made for

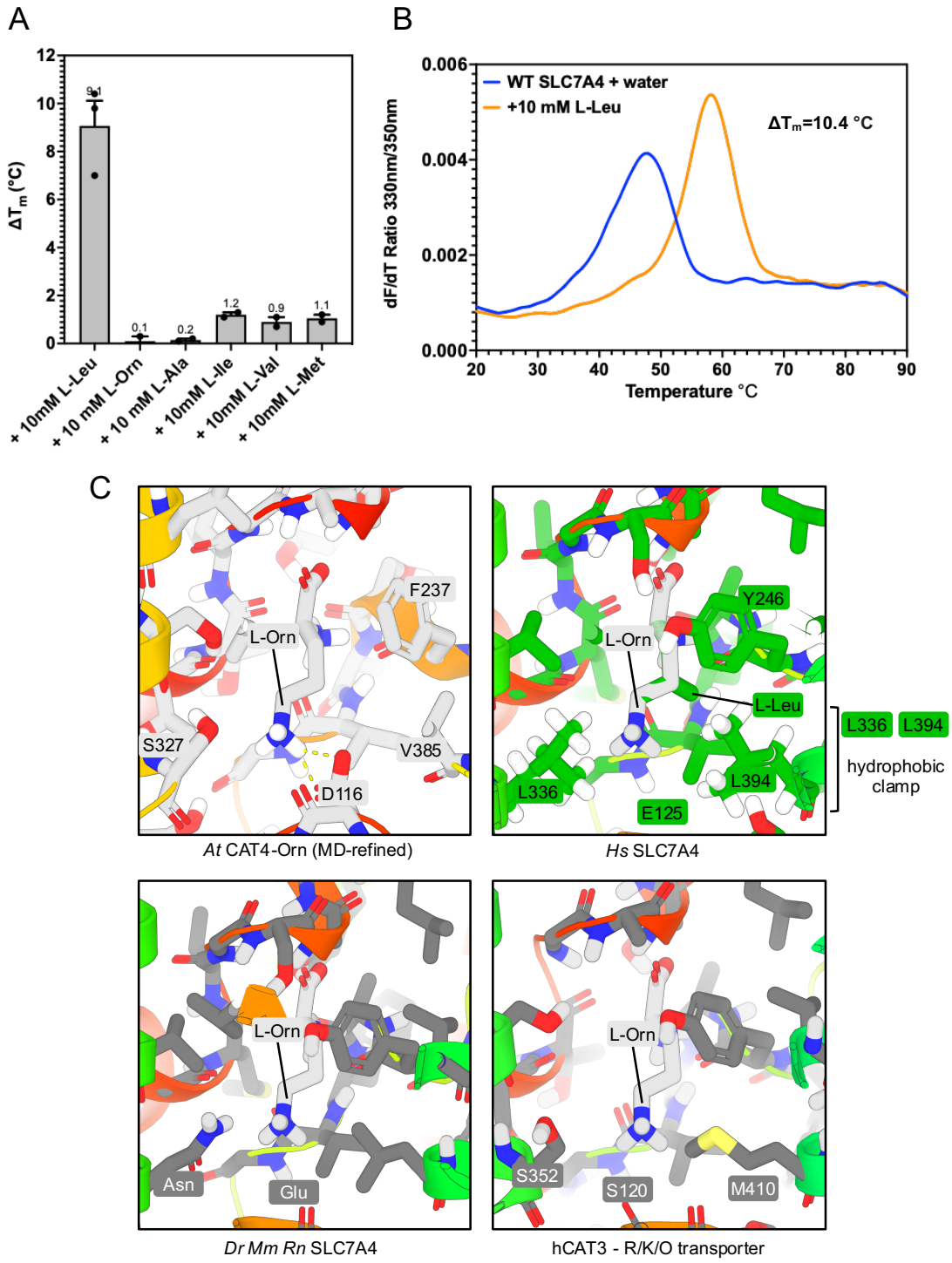


Figure 5.31 Amino acid binding assay on purified WT human SLC7A4. (A) DSF binding assay results upon addition of 10 mM of the α-amino acids shown to 0.5 mg/ml SLC7A4 (7.3 μM). (B) Representative thermal shift of SLC7A4 by L-leucine. (C) Binding site structural superpositions of the MD-relaxed CAT4-L-ornithine complex (t=500 ns), docked Orn and Leu to SLC7A4 (AF2 model); docked L-ornithine to *Dr Mm Rn* SLC7A4 (AF2 models); docked L-ornithine to hCAT3 (AF2 model).

SLC7A4 and leucine transport, the assay would need to be further optimised to

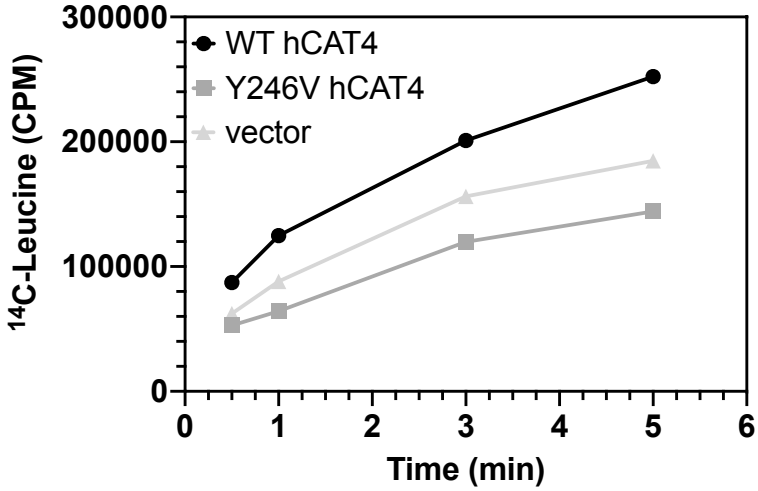


Figure 5.32 Time-resolved uptake of 10 μM L-leucine (0.33 μM hot 14C-L-leucine) uptake by WT, Y246V SLC7A4 and vector-only control. Y-axis is in counts per minute taken over 5 minutes per time point (CPM).

diminish the high background. An alternative strategy would be to subtract the background uptake of vector-only-transfected cells, as done in other whole-cell uptake studies of α-amino acids (Rodriguez *et al.*, 2021; Y. Lee *et al.*, 2022).

L-leucine is the most common proteinogenic amino acid with major metabolic roles; several transporters are known to mediate its uptake by cells (Gauthier-Coles *et al.*, 2021). LAT1, LAT2 and y^+LAT2 are SLC7 members that have transport activity for L-leucine, the knockout of which may improve signal for SLC7A4. Although there was not sufficient time towards the end of this thesis to optimise the transport assay for human SLC7A4 it is suggested that this investigation can proceed in future work, by:

- obtaining ΔLAT1 knockout HeLa cells,
- siRNA silencing LAT1 in a stable an SLC7A4-expressing HeLa cell line,
- substituting HeLa cells for HEK293T cells in which LAT1 has minimal mRNA transcript levels and native expression (Khunweeraphong *et al.*,

2012). A poly-D-lysine modification of the assay plate surface has been shown to immobilise HEK293 cells similar to HeLa (Geier *et al.*, 2013).

- substituting HeLa cells for glioblastoma U87-MG cells the amino acid transporters of which were recently characterised (Gauthier-Coles *et al.*, 2021), including L-leucine transporters. In U87-MG cells, LAT1, B^0AT2, y^+LAT2, and SNAT2, can be inhibited simultaneously by combining JPH203 in sodium-free transport buffer.
- transferring the radioactive uptake assay setup to *X.* oocytes expressing SLC7A4 (verified to express, Figure 5.28-D), and which have shown minimal background noise in L-leucine and L-alanine transport by LAT1 (Lee *et al.*, 2019b).

5.5 Chapter Conclusions

The original purpose of the Syb binder was to augment crystallisation; an attempt hard to achieve likely due to the intrinsic mobility of IL5 and the Myc-His-tagged SybB5. Data from SPA resulted in two cryoEM structures of CAT4-SybB5 which turned out to be governed by the same conformational state, even upon substrate incubation and relipidation, a testimony to the inherent stability of the state resolved and analysed within this chapter. The high purification yield of *At*CAT4 allowed crystallisation attempts with Syb, cryoEM grid screening and optimisation, native MS analysis, bottom-up tandem MS and setup, optimisation, and screening of α-amino acids in the DSF binding assay. Other lower-yield homologues would have been more restrictive in these applications.

The solution of the cryoEM structure of CAT4 in a new outward-open state

is stabilised by an extended interaction network of the thick gate, rich in charged residues, which to the best of our knowledge, has not been mentioned previously in the SLC7 literature. The thin gate F237 is swung-open in the outward-open state and is escorted by a corkscrew movement of TM1 leading V46 to occupy the binding site in the absence of substrate. Upon substrate binding V46 is displaced inducing the structural rearrangement of TM1. The electronegative binding site entrance of *At*CAT4 provides an electro-attraction component to substrate selectivity. *At*CAT4 has a binding site organisation similar to the L-arginine transporter M321S *Gk*ApcT.

An important goal was to understand the substrate preferences of mammalian SLC7A4. *At*CAT4 in LMNG requires relipidation of its surrounding environment to unlock binding site plasticity to amino acid substrates. The developed DSF binding assay showed a high preference of *At*CAT4 for cationic amino acids. We found the binding response to be concentration dependent for CAAs. The $K_{d,app}$ suggested the substrate preference profile: L-lysine > L-arginine > L-ornithine. Mutating binding site residues, F237, D116 and S327, abrogated L-ornithine binding. CAA interactions with *At*CAT4, explored with MD and DSF, uncovered similarities in *At*CAT4's substrate selectivity mechanism to the L-cystine/L-glutamate exchanger xCT; mainly involving the D116-D243 anionic clamp's similarity to R135-R396 cationic clamp in xCT. TEVC analysis of human SLC7A4 in *X.* oocytes failed to yield detectable influx currents with the tested α-amino acids, reproducing the literature results (Wolf *et al.*, 2002). It is likely that SLC7A4 is an amino acid transporter but is not H^+-coupled. Cell-based transport characterisation proved challenging. *At*CAT4 did not express in HeLa cells, converting us to the characterisation of human SLC7A4. Surprisingly the latter does not

transport L-arginine or L-ornithine. Further testing this we uncovered that purified human SLC7A4 does not bind L-arginine and L-ornithine but binds L-leucine in DSF. Preliminary transport data suggest SLC7A4-dependent transport of L-leucine in HeLa cells. Collectively, our characterisation uncovered that plant CAT4 and human SLC7A4 are not functional orthologues but display different substrate binding preferences, although a weaker thermal shift of CAT4 with L-leucine was recorded, and some level of functional homology might be shared. Due to the substrate discrepancy, it is suggested that *At*CAT4 is an orthologue of hCATs 1-3.

There still remains much to do to improve our understanding of SLC7A4's mechanisms of action. This includes a potential role for TM11-12 in homodimerisation, which we detected in 2DCAVGs upon relipidation. Regarding lipid dependence, a sterol density which we modelled as cholesterol was solved in the *At*CAT4 cryoEM structure and will be discussed in **chapter 6**.

Cholesterol Interactions with SLC7 Transporters

CHAPTER 6

Cholesterol comprises 20-40% of eukaryotic membranes (Mouritsen and Zuckermann, 2004). Due to its smaller molecular weight compared to other lipid classes, almost half of the lipid molecules in the plasma membranes of human cells are cholesterol molecules (Das *et al.*, 2014). In chapter 1, a review of cholesterol interactions with SLC7 members was presented (§1.4.2). Notably, this highlighted that member of the HAT subfamily, which function as heterodimers in the plasma membrane, were solved with steroid densities in their cryoEM structures (rBAT/$b^{0,+}$AT, 4F2hc/LAT1, 4F2hc/xCT) while functional characterisations of 4F2hc/LAT1 show cholesterol-modulating effects in transport activity. It is also possible that cholesterol forms part of the functional transporter unit, i.e., HATs may have co-evolved to function with cholesterol.

Molecular simulation methods are well suited for the identification of lipid sites in membrane proteins (Corradi, Besian I Sejdiu, *et al.*, 2019; Marrink *et al.*, 2019; Corey, Stansfeld and Sansom, 2020). In this chapter, we have utilised CG and AA simulations, umbrella sampling, and sophisticated trajectory analyses to explore the cholesterol binding sites in the heterodimers, rBAT/$b^{0,+}$AT, 4F2hc/LAT1, and 4F2hc/xCT. The computational data informed transport assays of cells expressing 4F2hc/LAT1 with cholesterol site mutations. The aim of

this work is to study cholesterol binding patterns on the three solved human transporters of the family and propose physiological binding sites.

6.1 Investigation & Comparison of Cholesterol Sites on HATs

6.1.1 Computational Pipeline

An outline of the computational analysis pipeline used for the investigation of SLC7 cholesterol interaction sites is shown (Figure 6.1). The multiscale approach, i.e. from CG to AT chemical resolution (Noid *et al.*, 2008; Stansfeld and Sansom, 2011; Ingólfsson *et al.*, 2014) to identify lipid binding sites, provides unique advantages to collect binding/unbinding statistics at the enhanced speed provided from CG before validating the results at AT resolution. Applications of the method in protein-lipid interactions have been reported with GPCRs (Hedger *et al.*, 2019; Song *et al.*, 2019), channels (Stansfeld, Jefferys and Sansom, 2013; Aryal *et al.*, 2015), and transporters (Domicevica, Koldsø and Biggin, 2018; Zeppelin *et al.*, 2018).

6.1.2 Cholesterol Enrichment in the 4F2hc/LAT1 Vicinity in Binary & Complex Membranes

In order to understand whether cholesterol displays any local lipidic preferences for SLC7s, LAT1 in the inward-open state was used as a model system to explore cholesterol distribution in: (a) a complex membrane setup mimicking a plasma membrane (PM; with an inner leaflet containing PC:PE:Chol:PS:PIP2

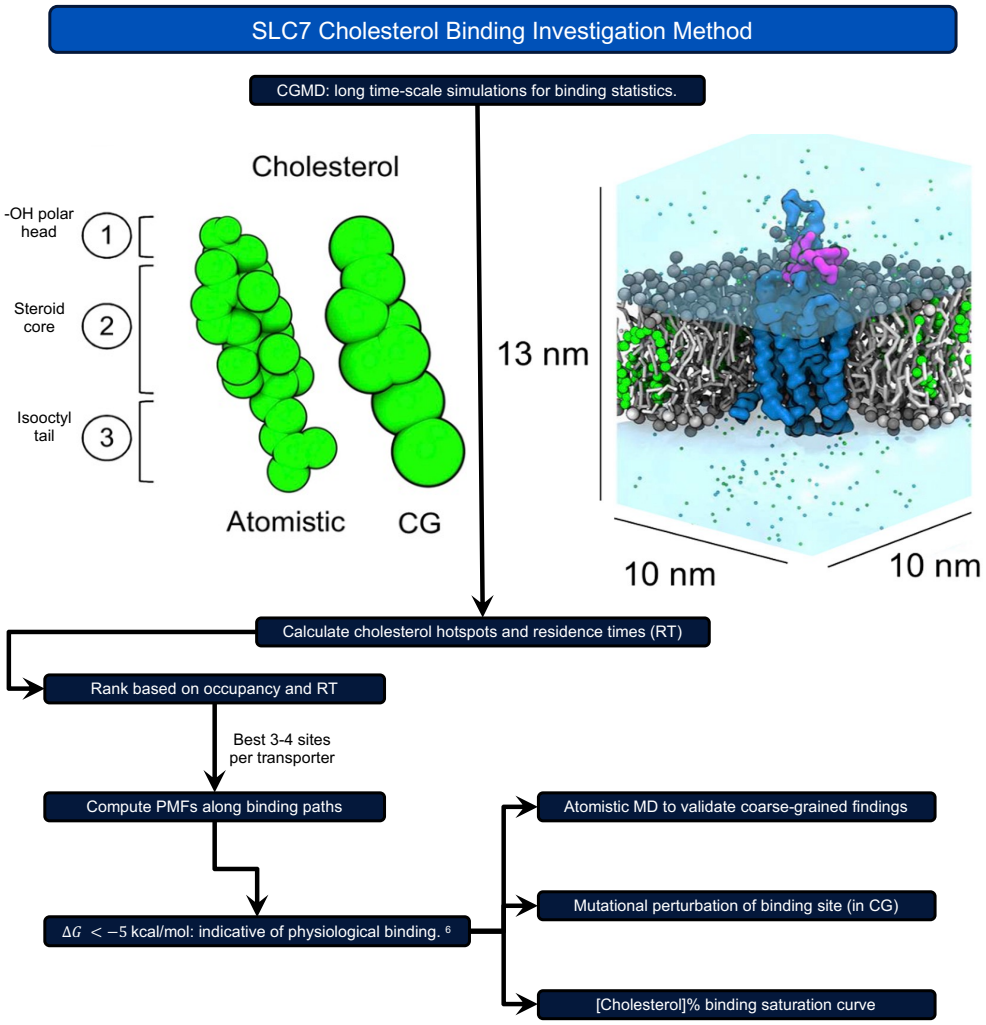


Figure 6.1 Schematic representation of the multiscale MD analysis pipeline used in this project.

and an outer leaflet containing PC:PE:Chol:Sph:GM3) and, (b) a simple binary PC:Cholesterol membrane. The results of the 5×10 μs simulations are presented as a 2D cholesterol density heatmap (**Figure 6.2**). Cholesterol is enriched in the vicinity of 4F2hc/LAT1 heterodimer – evidenced by the increased local cholesterol density coloured green/white compared to bulk density levels in light blue (**Figure 6.2-A**). Distinct hotspots are identified as one traverses from the top to the middle to the bottom leaflet of the membrane (**Figure 6.2-A,B,D**). The increased enrichment is present both in the simple binary lipid mixture and the

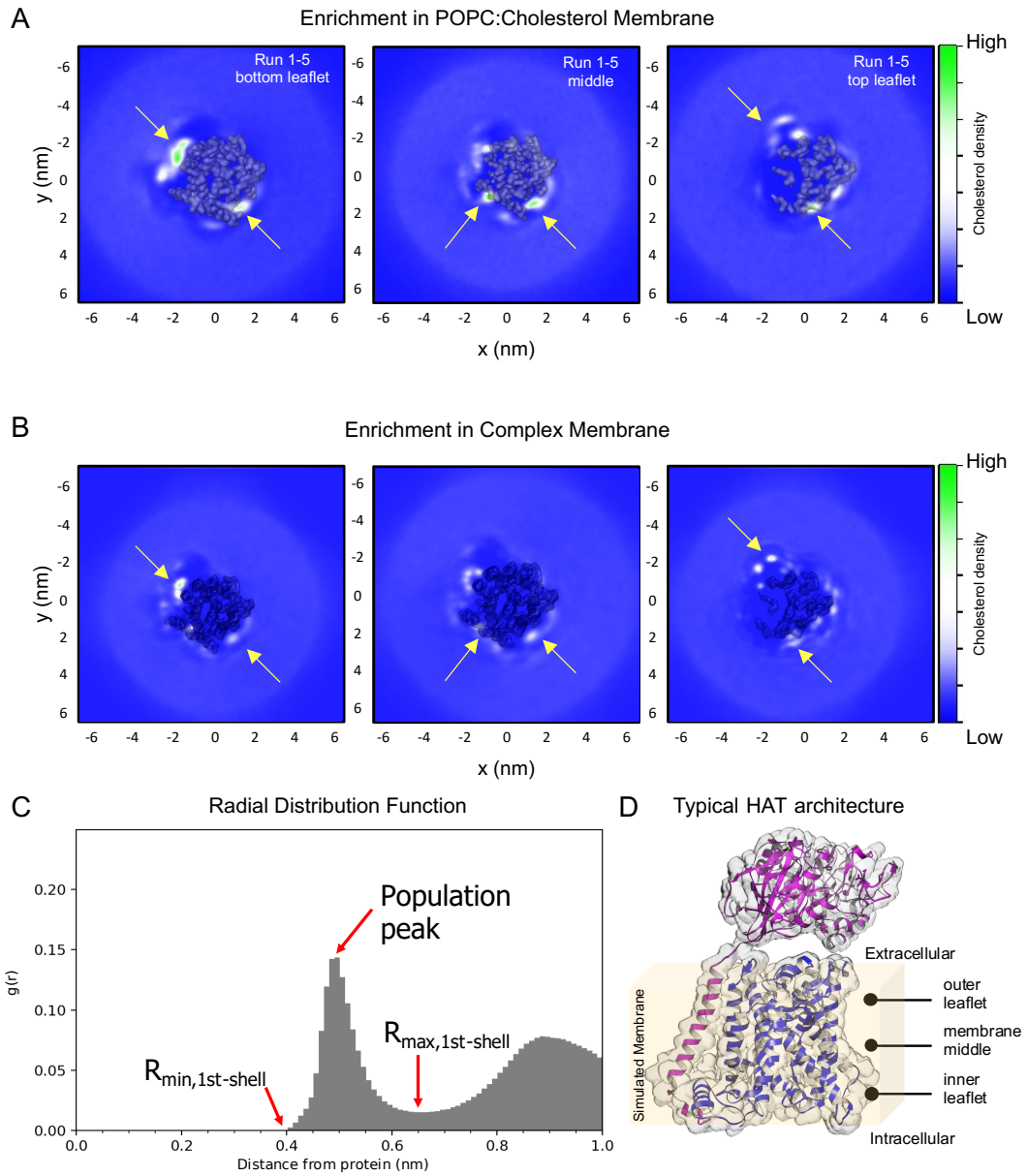


Figure 6.2 Simulation of LAT1 and cholesterol distributions within the membrane. (A) The repeat- and time-averaged cholesterol density heatmap showing cholesterol hotspot projections in the *xy* plane at the bottom, middle and top leaflet of a binary POPC:Cholesterol membrane (4:1 mol). (B) As in (A) but in a complex plasma membrane mimetic model described in the main text. Equivalent cholesterol hotspots between (A) and (B) are marked with a yellow arrow. (C) Normalised radial distribution of cholesterol (bin width=0.1 nm) showing an annular cholesterol solvation shell between R_{min}-R_{max}. (D) Heteromeric amino acid transporter (HAT) architecture in the simulated membranes.

plasma membrane mimetic model (**Figure 6.2-B**). A relative drop in density levels is seen in the PM model compared to the binary mixture when one compares the

equivalent hotspots marked with yellow arrows in (Figure 6.2-A,B) suggesting that some degree of annular site competition is nevertheless present in both leaflets, although the densities remain above bulk levels demonstrating local cholesterol enrichment nevertheless. Visual inspection of the trajectories showed that the other lipid species do not occupy the cholesterol hotspots and that the decrease of cholesterol density in the PM model is a result of dynamic lipid mixing in the membrane. This suggests that cholesterol associates at specific locations on the transporter surface.

The normalised cholesterol radial distribution extracted from the binary system (Figure 6.2-C) shows a distinguished cholesterol solvation shell in the range 0.4-0.6 nm from the transporter surface which tails-off at 0.65 nm. As such, cholesterols existing within 0.6 nm of the transporter may have specific interactions.

6.1.3 Identification of Cholesterol Binding Sites in Human HATs

Examination of 4F2hc/LAT1 Interactions with Cholesterol

The cholesterol local enrichment at seemingly specific hotspots introduced the question of whether these hotspots are areas of increased non-specific cholesterol traffic (annular sites), or whether they are specific binding pockets. Specific cholesterol sites in GPCRs have been found with CGMD to be enriched in Ala, Gly and Leu residues (Song *et al.*, 2022). Analysis of LAT1 in POPC/Cholesterol revealed increased interaction frequencies (>80%) over repeat simulations in sporadic residues mainly located in TM2, 7, 9, 11, 12 and lateral helix 4 (C-terminal amphipathic helix) (Figure 6.3-A). By mapping the interaction frequencies on the

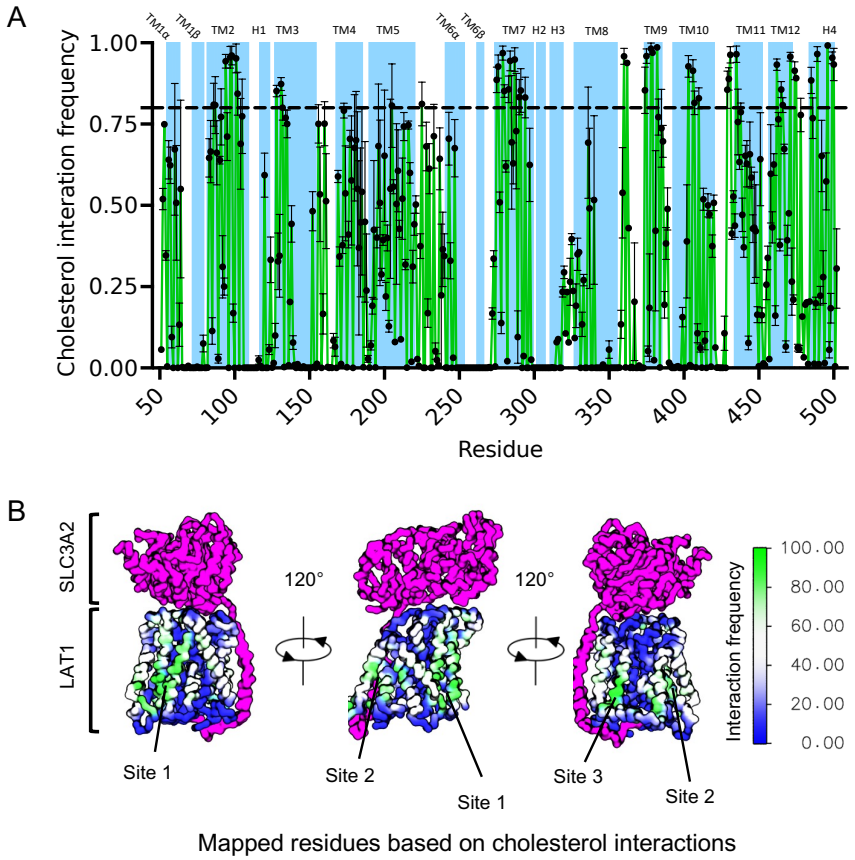


Figure 6.3 Cholesterol-LAT1 interactions on a per-residue basis. (A) Residue interaction frequencies with cholesterol averaged over repeats. Error bars report standard error of the mean between repeat simulation. Dashed line shows an 80% interaction cutoff. An interaction was counted if a cholesterol bead was detected within 6 Å of a LAT1 residue. (B) Mapping of mean frequencies of each residue on the LAT1 transport domain structure (PDB ID: 6IRT); colour-coded between 0-100% (blue-to-green). Highly-interacting residues cluster in three sites (Site 1-3).

structure, it was identified that sequence-distant residues assemble in 3D to create binding pockets (**Figure 6.3-B**). Specifically, TM2-TM7 create binding Site 1, TM11-TM12 create Site 2 and TM9 (with TM4 to a lesser extent and the SLC3A2 TM helix) create Site 3. Therefore, the cholesterol interacting residues within the first solvation shell are dispersed throughout the length of the LAT1 sequence (along residues 51-507); but are topologically adjacent in the structure, creating resolved cholesterol pockets. With the simple contact analysis three pockets were unveiled with high interaction frequencies for cholesterol exist in 4F2hc/LAT1.

6.1.3.1. Per-residue Cholesterol Binding Residence Times of HATs

We proceeded with comparison of cholesterol binding interaction times in inward-open $b^{0,+}AT$, xCT and LAT1, and outward-open LAT1 in the binary membrane POPC:Cholesterol (4:1). LAT1 outward-open resulting from unbiased MD of the outward-occluded 3,5-diido-L-tyrosine-bound structure after 3,5-diido-L-tyrosine deletion. An initial broad interaction cutoff was defined based on the RDF (**Figure 6.2-C**), ranging 0.55-1.0 nm, to allow binding event activation by most first-solvation shell cholesterols until they are exchanged with the second-shell (1.0 nm).

The residence time (RT) of cholesterols binding with a specific residue – i.e. the calculated amount of simulation time that a cholesterol bead is in contact with a given residue – is a good descriptor of binding kinetic information (Song *et al.*, 2022). The inward-open structures have comparable per-residue RT with cholesterols (**Figure 6.4**). If one compares the residues with high RT (> 4.0 μs) between Interfaces A, B and C, one can observe that the white-blue areas are in conserved locations with some variability in the "blueness". An additional cluster of high RT residues is present in LAT1 inward-open Interface C which is lacking in $b^{0+}AT$ and xCT. Therefore, even though there is great commonality in the regions on the transporter architecture where high RT residues are, there is some variability with regards to the exact RT value of equivalent residues. This suggests that general sites for cholesterol binding in HATs are present but are customised according to the exact transporter identity. Additionally, cholesterol-interacting residues are dispersed and close to the centre of the membrane, as has been found in CG study of lipid species binding in Kir2.2 (Duncan, Corey and Sansom, 2020).

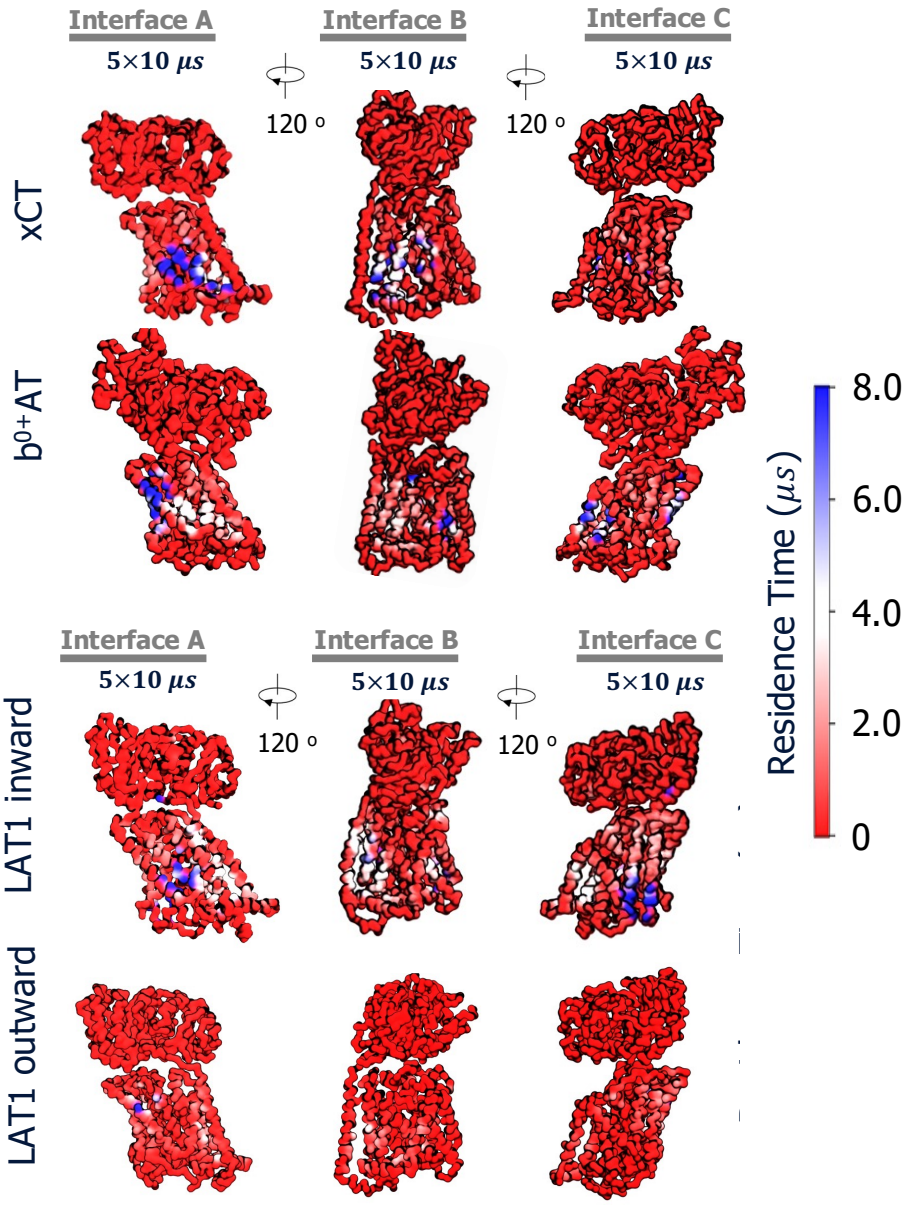


Figure 6.4 Per-residue residence times calculated at the 0.55-1.0 nm dual cutoff for the HATs under study and mapped to the structure. Averaged over 5×10 μs repeats with the initial 2 μs discarded as equilibration time.

Interestingly, while high RT clusters are available in inward-open LAT1 the same is not true for LAT1 in the outward-open state. Extended residue clusters with high RTs (> 4.0 μs) are absent in the outward-open state (**Figure 6.4**). This result highlights that there are differences in cholesterol binding RTs in HATs

depending on the conformational state of the transporter, further suggesting conformational selection of cholesterol binding.

Dual-Cutoff Optimisation for Community-Network Analysis

To further quantitatively compare cholesterol binding across different HAT members, we applied the community-network analysis procedure (described in §2.7.5) to discern binding sites. The dual interaction cutoff method (Song *et al.*, 2022) was employed, specifically tailored towards CG simulations with the MARTINI forcefield, in which a binding event during the simulation is initialised when a cholesterol bead comes into close contact with the protein (at a distance $r \leq R_{cutoff,low}$) and remains active as long as the cholesterol remains within a specific range (at a distance $r \leq R_{cutoff,high}$; $R_{cutoff,low} \leq R_{cutoff,high}$).

For the HATs, these cutoff values needed optimisation, and thus different metrics – i.e., the number of binding sites, interacting residues and interaction mean durations – were calculated for different dual-cutoff pairs (Figure 6.5). The optimal cutoff pair was decided to be 0.45-0.65 nm since, (a) the lower cutoff 0.45 nm increases the number of identified sites drastically from 0.425 nm (Figure 6.5-A); showing that this value better activates a binding event, (b) selection of a larger lower cutoff, e.g. 0.475 nm increases the number of binding sites but overall decreases binding mean durations (Figure 6.5-C); suggesting that larger sites which can accommodate more than one possible pose are being split artificially, and (c) a further increase of the higher cutoff above 0.65 nm was not justified since it causes "jiggling" in binding site number (Figure 6.5-A) and increases the number of lipids per binding site above 1.0 (not shown). Additionally, the dual pair 0.45-0.65 nm is consistent with the cholesterol first solvation shell in the RDF (Figure

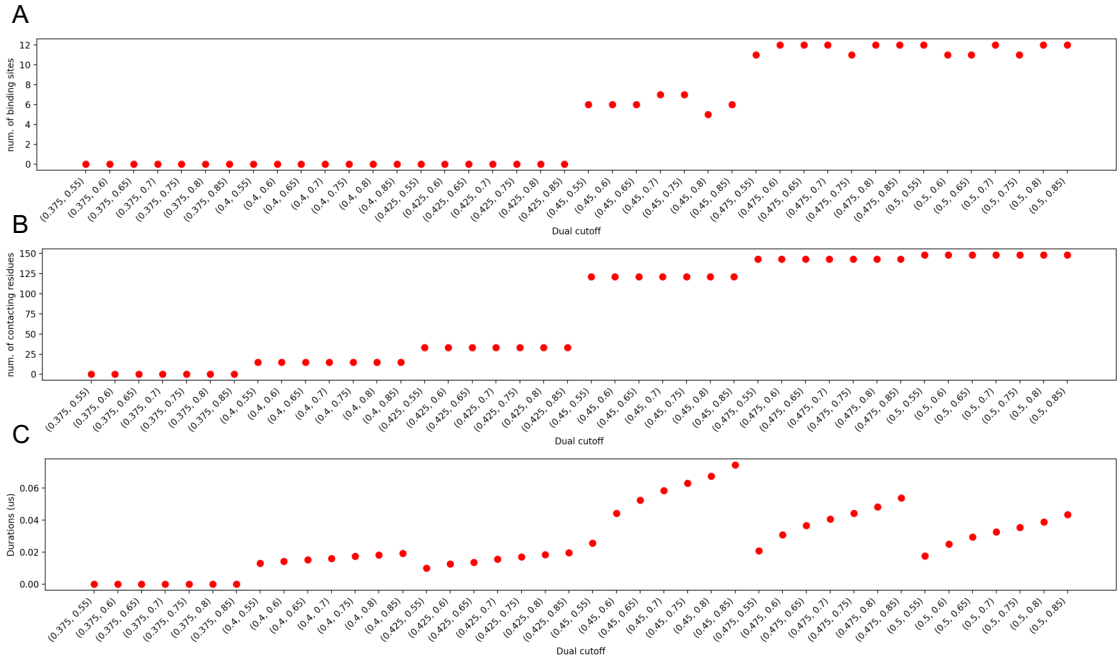


Figure 6.5 Dual cutoff optimisation for HATs. Pairs of dual cutoffs are shown in ($R_{cuttoff,low}$, $R_{cuttoff,high}$) and calculation results (5 repeat averages) are shown for three metrics (A) number of binding sites, (B) numbers of contacting residues (C) mean duration of cholesterol interactions for all sites. Example shown is from system xc^- calculations. Similar results were generated for LAT1 and $b^{0,+}AT$.

6.2-C). Thus, this cutoff was selected for further calculations.

6.1.3.2. Classification & Comparison of Cholesterol Binding Sites

The graph-theoretical approach to computing nodes with high modularity translates to a binding site definition that consists of concurrently cholesterol-interacting protein residues (Song *et al.*, 2022). Using this approach, distinct binding sites were identified for cholesterols in HAT surfaces, and by calculating the collective RT at each site, we recorded similarities and differences in binding kinetic profiles. The kinetic analysis was performed on 100 μs runs of each transporter to allow uncapped estimation of binding interactions times, as the shorter repeats are capped at 10 μs. The number of binding sites afforded were 18 ($b^{0,+}AT$) to 19 (LAT1 and xCT). To filter out false positives - likely the result

of kinetic trapping (high RT, low occupancy) or transient short-lived interactions (low RT, high occupancy) resulting from cholesterol exchange – a criterium cutoff of 50% occupancy and >1.0 μs RT was decided on the merit of two factors:

- Sites with high occupancy and low RT have been suggested to be annular for cholesterol (Song *et al.*, 2022). We reproduced this by benchmarking cholesterol binding in LAT1 (inward- and outward-open) against a bacterial APC-fold nucleobase-cation symporter, Mhp1, which should not bind cholesterol [negative control, **Appendix A – Benchmarking Cholesterol Binding**].
- This combination provided good agreement with experimentally determined cholesterol binding sites determined by cryoEM/crystallography of APC fold structures (including HATs) (**Figure 6.6**).

The calculated binding sites are shown in **Figure 6.6**.

Common HAT cholesterol binding sites include the "green" site – formed by TM4-TM9 and the SLC3A2 TM helix - , and the "yellow" site – formed by TM4-TM5-TM8. The green site has been experimentally observed to be occupied by steroid densities in both the inward- and outward-facing LAT1 cryoEM structures. The kinetic profiles of binding differ within the green site showing the preference order LAT1 (16.8 μs) > $b^{0,+}$AT (4.7 μs) > xCT (1.2 μs) and for the yellow site LAT1 (49.0 μs) >> xCT (9.4 μs) > $b^{0,+}$AT (3.5 μs). Two of three members share the "red" site, LAT1 (1.2 μs) and $b^{0,+}$AT (1.8 μs) – a site of likely functional implications formed by helices involved in the rocking bundle mechanism; TM2-TM6α-TM7-TM11. The red site interactions range on the lower side, between 1.2-1.8 μs. Nevertheless it is experimentally verified by cryoEM in the outward-facing structures of plant *At*CAT4 and the SLC12 member NKCC1 (Neumann *et al.*, 2022).

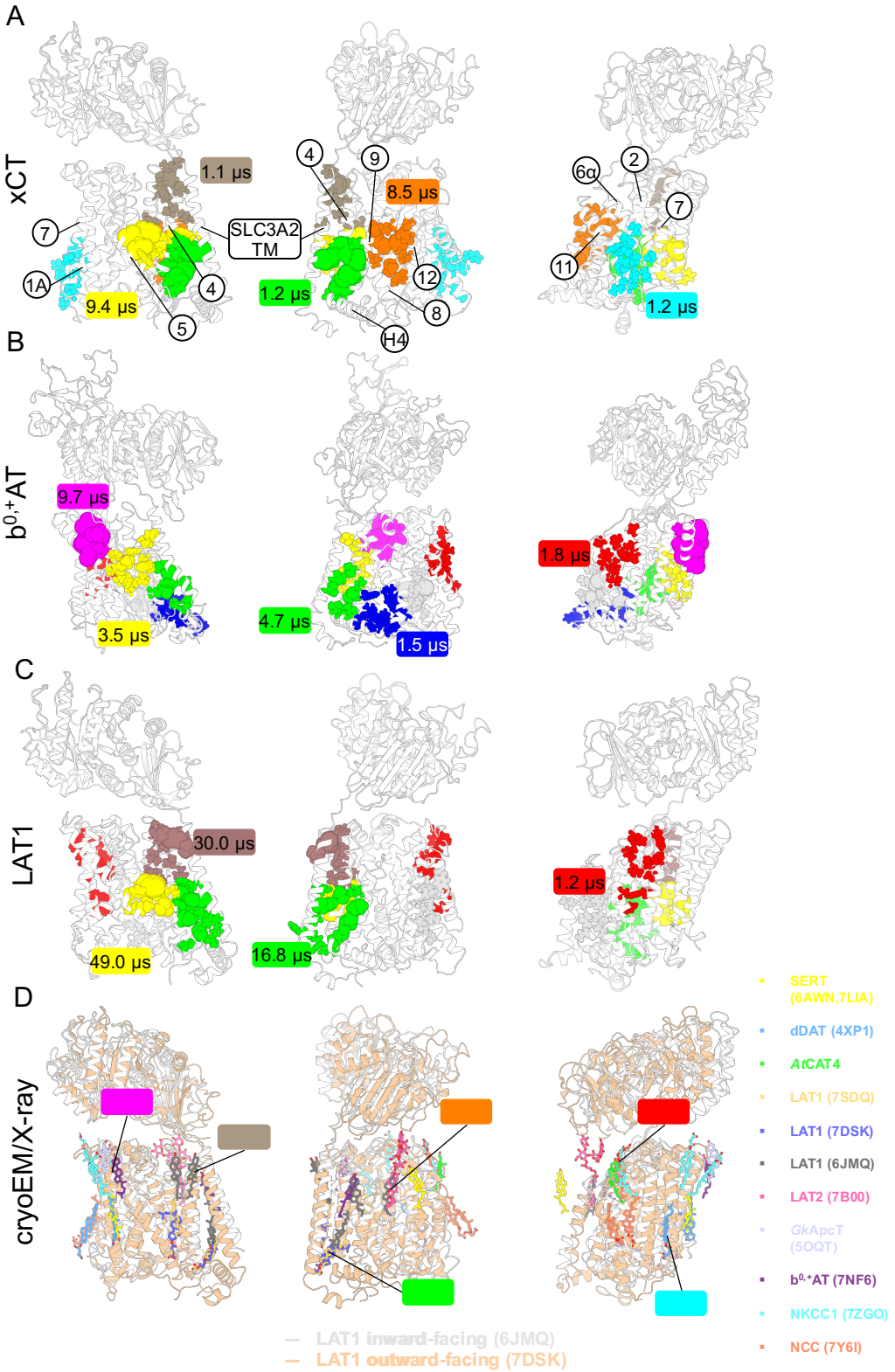


Figure 6.6 Comparison of inward-open cholesterol sites from 100 µs CG MD of (A) xCT, (B) $b^{0,+}$AT, (C) LAT1. Each site binds cholesterol with the RT shown. Topologically common sites are coloured identically. (D) Cholesterol and CHS models from APC fold experimental structures with PDB IDs shown in parentheses.

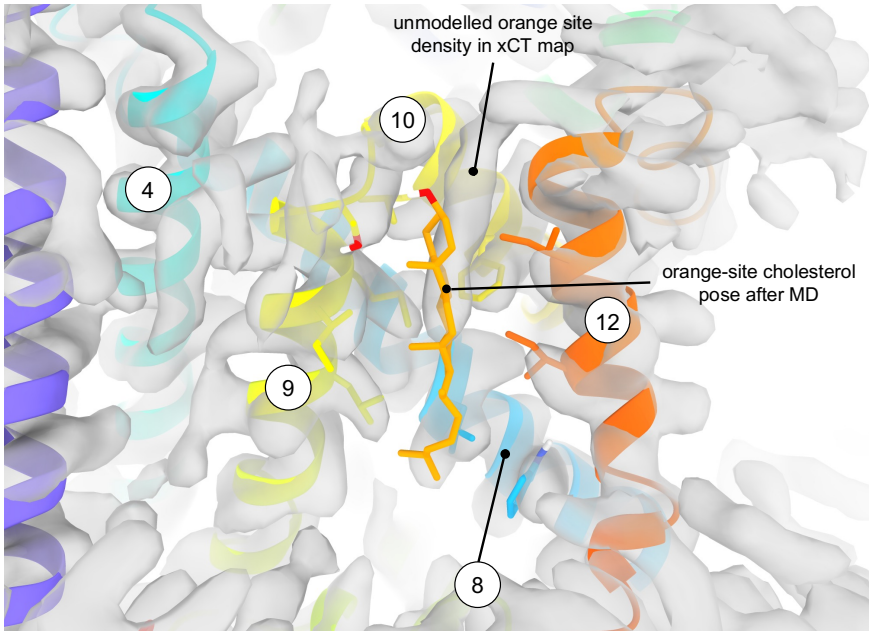


Figure 6.7 Comparison of unmodelled density equivalent to the orange site and the structural model of xCT inward-open in complex with orange-site cholesterol. CGMD model backmapped to atomistic representation with CG2AT2 (Vickery and Stansfeld, 2021).

LAT1 (30.0 μs) and xCT (1.1 μs) share the "grey" site – formed on the top leaflet between TM4 and the SLC3A2 TM helix. The grey site shows higher specificity for LAT1 than for xCT and, has been observed to be occupied in the inward-open LAT1 structure (Lee *et al.*, 2019b), but not the outward-facing inhibitor-bound LAT1 structures or the xCT inward-facing structure (**Figure 6.6-D**), suggesting its specificity for inward-open LAT1. Indeed, both the red and grey sites showed very low occupancies in outward-open LAT1 simulations (32.4%, 14.8%, respectively) and accordingly low RT (<1.0 μs).

Overall, the calculated binding sites show great overlap between LAT1, $b^{0,+}$AT and xCT, although their residence times differ. Other than shared binding sites, two sites with high residence times were unique. The "magenta" site for $b^{0,+}$AT (**Figure 6.6-B**) with experimental validation from, (a) the ovine $b^{0,+}$AT inward-open cryoEM structure (Y. Lee *et al.*, 2022) and, (b) the bacterial

*Gk*ApcT crystallographic structure (K. E. J. Jungnickel, Parker and Newstead, 2018). The other is the "orange" site, unique for xCT (Figure 6.6-A). This site was not modelled with cholesterol in the xCT inward-open structure from our group (Parker, Deme, Kolokouris, *et al.*, 2021). However, we noticed an unassigned steroid-like density is present in the electron density map [EMD: 13267], although lacking density for the isooctyl tail, explaining our reluctancy to model cholesterol (Figure 6.7). The LAT2 structure was solved with a clearer density corresponding to the orange site, which was modelled as digitonin (Figure 6.6-D) (Rodriguez *et al.*, 2021), a cholesteryl-containing surfactant composing the LAT2 detergent belt. The orange site was additionally modelled with cholesterol in the LAT1 inward-open structure (Lee *et al.*, 2019b) and nanodisc-reconstituted SERT [(Yang *et al.*, 2022), termed "CHOL3" site in the SLC6 literature] (Figure 6.6-D).

The overlap of experimentally resolved structural models of APC transporters with the cholesterol binding sites generated with unbiased CGMD provided confidence that the predictive model accurately represents cholesterol binding preferences in HATs. The binding sites were primarily lined with hydrophobic residues (7-11 per site), and secondarily by polar and basic residues (1-2 per site) (Appendix A – HAT Binding Site Residue Comparison). As our interest in cholesterol stemmed from the lack of molecular understanding for cholesterol binding preferences in the SLC7s, the focus was shifted to sites that are shared between HATs for further investigation.

6.1.4 Energetics of Cholesterol Sites in HATs by US/PMF Calculations

To assess the affinity of cholesterol interactions, potential of mean force

(PMF) profiles were calculated at CG resolution with umbrella sampling. A PMF profile, sometimes called the free energy profile integrates-out the subset of system coordinates that is not of interest from the total *N*-dimensional particle distribution, yielding a low-dimensional distribution along the reaction coordinate (Porohille and Chipot, 2008; Decherchi and Cavalli, 2020). The reaction coordinate can be as simple as a distance between particles and as sophisticated as non-linear combinations of dimensions. The PMF between two molecules describes the free energy change along the desired dimensions. In our case this was constructed as the distance between a cholesterol and the centre-of-mass of a binding site. The distance approach, although simple, has shown quantitatively consistent results with experimental affinities of lipids, including the Martini cholesterol description (Corey, Stansfeld and Sansom, 2020). The PMF profiles for the sites were calculated and constructed as described in **chapter 2**.

The profiles revealed both tight high-affinity and looser cholesterol sites (**Figure 6.8**). In particular the green site shows high affinity for the three HATs, ranging from –18 to –27 kJ/mol, being in the upper range of cholesterol affinities calculated from CG and AT simulations (Corey, Stansfeld and Sansom, 2020). This result is in agreement with multiple detergent-solubilised HAT structures being occupied by cholesterol at this site. Additionally, the site lies at the interface between the HAT SLC3 and SLC7 domain, tightly bridging the interface at the lower leaflet while the upper leaflet interaction is being held covalently by the interdomain disulphide bond. Interestingly, while the yellow site affinities for $b^{0,+}$AT and xCT are on the typical, low-affinity end that is expected for cholesterol binding (Corey, Stansfeld and Sansom, 2020), there is an appreciable affinity increase for LAT1 (~ –29 kJ/mol), in good agreement with the enhanced

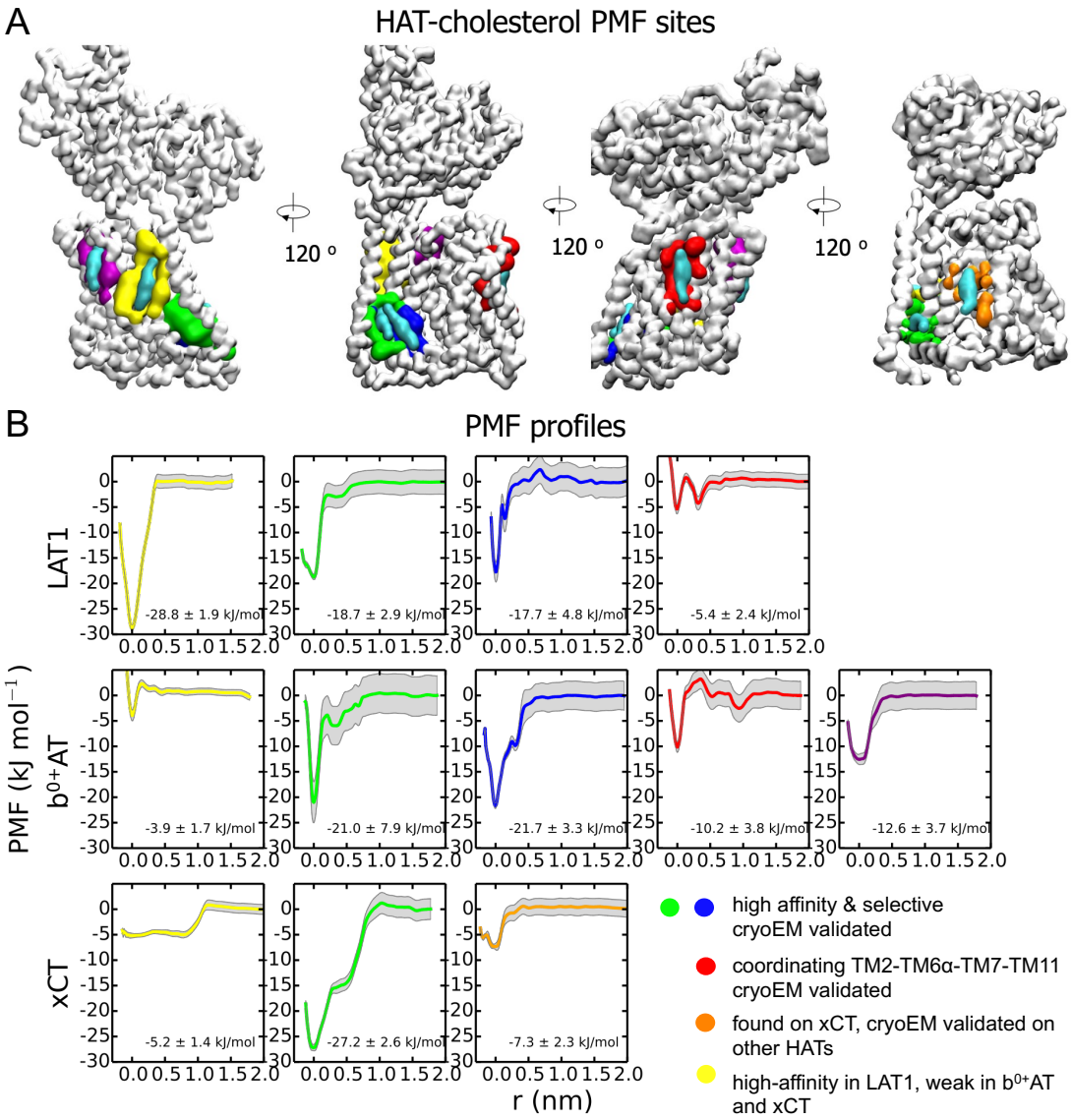


Figure 6.8 Potentials of mean force for cholesterol binding to HATs. (A) Side views of the HAT sites mapped on a CG resolution structure of $b^{0,+}$AT. Representative cholesterol poses are shown as cyan surfaces, encapsulated within the color-coded binding sites. (B) PMF profiles of cholesterol binding to the yellow, green, blue, red, purple, and orange sites.

residence time reported for LAT1 in the previous section. However, as it will be seen in the next section, the yellow site cholesterol interactions are not stable in atomistic resolution MD (AT). The calculated –5.4 kJ/mol for the LAT1 red site shows a lower affinity for this site, but large enough to be more than twice the thermal noise, RT (=2.479 kJ/mol). The xCT orange and $b^{0,+}$AT red interaction

affinities lie in between the lower and higher affinity values, suggesting that these sites may be of importance but perhaps to a lower extent than the green site. Results for the blue site, kinetically ranked for $b^{0,+}$AT as a site with lower-end residence time (1.5 μs), showed high affinity binding for $b^{0,+}$AT (~ –22 kJ/mol). Although not reported in the 100 μs LAT1 binding sites (**Figure 6.6**), a similar cholesterol pose was seen in the 5×10 μs LAT1 trajectories and its binding affinity was comparably high to the $b^{0,+}$AT blue site (~ –18 kJ/mol).

Collectively two discrepancies were observed, (a) the LAT1 blue site high energetics that were not expected according to binding site kinetic ranking (**Figure 6.6-A,B,C**) and (b) the LAT1 yellow site high energetics and RTs not found in other members. In MARTINI simulations the secondary structure is restrained, while the CG nature of the simulation reduces the degrees of freedom and as a result scales the kinetics; the latter especially true for lipids (Marrink and Tieleman, 2013). The CG PMF calculations presented are not subject to kinetic alteration and are generally comparable to AT values in apolar membrane environments (Marrink and Tieleman, 2013). They can be interpreted as an end-point affinity ranking of the lipid sites.

The previous analyses improved our confidence in cholesterol binding the human HATs in unbiased short-repeat and long simulations by forming separated clear binding sites with variable degree of interaction durations and binding affinity. The affinities reported are within the range for cholesterol binding, also reported from STD-NMR (~ –10 kJ/mol) and thermal shift assays (~ –50 kJ/mol) for the β_2AR adrenergic receptor (Gater *et al.*, 2014), showing that two

affinity modes are available for cholesterol binding in HATs. Subsequently, atomistic backmapping of the CG systems was performed to examine site stabilities upon resolution increase.

6.1.5 Atomistic Refinement Analysis of HAT Cholesterol Sites

A serial multiscale approach (Stansfeld and Sansom, 2011) in investigating protein-cholesterol interactions was adopted. CG simulation snapshots of the inward-open HATs in which all of the binding sites reported in Figure 6.6 were occupied. The snapshots were extracted and converted to AT resolution. Each system was simulated for 3×200 ns after careful restrained/unrestrained relaxation cycle of the lipid membrane (chapter 2). In AT, cholesterols were defined as retaining their binding if the distance between the binding site centre-of-mass and cholesterol centre-of-mass remained within 6.0 Å. At 6.0-10.0 Å, cholesterols could be refining their CG pose or starting to dissociate. At $>$ 10.0 Å, cholesterols are diffusing away from the CG site.

In accordance with the binding site, kinetics, and affinity calculations, reported in the previous sections, all three transporters remain fully occupied at the green site for the full duration of the repeats (Figure 6.9). Verifying the affinity PMF calculations, cholesterol in both the $b^{0,+}$AT and LAT1 blue sites remains bound (Figure 6.9-A,B), with the exception of a single cholesterol unbinding LAT1 ($R_{\text{COM-COM}} > 6$ Å) within the initial 25 ns (Figure 6.9-A). LAT1, $b^{0,+}$AT and xCT quickly dissociate their yellow site cholesterols, suggesting this to have been a false positive interaction site in CG, filtered-out in AT, likely a result of cholesterol CG trapping. The xCT-specific orange and cyan sites found in CG remained

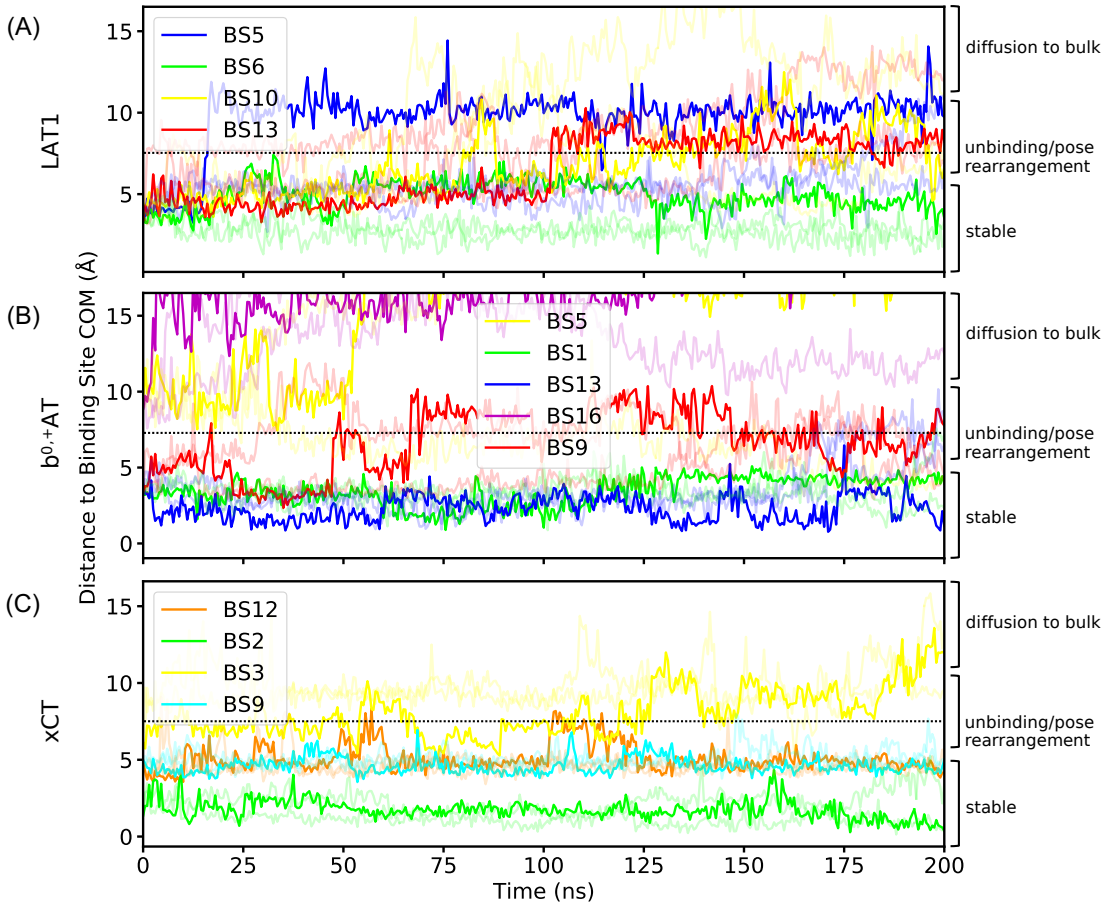


Figure 6.9 Time-courses of cholesterol distances to binding site COM from atomistic MD simulations. HAT transporters are initialised as cholesterol-occupied at the indicated sites. (A) LAT1 inward-open, (B) $b^{0,+}$AT inward-open, (C) xCT inward-open. Refer to **Figure 6.6** for colouring of sites.

bound within all three repeat AT simulations in good agreement with the detergent-solubilised LAT1/LAT2 cryoEM and dDAT crystal structures, suggesting that the steroid density in our inward-open xCT map may be modelled as cholesterol (**Figure 6.7**). The red site cholesterols bound to LAT1 and $b^{0,+}$AT retain $5 < R_{\text{COM-COM}} < 10$ Å, suggesting a reorientation of cholesterol in the site instead of total unbinding to the bulk . Indeed, instead of dissociation, trajectory inspection revealed a conformational rearrangement of the loop coordinating red site

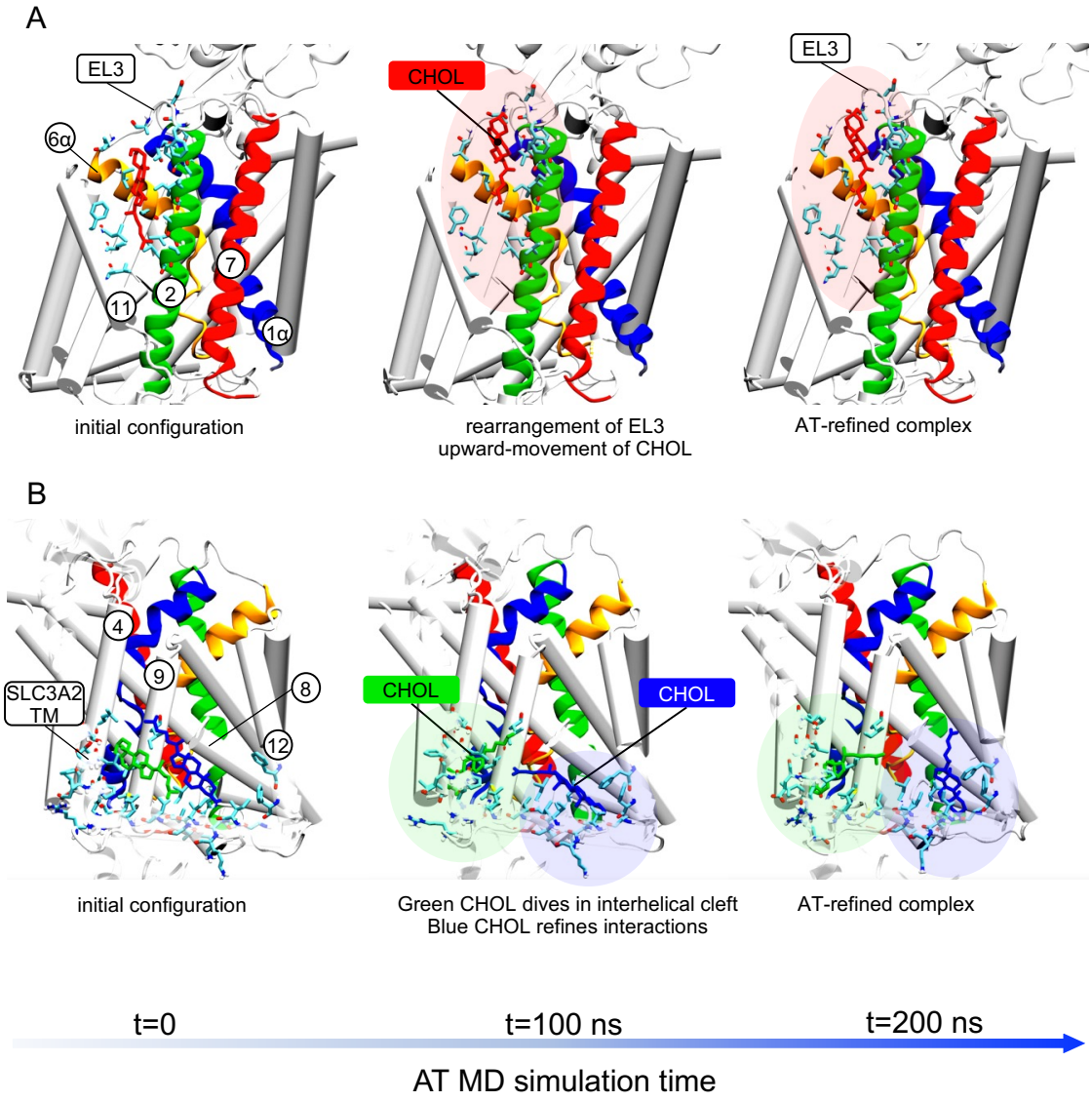


Figure 6.10 AT resolution time-course rearrangements of cholesterols in the sites defined in CG- $b^{0,+}$AT. (A) Visualisation of the red site cholesterol remaining bound and shifting upwards towards extracellular loop EL3. (B) Visualisation of the green and blue site cholesterols; the green site cholesterol is AT refined to "dive" between the interprotein space; the blue site undergoes further, more drastic reorientation of the resident cholesterol from a quasi-lateral to a longitudinal pose. Bundle domain helices are coloured, TM1 – blue, TM2 – green, TM6 – orange, TM7 – red.

cholesterols in $b^{0,+}$AT such that interactions of the headgroup hydroxyl with extracellular loop 3 (connecting TM5-TM6α) are increased and cholesterol situates higher in the membrane (**Figure 6.10-A**). Lastly, the purple site [identified for $b^{0,+}$AT with experimental observations of cholesterols in its vicinity (TM2-TM7

interface) from *Gk*ApcT and $b^{0,+}$AT structures and a medium affinity (~ –13 kJ/mol)] quickly dissociated in AT resolution and escaped to the bulk membrane suggesting this to be a less specific site. The rearrangement and/or unbinding of cholesterols for some sites demonstrates the importance of verifying CG results at AT resolution.

Collectively, the multiscale analysis revealed that the green site is common in HATs and provides a high-affinity, high-RT, stable interaction for with cholesterols which dive in the SLC3-SLC7 interprotein space, bridging the two domains. The red site, despite showing less clear indications of high-affinity, provided stable interactions with cholesterol in AT MD of LAT1 and $b^{0,+}$AT, but not xCT. Its composition of residues in bundle domain helices led us to investigate this further with the green site in LAT1.

6.2 Characterisation of LAT1 Cholesterol Sites

From this point, we focused our analyses on LAT1 as both inward- and outward-facing structures were available to test conformational-dependent cholesterol binding.

6.2.1 Cholesterol Concentration Effects on Red & Green Site Occupancies

The presented simulations and results have been conducted at 20% membrane cholesterol in a binary membrane with POPC. Additionally, the per-residue residence times mapped on the simulation-induced outward-open LAT1 structure showed a marked decrease in cholesterol interactions suggesting a conformational

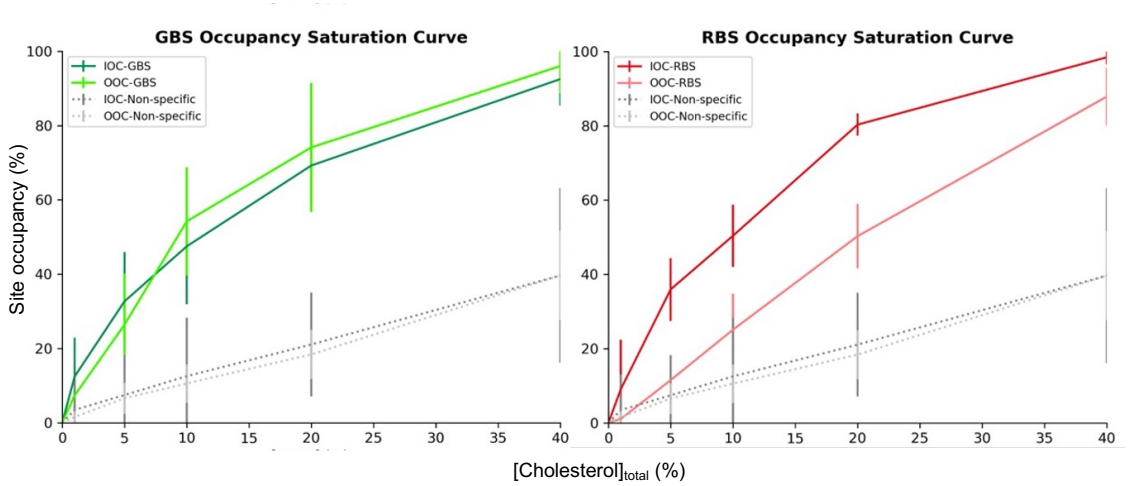


Figure 6.11 Cholesterol binding saturation curves for green and red sites reporting the mean occupancy and standard deviation at each concentration point. IOC (inward-open conformation), OOC (outward-open conformation) LAT1 structures, GBS (green binding site), RBS (red binding site). The grey lines represent non-specific interactions defined as mean occupancy of residues which, in the 40% cholesterol system, had occupancies in the 30-50 % cholesterol range (40 % cholesterol residues). Each data point is the mean ± SD from 5×10 μs CG MD repeats.

bias (§6.1.3.2). Thus, computational titration of the red and green sites was performed with differential cholesterol concentrations in a POPC membrane, to examine whether the inward and outward-open states induce cholesterol occupancy differences, according to a recently published method (Ansell *et al.*, 2021).

Membrane cholesterol concentration was varied, by increasing it in discrete steps of 1%, 5%, 10%, 20%, and 40% and dose-response curves that mimics an experimental ligand binding assay were constructed. As binding site definition extracted with PyLipID is dependent on concurrent residue interactions, this was expected to be perturbed between different cholesterol concentrations. To consistently define sites at different cholesterol concentrations, six residues that were found to be conserved in the site definitions across the concentration range were selected (Ansell *et al.*, 2021) (**Figure 6.11**).

	$K_{d,app}$ (%)		B_{max} (%) – max. occupancy	
Conformation	inward	outward	inward	outward
Green site	15.59 ± 2.46	17.63 ± 2.60	126.53 ± 8.60	139.27 ± 9.28
Red site	15.21 ± 1.88	172.10 ± 50.79	136.79 ± 7.23	467.78 ± 116.19

Table 6.1 Cholesterol binding statistics obtained from fitting the cholesterol saturation curves shown in **Figure 6.11**. The equation used to extract the parameters reported was occupancy=$B_{max}{}^{app}\frac{[CHOL]_{free}}{K_d{}^{app}+[CHOL]_{free}}$, as described in Ansell *et al.*, 2021, under the assunption that [CHOL]=$[CHOL]_{free}$. Fitting restrained such that [CHOL]=0 % gives 0 % site occupancy.

The conformational change from inward to outward-open did not result in any appreciable changes in the saturation curves for the cholesterol green site (**Figure 6.11**). This suggests that binding in this location is not dependent on conformational change. Indeed, between the two LAT1 conformations, the RMSD of the residue backbones lining the green site is <1.0 Å showing conservation of pocket geometry irrespective of conformation ($RMSD_{global}$ = 2.1 Å). Collectively with the high affinity data and the stable binding in AT, the conformational- independence of green site occupancy suggests that cholesterol plays a structural role in this site, i.e., bridging the SLC3 and SLC7 domains, rather than a transport-regulating role.

In contrast, the red site shows conformational-dependency in cholesterol binding, with a marked "gap" separating the inward and outward-facing structures saturation curves (**Figure 6.11**). In the range 0-20% [CHOL], the red site is occupied at least twice as much in the inward state as in the outward state. Between 20-40% [CHOL] – i.e., at physiological plasma membrane [CHOL] – the gap starts to shrink. However, the inward-open state still displays higher occupancy suggesting marked preference of cholesterol binding to this state than outward-open.

At 40% [CHOL] both LAT1 conformations reach > 80 % red site occupancy. For both binding sites, a linear trend in the traces at [CHOL]>20 % is seen, parallel to the non-specific linear curves. This demonstrates that at concentrations higher than 20 % total membrane cholesterol, both sites are subject to some degree of non-specific binding.

With the saturation method, the apparent K_d extracted for both red and green sites of the inward-open LAT1 are within error comparably similar (Table 6.1). On the other side, while the green site retains a similar $K_{d,app}$ in the outward-open state (15.59% versus 15.21%), this is decreased by an order of magnitude for the red site (17.63% versus 172.10%), verifying that the red site is specific for inward-facing LAT1.

6.2.2 Computational Mutagenesis of LAT1 Red & Green Sites

In order to understand the nature of cholesterol attraction in the two sites, we proceeded to *in silico* mutate site residues (Table 6.2). We mutated residues coordinating the cholesterol -OH group in AT simulations and hydrophobic residues in proximity to the steroid group (Figure 6.12). 5×10 μs CG simulations (in 20% cholesterol binary membranes) of the mutants were performed similar to the WT LAT1. The hydroxyl residue in the green site is coordinated by polar residues R183 (of SLC3A2) and H367 while a hydrophobic residue, V378, was mutated to a bulkier Phe to test whether the site would be inaccessible (Figure 6.12-B). In the red site accordingly, polar K237 and D239 in proximity to -OH group and two Phe, F95 and F445, lining cholesterol were mutated (Figure 6.12-B).

Green site	Mutants
R183	R183A (in SLC3A2)
H367	H367A
V378	V378F
	H367A/R183A (DM)
Red site	**Mutants**
F95	F95A
K237	K237A
D239	D239A
F445	F445A

Table 6.2 Mutants used in computational mutagenesis of LAT1.

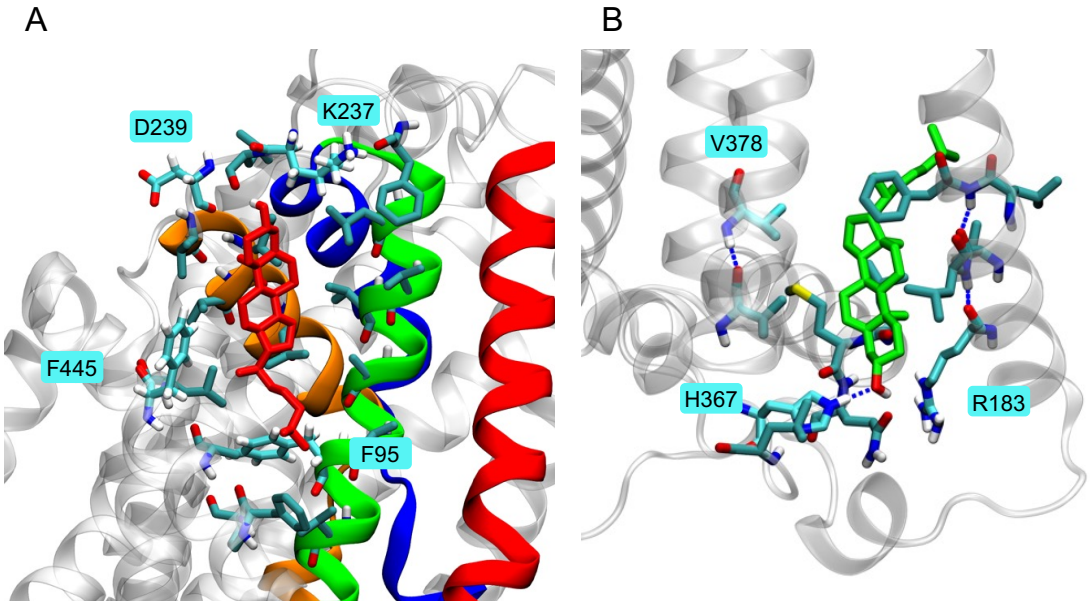


Figure 6.12 AT representations of the (A) red and (B) green sites occupied by cholesterols. The has domain helices are coloured (TM1 – blue, TM2 – green, TM6 – orange, TM7 - red). Residues within 5 Å are shown and the mutagenesis selections are labelled. Snapshots are representative from atomistic simulation. Bundle domain helices are coloured, TM1 – blue, TM2 – green, TM6 – orange, TM7 – red.

Binding site occupancies were calculated for the selected mutants, with the binding sites defined as those topologically manifested in the red and green site

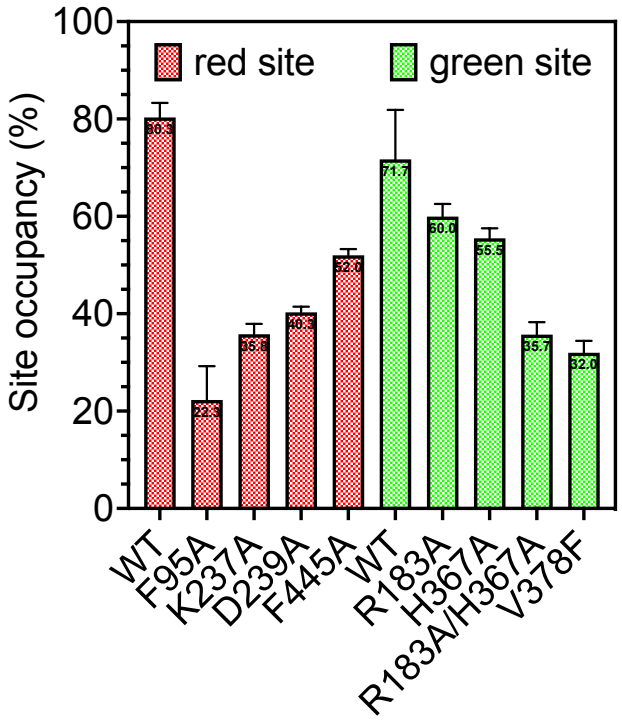


Figure 6.13 Change of red and green site occupancies after mutating the indicated residues. Data is show in bar plots as mean ± SEM from 5×10 μs MD in CG resolution for each mutant.

locales and matched based upon residue conservation. All green site mutants showed a marked reduction of cholesterol occupancies compared to the WT value, 71.7 ± 7.28 % (**Figure 6.13**). The polar residues coordinating the cholesterol -OH group showed a moderate reduction, namely –11.7 ± 1.86 and –12.4 ± 1.48 % for R183A and H367A, respectively. The effect was intensified by combining R183A/H367A, reducing the occupancy by 36.0 % to a final 35.7 ± 1.73 %. This highlights that both residues are important in the coordination of the cholesterol polar headgroup, and while at least one is available it can compensate for the lack of the other. When both are mutated, cholesterol binding is depleted. Interestingly, the V378F mutation which increases the steric hindrance of the green site, reduced the cholesterol occupancy to 32.0 ± 1.73 %, impairing binding comparably to the R183A/H367A double mutant. A Val to Phe mutation abrogating cholesterol binding by a steric hindrance mechanism has also been suggested for

the cholesterol-binding lysosomal Niemann–Pick C2 protein, which showed a drop in $K_{d,app}$ from ~30–50 nM (WT) to non-detectable (V96F) in [^{3}H]-cholesterol saturation (Ko *et al.*, 2003).

In the red site, F95A, K237A, D239A, and F445A all reported a drop in occupancy with the highest occupancy reduction found for the F95A mutant (22.3% occupancy). This residue coordinates the cholesterol isooctyl tail in the red site (Figure 6.13). It is equivalent to V329 (TM2), a residue lining the CHS-bound red site found on NKCC1 (SLC12A2) (Neumann *et al.*, 2022) and T180 on NCC (SLC12A3) (Nan *et al.*, 2022). NKCC1-V329 likely behaves similarly to LAT1-F95. However, V329 was not functionally tested via uptake assays on NKCC1 by Neumann et al., 2022, even though the authors tested other vicinial CHS-coordinating residues which resulted in moderate decrease of transport activity (~ 50 % of WT). F445A seems to least affect cholesterol binding occupancy, whereas the -OH group coordinating residues mutations, K237A and D239A, reduced binding occupancy to approximately ~ 50 % of WT.

6.2.3 Experimental Testing of LAT1 Red & Green Sites by Transport Assays

The computational results demonstrate that LAT1 cholesterol binding to the red and green sites can be modulated by point mutations. We further subjected WT LAT1 and site-mutants to cell-based uptake assays with a physiological substrate, L-leucine, in a HeLa cell 12-well plate format. Each well was transfected with the LAT1-SLC3A2 heterodimer at a confluency 5×10^5 cells/ml. This assay setup has been described recently for 4F2hc/LAT2 (Rodriguez *et al.*, 2021) and

4F2hc/xCT (Parker, Deme, Kolokouris, *et al.*, 2021).

The uptake data measured by β-scintillation counting of [^{14}C]-L-leucine measured at t=3 min substrate incubation, are presented in Figure 6.14. Modulation of L-leucine uptake activity was seen for all mutants except K327A. H367A, i.e., the polar residue coordinating the cholesterol hydroxyl group belonging to LAT1 shows transport reduction compared to WT, reducing transport to 20%. The other polar residue at the green site, R183, belonging to SLC3A2 and not LAT1, has yet to be tested but a similar disruption of transport activity may be expected. The steric-hindrance-inducing mutation, V378F, reduced transport to approximately 55%, verifying its role in the green site experimentally. Compared to the occupancy data from CG MD, the binding occupancy and transport by V378F are both around 50 % of the maximum value. In contrast while H367A has a more moderate effect in CG MD (reduction by 16.2 %), it reduces transport in experiment by 80 %. It is suggested that the seeming discrepancy is a result of the CG force field which does not model hydrogen bonding explicitly (Marrink and Tieleman, 2013). The effect of both residues identified to be important for cholesterol binding in simulations is transferrable in experiment.

The situation is more complex regarding the red binding site residues of LAT1. The mutants, with the exception of K237A, decreased transport of L-leucine by $>$ 50% with the largest reduction reported for F445A (20% of WT). This mutation reduced site occupancy to 52% in CG MD (Figure 6.13). F95A reduced MD binding to 22% and transport to 35% of WT. Overall the Phe residues that coordinate the cholesterol steroid and isooctyl tails affect cholesterol binding to the red site. Similarly, D239, a residue in close proximity to the cholesterol -OH group

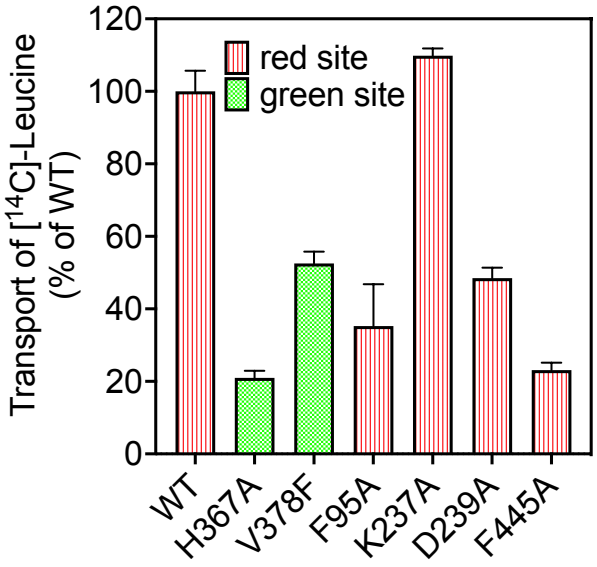


Figure 6.14 Whole-cell uptake of 10 μM [^{14}C]-L-leucine measured at 3 min of transport for LAT1-SLC3A2 WT and mutant-transfected HeLa cells. The uptake values are presented as % uptake compared to WT data and are presented as mean and SEM between N=3 repeats. The uptake data is baseline-corrected by subtracting the equivalent cell uptake in vector-only-transfected HeLa cells.

(**Figure 6.12-A**) also resulted in reduced transport activity, suggesting an important role in coordination of cholesterol. Interestingly, the other polar residue, K237, did not have the same effect as D237 in transport activity, as mutation of this residue into an Ala did not reduce transport. Both D237 and K239 showed an MD occupancy of 36 and 40 % in 20% cholesterol membranes, respectively (**Figure 6.13**).

Since K239A did not reduce transport, it is suggested that it is of lower significance in cholesterol coordination in the red site and that careful consideration of CG MD results in conjunction with experiment might be necessary to carefully derive conclusions.

6.3 Functional Implications & Relations to other SLC7s

6.3.1 Cholesterol-coordinating Residues Implicated in Disease

After discovering the selected cholesterol site residues that modulate CG cholesterol binding and transport experimentally, the physiological impact of these was investigated in the COSMIC (Catalogue of Somatic Mutations In Cancer) database for cancer mutations (Harry C. Jubb *et al.*, 2018). Notably, R183S in SLC3A2 (green site) is a missense mutation found in endometroid carcinoma. R183 mutations to A/E/L/K displayed a pronounced reduction in transport activity in proteoliposomes (Yan *et al.*, 2019a), suggesting it is involved in lipid coordination and is critical for transport activity. R183 belongs to the green site, a site equally accessible to cholesterol in both inward- and outward-open LAT1 which we denoted as a "structural" site. 4F2hc is suggested to stabilise the scaffold domain of LAT1 in the membrane, which may facilitate the local conformational shifts of the bundle domain elements during the transport cycle (Yan *et al.*, 2019a). Collectively, this site likely uses cholesterol to bridge the 4F2hc with LAT1 so that the latter may transport amino acids. F445L (red site) and V378M (green site) in LAT1 were also found in samples of malignant melanoma and adenocarcinoma, respectively (**Figure 6.15**, **Table 6.3**).

While there were no corresponding cancer-related mutations in any of the equivalent residue positions for H367, V378, or F445 in xCT and $b^{0,+}$AT, cancer mutation hotspots for an Arg residue located two amino acids downstream from the conserved H367 (**Figure 6.16**) were found (xCT: R361C, $b^{0,+}$AT: R347T, human SLC7A4: R363W), indicating that Arg enhances the role of H367 in coordinating cholesterol in these proteins.

Protein	Mutant	COSMIC ID	Effect
LAT1	F445L	COSU419	Malignant melanoma
LAT1	V378M	COSU376	Adenocarcinoma
xCT	R361C	COSP44021	Equiv. to H367-LAT1: cancer
$b^{0,+}$AT	R347T	COSU382	Equiv. to H367-LAT1: cancer
SLC3A2	R183S	COSU419	Endometroid carcinoma
SLC7A4	R363W/frameshift	COSU323	Japanese liver cancer;
		COSU376	Colon adenocarcinoma

Table 6.3 HAT mutants located in red and green site cholesterol sites found in cancer patients.

A structure-based sequence alignment of SLC7 members revealed that the aforementioned residues are highly conserved within the HAT subfamily (**Figure 6.16**). H367 is conserved in all HATs (except $b^{0,+}$AT), while V378 has conservative bulky hydrophobic residues in all SLC7s, with the sequence consensus being Leu or Ile (**Figure 6.12** & **Figure 6.15**). SLC3A2 is present as the extracellular subunit in at least 7 out of the 14 members of the SLC7 family and thus R183 is conserved in these complexes by default. The transporter $b^{0,+}$AT is coupled with rBAT, which conserves this Arg residue as R86. Overall, the green cholesterol site morphology and residual features are conserved amongst HATs, while human SLC7A4 also has a high charge density (H361, R363). The latter is not coupled with an SLC3 ancillary protein – as is the case for CATs. Similarly, the red site residue that caused the largest decrease in transport, F445, is strictly conserved in HATs, and is semiconserved as a Tyr in hCATs, plant CAT4 and SLC7A14 (**Figure 6.16**).

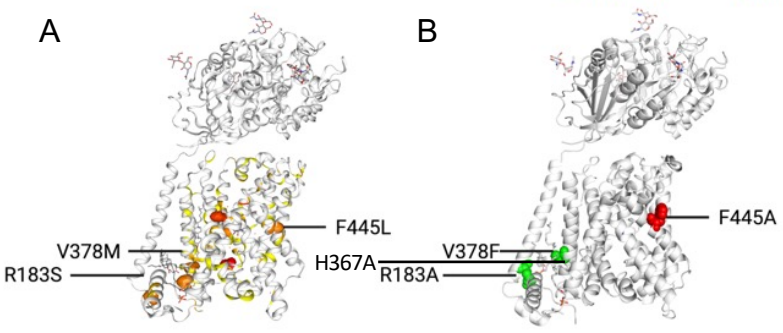


Figure 6.15 LAT1-SLC3A2 heterodimer mutants that have been connected to cancer. (A) Mutants in the COSMIC database and (B) equivalent mutants that were identified from the multiscale CG-AT MD simulations reported in this chapter in the green and red sites.

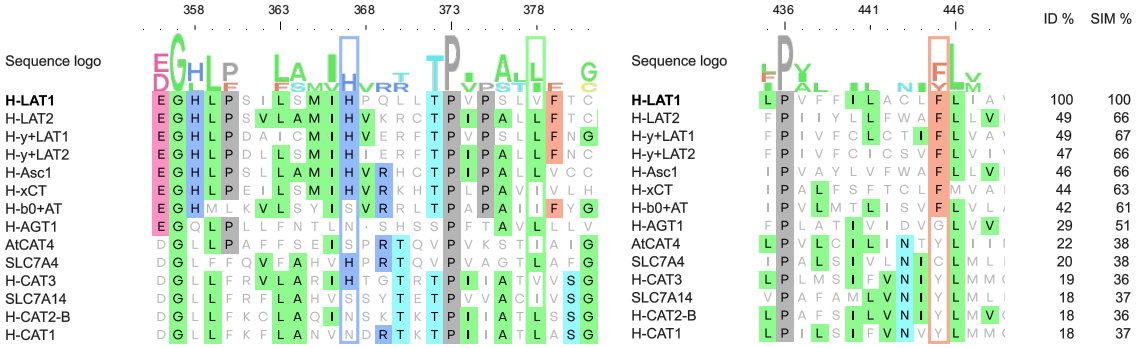


Figure 6.16 Structure-guided multiple sequence alignment after all-atom superposition of available eukaryotic SLC7 structures from the PDB (LAT1 – 6IRS; LAT2 – 7CMI; xCT – 7P9V; $b^{0,+}$AT – 6LI9; *At*CAT4 – chapter 5) and Alphafold2 structure predictions. Residue colouring according to side chain chemistry. Residues that are identical to the consensus sequence logo of the alignment are coloured according to side chain properties (sky blue – cationic; cyan – polar; green – alkyl; grey – proline; carnation pink – anionic). Sequence disagreements are coloured white. Residue numbering displayed is according to human LAT1. Residues equivalent to H367, V378 and F445 are highlighted in a consensus-coloured box. Global sequence identities and similarities to human LAT1) are displayed.

6.3.2 Comparison of AtCAT4 and LAT1 Cholesterol Sites

When solving plant CAT4 in the outward-open state, we modelled a cholesterol in a steroid density located at a location similar to the red site found for

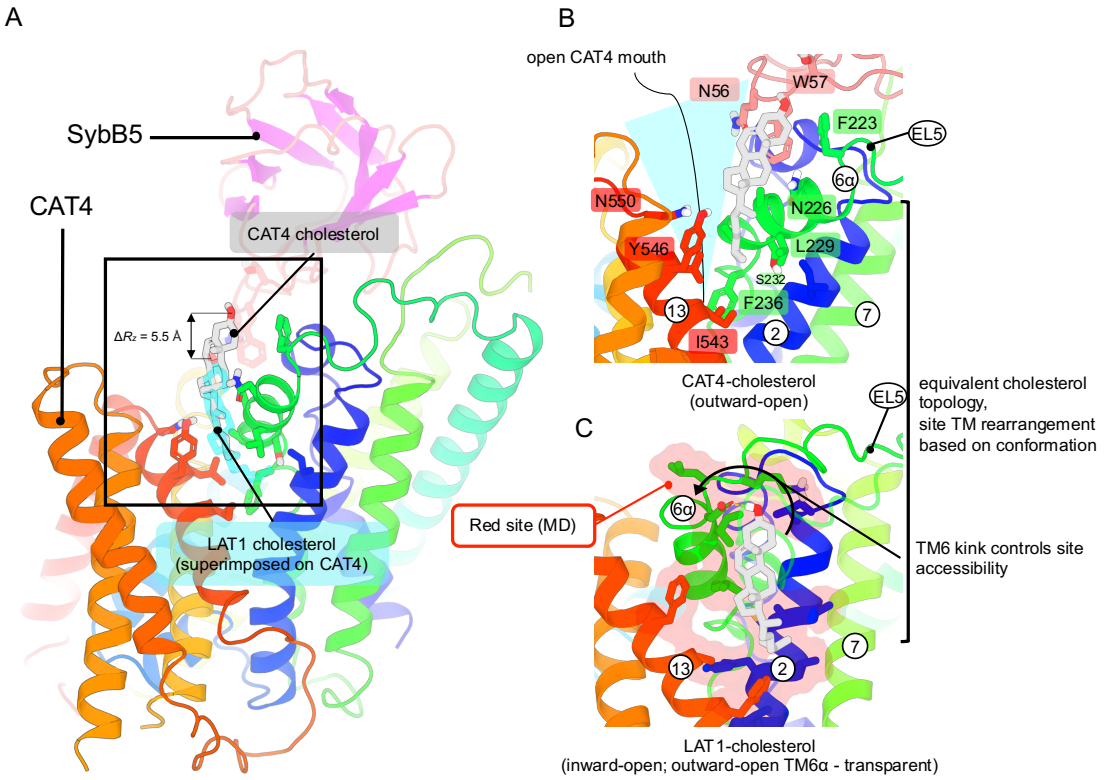


Figure 6.17 Structural comparison of outward-open CAT4-cholesterol complex and inward-open LAT1-cholesterol complex (at red site; MD simulation). (A) Both cholesterols bind the same *xy* location and have a *z*-distance 5.5 Å, after superimposing CAT4 and LAT1 on backbone atoms excluding bundle domain helix residues. (B) Cholesterol binding site in CAT4. (C) Cholesterol binding site (red) in inward-open LAT1 and the TM6α helix outward-open configuration shown in transparent green, highlighting steric hindrance in cholesterol binding upon conformational change.

inward-open LAT1 and $b^{0,+}AT$ in MD (**Figure 6.17-A**). Direct interactions between cholesterol and the F445-equivalent – Y546 in plant CAT4 – are seen in the CAT4 structure (**Figure 6.17-B**). Interestingly, the helices lining the binding site have a notable difference compared to the LAT1 red site. Due to the opening motion of TM6α in the outward-open conformation of CAT4, the cholesterol site boundaries are shifted. As a result, TM2 and EL5 do not coordinate cholesterol; the latter is "sandwiched" between TM6α and TM13 – with some auxiliary interactions with SybB5 (see **chapter 5**, **Figure 6.17-B**). A notable CAT4 residue within the cholesterol site is F236, the residue next to the extracellular thin gate F237,

suggesting cholesterol interactions with TM6α may be stabilising the outward-open conformation in CAT4. The nMS spectrum for CAT4 (chapter 5) did not uncover an intact CAT4-cholesterol complex, likely the result of cholesterol producing small m/z shifts in an otherwise noisy baseline.

In LAT1 cholesterol is captured by TM2-TM6α-TM13 residues whereas in CAT4 TM6α-TM13 (Figure 6.17-C). The structural comparison suggests that, as TM6α rearranges during the inward-to-outward transition, it causes a steric clash with the bound cholesterol. Although this comparison is between a CAT and a HAT, it suggests that cholesterol binds the red site in the open states of the transport cycle. Without additional structural information at different conformational state, the current data suggests that HATs have a preference for the red site in the inward-open state while CAT4 has the capacity do this in outward-open state.

6.4 Chapter Conclusions

HATs are predominantly solved in their inward-open state in the absence of inhibitors. Cholesterol is the predominant lipid bound in HAT cryoEM maps. In this chapter, hypotheses deriving from structural and functional literature on HAT-cholesterol interactions were investigated.

- LAT1 forms an enriched annular cholesterol solvation shell in both binary POPC:Cholesterol and in a plasma membrane mimetic membrane.
- CG MD simulations identified cholesterol binding sites to be topologically conserved in inward-open HATs. Sites showed differences in their interaction residence times ($RT = {}^{1}/_{k_{off}}$) and binding affinities (K_d). Similarities

in RTs and K_d were detected for the green and red sites amongst xCT, b^{0+}AT and LAT1. xCT lacked the red site, but had two unique sites, cyan and orange. Comparison of the HAT binding sites with experimental APC transporter-cholesterol complexes verified the binding site calculations.

- The green, blue, orange, and cyan sites were stable (distance <5 Å) in atomistic simulation, further verifying CG calculations for these sites. The red site cholesterols in $b^{0,+}$AT and LAT1 rearranged upon AT refinement, displacing EL3 (TM6α-TM7 loop).
- The conformational change, outward ⟶ inward in LAT1 resulted in no affinity change for the green site, but increased affinity for the red site. Indeed, the cryoEM structures of the inward-facing NKCC1 and NCC are solved with a red site cholesterol, stabilising this conformation (Nan *et al.*, 2022; Neumann *et al.*, 2022).
- Cholesterol occupancy in the red and green sites was reduced in site mutants of the inward-open LAT1. The *in silico* mutagenesis data is supported by mutations of the green and red sites in *in vitro* transport assays, which resulted in transport reduction (with the exception of K237A). This data supports the cholesterol-dependency of LAT1 uptake function reported previously (Dickens *et al.*, 2017; Cosco *et al.*, 2020).
- Residues involved in coordination of green and red site cholesterols are conserved in HATs and their mutations have been detected in cancer patients, suggesting a link between correct plasma membrane trafficking and/or transport of HATs in the membrane and cholesterol binding.

The results here, in conjunction with the literature, provide structural insight

by which cholesterol binding at discrete binding sites stabilises HATs in the membrane. We suggest the green binding site serves to accommodate cholesterol as a structural ingredient of the heterodimer, situating it between the SLC7 and SLC3 domains in the membrane, previously suggested from a structural standpoint (Lee *et al.*, 2019b; Yan *et al.*, 2019a). On the other side, binding to the red site seems to be preferred in the inward-open state and would stabilise this conformation. Indeed, presence of cholesterol increases LAT1 affinity for internal substrates in proteoliposomes (Cosco *et al.*, 2020), likely by stabilising this conformation, leading to increased accessibility of the inward-open state to substrates.

6.5 Limitations & Future Directions

Simulation studies using the MARTINI force field have provided valuable insights into protein-lipid interactions (Corradi *et al.*, 2018; Corradi, Besian I. Sejdiu, *et al.*, 2019; Song *et al.*, 2022). MARTINI 2, which was used in this thesis has limitations (Marrink and Tieleman, 2013; Javanainen, Martinez-Seara and Vattulainen, 2017; Alessandri *et al.*, 2019; Jarin, Newhouse and Voth, 2021). Most importantly, while binding affinities are not as affected, the smoothing of the free energy landscape and the increased diffusion rates for hydrocarbons, including lipids, may have resulted in unpredictable boosts in binding kinetics. However, the general consensus of the calculated cholesterol binding sites with the structural data support that usage of the model to capture binding sites is likely justified. Future studies of HAT-cholesterol interactions using MARTINI 3 (Souza *et al.*, 2021) may be useful to compare with the results presented here, although during this investigation, cholesterol parameters were not published.

In the future, it is suggested that more biological repetitions of the mutant cell-based transport data be performed (currently N=3), which are presently limited due to time restrictions this thesis project. The suggestions can include more red and green site mutants on LAT1. As this investigation started with comparisons between LAT1 and other HAT transporters, a future direction may be to include a more thorough investigation of other HATs. A starting point would certainly be 4F2hc/xCT, which specifically uptakes L-cystine, making transport assays robust with higher signal-to-noise.

Final Remarks & Future Perspectives

CHAPTER 7

In the field of membrane biology a lot remains unresolved, and thus fundamental science research remains important. This thesis describes a multidisciplinary effort, connecting structure to function and molecular dynamics, to understanding previously untapped aspects of the SLC7 amino acid transporter physiology.

Application of cryoEM is still a new avenue for many membrane protein structural biologists, with models < 5.0 Å arising only after 2014 and the shift in resolution from subnanometer to near atomic resolution being evident as late as 2016 (Chiu *et al.*, 2021). On the onset of my D.Phil in 2018, eukaryotic SLC7 transporter structures were lacking and prokaryotic structures were the main models for rationalising function. In 2019, the first high-resolution human HAT transporter was solved by cryoEM (Yan *et al.*, 2019b) and other members have started to follow. Interestingly, all the structural-functional studies concerned HATs, while CATs remained elusive, most likely due to their difficult purification, resistance to crystallisation and small size in cryoEM (< 70 kDa). Nevertheless, the connection of CATs with arginine-NO/polyamine metabolism and inflammation (K. E. J. Jungnickel, Parker and Newstead, 2018) necessitates their further study.

7.1 Why Was SLC7A4 Studied?

The selection of a eukaryotic CAT was decided. The high binding site similarity between prokaryotic H^+-coupled *Gk*ApcT and the functionally elusive SLC7 member 4 provided interesting hypotheses. The first was that SLC7A4 homologues are H^+-coupled transporters by conserving the H^+-coupling residue found in *Gk*ApcT, unlike other SLC7s. This hypothesis presented an interesting turning point in our mechanistic taxonomy. Secondly, a similar substrate binding profile to *Gk*ApcT M321S was expected for SLC7A4, namely L-arginine, L-lysine, and L-ornithine. It was therefore surprising that SLC7A4 was not found to transport amino acids and polyamines (Wolf *et al.*, 2002) in oocytes. Could SLC7A4 be a non-transporter? An example of an MFS-transporter-fold protein, UNC93B1, being a TLR chaperone instead of a solute carrier (Ishida *et al.*, 2021), raised the question of whether an orphan SLC7 member can adopt similar roles.

With the use of heterologous overexpression and protein quality control a suitable homologue from *Arabidopsis thaliana* of the substrate-orphan SLC7A4 transporter was selected (chapter 3). A novel entry in EM fiducials, synthetic nanobody technology, greatly expediated structure solution by cryoEM since crystallisation proved recalcitrant (chapter 4 and 5). Retrospectively, the reason behind CAT4 resistance to crystallisation seems to be its long, unstructured intracellular loop IL5; a feature only present in eukaryotic SLC7A1-4 and SLC7A14.

7.2 Is SLC7A4 an Amino Acid Transporter?

AtCAT4 Structure & Substrate Recognition

The obtained structure testifies that *At*CAT4 conserves the necessary binding site "ingredients" to recognise α-amino acids, in accordance with other APC-fold transporters. Specifically, it has the TM1 and TM6 architecture which serves to coordinate α-amino and carboxy substrate groups and a polar distal cavity for substrate side chain selectivity (salt-bridge, H-bond interactions) via its D116, D243 and S327. The electronegative binding site entrance of *At*CAT4 provides an electrical component to cationic amino acid selectivity. We uncovered an extended electrostatic interaction network (thick gate), providing a structural basis for the outward-open conformation, which is rearranged in the outward-facing substrate-occluded conformation induced by binding site conformational changes. This is likely to occur during substrate entrance, whereupon V46 (TM1) is displaced inducing TM1 structural rearrangement to both (a) capture the substrate and (b) rearrange the thick gate residues.

The development of an amino acid binding assay helped us elucidate the site-dependent binding of α-amino acids and elucidate the substrate selectivity profile of plant CAT4. We proved that there is preference for cationic amino acids, and, secondarily for L-leucine at neutral pH. In the future it will be important to investigate the effects of differential pH to test whether other amino acids with titratable side chains (e.g., L-histidine) can bind. Via its D116, D243 and S327 residues, CAT4 captures cationic amino acid side chains as suggested by MD simulations and functional mutants, while side chain length modulates the ap-

parent binding affinity to the transporter, with L-lysine having the highest affinity, $K_{d,app}$.

Since our assay reflects thermal stability of the protein, the differential B_{max} between CAAs reports differential stabilisation of the binding site. The highest stabilisation was with L-ornithine and not the longer counterpart, L-lysine. When considering its shorter length, a mechanism may be that L-ornithine "pulls" TM1 and TM6 closer to TM3 and TM8 via electrostatics and water-bridge network, respectively. This likely stabilises the protein. The other CAAs, being longer, do not need to exert a pulling force and are able to reach D116 without pulling TM3. Another example of the induced-fit mechanism driven by substrate recognition is L-cystine binding to xCT, which proceeds via kinking of TM8 (Parker, Deme, Kolokouris, *et al.*, 2021), raising a new paradigm in substrate recognition in amino acid transporters involving geometrical adaptation other than residue fingerprint. Similar to xCT which selects its substrates via a cationic arginine clamp, CAT4 uses the anionic residues to interact with CAAs, D116 and D243, suggesting an equivalent anionic clamp mechanism for selectivity in CATs. To understand this better, mutational transport assays and more structures of substrates in different conformations need to become available by obtaining new intermediates through nanodiscs or detergent selection.

*At*CAT4 has a subcellular localisation. In plants the vacuole serves as a storage unit for free amino acids and interestingly some amino acids have higher abundance in the vacuole than the cytosol (Tohge *et al.*, 2011). These amino acids require dedicated transporters to migrate them across the tonoplast membrane.

Does SLC7A4 Transport Cationic Amino Acids?

Although binding data shows plant CAT4 to recognise CAAs in its binding site, the protein could not be characterised by the means of a functional assay. Future directions include the optimisation of CAT4 proteoliposome reconstitution to verify the observations of the binding assay with transport assays. It would be interesting to test whether the CAT4-specific sybodies, selected for in LMNG, could act as inhibitors in transport assay similarly to nanobodies for a peptide transporter (Parker, Deme, Wu, et al., 2021). This may rationalise why a relipidated ornithine-bound CAT4-SybB5 structure was not realised.

Human SLC7A4 was assayed, with a similar substrate profile to the plant protein being expected. The results showed that it behaves differently to the plant protein by not binding CAAs. We verified this in cell-based transport assays with L-arginine and L-ornithine. Our results are in agreement with lack of L-arginine uptake by SLC7A4 reported by others (Wolf *et al.*, 2002; Vanoaica *et al.*, 2016). We uncovered that SLC7A4 unexpectedly binds L-leucine ($\Delta T_m > 10$ °C) and preliminary transport data suggest SLC7A4-dependent transport. As a GFP-fusion, SLC7A4 localised primarily in the plasma membrane. The need for a L-leucine transporter in the plasma membrane, other than LAT1 (Kandasamy *et al.*, 2018b), is unclear but may be linked with L-leucine regulation in growth control (Lawrence and Zoncu, 2019). More work in L-leucine transport assays is nevertheless necessary before concluding transport: (a) more biological repeats, (b) binding site mutants, (c) competition of transport by other amino acids and (d) characterisation of transport kinetics by K_M and V_{max}. We plan to complete these steps in the immediate future.

The structural and functional findings support an amino acid transporter role for both plant CAT4 and human SLC7A4 although substrate preferences differ. We selected plant CAT4 (*At*CAT4) as an SLC7A4 homologue based on its sequence conservation (chapter 3). Our reasoning to select *At*CAT4 was, that among the other *At*CATs, *At*CAT4 had a closer sequence-relation to SLC7A4. Specifically, amongst the *At*CATs, *At*CAT4 presented the highest sequence identity to SLC7A4 (albeit, limited to 34%), whereas the remaining *At*CATs were <30%, thus justifying our initial selection of *At*CAT4 as a model for SLC7A4. Retrospectively, since it binds L-arginine/ L-lysine/ L-ornithine, *At*CAT4 is better described as an orthologue of human CATs 1-3 (SLC7A1-3, binding site sequence similarity = 92 %) which transport these substrates, instead of SLC7A4 (binding site sequence similarity = 78 %).

7.3 Binding Site Modularity in SLC7s

Why is there a difference between the human and plant proteins and what can it teach about SLC7 substrate selectivity? The discrepancy seems to be a hydrophobic clamp composed of two Leu residues which protrude in the binding site of human SLC7A4, and which are absent in plant CAT4 and other human CATs. Similarly, two Arg residues were shown to have an equally significant role in determining the affinity and recognition of L-cystine (chapter 5) (Parker, Deme, Kolokouris, *et al.*, 2021). The results recommend that substrate preferences are subtle in eukaryotic SLC7s, as signified by the complete abolishing of CAA binding and transport by human SLC7A4 compared to plant CAT4. A similar L-Ala→L-Arg affinity switch was induced with M321S mutation in *Gk*ApcT (K. E. J. Jungnickel, Parker and Newstead, 2018). An interesting test of this theory

would be the mutation of the Leu clamp in human SLC7A4 to the equivalent residues in hCATs to test whether transport of CAAs can occur in cell-based assays and TEVC.

7.4 Cholesterol Interactions with SLC7s

Amino acid transporters in the SLC7 family and related families [SLC36, 38] seem to have a widespread distribution in the cell's membrane network (Schiöth *et al.*, 2013b), where cholesterol concentrations differ (Levental and Lyman, 2022). Still even in organelles with lower cholesterol levels, members of the APC family have shown interactions, such as SLC38A9 which senses lysosomal cholesterol (Castellano *et al.*, 2017b). These proteins share a common fold yet are localised in cellular contexts which differ in their local lipid environment. Similarly to GPCRs where binding sites are common among different members (Song *et al.*, 2022), the APC fold seems to be driving lipid interactions.

The plasma membrane (HAT residence location) is rich in cholesterol (up to 40 mol%) (Levental and Lyman, 2022). The effect of sterols on the activity and stability of LAT1-SLC3A2 heterodimer has been reported (Dickens *et al.*, 2017). We investigated the specificity of this interaction on selected HAT transporters with a computational pipeline (**chapter 6**). We have demonstrated that specific cholesterol interactions in HATs are not artifacts from purification or erroneously overinterpreted density data. Unbiased MD simulation identified the equivalent cholesterol sites found in a wide array of APC fold solute carrier structures suggesting common sites for different carriers. We discussed the similarities and differences between xCT, $b^{0,+}$AT and LAT1. They all had in common what we

termed as the green site, a cholesterol pocket between SLC3 and SLC7 interprotein space. For LAT1, the site was accessible in inward- and outward-facing conformations. HATs have the ancillary SLC3 domain to assist in trafficking, although how this domain assists in the trafficking from the Golgi to the PM remains unknown. Via the green site, cholesterol may play an indirect role in stabilising the heterodimers in the membrane to traffic from the Golgi to the plasma membrane. $b^{0,+}$AT and LAT1 shared the red site in the inward-facing conformation, a cholesterol binding pocket contacting mechanistically important TM6α. Cholesterol had a lower affinity for the same site in the outward-facing conformation in LAT1. A similar mechanism by which cholesterol binds mechanistically important TM1α in SERT was suggested to stabilise the outward-open conformation by sterically hindering the opening of TM1α (Zeppelin *et al.*, 2018). Our data suggest cholesterol would stabilise the inward conformation in LAT1. Importantly, assaying mutants of the green and red sites in LAT1 an influence in transport activity was found. Transport kinetics in membranes of defined cholesterol concentration combined with nMS analysis of photochemically-labelled cholesterol-bound HATs could be useful in deciphering which binding sites are important for transport.

TM6α coordination of a steroid molecule was seen in the CAT4 outward-open structure at a position similar to the red site. Plant CAT4 is physiologically in the vacuole, the acidified plant/fungal version of the human lysosome, a membrane organelle that has low steroid content (Tsuji and Fujimoto, 2018). Therefore, the interaction with steroid observed in our structure is likely a specific one, especially since CHS was not used in the purification. A CAT4 structure in an

inward-facing conformation would allow comparison of cholesterol site occupancies and binding energetics.

Whether cholesterol has different pocket preferences for CATs compared to HATs is an interesting future direction that could be addressed with a combined cryoEM and nMS analysis of SLC7s purified from native membranes in co-polymer nanodiscs. The binding information from nMS and structures from cryoEM would provide starting material for MD simulation to titrate site affinities and preferences arising in lipid compositions of different intracellular organelles. This approach would be the first step to deciphering the lipid code for SLC7s.

Bibliography

van 't Klooster, J. S. *et al.* (2020) 'Membrane Lipid Requirements of the Lysine Transporter Lyp1 from Saccharomyces cerevisiae', *Journal of Molecular Biology*. Academic Press, 432(14), pp. 4023–4031. doi: 10.1016/j.jmb.2020.04.029.

Abramson, J. *et al.* (2003) 'Structure and mechanism of the lactose permease of Escherichia coli.', *Science (New York, N.Y.)*, 301(5633), pp. 610–5. doi: 10.1126/science.1088196.

Abramson, J. and Wright, E. M. (2009) 'Structure and function of Na+-symporters with inverted repeats', *Current Opinion in Structural Biology*. Elsevier Current Trends, pp. 425–432. doi: 10.1016/j.sbi.2009.06.002.

Aimon, S. *et al.* (2014) 'Membrane Shape Modulates Transmembrane Protein Distribution', *Developmental Cell*. Dev Cell, 28(2), pp. 212–218. doi: 10.1016/j.devcel.2013.12.012.

Aldeghi, M. *et al.* (2017) 'Statistical Analysis on the Performance of Molecular Mechanics Poisson-Boltzmann Surface Area versus Absolute Binding Free Energy Calculations: Bromodomains as a Case Study', *Journal of Chemical Information and Modeling*, 57(9), pp. 2203–2221. doi: 10.1021/acs.jcim.7b00347.

Alessandri, R. *et al.* (2019) 'Pitfalls of the Martini Model', *Journal of Chemical Theory and Computation*, 15(10), pp. 5448–5460. doi: 10.1021/acs.jctc.9b00473.

Allen, J. A., Halverson-Tamboli, R. A. and Rasenick, M. M. (2007) 'Lipid raft microdomains and neurotransmitter signalling', *Nature Reviews Neuroscience*, 8(2), pp. 128–140. doi: 10.1038/nrn2059.

del Amo, E. M., Urtti, A. and Yliperttula, M. (2008) 'Pharmacokinetic role of L-type amino acid transporters LAT1 and LAT2', *European Journal of Pharmaceutical Sciences*, pp. 161–174. doi: 10.1016/j.ejps.2008.06.015.

Anderluh, A. *et al.* (2017) 'Direct PIP2 binding mediates stable oligomer formation of the serotonin transporter.', *Nature communications*, 8(May 2016), p. 14089. doi: 10.1038/ncomms14089.

Andersen, O. S. and Koeppe, R. E. (2007) 'Bilayer thickness and membrane protein function: An energetic perspective', *Annual Review of Biophysics and Biomolecular Structure*. Annu Rev Biophys Biomol Struct, pp. 107–130. doi: 10.1146/annurev.biophys.36.040306.132643.

Angelovici, R. *et al.* (2017) 'Network-guided GWAS improves identification of genes affecting free amino acids', *Plant Physiology*, 173(1), pp. 872–886. doi: 10.1104/pp.16.01287.

Ansell, T. B. *et al.* (2021) 'Relative Affinities of Protein-Cholesterol Interactions from Equilibrium Molecular Dynamics Simulations', *Journal of Chemical Theory and Computation*, 17(10), pp. 6548–6558. doi: 10.1021/acs.jctc.1c00547.

Arroyo, J. P., Kahle, K. T. and Gamba, G. (2013) 'The SLC12 family of electroneutral cation-coupled chloride cotransporters', *Molecular Aspects of Medicine*, 34(2–3), pp. 288–298. doi: 10.1016/j.mam.2012.05.002.

Aryal, P. *et al.* (2015) 'Influence of lipids on the hydrophobic barrier within the pore of the TWIK-1 K2P channel', *Channels*, 9(1), pp. 44–49. doi: 10.4161/19336950.2014.981987.

Asarnow, D., Palovcak, E., Cheng, Y. (2019) 'UCSF pyem v0.5', *Zenodo*. doi: 10.5281/zenodo.3576630.

Baker, N. A. *et al.* (2001) 'Electrostatics of nanosystems: Application to microtubules and the ribosome', *Proceedings of the National Academy of Sciences of the United States of America*, 98(18), pp. 10037–10041. doi: 10.1073/pnas.181342398.

Ban, B., Sharma, M. and Shetty, J. (2020) 'Optimization of Methods for the Production and Refolding of Biologically Active Disulfide Bond-Rich Antibody Fragments in Microbial Hosts.', *Antibodies (Basel, Switzerland)*. Switzerland, 9(3). doi: 10.3390/antib9030039.

Bannai, S. and Kitamura, E. (1981) 'Role of proton dissociation in the transport of cystine and glutamate in human diploid fibroblasts in culture.', *Journal of Biological Chemistry*. Â© 1981 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 256(11), pp. 5770–5772. doi: 10.1016/S0021-9258(19)69273-2.

Bartoccioni, P. *et al.* (2019) 'Functional characterization of the alanine-serine-cysteine exchanger of Carnobacterium sp AT7', *Journal of General Physiology*. Rockefeller University Press, 151(4), pp. 505–517. doi: 10.1085/jgp.201812195.

Bayburt, T. H., Grinkova, Y. V. and Sligar, S. G. (2002) 'Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins', *Nano Letters*. American Chemical Society, 2(8), pp. 853–856. doi: 10.1021/nl025623k.

Beale, J. H. *et al.* (2015) 'Crystal Structures of the Extracellular Domain from PepT1 and PepT2 Provide Novel Insights into Mammalian Peptide Transport.', *Structure (London, England : 1993)*. Elsevier, 23(10), pp. 1889–1899. doi: 10.1016/j.str.2015.07.016.

Bennett, J. L., Nguyen, G. T. H. and Donald, W. A. (2022) 'Protein-Small Molecule Interactions in Native Mass Spectrometry', *Chemical Reviews*, 122(8),

pp. 7327–7385. doi: 10.1021/acs.chemrev.1c00293.

Berendsen, H. J. C. *et al.* (1984) 'Molecular dynamics with coupling to an external bath', *The Journal of Chemical Physics*, 81(8), pp. 3684–3690. doi: 10.1063/1.448118.

Berendsen, H. J. C. and Van Gunsteren, W. F. (1986) 'Practical Algorithms for Dynamic Simulation , in Enrico Fermi Summer School , Varenna , Italy ', *Molecular-Dynamics Simulations of Statistical-Mechanical Systems*, pp. 43–65.

Betaneli, V., Petrov, E. P. and Schwille, P. (2012) 'The role of lipids in VDAC oligomerization', *Biophysical Journal*. The Biophysical Society, 102(3), pp. 523–531. doi: 10.1016/j.bpj.2011.12.049.

Betsunoh, H. *et al.* (2013) 'Increased expression of system large amino acid transporter (LAT)-1 mRNA is associated with invasive potential and unfavorable prognosis of human clear cell renal cell carcinoma', *BMC Cancer*, 13, p. 509. doi: 10.1186/1471-2407-13-509.

Bill, R. M. *et al.* (2011) 'Overcoming barriers to membrane protein structure determination', *Nature Biotechnology*. United States, 29(4), pp. 335–340. doi: 10.1038/nbt.1833.

Bloch, J. S. *et al.* (2021) 'Development of a universal nanobody-binding Fab module for fiducial-assisted cryo-EM studies of membrane proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 118(47). doi: 10.1073/pnas.2115435118.

Blondel, V. D. *et al.* (2008) 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*. doi: 10.1088/1742-5468/2008/10/P10008.

Bogdan, C. (2001) 'Nitric oxide and the immune response', *Nature Immunology*, 2(10), pp. 907–916. doi: 10.1038/ni1001-907.

Bolla, J. R. *et al.* (2019) 'Membrane protein-lipid interactions probed using mass spectrometry', *Annual Review of Biochemistry*. Annu Rev Biochem, pp. 85–111. doi: 10.1146/annurev-biochem-013118-111508.

Borst, P. and Oude Elferink, R. (2002) 'Mammalian ABC transporters in health and disease', *Annual Review of Biochemistry*. Annu Rev Biochem, pp. 537–592. doi: 10.1146/annurev.biochem.71.102301.093055.

Bosshart, P. D. and Fotiadis, D. (2019) *Secondary Active Transporters*, *Subcellular Biochemistry*. Springer International Publishing. doi: 10.1007/978-3-030-18768-2_9.

Botelho, A. V. *et al.* (2006) 'Curvature and Hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes', *Biophysical Journal*. The Biophysical Society, 91(12), pp. 4464–4477. doi: 10.1529/biophysj.106.082776.

Botte, M. *et al.* (2022) 'Cryo-EM structures of a LptDE transporter in complex with Pro-macrobodies offer insight into lipopolysaccharide translocation', *Nature Communications*. Springer US, 13(1), pp. 1–10. doi: 10.1038/s41467-022-29459-2.

Boudker, O. and Verdon, G. (2010) 'Structural perspectives on secondary active transporters', *Trends in Pharmacological Sciences*, 31(9), pp. 418–426. doi: 10.1016/j.tips.2010.06.004.

Bräuer, P. *et al.* (2019) 'Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor', *Science*. American Association for the Advancement of Science, 363(6431), pp. 1103–1107. doi: 10.1126/science.aaw2859.

Bräuer, P. (2019) *The KDEL Receptor: Structural & Functional Investigations*.

Braun, D. *et al.* (2011) 'Developmental and cell type-specific expression of thyroid hormone transporters in the mouse brain and in primary brain cells', *GLIA*, 59(3), pp. 463–471. doi: 10.1002/glia.21116.

Bröer, S. and Gether, U. (2012) 'The solute carrier 6 family of transporters', *British Journal of Pharmacology*. John Wiley & Sons, Ltd, 167(2), pp. 256–278. doi: 10.1111/j.1476-5381.2012.01975.x.

Brown, J. M. *et al.* (2014) 'In vitro characterization of a small molecule inhibitor of the alanine serine cysteine transporter -1 (SLC7A10)', *Journal of Neurochemistry*, 129(2), pp. 275–283. doi: 10.1111/jnc.12618.

Bussi, G., Donadio, D. and Parrinello, M. (2007) 'Canonical sampling through velocity rescaling', *Journal of Chemical Physics*, 126(1). doi: 10.1063/1.2408420.

Caesar, J. *et al.* (2020) 'SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time', *Journal of Structural Biology: X*. Elsevier, 4(November), p. 100040. doi: 10.1016/j.yjsbx.2020.100040.

Caffrey, M. (2008) 'On the Mechanism of Membrane Protein Crystallization in Lipidic Mesophases', *Crystal Growth & Design*. American Chemical Society, 8(12), pp. 4244–4254. doi: 10.1021/cg800693r.

Caffrey, M. (2015) 'IYCr crystallization series Structural Biology Communications A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes', *Acta Cryst*, 71, pp. 3–18. doi: 10.1107/S2053230X14026843.

Caplan, R. *et al.* (2008) 'Childhood absence epilepsy: Behavioral, cognitive, and linguistic comorbidities', *Epilepsia*. John Wiley & Sons, Ltd, 49(11), pp. 1838–1846. doi: 10.1111/j.1528-1167.2008.01680.x.

Carter, C. *et al.* (2004) 'The vegetative vacuole proteome of Arabidopsis thaliana reveals predicted and unexpected proteins', *Plant Cell*, 16(12), pp. 3285–

3303. doi: 10.1105/tpc.104.027078.

Castellano, B. M. *et al.* (2017a) 'Lysosomal cholesterol activates mTORC1 via an SLC38A9-Niemann-Pick C1 signaling complex', *Science*, 355(6331), pp. 1306–1311. doi: 10.1126/science.aag1417.

Castellano, B. M. *et al.* (2017b) 'Lysosomal cholesterol activates mTORC1 via an SLC38A9-Niemann-Pick C1 signaling complex', *Science*, 355(6331), pp. 1306–1311. doi: 10.1126/science.aag1417.

Celik, L. *et al.* (2008) 'Binding of serotonin to the human serotonin transporter. Molecular modeling and experimental validation', *Journal of the American Chemical Society*. J Am Chem Soc, 130(12), pp. 3853–3865. doi: 10.1021/ja076403h.

César-Razquin, A. *et al.* (2015) 'A Call for Systematic Research on Solute Carriers', *Cell*, pp. 478–487. doi: 10.1016/j.cell.2015.07.022.

Chadda, R. *et al.* (2021) 'Membrane transporter dimerization driven by differential lipid solvation energetics of dissociated and associated states', *eLife*, 10, pp. 1–28. doi: 10.7554/ELIFE.63288.

Chae, P. S. *et al.* (2010) 'Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins', *Nature Methods*, 7(12), pp. 1003–1008. doi: 10.1038/nmeth.1526.

Chantranupong, L. *et al.* (2016) 'The CASTOR Proteins Are Arginine Sensors for the mTORC1 Pathway.', *Cell*. Elsevier, 165(1), pp. 153–164. doi: 10.1016/j.cell.2016.02.035.

Chen, Y. *et al.* (2016) 'Structure of the STRA6 receptor for retinol uptake', *Science*, 353(6302). doi: 10.1126/science.aad8266.

Cheng, Y. (2018) 'Membrane protein structural biology in the era of single particle cryo-EM', *Current Opinion in Structural Biology*. Elsevier Ltd, 52, pp. 58–63. doi: 10.1016/j.sbi.2018.08.008.

Chien, H. C. *et al.* (2018) 'Reevaluating the Substrate Specificity of the L-Type Amino Acid Transporter (LAT1)', *Journal of Medicinal Chemistry*, 61(16), pp. 7358–7373. doi: 10.1021/acs.jmedchem.8b01007.

Chiu, W. *et al.* (2021) 'Evolution of standardization and dissemination of cryo-EM structures and data jointly by the community, PDB, and EMDB', *Journal of Biological Chemistry*. Elsevier B.V, 296, p. 100560. doi: 10.1016/j.jbc.2021.100560.

Christensen, H. N. (1990) 'Role of amino acid transport and countertransport in nutrition and metabolism.', *Physiological Reviews*, 70(1), pp. 43–77. doi: 10.1152/physrev.1990.70.1.43.

Chung, K. Y. *et al.* (2012) 'Role of Detergents in Conformational Exchange of a G Protein-coupled Receptor', *Journal of Biological Chemistry*. Â© 2012

ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 287(43), pp. 36305–36311. doi: 10.1074/jbc.M112.406371.

Closs, E. I., Gräf, P., *et al.* (1997) 'Human cationic amino acid transporters hCAT-1, hCAT-2A, and hCAT-2B: Three related carriers with distinct transport properties', *Biochemistry*, 36(21), pp. 6462–6468. doi: 10.1021/bi962829p.

Closs, E. I., Basha, F. Z., *et al.* (1997) 'Interference of L-arginine analogues with L-arginine transport mediated by the y+ carrier hCAT-2B', *Nitric Oxide - Biology and Chemistry*, 1(1), pp. 65–73. doi: 10.1006/niox.1996.0106.

Closs, E. I. *et al.* (2006) 'Structure and function of cationic amino acid transporters (CATs)', *Journal of Membrane Biology*, pp. 67–77. doi: 10.1007/s00232-006-0875-7.

Colas, C. (2020) 'Toward a Systematic Structural and Functional Annotation of Solute Carriers Transporters—Example of the SLC6 and SLC7 Families', *Frontiers in Pharmacology*. Frontiers, 11, p. 1229. doi: 10.3389/fphar.2020.01229.

Colas, C., Ung, P. M. U. and Schlessinger, A. (2016) 'SLC transporters: Structure, function, and drug discovery', *MedChemComm*. The Royal Society of Chemistry, pp. 1069–1081. doi: 10.1039/c6md00005c.

Coleman, J. A. *et al.* (2019) 'Serotonin transporter–ibogaine complexes illuminate mechanisms of inhibition and transport', *Nature*. Springer US, 569(7754), pp. 141–145. doi: 10.1038/s41586-019-1135-1.

Coleman, J. A., Green, E. M. and Gouaux, E. (2016) 'X-ray structures and mechanism of the human serotonin transporter', *Nature*. Nature Publishing Group, 532(7599), pp. 334–339. doi: 10.1038/nature17629.

Contreras, F. X. *et al.* (2012) 'Molecular recognition of a single sphingolipid species by a protein's transmembrane domain', *Nature*. Nature, 481(7382), pp. 525–529. doi: 10.1038/nature10742.

Corey, R. A. *et al.* (2019) 'Insights into Membrane Protein—Lipid Interactions from Free Energy Calculations', *J. Chem. Theory Comput*, 15, p. 5736. doi: 10.1021/acs.jctc.9b00548.

Corey, R. A. *et al.* (2021) 'Identification and assessment of cardiolipin interactions with E. coli inner membrane proteins', *Science Advances*, 7(34), pp. 1–10. doi: 10.1126/sciadv.abh2217.

Corey, R. A., Stansfeld, P. J. and Sansom, M. S. P. (2020) 'The energetics of protein–lipid interactions as viewed by molecular simulations', *Biochemical Society Transactions*. Portland Press, 48(1), pp. 25–37. doi: 10.1042/BST20190149.

Cornelius, F. (2001) 'Modulation of Na,K-ATPase and Na-ATPase activity by phospholipids and cholesterol. I. Steady-state kinetics', *Biochemistry*.

Biochemistry, 40(30), pp. 8842–8851. doi: 10.1021/bi010541g.

Corradi, V. *et al.* (2018) 'Lipid–Protein Interactions Are Unique Fingerprints for Membrane Proteins', *ACS Central Science*. American Chemical Society, 4(6), pp. 709–717. doi: 10.1021/acscentsci.8b00143.

Corradi, V., Sejdiu, Besian I., *et al.* (2019) 'Emerging Diversity in Lipid-Protein Interactions', *Chemical Reviews*, 119(9), pp. 5775–5848. doi: 10.1021/acs.chemrev.8b00451.

Corradi, V., Sejdiu, Besian I, *et al.* (2019) 'Emerging Diversity in Lipid–Protein Interactions', *Chemical Reviews*, 119(9), pp. 5775–5848. doi: 10.1021/acs.chemrev.8b00451.

Cosco, J. *et al.* (2019a) 'Regulatory aspects of the vacuolar cat2 arginine transporter of s. Lycopersicum: Role of osmotic pressure and cations', *International Journal of Molecular Sciences*, 20(4), p. 906. doi: 10.3390/ijms20040906.

Cosco, J. *et al.* (2019b) 'Regulatory aspects of the vacuolar cat2 arginine transporter of s. Lycopersicum: Role of osmotic pressure and cations', *International Journal of Molecular Sciences*. MDPI AG, 20(4). doi: 10.3390/ijms20040906.

Cosco, J. *et al.* (2020) 'ATP modulates SLC7A5 (LAT1) synergistically with cholesterol', *Scientific Reports*. Nature Publishing Group UK, 10(1), p. 16738. doi: 10.1038/s41598-020-73757-y.

Coste, B. *et al.* (2010) 'Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels', *Science*. Science, 330(6000), pp. 55–60. doi: 10.1126/science.1193270.

Crichton, P. G. *et al.* (2015) 'Trends in thermostability provide information on the nature of substrate, inhibitor, and lipid interactions with mitochondrial carriers', *Journal of Biological Chemistry*. Â© 2015 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 290(13), pp. 8206–8217. doi: 10.1074/jbc.M114.616607.

Crippen, G. M. (2001) 'A Gaussian statistical mechanical model for the equilibrium thermodynamics of barnase folding', *Journal of Molecular Biology*, 306(3), pp. 565–573. doi: 10.1006/jmbi.2000.4401.

Custódio, T. F. *et al.* (2020) 'Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2', *Nature Communications*. Springer US, 11(1). doi: 10.1038/s41467-020-19204-y.

Das, A. *et al.* (2014) 'Three pools of plasma membrane cholesterol and their relation to cholesterol homeostasis', *eLife*, 2014(3), pp. 1–16. doi: 10.7554/eLife.02882.

Decherchi, S. and Cavalli, A. (2020) 'Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation', *Chemical Reviews*, 120(23), pp. 12788–12833. doi: 10.1021/acs.chemrev.0c00534.

Denisov, I. G. *et al.* (2004) 'Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size', *Journal of the American Chemical Society*. American Chemical Society, 126(11), pp. 3477–3487. doi: 10.1021/ja0393574.

Dickens, D. *et al.* (2017) 'Modulation of LAT1 (SLC7A5) transporter activity and stability by membrane cholesterol', *Scientific Reports*. Nature Publishing Group, 7(1), pp. 1–13. doi: 10.1038/srep43580.

Dobrin, N. and Randall, N. W. (2006) *New and Emerging Proteomic Techniques*. New Jersey: Humana Press. doi: 10.1385/159745026X.

Domicevica, L., Koldsø, H. and Biggin, P. C. (2018) 'Multiscale molecular dynamics simulations of lipid interactions with P-glycoprotein in a complex membrane', *Journal of Molecular Graphics and Modelling*. Elsevier Inc., 80, pp. 147–156. doi: 10.1016/j.jmgm.2017.12.022.

Dreier, B. and Plückthun, A. (2011) 'Ribosome Display: A Technology for Selecting and Evolving Proteins from Large Libraries', in Park, D. J. (ed.) *Methods in molecular biology (Clifton, N.J.)*, pp. 283–306. doi: 10.1007/978-1-60761-944-4_21.

Drew, D. *et al.* (2008) 'GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae', *Nature Protocols*, 3(5), pp. 784–798. doi: 10.1038/nprot.2008.44.

Duan, L., Liu, X. and Zhang, J. Z. H. (2016) 'Interaction entropy: A new paradigm for highly efficient and reliable computation of protein-ligand binding free energy', *Journal of the American Chemical Society*, 138(17), pp. 5722–5728. doi: 10.1021/jacs.6b02682.

Duncan, A. L., Corey, R. A. and Sansom, M. S. P. (2020) 'Defining how multiple lipid species interact with inward rectifier potassium (Kir2) channels', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 117(14), pp. 7803–7813. doi: 10.1073/pnas.1918387117.

Edwards, N. *et al.* (2018) 'Resculpting the binding pocket of APC superfamily LeuT-fold amino acid transporters', *Cellular and Molecular Life Sciences*. Birkhauser Verlag AG, 75(5), pp. 921–938. doi: 10.1007/s00018-017-2677-8.

Elkins, M. R. *et al.* (2017) 'Cholesterol-binding site of the influenza M2 protein in lipid bilayers from solid-state NMR.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(49), pp. 12946–12951. doi: 10.1073/pnas.1715127114.

Elkins, M. R., Sergeyev, I. V. and Hong, M. (2018) 'Determining Cholesterol Binding to Membrane Proteins by Cholesterol ^{13} C Labeling in Yeast and Dynamic Nuclear Polarization NMR', *Journal of the American Chemical Society*. American Chemical Society, 140(45), pp. 15437–15449. doi: 10.1021/jacs.8b09658.

Emsley, P. *et al.* (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Enkavi, G. *et al.* (2019) 'Multiscale Simulations of Biological Membranes: The Challenge To Understand Biological Phenomena in a Living Substance', *Chemical Reviews*, 119(9), pp. 5607–5774. doi: 10.1021/acs.chemrev.8b00538.

Errasti-Murugarren, E. *et al.* (2019a) 'L amino acid transporter structure and molecular bases for the asymmetry of substrate interaction', *Nature Communications*. Nature Publishing Group, 10(1), p. 1807. doi: 10.1038/s41467-019-09837-z.

Errasti-Murugarren, E. *et al.* (2019b) 'L amino acid transporter structure and molecular bases for the asymmetry of substrate interaction', *Nature Communications*. Nature Publishing Group, 10(1), pp. 1–12. doi: 10.1038/s41467-019-09837-z.

Essmann, U. *et al.* (1995a) 'A smooth particle mesh Ewald method', *The Journal of Chemical Physics*. American Institute of Physics, 103(19), pp. 8577–8593. doi: 10.1063/1.470117.

Essmann, U. *et al.* (1995b) 'A smooth particle mesh Ewald method', *The Journal of Chemical Physics*. American Institute of Physics, 103(19), pp. 8577–8593. doi: 10.1063/1.470117.

Evans, D. J. and Holian, B. L. (1985) 'The Nose–Hoover thermostat', *The Journal of Chemical Physics*, 83(8), pp. 4069–4074. doi: 10.1063/1.449071.

Falzone, M. E. *et al.* (2022) 'TMEM16 scramblases thin the membrane to enable lipid scrambling', *Nature Communications*. Nature Publishing Group, 13(1), pp. 1–14. doi: 10.1038/s41467-022-30300-z.

Fan, S.-J. *et al.* (2016) 'PAT4 levels control amino-acid sensitivity of rapamycin-resistant mTORC1 from the Golgi and affect clinical outcome in colorectal cancer', *Oncogene*. Nature Publishing Group, 35(23), pp. 3004–3015. doi: 10.1038/onc.2015.363.

Fang, Y. *et al.* (2009) 'Structure of a prokaryotic virtual proton pump at 3.2 Å resolution', *Nature*. Nature Publishing Group, 460(7258), pp. 1040–1043. doi: 10.1038/nature08201.

Fantini, J. and Barrantes, F. J. (2013) 'How cholesterol interacts with membrane proteins: An exploration of cholesterol-binding sites including CRAC,

CARC, and tilted domains', *Frontiers in Physiology*, 4 FEB(February), pp. 1–9. doi: 10.3389/fphys.2013.00031.

Feng, L. and Frommer, W. B. (2015) 'Structure and function of SemiSWEET and SWEET sugar transporters', *Trends in Biochemical Sciences*. Elsevier Ltd, 40(8), pp. 480–486. doi: 10.1016/j.tibs.2015.05.005.

Ferraro, M. *et al.* (2016) 'Mapping Cholesterol Interaction Sites on Serotonin Transporter through Coarse-Grained Molecular Dynamics', *PLOS ONE*. Edited by G. Colombo. Public Library of Science, 11(12), p. e0166196. doi: 10.1371/journal.pone.0166196.

Fitzner, D. *et al.* (2020) 'Cell-Type- and Brain-Region-Resolved Mouse Brain Lipidome', *Cell Reports*. Cell Press, 32(11), p. 108132. doi: 10.1016/j.celrep.2020.108132.

Flegler, V. J. (2021) *Application of electron cryomicroscopy for structural and functional studies on the mechanosensitive channels of small conductance, PhD thesis*.

Flegler, V. J. *et al.* (2021) 'Mechanosensitive channel gating by delipidation', *Proceedings of the National Academy of Sciences of the United States of America*, 118(33), pp. 1–8. doi: 10.1073/pnas.2107095118.

Focht, D. *et al.* (2021) 'A non-helical region in transmembrane helix 6 of hydrophobic amino acid transporter MhsT mediates substrate recognition', *The EMBO Journal*, 40(1), pp. 1–13. doi: 10.15252/embj.2020105164.

Forrest, L. R. *et al.* (2008) 'Mechanism for alternating access in neurotransmitter transporters', *Proceedings of the National Academy of Sciences of the United States of America*, 105(30), pp. 10338–10343. doi: 10.1073/pnas.0804659105.

Forrest, L. R. and Rudnick, G. (2009) 'The rocking bundle: A mechanism for ion-coupled solute flux by symmetrical transporters', *Physiology*. Physiology (Bethesda), 24(6), pp. 377–386. doi: 10.1152/physiol.00030.2009.

Fotiadis, D., Kanai, Y. and Palacín, M. (2013) 'The SLC3 and SLC7 families of amino acid transporters', *Molecular Aspects of Medicine*, pp. 139–158. doi: 10.1016/j.mam.2012.10.007.

Frenkel, D. and Smit, B. (1996) 'Understanding molecular simulation: From algorithms to applications', *Understanding molecular simulation: From algorithms to applications*. doi: 10.1063/1.881812.

Friesner, R. A. *et al.* (2006) 'Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes', *Journal of Medicinal Chemistry*, 49, pp. 6177–6196. doi: 10.1021/jm051256o.

Fuchs, B. C. and Bode, B. P. (2005) 'Amino acid transporters ASCT2 and LAT1 in cancer: Partners in crime?', *Seminars in Cancer Biology*, 15(4), pp. 254–

266. doi: 10.1016/j.semcancer.2005.04.005.

Furuya, M. *et al.* (2012) 'Correlation of L-type amino acid transporter 1 and CD98 expression with triple negative breast cancer prognosis', *Cancer Science*, 103(2), pp. 382–389. doi: 10.1111/j.1349-7006.2011.02151.x.

Gao, X. *et al.* (2009a) 'Structure and mechanism of an amino acid antiporter', *Science*. Science, 324(5934), pp. 1565–1568. doi: 10.1126/science.1173654.

Gao, X. *et al.* (2009b) 'Structure and mechanism of an amino acid antiporter', *Science*, 324(5934), pp. 1565–1568. doi: 10.1126/science.1173654.

Gao, X. *et al.* (2010) 'Mechanism of substrate recognition and transport by an amino acid antiporter', *Nature*. Nature Publishing Group, 463(7282), pp. 828–832. doi: 10.1038/nature08741.

García-Nafría, J. *et al.* (2018) 'Cryo-EM structure of the adenosine A2A receptor coupled to an engineered heterotrimeric G protein', *eLife*, 7, pp. 1–19. doi: 10.7554/eLife.35946.001.

Gater, D. L. *et al.* (2014) 'Two classes of cholesterol binding sites for the β2AR revealed by thermostability and NMR', *Biophysical Journal*, 107(10), pp. 2305–2312. doi: 10.1016/j.bpj.2014.10.011.

Gauthier-Coles, G. *et al.* (2021) 'Quantitative modelling of amino acid transport and homeostasis in mammalian cells', *Nature Communications*. Springer US, 12(1), pp. 1–18. doi: 10.1038/s41467-021-25563-x.

Geier, E. G. *et al.* (2013) 'Structure-based ligand discovery for the Large-neutral Amino Acid Transporter 1, LAT-1', *Proceedings of the National Academy of Sciences of the United States of America*, 110(14), pp. 5480–5485. doi: 10.1073/pnas.1218165110.

Giacomini, K. M. and Huang, S. M. (2022) 'More Than Pharmacokinetics: Transporters in Clinical Pharmacology', *Clinical Pharmacology & Therapeutics*. United States, 112(3), pp. 423–426. doi: 10.1002/cpt.2710.

Glaeser, R. M., Nogales, E. and Chiu, W. (2021) *Single-particle Cryo-EM of Biological Macromolecules*. Edited by R. M. Glaeser, E. Nogales, and W. Chiu. IOP Publishing. doi: 10.1088/978-0-7503-3039-8.

Goberdhan, D. C. I. *et al.* (2005) 'PAT-related amino acid transporters regulate growth via a novel mechanism that does not require bulk transport of amino acids.', *Development (Cambridge, England)*, 132(10), pp. 2365–75. doi: 10.1242/dev.01821.

Godoy-hernandez, A. *et al.* (2022) 'Autoinsertion (LAiR): rapid functional reconstitution of integral membrane proteins into lipid bilayers', *bioRxiv*.

Gonen, T. *et al.* (2005) 'Lipid-protein interactions in double-layered two-dimensional AQP0 crystals', *Nature*. Nature, 438(7068), pp. 633–638. doi: 10.1038/nature04321.

Gowers, R. *et al.* (2016) 'MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations', in *Proceedings of the 15th Python in Science Conference*. SciPy, pp. 98–105. doi: 10.25080/majora-629e541a-00e.

Green, E. M., Coleman, J. A. and Gouaux, E. (2015) 'Thermostabilization of the Human Serotonin Transporter in an Antidepressant-Bound Conformation', *PLOS ONE*. Edited by D. Fotiadis. Public Library of Science, 10(12), p. e0145688. doi: 10.1371/journal.pone.0145688.

Greene, D. *et al.* (2019) 'Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies', *Journal of Chemical Information and Modeling*, 59(6), pp. 3041–3056. doi: 10.1021/acs.jcim.9b00363.

Greenfield, N. J. (2007) 'Using circular dichroism spectra to estimate protein secondary structure', *Nature Protocols*. Nature Publishing Group, 1(6), pp. 2876–2890. doi: 10.1038/nprot.2006.202.

Grote, F. and Lyubartsev, A. P. (2020) 'Optimization of Slipids Force Field Parameters Describing Headgroups of Phospholipids', *Journal of Physical Chemistry B*, 124(40), pp. 8784–8793. doi: 10.1021/acs.jpcb.0c06386.

Grouleff, J. *et al.* (2015) 'Monoamine transporters: Insights from molecular dynamics simulations', *Frontiers in Pharmacology*, 6(OCT), pp. 1–15. doi: 10.3389/fphar.2015.00235.

Guigas, G. and Weiss, M. (2016) 'Effects of protein crowding on membrane systems', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(10), pp. 2441–2450. doi: 10.1016/j.bbamem.2015.12.021.

Guo, J. *et al.* (2016) 'Structure of the voltage-gated two-pore channel TPC1 from Arabidopsis thaliana', *Nature*. Nature Publishing Group, 531(7593), pp. 196–201. doi: 10.1038/nature16446.

Gupta, K. *et al.* (2017) 'The role of interfacial lipids in stabilizing membrane protein oligomers', *Nature*, 541(7637), pp. 421–424. doi: 10.1038/nature20820.

Harry C. Jubb *et al.* (2018) 'COSMIC-3D provides structural perspectives on cancer genetics for drug discovery', *Nature Genetics*, 50(September).

Heckman, K. L. and Pease, L. R. (2007) 'Gene splicing and mutagenesis by PCR-driven overlap extension', *Nature Protocols*, 2(4), pp. 924–932. doi: 10.1038/nprot.2007.132.

Hedger, G. *et al.* (2019) 'Cholesterol Interaction Sites on the Transmembrane Domain of the Hedgehog Signal Transducer and Class F G Protein-Coupled Receptor Smoothened', *Structure*. Elsevier Ltd., 27(3), pp. 549-559.e2. doi: 10.1016/j.str.2018.11.003.

Hediger, M. A. *et al.* (2004) 'The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins', *Pflügers Archiv European Journal of Physiology*. Springer-Verlag,

447(5), pp. 465–468. doi: 10.1007/s00424-003-1192-y.

Hediger, M. A. *et al.* (2013) 'The ABCs of membrane transporters in health and disease (SLC series): Introduction', *Molecular Aspects of Medicine*. Pergamon, pp. 95–107. doi: 10.1016/j.mam.2012.12.009.

Hernández, H. and Robinson, C. V. (2007) 'Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry', *Nature Protocols*, 2(3), pp. 715–726. doi: 10.1038/nprot.2007.73.

Herzik, M. A., Wu, M. and Lander, G. C. (2019) 'High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM', *Nature Communications*. Springer US, 10(1), pp. 1–9. doi: 10.1038/s41467-019-08991-8.

Hess, B. *et al.* (1997) 'LINCS: A linear constraint solver for molecular simulations', *Journal of Computational Chemistry*. John Wiley & Sons, Ltd, 18(12), pp. 1463–1472. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

Hess, B. *et al.* (2008) 'GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation', *Journal of Chemical Theory and Computation*, 4(3), pp. 435–447. doi: 10.1021/ct700301q.

Heublein, S. *et al.* (2010) 'Proton-assisted amino-acid transporters are conserved regulators of proliferation and amino-acid-dependent mTORC1 activation.', *Oncogene*, 29(28), pp. 4068–79. doi: 10.1038/onc.2010.177.

Hinz, K. M. *et al.* (2015) 'Structural Insights Into Thyroid Hormone Transport Mechanisms of the L-Type Amino Acid Transporter 2', *Molecular Endocrinology*. Endocrine Society, 29(6), pp. 933–942. doi: 10.1210/me.2015-1044.

Hofmann, S. *et al.* (2019) 'Conformation space of a heterodimeric ABC exporter under turnover conditions', *Nature*. Springer US, 571(7766), pp. 580–583. doi: 10.1038/s41586-019-1391-0.

Höglund, P. J. *et al.* (2011) 'The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species.', *Molecular biology and evolution*. Oxford University Press, 28(4), pp. 1531–41. doi: 10.1093/molbev/msq350.

Holsbeeks, I. *et al.* (2004) 'The eukaryotic plasma membrane as a nutrient-sensing device', *Trends in Biochemical Sciences*, 29(10), pp. 556–564. doi: https://doi.org/10.1016/j.tibs.2004.08.010.

Hu, B. *et al.* (2020) 'Photoactivable Cholesterol as a Tool to Study Interaction of Influenza Virus Hemagglutinin with Cholesterol.', *Bio-protocol*. United States, 10(4), p. e3523. doi: 10.21769/BioProtoc.3523.

Huang, S. K. *et al.* (2022) 'Allosteric modulation of the adenosine A2A receptor by cholesterol', *eLife*. eLife Sciences Publications Ltd, 11. doi: 10.7554/ELIFE.73901.

Huang, Y. *et al.* (2003) 'Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli.', *Science (New York, N.Y.)*, 301(5633), pp. 616–20. doi: 10.1126/science.1087619.

Hub, J. S., de Groot, B. L. and van der Spoel, D. (2010) 'g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates', *Journal of Chemical Theory and Computation*, 6(12), pp. 3713–3720. doi: 10.1021/ct100494z.

Hulce, J. J. *et al.* (2013) 'Proteome-wide mapping of cholesterol-interacting proteins in mammalian cells', *Nature Methods*, 10(3), pp. 259–264. doi: 10.1038/nmeth.2368.

Humphrey, W., Dalke, A. and Schulten, K. (1996) 'VMD: Visual Molecular Dynamics', *Journal of Molecular Graphics*. Elsevier, 14(1), pp. 33–38. doi: 10.1016/0263-7855(96)00018-5.

Hundal, H. S. and Taylor, P. M. (2009) 'Amino acid transceptors: gate keepers of nutrient exchange and regulators of nutrient signaling', *American Journal of Physiology-Endocrinology and Metabolism*. United States, 296(4), pp. E603–E613. doi: 10.1152/ajpendo.91002.2008.

Ilgü, H. *et al.* (2016) 'Insights into the molecular basis for substrate binding and specificity of the wild-type L-arginine/agmatine antiporter AdiC', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(37), pp. 10358–10363. doi: 10.1073/pnas.1605442113.

Ilgü, H. *et al.* (2021) 'High-resolution structure of the amino acid transporter AdiC reveals insights into the role of water molecules and networks in oligomerization and substrate binding', *BMC Biology*. BMC Biology, 19(1), pp. 1–13. doi: 10.1186/s12915-021-01102-4.

Ingólfsson, H. I. *et al.* (2014) 'The power of coarse graining in biomolecular simulations', *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(3), pp. 225–248. doi: 10.1002/wcms.1169.

Ishida, H. *et al.* (2021) 'Cryo-EM structures of Toll-like receptors in complex with UNC93B1', *Nature Structural and Molecular Biology*. Springer US, 28(2), pp. 173–180. doi: 10.1038/s41594-020-00542-w.

Ito, K. and Groudine, M. (1997) 'A new member of the cationic amino acid transporter family is preferentially expressed in adult mouse brain', *Journal of Biological Chemistry*. Â© 1997 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 272(42), pp. 26780–26786. doi: 10.1074/jbc.272.42.26780.

Jacobus, A. P. and Gross, J. (2015) 'Optimal Cloning of PCR Fragments by Homologous Recombination in Escherichia coli', *PLOS ONE*. Edited by S.

Korolev, 10(3), p. e0119221. doi: 10.1371/journal.pone.0119221.

Jardetzky, O. (1966) 'Simple allosteric model for membrane pumps [27]', *Nature*. Nature Publishing Group, pp. 969–970. doi: 10.1038/211969a0.

Jarin, Z., Newhouse, J. and Voth, G. A. (2021) 'Coarse-Grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding of the Origins of a MARTINI Hangover', *Journal of Chemical Theory and Computation*, 17(2), pp. 1170–1180. doi: 10.1021/acs.jctc.0c00638.

Javanainen, M., Martinez-Seara, H. and Vattulainen, I. (2017) 'Excessive aggregation of membrane proteins in the Martini model', *PLOS ONE*. Public Library of Science, 12(11), p. e0187936. doi: 10.1371/JOURNAL.PONE.0187936.

Jo, S. *et al.* (2008) 'CHARMM-GUI: A web-based graphical user interface for CHARMM', *Journal of Computational Chemistry*, 29(11), pp. 1859–1865. doi: 10.1002/jcc.20945.

Jones, K. T., Zhen, J. and Reith, M. E. A. (2012) 'Importance of cholesterol in dopamine transporter function', *Journal of Neurochemistry*, 123(5), pp. 700–715. doi: 10.1111/jnc.12007.

de Jong, D. H. *et al.* (2013) 'Improved Parameters for the Martini Coarse-Grained Protein Force Field', *Journal of Chemical Theory and Computation*. American Chemical Society, 9(1), pp. 687–697. doi: 10.1021/ct300646g.

De Jong, D. H. *et al.* (2013) 'Improved parameters for the martini coarse-grained protein force field', *Journal of Chemical Theory and Computation*. American Chemical Society, 9(1), pp. 687–697. doi: 10.1021/ct300646g.

De Jong, D. H. *et al.* (2016) 'Martini straight: Boosting performance using a shorter cutoff and GPUs', *Computer Physics Communications*. Elsevier B.V., 199, pp. 1–7. doi: 10.1016/j.cpc.2015.09.014.

Jørgensen, M. E. *et al.* (2017) 'Origin and evolution of transporter substrate specificity within the NPF family', *eLife*. eLife Sciences Publications, Ltd, 6. doi: 10.7554/elife.19466.

Jorgensen, W. L. *et al.* (1983) 'Comparison of simple potential functions for simulating liquid water', *The Journal of Chemical Physics*, 79(2), pp. 926–935. doi: 10.1063/1.445869.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*. Springer US, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.

Jungnickel, K. E. J., Parker, J. L. and Newstead, S. (2018) 'Structural basis for amino acid transport by the CAT family of SLC7 transporters', *Nature Communications*. Nature Publishing Group, 9(1), p. 550. doi: 10.1038/s41467-018-03066-6.

Jungnickel, K., Parker, J. and Newstead, S. (2018) *Structural and functional*

studies on a prokaryotic homologue of the mammalian SLC7 cationic amino acid transporters. doi: 10.1107/s2053273318094706.

Kabsch, W. and Sander, C. (1983) 'Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features', *Biopolymers*, 22(12), pp. 2577–2637. doi: 10.1002/bip.360221211.

Kaira, K. *et al.* (2008) 'L-type amino acid transporter 1 and CD98 expression in primary and metastatic sites of human neoplasms', *Cancer Science*, 99(12), pp. 2380–2386. doi: 10.1111/j.1349-7006.2008.00969.x.

Kanai, Y. *et al.* (1998) 'Expression cloning and characterization of a transporter for large neutral amino acids activated by the heavy chain of 4F2 antigen (CD98)', *Journal of Biological Chemistry*, 273(37), pp. 23629–23632. doi: 10.1074/jbc.273.37.23629.

Kandasamy, P. *et al.* (2018a) 'Amino acid transporters revisited: New views in health and disease', *Trends in Biochemical Sciences*. Elsevier Ltd, pp. 752–789. doi: 10.1016/j.tibs.2018.05.003.

Kandasamy, P. *et al.* (2018b) 'Amino acid transporters revisited: New views in health and disease', *Trends in Biochemical Sciences*. Elsevier Ltd, 43(10), pp. 752–789. doi: 10.1016/j.tibs.2018.05.003.

Kang, Y. *et al.* (2015) 'Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser', *Nature*, 523(7562), pp. 561–567. doi: 10.1038/nature14656.

Kantipudi, S. and Fotiadis, D. (2021) 'Yeast Cell-Based Transport Assay for the Functional Characterization of Human 4F2hc-LAT1 and -LAT2, and LAT1 and LAT2 Substrates and Inhibitors', *Frontiers in Molecular Biosciences*, 8(May), pp. 1–11. doi: 10.3389/fmolb.2021.676854.

Kariuki, C. K. and Magez, S. (2021) 'Improving the yield of recalcitrant Nanobodies® by simple modifications to the standard protocol.', *Protein expression and purification*. United States, 185, p. 105906. doi: 10.1016/j.pep.2021.105906.

Kato, T. *et al.* (2019) 'Crystal structure of plant vacuolar iron transporter VIT1', *Nature Plants*. Springer US, 5(3), pp. 308–315. doi: 10.1038/s41477-019-0367-2.

Kavanaugh, M. P. (1993) 'Voltage Dependence of Facilitated Arginine Flux Mediated by the System y+ Basic Amino Acid Transporter', *Biochemistry*, 32(22), pp. 5781–5785. doi: 10.1021/bi00073a009.

Kawate, T. and Gouaux, E. (2006a) 'Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins.', *Structure (London, England : 1993)*, 14(4), pp. 673–81. doi: 10.1016/j.str.2006.01.013.

Kawate, T. and Gouaux, E. (2006b) 'Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins', *Structure*, 14(4), pp. 673–681. doi: 10.1016/j.str.2006.01.013.

Kazmier, K. *et al.* (2014) 'Conformational dynamics of ligand-dependent alternating access in LeuT', *Nature Structural and Molecular Biology*. Nat Struct Mol Biol, 21(5), pp. 472–479. doi: 10.1038/nsmb.2816.

Keener, J. E., Zhang, G. and Marty, M. T. (2021) 'Native Mass Spectrometry of Membrane Proteins', *Analytical Chemistry*, 93(1), pp. 583–597. doi: 10.1021/acs.analchem.0c04342.

Keller, R., Ziegler, C. and Schneider, D. (2014) 'When two turn into one: Evolution of membrane transporters from half modules', *Biological Chemistry*, 395(12), pp. 1379–1388. doi: 10.1515/hsz-2014-0224.

Khunweeraphong, N. *et al.* (2012) 'Establishment of stable cell lines with high expression of heterodimers of human 4F2hc and human amino acid transporter LAT1 or LAT2 and delineation of their differential interaction with α-alkyl moieties', *Journal of Pharmacological Sciences*, 119(4), pp. 368–380. doi: 10.1254/jphs.12124FP.

Kim, C. H. *et al.* (2006) 'The RNA interference of amino acid transporter LAT1 inhibits the growth of KB human oral cancer cells', *Anticancer Research*, 26(4 B), pp. 2943–2948. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16886618 (Accessed: 23 April 2020).

Kim, J. W. *et al.* (1991) 'Transport of cationic amino acids by the mouse ecotropic retrovirus receptor', *Nature*. Nature, 352(6337), pp. 725–728. doi: 10.1038/352725a0.

Ko, D. C. *et al.* (2003) 'The integrity of a cholesterol-binding pocket in Niemann-Pick C2 protein is necessary to control lysosome cholesterol levels', *Proceedings of the National Academy of Sciences of the United States of America*, 100(5), pp. 2518–2525. doi: 10.1073/pnas.0530027100.

Kobayashi, K. *et al.* (2008) 'Enhanced tumor growth elicited by L-type amino acid transporter 1 in human malignant glioma cells', *Neurosurgery*, 62(2), pp. 493–503. doi: 10.1227/01.neu.0000316018.51292.19.

Kotov, V. *et al.* (2019) 'High-throughput stability screening for detergent-solubilized membrane proteins', *Scientific Reports*, 9(1), p. 10379. doi: 10.1038/s41598-019-46686-8.

Kotyk, A. (1997) 'How to name various transport mechanisms and their vehicles', *Biochemical Education*. John Wiley & Sons, Ltd, pp. 86–87. doi: 10.1016/S0307-4412(96)00177-X.

Kowalczyk, L. *et al.* (2011a) 'Molecular basis of substrate-induced permeation by an amino acid antiporter', *Proceedings of the National Academy of Sciences*

of the United States of America. Proc Natl Acad Sci U S A, 108(10), pp. 3935–3940. doi: 10.1073/pnas.1018081108.

Kowalczyk, L. *et al.* (2011b) 'Molecular basis of substrate-induced permeation by an amino acid antiporter', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(10), pp. 3935–3940. doi: 10.1073/pnas.1018081108.

Krishnamurthy, H. and Gouaux, E. (2012) 'X-ray structures of LeuT in substrate-free outward-open and apo inward-open states', *Nature*. Nature, 481(7382), pp. 469–474. doi: 10.1038/nature10737.

Krishnamurthy, H., Piscitelli, C. L. and Gouaux, E. (2009) 'Unlocking the molecular secrets of sodium-coupled transporters.', *Nature*, 459(7245), pp. 347–55. doi: 10.1038/nature08143.

Krishnan, K. *et al.* (2021) 'Validation of Trifluoromethylphenyl Diazirine Cholesterol Analogues As Cholesterol Mimetics and Photolabeling Reagents', *ACS Chemical Biology*, 16(8), pp. 1493–1507. doi: 10.1021/acschembio.1c00364.

Kristensen, A. S. *et al.* (2011) 'SLC6 neurotransmitter transporters: Structure, function, and regulation', *Pharmacological Reviews*. American Society for Pharmacology and Experimental Therapeutics, 63(3), pp. 585–640. doi: 10.1124/pr.108.000869.

Kühlbrandt, W. (2022) 'Forty years in cryoEM of membrane proteins', *Microscopy*, 71(October 2021), pp. I30–I50. doi: 10.1093/jmicro/dfab041.

Kulkarni, C. V *et al.* (2011) 'Monoolein: a magic lipid?w', *Phys. Chem. Chem. Phys*, 13, pp. 3004–3021. doi: 10.1039/c0cp01539c.

Kumar, S. *et al.* (1992) 'THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method', *Journal of Computational Chemistry*, 13(8), pp. 1011–1021. doi: 10.1002/jcc.540130812.

Kwik, J. *et al.* (2003) 'Membrane cholesterol, lateral mobility, and the phosphatidylinositol 4,5-bisphosphate-dependent organization of cell actin', *Proceedings of the National Academy of Sciences*, 100(24), pp. 13964–13969. doi: 10.1073/pnas.2336102100.

Lagarias, P. *et al.* (2019) 'Insights to the Binding of a Selective Adenosine A3 Receptor Antagonist Using Molecular Dynamic Simulations, MM-PBSA and MM-GBSA Free Energy Calculations, and Mutagenesis', *Journal of Chemical Information and Modeling*, 59(12), pp. 5183–5197. doi: 10.1021/acs.jcim.9b00751.

Landau, E. M. and Rosenbusch, J. P. (1996) 'Lipidic cubic phases: A novel concept for the crystallization of membrane proteins', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 93(25), pp. 14532–14535. doi: 10.1073/pnas.93.25.14532.

Laursen, L. *et al.* (2018) 'Cholesterol binding to a conserved site modulates

the conformation, pharmacology, and transport kinetics of the human serotonin transporter', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 293(10), pp. 3510–3523. doi: 10.1074/jbc.M117.809046.

Lawrence, R. E. and Zoncu, R. (2019) 'The lysosome as a cellular centre for signalling, metabolism and quality control', *Nature Cell Biology*, 21, pp. 133–142. doi: 10.1038/s41556-018-0244-7.

Leach, A. R. (2001) *Molecular Modelling: Principles and Applications*. 2nd edn. Prentice Hall. Available at: http://www.amazon.com/Molecular-Modelling-Principles-Applications-2nd/dp/0582382106.

Lee, A. G. (2005) 'How lipids and proteins interact in a membrane: a molecular approach', *Molecular BioSystems*, 1(3), p. 203. doi: 10.1039/b504527d.

Lee, H. J. *et al.* (2022) 'Impact of novel detergents on membrane protein studies', *Chem*. Elsevier Inc., 8(4), pp. 980–1013. doi: 10.1016/j.chempr.2022.02.007.

Van Der Lee, R. *et al.* (2014) 'Classification of intrinsically disordered regions and proteins', *Chemical Reviews*, 114(13), pp. 6589–6631. doi: 10.1021/cr400525m.

Lee, Y. *et al.* (2019a) 'Cryo-EM structure of the human L-type amino acid transporter 1 in complex with glycoprotein CD98hc', *Nature Structural & Molecular Biology*. Nat Struct Mol Biol, 26(6), pp. 510–517. doi: 10.1038/s41594-019-0237-7.

Lee, Y. *et al.* (2019b) 'Cryo-EM structure of the human L-type amino acid transporter 1 in complex with glycoprotein CD98hc', *Nature Structural and Molecular Biology*. Nature Publishing Group, 26(6), pp. 510–517. doi: 10.1038/s41594-019-0237-7.

Lee, Y. *et al.* (2022) 'Ca2+-mediated higher-order assembly of heterodimers in amino acid transport system b0,+ biogenesis and cystinuria', *Nature Communications*. Springer US, 13(1). doi: 10.1038/s41467-022-30293-9.

Lei, H.-T. *et al.* (2018) 'Crystal structure of arginine-bound lysosomal transporter SLC38A9 in the cytosol-open state', *Nature Structural & Molecular Biology*. Nature Publishing Group, 25(6), pp. 522–527. doi: 10.1038/s41594-018-0072-2.

Levental, I., Levental, K. R. and Heberle, F. A. (2020) 'Lipid Rafts: Controversies Resolved, Mysteries Remain.', *Trends in cell biology*, 30(5), pp. 341–353. doi: 10.1016/j.tcb.2020.01.009.

Levental, I. and Lyman, E. (2022) 'Regulation of membrane protein structure and function by their lipid nano-environment', *Nature Reviews Molecular Cell Biology*. Springer US, 0123456789. doi: 10.1038/s41580-022-00524-4.

Levitt, M. (2001) 'The birth of computational structural biology', *Nature Structural Biology*, 8(5), pp. 392–393. doi: 10.1038/87545.

Liang, Z. *et al.* (2011) 'Potential Biomarker of L-type Amino Acid Transporter 1 in Breast Cancer Progression', *Nuclear Medicine and Molecular Imaging*, 45(2), pp. 93–102. doi: 10.1007/s13139-010-0068-2.

Liebschner, D. *et al.* (2019) 'Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix', *Acta Crystallographica Section D: Structural Biology*, 75, pp. 861–877. doi: 10.1107/S2059798319011471.

Lin, L. *et al.* (2015) 'SLC transporters as therapeutic targets: emerging opportunities', *Nature Reviews Drug Discovery*. Nature Publishing Group, 14(8), pp. 543–560. doi: 10.1038/nrd4626.

Lin, W. Y. *et al.* (2015) 'The SLC36 transporter pathetic is required for extreme dendrite growth in Drosophila sensory neurons', *Genes and Development*. Genes Dev, 29(11), pp. 1120–1135. doi: 10.1101/gad.259119.115.

Linding, R. *et al.* (2003) 'Protein disorder prediction: Implications for structural proteomics', *Structure*, 11(11), pp. 1453–1459. doi: 10.1016/j.str.2003.10.002.

Lindorff-Larsen, K. *et al.* (2010) 'Improved side-chain torsion potentials for the Amber ff99SB protein force field', *Proteins: Structure, Function and Bioinformatics*, 78(8), pp. 1950–1958. doi: 10.1002/prot.22711.

Liu, X. *et al.* (2018) 'Computational Alanine Scanning with Interaction Entropy for Protein–Ligand Binding Free Energies', *Journal of Chemical Theory and Computation*, 14(3), pp. 1772–1780. doi: 10.1021/acs.jctc.7b01295.

Livesay, D. R. *et al.* (2007) 'Assessing the ability of sequence-based methods to provide functional insight within membrane integral proteins: A case study analyzing the neurotransmitter/Na+ symporter family', *BMC Bioinformatics*, 8. doi: 10.1186/1471-2105-8-397.

Löbel, M. *et al.* (2022) 'Structural basis for proton coupled cystine transport by cystinosin', *Nature Communications*. Springer US, 13(1), pp. 1–12. doi: 10.1038/s41467-022-32589-2.

Locascio, A. *et al.* (2019) 'Saccharomyces cerevisiae as a tool to investigate plant potassium and sodium transporters', *International Journal of Molecular Sciences*, 20(9). doi: 10.3390/ijms20092133.

Lomize, A. L., Todd, S. C. and Pogozheva, I. D. (2022) 'Spatial arrangement of proteins in planar and curved membranes by PPM 3.0', *Protein Science*, 31(1), pp. 209–220. doi: 10.1002/pro.4219.

Lomize, M. A. *et al.* (2012) 'OPM database and PPM web server: Resources for positioning of proteins in membranes', *Nucleic Acids Research*, 40(D1), pp.

370–376. doi: 10.1093/nar/gkr703.

Long, F. *et al.* (2017) 'AceDRG: A stereochemical description generator for ligands', *Acta Crystallographica Section D: Structural Biology*, 73(2), pp. 112–122. doi: 10.1107/S2059798317000067.

Long, S. B. *et al.* (2007) 'Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment', *Nature*. Nature, 450(7168), pp. 376–382. doi: 10.1038/nature06265.

Loo, D. D. F. *et al.* (1999) 'Passive water and ion transport by cotransporters', *Journal of Physiology*. John Wiley & Sons, Ltd, 518(1), pp. 195–202. doi: 10.1111/j.1469-7793.1999.0195r.x.

Loura, L. M. S., Prieto, M. and Fernandes, F. (2010) 'Quantification of protein-lipid selectivity using FRET', *European Biophysics Journal*. Springer, pp. 565–578. doi: 10.1007/s00249-009-0532-z.

Luckey, M. (2014) *Membrane Structural Biology: With Biochemical and Biophysical Foundations*. Second. Cambridge University Press.

Maier, J. A. *et al.* (2015) 'ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB', *Journal of Chemical Theory and Computation*, 11(8), pp. 3696–3713. doi: 10.1021/acs.jctc.5b00255.

Majd, H. *et al.* (2018) 'Screening of candidate substrates and coupling ions of transporters by thermostability shift assays', *eLife*, 7, pp. 1–17. doi: 10.7554/eLife.38821.

Manea, M. *et al.* (2007) 'Mass spectrometric identification of the trypsin cleavage pathway in lysyl-proline containing oligotuftsin peptides', *Journal of Peptide Science*, 13(4), pp. 227–236. doi: 10.1002/psc.836.

Mann, P. (2018) 'Lagrangian and Hamiltonian dynamics', *Lagrangian and Hamiltonian Dynamics*, pp. 1–538. doi: 10.1093/oso/9780198822370.001.0001.

Marrink, S. J. *et al.* (2019) 'Computational Modeling of Realistic Cell Membranes', *Chemical Reviews*, 119(9), pp. 6184–6226. doi: 10.1021/acs.chemrev.8b00460.

Marrink, S. J. and Tieleman, D. P. (2013) 'Perspective on the martini model', *Chemical Society Reviews*, 42(16), pp. 6801–6822. doi: 10.1039/c3cs60093a.

Marsh, D. (2008) 'Protein modulation of lipids, and vice-versa, in membranes', *Biochimica et Biophysica Acta - Biomembranes*. Biochim Biophys Acta, pp. 1545–1575. doi: 10.1016/j.bbamem.2008.01.015.

Marshall, A. D. *et al.* (2016) 'LAT1 is a putative therapeutic target in endometrioid endometrial carcinoma', *International Journal of Cancer*. Wiley-Liss Inc., 139(11), pp. 2529–2539. doi: 10.1002/ijc.30371.

Mavridou, V. *et al.* (2022) 'Substrate binding in the mitochondrial ADP/ATP carrier is a step-wise process guiding the structural changes in the transport

cycle', *Nature Communications*. Springer US, 13(1), pp. 1–12. doi: 10.1038/s41467-022-31366-5.

McMahon, C. *et al.* (2018) 'Yeast surface display platform for rapid discovery of conformationally selective nanobodies', *Nature Structural & Molecular Biology*, 25(3), pp. 289–296. doi: 10.1038/s41594-018-0028-6.

Melo, M. N., Ingólfsson, H. I. and Marrink, S. J. (2015) 'Parameters for Martini sterols and hopanoids based on a virtual-site description', *Journal of Chemical Physics*. American Institute of Physics Inc., 143(24), p. 243152. doi: 10.1063/1.4937783.

Meury, M. *et al.* (2014) 'Detergent-Induced Stabilization and Improved 3D Map of the Human Heteromeric Amino Acid Transporter 4F2hc-LAT2', *PLoS ONE*. Edited by D. Boudko. Public Library of Science, 9(10), p. e109882. doi: 10.1371/journal.pone.0109882.

Michael Green, N. (1990) '[5] Avidin and streptavidin', in Wilchek, M. and Bayer, E. A. B. T.-M. in E. (eds) *Avidin-Biotin Technology*. Academic Press, pp. 51–67. doi: 10.1016/0076-6879(90)84259-J.

Michaud-Agrawal, N. *et al.* (2011) 'MDAnalysis: A toolkit for the analysis of molecular dynamics simulations', *Journal of Computational Chemistry*. John Wiley & Sons, Ltd, 32(10), pp. 2319–2327. doi: 10.1002/jcc.21787.

Miller, B. R. *et al.* (2012) 'MMPBSA.py : An Efficient Program for End-State Free Energy Calculations', *Journal of Chemical Theory and Computation*, 8(9), pp. 3314–3321. doi: 10.1021/ct300418h.

Minhas, G. S. and Newstead, S. (2019) 'Structural basis for prodrug recognition by the SLC15 family of proton-coupled peptide transporters', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(3), pp. 804–809. doi: 10.1073/pnas.1813715116.

Mirdita, M. *et al.* (2022) 'ColabFold: making protein folding accessible to all', *Nature Methods*. Springer US, 19(6), pp. 679–682. doi: 10.1038/s41592-022-01488-1.

Mitchell, L. S. and Colwell, L. J. (2018) 'Comparative analysis of nanobody sequence and structure data', *Proteins: Structure, Function and Bioinformatics*, 86(7), pp. 697–706. doi: 10.1002/prot.25497.

Mitra, K. *et al.* (2004) 'Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 101(12), pp. 4083–4088. doi: 10.1073/pnas.0307332101.

Monticelli, L. *et al.* (2008) 'The MARTINI Coarse-Grained Force Field: Extension to Proteins', *Journal of Chemical Theory and Computation*, 4(5), pp.

819–834. doi: 10.1021/ct700324x.

Moraes, I. *et al.* (2014) 'Membrane protein structure determination — The next generation', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838(1), pp. 78–87. doi: 10.1016/j.bbamem.2013.07.010.

Morris, S. M. (2016) 'Arginine Metabolism Revisited', *The Journal of Nutrition*, 146(12), pp. 2579S-2586S. doi: 10.3945/jn.115.226621.

Mouritsen, O. G. and Zuckermann, M. J. (2004) 'What's so special about cholesterol?', *Lipids*, 39(11), pp. 1101–1113. doi: 10.1007/s11745-004-1336-x.

Moutel, S. *et al.* (2016) 'NaLi-H1: A universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies', *eLife*, 5. doi: 10.7554/eLife.16228.

Munder, M. (2009) 'Arginase: an emerging key player in the mammalian immune system', *British Journal of Pharmacology*, 158(3), pp. 638–651. doi: 10.1111/j.1476-5381.2009.00291.x.

Nakamura, E. *et al.* (1999) '4F2 (CD98) Heavy Chain Is Associated Covalently with an Amino Acid Transporter and Controls Intracellular Trafficking and Membrane Topology of 4F2 Heterodimer', *Journal of Biological Chemistry*. Â© 1999 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 274(5), pp. 3009–3016. doi: 10.1074/jbc.274.5.3009.

Nakanishi, T. and Tamai, I. (2011) 'Solute carrier transporters as targets for drug delivery and pharmacological intervention for chemotherapy', *Journal of Pharmaceutical Sciences*. John Wiley and Sons Inc., pp. 3731–3750. doi: 10.1002/jps.22576.

Nan, J. *et al.* (2022) 'Cryo-EM structure of the human sodium-chloride cotransporter NCC', *Science Advances*, 8(45), pp. 1–11. doi: 10.1126/sciadv.add7176.

Napolitano, L. *et al.* (2015) 'LAT1 is the transport competent unit of the LAT1/CD98 heterodimeric amino acid transporter', *International Journal of Biochemistry and Cell Biology*. Elsevier Ltd, 67, pp. 25–33. doi: 10.1016/j.biocel.2015.08.004.

Nawashiro, H. *et al.* (2006) 'L-type amino acid transporter 1 as a potential molecular target in human astrocytic tumors', *International Journal of Cancer*, 119(3), pp. 484–492. doi: 10.1002/ijc.21866.

Nawrath, H. *et al.* (2000) 'Voltage dependence of L-arginine transport by hCAT-2A and hCAT-2B expressed in oocytes from Xenopus laevis', *American Journal of Physiology - Cell Physiology*, 279(5 48-5), pp. 1336–1344. doi: 10.1152/ajpcell.2000.279.5.c1336.

Neumann, C. *et al.* (2022) 'Cryo-EM structure of the human NKCC1

transporter reveals mechanisms of ion coupling and specificity', *The EMBO Journal*, pp. 1–15. doi: 10.15252/embj.2021110169.

Newby, Z. E. R. *et al.* (2009) 'A general protocol for the crystallization of membrane proteins for X-ray structural investigation', *Nature Protocols*, 4(5), pp. 619–637. doi: 10.1038/nprot.2009.27.

Newstead, S., Kim, H., von Heijne, G., *et al.* (2007) 'High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(35), pp. 13936–41. doi: 10.1073/pnas.0704546104.

Newstead, S., Kim, H., Von Heijne, G., *et al.* (2007) 'High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(35), pp. 13936–13941. doi: 10.1073/pnas.0704546104.

Nicolàs-Aragó, A. *et al.* (2021) 'Rush hour of LATs towards their transport cycle', *Membranes*, pp. 1–13. doi: 10.3390/membranes11080602.

Niu, L., Kim, J. M. and Khorana, H. G. (2002) 'Structure and function in rhodopsin: Asymmetric reconstitution of rhodopsin in liposomes', *Proceedings of the National Academy of Sciences of the United States of America*, 99(21), pp. 13409–13412. doi: 10.1073/pnas.212518899.

Niu, Y. *et al.* (2021) 'Structural basis of inhibition of the human SGLT2–MAP17 glucose transporter', *Nature*. Springer US, 601(September 2021). doi: 10.1038/s41586-021-04212-9.

Nji, E. *et al.* (2018) 'An engineered thermal-shift screen reveals specific lipid preferences of eukaryotic and prokaryotic membrane proteins', *Nature Communications*. Nature Publishing Group, 9(1), pp. 1–12. doi: 10.1038/s41467-018-06702-3.

Noack, L. C. and Jaillais, Y. (2020) 'Functions of Anionic Lipids in Plants', *Annual Review of Plant Biology*, 71, pp. 71–102. doi: 10.1146/annurev-arplant-081519-035910.

Noid, W. *et al.* (2008) 'The Multiscale Coarse- Graining Method', *Coarse-Graining of Condensed Phase and Biomolecular Systems*, 149, pp. 21–39. doi: 10.1201/9781420059564.ch3.

Norel, R. *et al.* (1994) 'Shape complementarity at protein–protein interfaces', *Biopolymers*, pp. 933–940. doi: 10.1002/bip.360340711.

Noskov, S. Y. and Roux, B. (2008) 'Control of Ion Selectivity in LeuT: Two Na+ Binding Sites with Two Different Mechanisms', *Journal of Molecular Biology*. J Mol Biol, 377(3), pp. 804–818. doi: 10.1016/j.jmb.2008.01.015.

O'Brien, D. F., Costa, L. F. and Ott, R. A. (1977) 'Photochemical Functionality of Rhodopsin-Phospholipid Recombinant Membranes', *Biochemistry*, 16(7), pp. 1295–1303. doi: 10.1021/bi00626a009.

Pardon, E. *et al.* (2014) 'A general protocol for the generation of Nanobodies for structural biology', *Nature Protocols*, 9(3), pp. 674–693. doi: 10.1038/nprot.2014.039.

Parker, J. L., Deme, J. C., Kolokouris, D., *et al.* (2021) 'Molecular basis for redox control by the human cystine/glutamate antiporter system xc−', *Nature Communications*. Springer US, 12(1), pp. 1–11. doi: 10.1038/s41467-021-27414-1.

Parker, J. L., Deme, J. C., Kuteyi, G., *et al.* (2021) 'Structural basis of antifolate recognition and transport by PCFT', *Nature*. Springer US, (December 2020). doi: 10.1038/s41586-021-03579-z.

Parker, J. L. and Newstead, S. (2012) 'Current trends in a-helical membrane protein crystallization: An update'. doi: 10.1002/pro.2122.

Parker, Joanne L and Newstead, S. (2014) 'Method to increase the yield of eukaryotic membrane protein expression in saccharomyces cerevisiae for structural and functional studies', *Protein Science*. Wiley-Blackwell, 23(9), pp. 1309–1314. doi: 10.1002/pro.2507.

Parker, Joanne L. and Newstead, S. (2014) 'Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1', *Nature*. Nature Publishing Group, 507(7490), pp. 68–72. doi: 10.1038/nature13116.

Parker, J. L. and Newstead, S. (2016) 'Membrane Protein Crystallisation: Current Trends and Future Perspectives', in *Advances in experimental medicine and biology*, pp. 61–72. doi: 10.1007/978-3-319-35072-1_5.

Parker, J. L. and Newstead, S. (2017) 'Structural basis of nucleotide sugar transport across the Golgi membrane', *Nature*. Nature, 551(7681), pp. 521–524. doi: 10.1038/nature24464.

Parrinello, M. and Rahman, A. (1981) 'Polymorphic transitions in single crystals: A new molecular dynamics method', *Journal of Applied Physics*, 52(12), pp. 7182–7190. doi: 10.1063/1.328693.

Payandeh, J. and Volgraf, M. (2021) 'Ligand binding at the protein–lipid interface: strategic considerations for drug design', *Nature Reviews Drug Discovery*. Nat Rev Drug Discov, pp. 710–722. doi: 10.1038/s41573-021-00240-2.

Penmatsa, A. and Gouaux, E. (2014) 'How LeuT shapes our understanding of the mechanisms of sodium-coupled neurotransmitter transporters', *Journal of Physiology*. Blackwell Publishing Ltd, 592(5), pp. 863–869. doi: 10.1113/jphysiol.2013.259051.

Penmatsa, A., Wang, K. H. and Gouaux, E. (2013) 'X-ray structure of dopamine transporter elucidates antidepressant mechanism', *Nature*. Nature

Publishing Group, 503(7474), pp. 85–90. doi: 10.1038/nature12533.

Periole, X. *et al.* (2009) 'Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition', *Journal of Chemical Theory and Computation*, 5(9), pp. 2531–2543. doi: 10.1021/ct9002114.

Periole, X., Zeppelin, T. and Schiøtt, B. (2018) 'Dimer Interface of the Human Serotonin Transporter and Effect of the Membrane Composition', *Scientific Reports*, 8(1), pp. 1–15. doi: 10.1038/s41598-018-22912-7.

Perland, E. and Fredriksson, R. (2017) 'Classification Systems of Secondary Active Transporters', *Trends in Pharmacological Sciences*, 38(3), pp. 305–315. doi: 10.1016/j.tips.2016.11.008.

Pettersen, E. F. *et al.* (2021) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', *Protein Science*, 30(1), pp. 70–82. doi: 10.1002/pro.3943.

Piscitelli, C. L. and Gouaux, E. (2012) 'Insights into transport mechanism from LeuT engineered to transport tryptophan', *EMBO Journal*. Nature Publishing Group, 31(1), pp. 228–235. doi: 10.1038/emboj.2011.353.

Porohille, A. and Chipot, C. (2008) 'Free Energy Calculations - Theory and Applications in Chemistry and Biology', in Porohille, A. and Chipot, C. (eds) *SPRINGER SERIES IN CHEMICAL PHYSICS 86*. Springer International Publishing.

Poulos, S. *et al.* (2015) 'Bicelles coming of age: an empirical approach to bicelle crystallization.', *Methods in enzymology*. United States, 557, pp. 393–416. doi: 10.1016/bs.mie.2014.12.024.

Prasad, R., Sliwa-Gonzalez, A. and Barral, Y. (2020) 'Mapping bilayer thickness in the ER membrane', *Science Advances*. Sci Adv, 6(46). doi: 10.1126/sciadv.aba5130.

Prejanò, M. *et al.* (2021) 'Computational Study Reveals the Role of Water Molecules in the Inhibition Mechanism of LAT1 by 1,2,3-Dithiazoles', *Journal of Chemical Information and Modeling*, 61(12), pp. 5883–5892. doi: 10.1021/acs.jcim.1c01012.

Privé, G. G. (2007) 'Detergents for the stabilization and crystallization of membrane proteins', *Methods*. Academic Press, 41(4), pp. 388–397. doi: 10.1016/j.ymeth.2007.01.007.

Pronk, S. *et al.* (2013) 'GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit', *Bioinformatics*, 29(7), pp. 845–854. doi: 10.1093/bioinformatics/btt055.

Punjani, A. *et al.* (2017) 'cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination', *Nature Methods*, 14(3), pp. 290–296. doi:

10.1038/nmeth.4169.

Rabanal-Ruiz, Y. and Korolchuk, V. I. (2018) 'mTORC1 and Nutrient Homeostasis: The Central Role of the Lysosome', *International Journal of Molecular Sciences*, 19(3), p. 818. doi: 10.3390/ijms19030818.

Rasmussen, S. G. F., DeVree, B. T., *et al.* (2011) 'Crystal structure of the β2 adrenergic receptor–Gs protein complex', *Nature*, 477(7366), pp. 549–555. doi: 10.1038/nature10361.

Rasmussen, S. G. F., Choi, H.-J., *et al.* (2011) 'Structure of a nanobody-stabilized active state of the β2 adrenoceptor', *Nature*, 469(7329), pp. 175–180. doi: 10.1038/nature09648.

Raunser, S. and Walz, T. (2009) 'Electron crystallography as a technique to study the structure on membrane proteins in a lipidic environment', *Annual Review of Biophysics*. Annu Rev Biophys, pp. 89–105. doi: 10.1146/annurev.biophys.050708.133649.

Rebsamen, M. *et al.* (2015) 'SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1', *Nature*. Nature Publishing Group, 519(7544), pp. 477–481. doi: 10.1038/nature14107.

Reig, N. *et al.* (2002) 'The light subunit of system b(o,+) is fully functional in the absence of the heavy subunit.', *The EMBO journal*. England, 21(18), pp. 4906–14. doi: 10.1093/emboj/cdf500.

Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*, 42(W1), pp. 320–324. doi: 10.1093/nar/gku316.

Rodríguez-Banqueri, A. *et al.* (2016) 'Stabilization of a prokaryotic LAT transporter by random mutagenesis', *Journal of General Physiology*. Rockefeller University Press, 147(4), pp. 353–368. doi: 10.1085/jgp.201511510.

Rodriguez, C. F. *et al.* (2021) 'Structural basis for substrate specificity of heteromeric transporters of neutral amino acids', *Proceedings of the National Academy of Sciences of the United States of America*, 118(49), pp. 1–10. doi: 10.1073/pnas.2113573118.

Rudnick, G. (2011) 'Cytoplasmic permeation pathway of neurotransmitter transporters', *Biochemistry*. Biochemistry, pp. 7462–7475. doi: 10.1021/bi200926b.

Ruivo, R. *et al.* (2012) 'Mechanism of proton/substrate coupling in the heptahelical lysosomal transporter cystinosin', *Proceedings of the National Academy of Sciences of the United States of America*, 109(5). doi: 10.1073/pnas.1115581109.

Ruprecht, J. J. *et al.* (2019) 'The Molecular Mechanism of Transport by the Mitochondrial ADP/ATP Carrier', *Cell*, 176(3), pp. 435-447.e15. doi:

10.1016/j.cell.2018.11.025.

Ryckaert, J.-P., Ciccotti, G. and Berendsen, H. J. . (1977) 'Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes', *Journal of Computational Physics*. Academic Press, 23(3), pp. 327–341. doi: 10.1016/0021-9991(77)90098-5.

Sáenz, J. P. *et al.* (2012) 'Functional convergence of hopanoids and sterols in membrane ordering', *Proceedings of the National Academy of Sciences of the United States of America*, 109(35), pp. 14236–14240. doi: 10.1073/pnas.1212141109.

Salema, V. and Fernández, L. Á. (2013) 'High yield purification of nanobodies from the periplasm of E. coli as fusions with the maltose binding protein', *Protein Expression and Purification*, 91(1), pp. 42–48. doi: https://doi.org/10.1016/j.pep.2013.07.001.

Sancak, Y. *et al.* (2008) 'The Rag GTPases Bind Raptor and Mediate Amino Acid Signaling to mTORC1', *Science*, 320(5882), pp. 1496–1501. doi: 10.1126/science.1157535.

Sanders, C. R. and Prestegard, J. H. (1990) 'Magnetically orientable phospholipid bilayers containing small amounts of a bile salt analogue, CHAPSO', *Biophysical Journal*, 58(2), pp. 447–460. doi: https://doi.org/10.1016/S0006-3495(90)82390-0.

Saxton, R. A. *et al.* (2016) 'Mechanism of arginine sensing by CASTOR1 upstream of mTORC1', *Nature*. Nature Publishing Group, 536(7615), pp. 229–233. doi: 10.1038/nature19079.

Saxton, R. A. and Sabatini, D. M. (2017) 'mTOR Signaling in Growth, Metabolism, and Disease', *Cell*, 169(2), pp. 361–371. doi: 10.1016/j.cell.2017.03.035.

Scanlon, S. M., Williams, D. C. and Schloss, P. (2001) 'Membrane cholesterol modulates serotonin transporter activity', *Biochemistry*. American Chemical Society , 40(35), pp. 10507–10513. doi: 10.1021/bi010730z.

Schachter, I. *et al.* (2022) 'The Two Faces of the Liquid Ordered Phase', *Journal of Physical Chemistry Letters*. American Chemical Society, 13(5), pp. 1307–1313. doi: 10.1021/acs.jpclett.1c03712.

Schiöth, H. B. *et al.* (2013a) 'Evolutionary origin of amino acid transporter families SLC32, SLC36 and SLC38 and physiological, pathological and therapeutic aspects', *Molecular Aspects of Medicine*, 34(2–3), pp. 571–585. doi: 10.1016/j.mam.2012.07.012.

Schiöth, H. B. *et al.* (2013b) 'Evolutionary origin of amino acid transporter families SLC32, SLC36 and SLC38 and physiological, pathological and therapeutic aspects', *Molecular Aspects of Medicine*. Pergamon, pp. 571–585. doi:

10.1016/j.mam.2012.07.012.

Schweikhard, E. S. and Ziegler, C. M. (2012) 'Amino acid secondary transporters: Toward a common transport mechanism', in *Current Topics in Membranes*, pp. 1–28. doi: 10.1016/B978-0-12-394316-3.00001-6.

Seiferth, D., Biggin, P. C. and Tucker, S. J. (2022) 'When is a hydrophobic gate not a hydrophobic gate?', *The Journal of general physiology*, 154(11), pp. 1–7. doi: 10.1085/jgp.202213210.

Sezgin, E. *et al.* (2017) 'The mystery of membrane organization: Composition, regulation and roles of lipid rafts', *Nature Reviews Molecular Cell Biology*. Nat Rev Mol Cell Biol, pp. 361–374. doi: 10.1038/nrm.2017.16.

Shaffer, P. L. *et al.* (2009) 'Structure and mechanism of a Na+-independent amino acid transporter.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 325(5943), pp. 1010–4. doi: 10.1126/science.1176088.

Shaikh, S. A. and Tajkhorshid, E. (2010) 'Modeling and dynamics of the inward-facing state of a Na+/Cl- dependent neurotransmitter transporter homologue', *PLoS Computational Biology*. PLoS Comput Biol, 6(8). doi: 10.1371/journal.pcbi.1000905.

Shi, L. *et al.* (2008) 'The Mechanism of a Neurotransmitter:Sodium Symporter-Inward Release of Na+ and Substrate Is Triggered by Substrate in a Second Binding Site', *Molecular Cell*. Cell Press, 30(6), pp. 667–677. doi: 10.1016/j.molcel.2008.05.008.

Shi, L. and Harel, W. (2010) 'Conformational rearrangements to the intracellular open states of the LeuT and ApcT transporters are modulated by common mechanisms', *Biophysical Journal*. Biophysical Society, 99(12), pp. L103–L105. doi: 10.1016/j.bpj.2010.10.003.

Shimamura, T. *et al.* (2010) 'Molecular Basis of Alternating Access Membrane Transport by the Sodium-Hydantoin Transporter Mhp1', *Science*, 328(5977), pp. 470–473. doi: 10.1126/science.1186303.

Sinensky, M. (1974) 'Homeoviscous adaptation: a homeostatic process that regulates the viscosity of membrane lipids in Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 71(2), pp. 522–525. doi: 10.1073/pnas.71.2.522.

Singh, S. K. (2008) 'LeuT: A prokaryotic stepping stone on the way to a eukaryotic neurotransmitter transporter structure', *Channels*. Taylor and Francis Inc., pp. 380–389. doi: 10.4161/chan.2.5.6904.

Sircar, A. *et al.* (2011) 'Analysis and Modeling of the Variable Region of Camelid Single-Domain Antibodies', *The Journal of Immunology*, 186(11), pp. 6357–6367. doi: 10.4049/jimmunol.1100116.

Søndergaard, C. R. *et al.* (2011) 'Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values', *Journal of Chemical Theory and Computation*, 7(7), pp. 2284–2295. doi: 10.1021/ct200133y.

Song, W. *et al.* (2019) 'State-dependent Lipid Interactions with the A2a Receptor Revealed by MD Simulations Using In Vivo-Mimetic Membranes', *Structure*. Elsevier Ltd., 27(2), pp. 392-403.e3. doi: 10.1016/j.str.2018.10.024.

Song, W. *et al.* (2022) 'PyLipID: A Python Package for Analysis of Protein-Lipid Interactions from Molecular Dynamics Simulations', *Journal of Chemical Theory and Computation*, 18(2), pp. 1188–1201. doi: 10.1021/acs.jctc.1c00708.

Souza, P. C. T. *et al.* (2021) 'Martini 3: a general purpose force field for coarse-grained molecular dynamics', *Nature Methods*. Springer US, 18(4), pp. 382–388. doi: 10.1038/s41592-021-01098-3.

Sperandeo, M. P. *et al.* (1998) 'The gene encoding a cationic amino acid transporter (SLC7A4) maps to the region deleted in the velocardiofacial syndrome', *Genomics*, 49(2), pp. 230–236. doi: 10.1006/geno.1998.5252.

Stansfeld, P. J., Jefferys, E. E. and Sansom, M. S. P. (2013) 'Multiscale Simulations Reveal Conserved Patterns of Lipid Interactions with Aquaporins', *Structure*. Elsevier Ltd, 21(5), pp. 810–819. doi: 10.1016/j.str.2013.03.005.

Stansfeld, P. J. and Sansom, M. S. P. (2011) 'From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations', *Journal of Chemical Theory and Computation*, 7(4), pp. 1157–1166. doi: 10.1021/ct100569y.

Stetsenko, A. and Guskov, A. (2017) 'An overview of the top ten detergents used for membrane protein crystallization', *Crystals*, 7(7). doi: 10.3390/cryst7070197.

Straub, M. S. *et al.* (2021) 'Cryo-EM structures of the caspase-activated protein XKR9 involved in apoptotic lipid scrambling', *eLife*. Edited by J. L. Robertson, O. Boudker, and A. K. Menon. eLife Sciences Publications, Ltd, 10, p. e69800. doi: 10.7554/eLife.69800.

Su, Y.-H., Frommer, W. B. and Ludewig, U. (2004) 'Molecular and functional characterization of a family of amino acid transporters from Arabidopsis.', *Plant physiology*. American Society of Plant Biologists, 136(2), pp. 3104–13. doi: 10.1104/pp.104.045278.

Su, Y. H., Frommer, W. B. and Ludewig, U. (2004) 'Molecular and functional characterization of a family of amino acid transporters from arabidopsis', *Plant Physiology*, 136(2), pp. 3104–3113. doi: 10.1104/pp.104.045278.

Sun, P., Tropea, J. E. and Waugh, D. S. (2011) 'Enhancing the solubility of recombinant proteins in Escherichia coli by using hexahistidine-tagged maltose-

binding protein as a fusion partner.', *Methods in molecular biology (Clifton, N.J.)*. United States, 705, pp. 259–274. doi: 10.1007/978-1-61737-967-3_16.

Taghon, G. J. *et al.* (2021) 'Predictable cholesterol binding sites in GPCRs lack consensus motifs', *Structure*. Elsevier Ltd., 29(5), pp. 499-506.e3. doi: 10.1016/j.str.2021.01.004.

Tao, X. and Mackinnon, R. (2019) 'Molecular structures of the human slo1 k+ channel in complex with b4', *eLife*, 8, pp. 1–27. doi: 10.7554/eLife.51409.

Tao, Y. *et al.* (2015) 'Structure of a eukaryotic SWEET transporter in a homotrimeric complex', *Nature*. Nature Publishing Group, 527(7577), pp. 259–263. doi: 10.1038/nature15391.

Tate, C. G. (2012) 'A crystal clear solution for determining G-protein-coupled receptor structures', *Trends in Biochemical Sciences*. Elsevier, pp. 343–352. doi: 10.1016/j.tibs.2012.06.003.

Tavoulari, S. *et al.* (2016) 'Two Na+ sites control conformational change in a neurotransmitter transporter homolog', *Journal of Biological Chemistry*. J Biol Chem, 291(3), pp. 1456–1471. doi: 10.1074/jbc.M115.692012.

Terry, D. S. *et al.* (2018) 'A partially-open inward-facing intermediate conformation of LeuT is associated with Na+ release and substrate transport', *Nature Communications*. Springer US, 9(1), p. 230. doi: 10.1038/s41467-017-02202-y.

Theodoulou, F. L. and Kerr, I. D. (2015) 'ABC transporter research: going strong 40 years on', *Biochemical Society Transactions*, 43(5), pp. 1033–1040. doi: 10.1042/BST20150139.

Thompson, A. A. *et al.* (2011) 'GPCR stabilization using the bicelle-like architecture of mixed sterol-detergent micelles', *Methods*. NIH Public Access, 55(4), pp. 310–317. doi: 10.1016/j.ymeth.2011.10.011.

Thwaites, David T and Anderson, C. M. (2011) 'The SLC36 family of proton-coupled amino acid transporters and their potential role in drug transport', *British Journal of Pharmacology*, 164(7), pp. 1802–1816. doi: 10.1111/j.1476-5381.2011.01438.x.

Thwaites, David T. and Anderson, C. M. H. (2011) 'The SLC36 family of proton-coupled amino acid transporters and their potential role in drug transport', *British Journal of Pharmacology*. Br J Pharmacol, pp. 1802–1816. doi: 10.1111/j.1476-5381.2011.01438.x.

Tironi, I. G., Sperb, R., Smith, P. E. and van Gunsteren, W. F. (1995) 'A generalized reaction field method for molecular dynamics simulations', *The Journal of Chemical Physics*, 102(13), pp. 5451–5459. doi: 10.1063/1.469273.

Tironi, I. G., Sperb, R., Smith, P. E. and Van Gunsteren, W. F. (1995) 'A generalized reaction field method for molecular dynamics simulations', *The*

Journal of Chemical Physics, 102(13), pp. 5451–5459. doi: 10.1063/1.469273.

Tohge, T. *et al.* (2011) 'From models to crop species: Caveats and solutions for translational metabolomics', *Frontiers in Plant Science*, 2(OCT), pp. 1–15. doi: 10.3389/fpls.2011.00061.

Tripathi, Parul *et al.* (2007) 'The role of nitric oxide in inflammatory reactions', *FEMS Immunology & Medical Microbiology*. Narnia, 51(3), pp. 443–452. doi: 10.1111/j.1574-695X.2007.00329.x.

Tsai, M. F. and Miller, C. (2013) 'Substrate selectivity in arginine-dependent acid resistance in enteric bacteria', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 110(15), pp. 5893–5897. doi: 10.1073/pnas.1301442110.

Tsuji, T. and Fujimoto, T. (2018) 'Lipids and lipid domains of the yeast vacuole', *Biochemical Society Transactions*, 46(5), pp. 1047–1054. doi: 10.1042/BST20180120.

Tusnády, G. E., Dobson, L. and Tompa, P. (2015) 'Disordered regions in transmembrane proteins', *Biochimica et Biophysica Acta - Biomembranes*. Elsevier B.V., 1848(11), pp. 2839–2848. doi: 10.1016/j.bbamem.2015.08.002.

Uchański, T. *et al.* (2021) 'Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM', *Nature Methods*. Springer US, 18(1), pp. 60–68. doi: 10.1038/s41592-020-01001-6.

Valdés-Tresanco, M. S. *et al.* (2021) 'gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS', *Journal of Chemical Theory and Computation*, 17(10), pp. 6281–6291. doi: 10.1021/acs.jctc.1c00645.

Valitova, J. N., Sulkarnayeva, A. G. and Minibayeva, F. V. (2016) 'Plant sterols: Diversity, biosynthesis, and physiological functions', *Biochemistry (Moscow)*, 81(8), pp. 819–834. doi: 10.1134/S0006297916080046.

Valleau, J. P. and J.M., T. (1977) 'Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling', *Journal of Computational Physics*, 23(2), pp. 187–199.

Vanaken, T. *et al.* (1986) 'Alkyl glycoside detergents: Synthesis and applications to the study of membrane proteins', in *Biomembranes Part M*. Academic Press, pp. 27–35. doi: https://doi.org/10.1016/S0076-6879(86)25005-3.

Vanoaica, L. *et al.* (2016) 'Real-time functional characterization of cationic amino acid transporters using a new FRET sensor', *Pflugers Archiv European Journal of Physiology*, 468(4), pp. 563–572. doi: 10.1007/s00424-015-1754-9.

Västermark, Å. and Saier, M. H. (2014) 'Evolutionary relationship between 5+5 and 7+7 inverted repeat folds within the amino acid-polyamine-organocation superfamily', *Proteins: Structure, Function, and Bioinformatics*. John Wiley & Sons, Ltd, 82(2), pp. 336–346. doi: 10.1002/prot.24401.

Vékony, N. *et al.* (2001) 'Human cationic amino acid transporter hCAT-3 is preferentially expressed in peripheral tissues', *Biochemistry*, 40(41), pp. 12387–12394. doi: 10.1021/bi011345c.

Verrey, F., Closs, Ellen I, *et al.* (2004) 'CATs and HATs: the SLC7 family of amino acid transporters', *Pflugers Arch-Eur J Physiol*, 447, pp. 532–542. doi: 10.1007/s00424-003-1086-z.

Verrey, F., Closs, Ellen I., *et al.* (2004) 'CATs and HATs: The SLC7 family of amino acid transporters', *Pflugers Archiv European Journal of Physiology*. Pflugers Arch, pp. 532–542. doi: 10.1007/s00424-003-1086-z.

Vickery, O. N. and Stansfeld, P. J. (2021) 'CG2AT2: An Enhanced Fragment-Based Approach for Serial Multi-scale Molecular Dynamics Simulations', *Journal of Chemical Theory and Computation*, 17(10), pp. 6472–6482. doi: 10.1021/acs.jctc.1c00295.

Voet, D. and Voet, J. (2011) *Biochemistry*. 4th edn. John Wiley & Sons, Inc.

Voth, G. A. (2008) *Coarse-Graining of Condensed Phase and Biomolecular Systems*, *CRC Press*. Available at: https://www.crcpress.com/Coarse-Graining-of-Condensed-Phase-and-Biomolecular-Systems/Voth/p/book/9781420059557.

Wang, H. *et al.* (1991) 'Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter', *Nature*, 352(6337), pp. 729–731. doi: 10.1038/352729a0.

Wang, J. *et al.* (2004) 'Development and testing of a general Amber force field', *Journal of Computational Chemistry*, 25(9), pp. 1157–1174. doi: 10.1002/jcc.20035.

Wang, K. H., Penmatsa, A. and Gouaux, E. (2015) 'Neurotransmitter and psychostimulant recognition by the dopamine transporter', *Nature*. Nature Publishing Group, 521(7552), pp. 322–327. doi: 10.1038/nature14431.

Wang, L., Chen, K. and Zhou, M. (2021) 'Structure and function of an Arabidopsis thaliana sulfate transporter', *Nature Communications*. Springer US, 12(1), pp. 1–8. doi: 10.1038/s41467-021-24778-2.

Wassenaar, T. A. *et al.* (2015) 'Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations', *Journal of Chemical Theory and Computation*, 11(5), pp. 2144–2155. doi: 10.1021/acs.jctc.5b00209.

Wei, Z. *et al.* (2022) 'LACC1 bridges NOS2 and polyamine metabolism in inflammatory macrophages', *Nature*. Springer US, 609(7926), pp. 348–353. doi: 10.1038/s41586-022-05111-3.

White, J. F. *et al.* (2012) 'Structure of the agonist-bound neurotensin receptor', *Nature*. Nature Publishing Group, 490(7421), pp. 508–513. doi: 10.1038/nature11558.

Wimley, W. C., Creamer, T. P. and White, S. H. (1996) 'Solvation energies of amino acid side chains and backbone in a family of host - Guest pentapeptides', *Biochemistry*, 35(16), pp. 5109–5124. doi: 10.1021/bi9600153.

Wolf, S. *et al.* (2002) 'Expression of solute carrier 7A4 (SLC7A4) in the plasma membrane is not sufficient to mediate amino acid transport activity', *Biochemical Journal*, 364(3), pp. 767–775. doi: 10.1042/BJ20020084.

Wu, D. *et al.* (2020) 'Structural basis for amino acid exchange by a human heteromeric amino acid transporter', *Proceedings of the National Academy of Sciences of the United States of America*, 117(35), pp. 21281–21287. doi: 10.1073/pnas.2008111117.

Wu, S. *et al.* (2012) 'Fabs enable single particle cryoEM studies of small proteins', *Structure*, 20(4), pp. 582–592. doi: 10.1016/j.str.2012.02.017.

Wu, X. and Rapoport, T. A. (2021) 'Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies)', *Proceedings of the National Academy of Sciences of the United States of America*, 118(41), pp. 1–9. doi: 10.1073/pnas.2115001118.

Wu, Z. *et al.* (2019) 'Proton Control of Transitions in an Amino Acid Transporter', *Biophysical Journal*. doi: 10.1016/j.bpj.2019.07.056.

Xu, L., Song, X. and Jia, L. (2017) 'A camelid nanobody against EGFR was easily obtained through refolding of inclusion body expressed in Escherichia coli.', *Biotechnology and applied biochemistry*. United States, 64(6), pp. 895–901. doi: 10.1002/bab.1544.

Yamashita, A. *et al.* (2005a) 'Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters', *Nature*. Nature, 437(7056), pp. 215–223. doi: 10.1038/nature03978.

Yamashita, A. *et al.* (2005b) 'Crystal structure of a bacterial homologue of Na+/Cl --dependent neurotransmitter transporters', *Nature*, 437(7056), pp. 215–223. doi: 10.1038/nature03978.

Yamashita, A. *et al.* (2005c) 'Crystal structure of a bacterial homologue of Na+/Cl --dependent neurotransmitter transporters', *Nature*. Nature Publishing Group, 437(7056), pp. 215–223. doi: 10.1038/nature03978.

Yan, J. *et al.* (2014) 'Construction of a synthetic phage-displayed Nanobody library with CDR3 regions randomized by trinucleotide cassettes for diagnostic applications', *Journal of Translational Medicine*, 12(1), p. 343. doi: 10.1186/s12967-014-0343-6.

Yan, R. *et al.* (2019a) 'Structure of the human LAT1–4F2hc heteromeric amino acid transporter complex', *Nature*. doi: 10.1038/s41586-019-1011-z.

Yan, R. *et al.* (2019b) 'Structure of the human LAT1–4F2hc heteromeric amino acid transporter complex', *Nature*. Nature Publishing Group, 568(7750),

pp. 127–130. doi: 10.1038/s41586-019-1011-z.

Yan, R., Li, Y., *et al.* (2020) 'Cryo-EM structure of the human heteromeric amino acid transporter b0,+AT-rBAT', *Science Advances*. American Association for the Advancement of Science, 6(16), p. eaay6379. doi: 10.1126/sciadv.aay6379.

Yan, R., Zhou, J., *et al.* (2020) 'Structural insight into the substrate recognition and transport mechanism of the human LAT2–4F2hc complex', *Cell Discovery*. Springer US, 6(1), pp. 20–23. doi: 10.1038/s41421-020-00207-4.

Yan, R. *et al.* (2021) 'Mechanism of substrate transport and inhibition of the human LAT1-4F2hc amino acid transporter', *Cell Discovery*. Springer US, 7(1), pp. 0–7. doi: 10.1038/s41421-021-00247-4.

Yanagida, O. *et al.* (2001) 'Human L-type amino acid transporter 1 (LAT1): Characterization of function and expression in tumor cell lines', *Biochimica et Biophysica Acta - Biomembranes*, 1514(2), pp. 291–302. doi: 10.1016/S0005-2736(01)00384-4.

Yang, D. *et al.* (2022) 'Structures and membrane interactions of native serotonin transporter in complexes with psychostimulants', *bioRxiv*, p. 2022.09.03.506477. doi: 10.1101/2022.09.03.506477.

Yang, D. and Gouaux, E. (2021) 'Illumination of serotonin transporter mechanism and role of the allosteric site', *Science Advances*, 7(49), pp. 1–12. doi: 10.1126/sciadv.abl3857.

Yang, H. *et al.* (2014) 'Characterization of the putative amino acid transporter genes AtCAT2, 3 &4: the tonoplast localized AtCAT2 regulates soluble leaf amino acids.', *Journal of plant physiology*. Elsevier GmbH., 171(8), pp. 594–601. doi: 10.1016/j.jplph.2013.11.012.

Yang, H., Stierhof, Y. D. and Ludewig, U. (2015) 'The putative cationic amino acid transporter 9 is targeted to vesicles and may be involved in plant amino acid homeostasis', *Frontiers in Plant Science*, 6(APR), pp. 1–10. doi: 10.3389/fpls.2015.00212.

Yang, X., Wang, Q. and Cao, E. (2020) 'Structure of the human cation–chloride cotransporter NKCC1 determined by single-particle electron cryo-microscopy', *Nature Communications*. Springer US, 11(1), pp. 1–11. doi: 10.1038/s41467-020-14790-3.

Yesylevskyy, S. O. *et al.* (2010) 'Polarizable Water Model for the Coarse-Grained MARTINI Force Field', *PLoS Computational Biology*. Edited by M. Levitt, 6(6). doi: 10.1371/journal.pcbi.1000810.

Yoshida, T. *et al.* (2019) 'Differential scanning fluorimetric analysis of the amino-acid binding to taste receptor using a model receptor protein, the ligand-binding domain of fish T1r2a/T1r3', *PLoS ONE*, 14(10), pp. 1–14. doi: 10.1371/journal.pone.0218909.

Zeppelin, T. *et al.* (2018) 'A direct interaction of cholesterol with the dopamine transporter prevents its out-to-inward transition', *PLoS Computational Biology*. Public Library of Science, 14(1), p. e1005907. doi: 10.1371/journal.pcbi.1005907.

Zhang, Y. *et al.* (2013) 'Protein Analysis by Shotgun/Bottom-up Proteomics', *Chemical Reviews*, 113(4), pp. 2343–2394. doi: 10.1021/cr3003533.

Zhao, Y. *et al.* (2010) 'Single-molecule dynamics of gating in a neurotransmitter transporter homologue', *Nature*. Nature, 465(7295), pp. 188–193. doi: 10.1038/nature09057.

Zheng, S. Q. *et al.* (2017) 'MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy', *Nature Methods*. Nature Publishing Group, 14(4), pp. 331–332. doi: 10.1038/nmeth.4193.

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

Zimmermann, I. *et al.* (2020a) 'Generation of synthetic nanobodies against delicate proteins', *Nature Protocols*. Nature Research, 15(5), pp. 1707–1741. doi: 10.1038/s41596-020-0304-x.

Zimmermann, I. *et al.* (2020b) 'Generation of synthetic nanobodies against delicate proteins', *Nature Protocols*. Nature Research, 15(5), pp. 1707–1741. doi: 10.1038/s41596-020-0304-x.

Zivanov, J. *et al.* (2018) 'New tools for automated high-resolution cryo-EM structure determination in RELION-3', *eLife*, 7, pp. 1–22. doi: 10.7554/eLife.42166.

Supplementary Information

APPENDIX A

Chapter 3 Supplementary Information

Comparison of *Hs* and *Dr* SLC7A4 Purification from *S. cerevisiae*

Purification from *S. cerevisiae* and biophysical characterisation was repeated

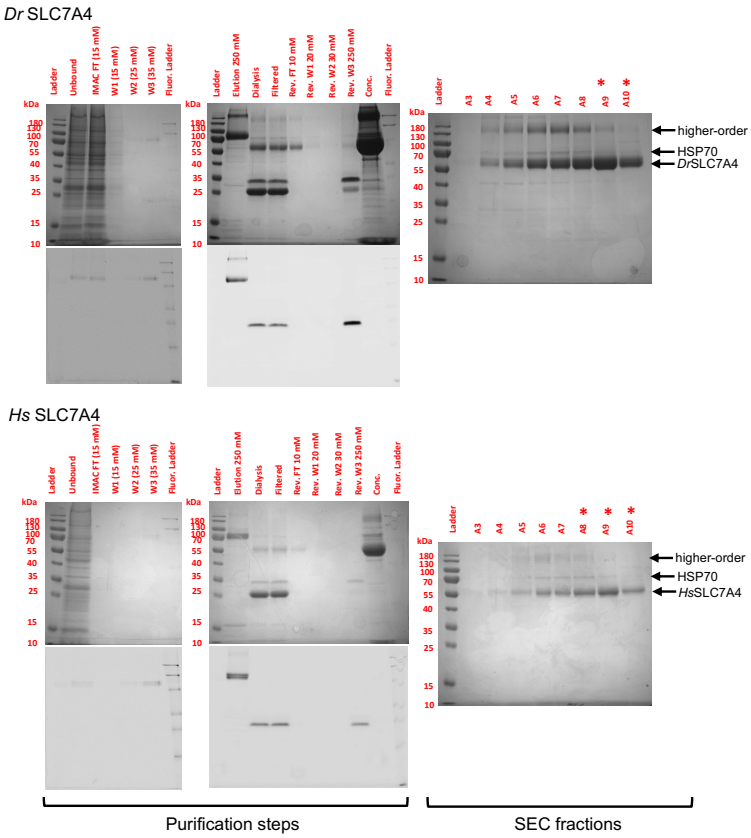


Figure 0.1 SDS-PAGE (coomasie-stained and in-gel GFP fluorescence) purification steps and SEC fraction comparison between *Hs* and *Dr* SLC7A4 constructs produced from *S. cerevisiae* overexpression host. Both constructs are enriched after Ni-affinity IMAC, cleave with TEV protease, and are significantly enriched in the concentrated samples prior to injection down a size-exclusion column. Pooled fractions used for thermal shift analysis are highlighted in *.

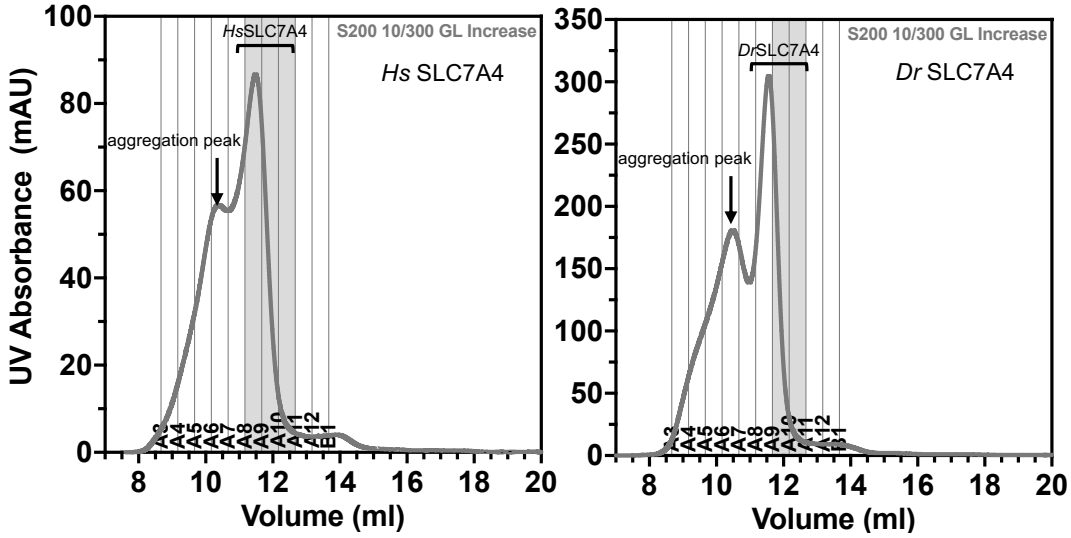


Figure 0.2 Comparison of gel-filtration traces from large-scale purification of Hs and Dr SLC7A4 constructs with the usual purification protocol from S. cerevisiae. Pooled fractions for analysis are indicated greyed.

for human SLC7A4 in LMNG-CHS and the DSF thermal shift profiles were similarly unresponsive as the zebrafish protein (**Figure 3.10-C**). Comparison of zebrafish and human SLC7A4 purification SDS-PAGE electrophoresis (**Figure 0.1**), size-exclusion chromatograms (**Figure 0.2**) are shown.

Rationalisation of *Hs* and *Dr* SLC7A4 Biochemical Quality

It is possible that the overexpressing system, *S. cerevisiae*, is incompatible with the correct, well-folded overexpression of higher SLC7A4 homologues. *S. cerevisiae* share the organisational hierarchy of more evolved hosts at the cellular level, but they do not possess the same post-translational machinery as insect or mammalian tissue cultures. In example, yeast expression systems have a different glycosylation pattern from mammalian cells, polymerising only mannose-containing glycans which would lead in a different molecular shape for SLC7A4 in yeast compared to HEK293 cells. The glycosylation predicted sites for human SLC7A4

are N151, N195, N221, N500 and N601 although only N151, N221 and N500 are extracellular glycosylation sites (uniprot entry O43246). Human, mouse, rat and SLC7A4 constructs conserve these asparagine residues in sequence alignment. Zebrafish conserves 2/3 (N151, N221). Plant conserves 1/3 (human N151 as plant N141) which is not obvious from a multiple sequence alignment but by superimposing human and plant predicted structures (Figure 0.3). Consequently, the con-

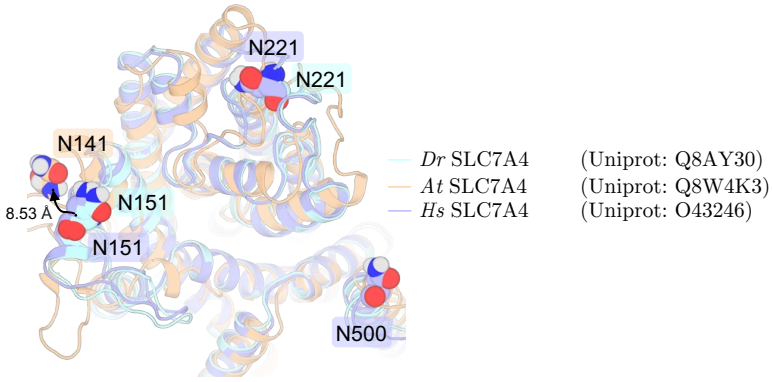


Figure 0.3 Structural superposition of SLC7A4 homologue Alphafold2 predicted structures with emphasis on the extracellular N-glycosylation-predicted sites. Asparagine residues are shown in vdW spheres, colour-coded per homologue (zebrafish – cyan, plant – wheat, human – light blue). The plant SLC7A4 N141 site was considered structurally homologous to human and zebrafish N151, both residing on extracellular loop EL1 and having a C_α distance 8.53 Å.

struct which would be least affected by the glycosylation machinery of the overexpression host would be the plant construct followed by zebrafish, and indeed the overexpression mg/l in *S. cerevisiae* follows this trend.

Human and zebrafish SLC7A4 homologues may require more stabilising purification conditions by (a) addition of protease inhibitors (e.g., aprotinin, PMSF, leupeptin, pepstatin A etc.), (b) pH changes and (c) substrate supplementation. 10 mM L-Arg/L-Lys supplementation to purified zebrafish and human SLC7A4 did not affect the DSF profile at pH 5.5, 6.5 or 7.5. Zebrafish purification at pH 6.5 (with 20 mM MES·NaOH) was tested with DDM and LMNG, but increased

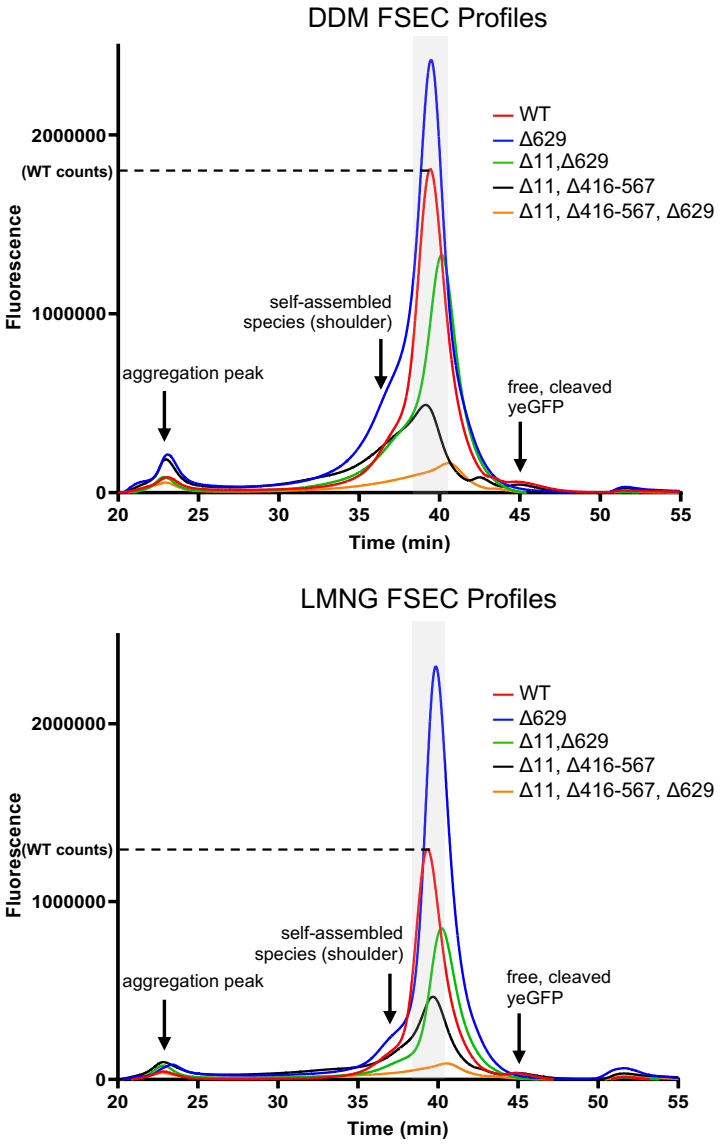


Figure 0.4 The behaviour of zebrafish SLC7A4 truncations in FSECs with DDM and LMNG. FSEC features are annotated with arrows and the grey highlighted region corresponds to the width of the WT chromatogram.

protein precipitation during dialysis was recorded for both detergents. Additional efforts for optimisation were abandoned.

Truncations to improve protein stability were tested on zebrafish SLC7A4, as it expressed better than the human homologue, yet, was evolutionary closer to human than the plant homologue (**Figure 3.1** & **Figure 3.2**). The truncations were

designed to ablate the (a) N-terminus (Δ11), (b) the C-terminus (Δ629) or (c) the additional amino acid stretch found in mammalian CATs which is absent from *Gk*ApcT (Δ416-567). Constructs were subjected to FSEC analysis (**Figure 0.4**). All constructs except the C-terminal truncation had whole-cell GFP expression counts <10,000 RFU, considerably less than WT (~ 15,000-16,000 RFU ≈ 0.08 mg SLC7A4 per l culture). The C-terminal truncation (~ 11,470 RFU ≈ 0.06 mg/L) produced an increased aggregation peak with a pronounced shoulder in both DDM and LMNG profiles, reflecting species aggregation. The dual Δ11,629 truncation FSEC was as monodisperse as WT but with only 30-50% expression levels. Any truncation involving the Δ416-567 resulted in hampered expression and broadened FSEC profiles in both detergents suggesting construct instability. Additional efforts for truncation optimisation were abandoned.

Chapter 4 Supplementary Information

AtCAT4-Avi Cloning Troubleshooting

Troubleshooting begun with assessing the sequence of the purified amplicon, i.e., the first step of the process. To sequence the amplicon, it was ligated into the commercial pJet1.2/blunt (CloneJet, ThermoFischer), a promiscuous host vector. Indeed, the CAT4-avi/pJet1.2 plasmid was sequence verified. It was hypothesised that using WT-CAT4/pLeu2D as a PCR template may be yielding non-specific PCR products that are small in size and are being cloned more efficiently in the yeast HR system than the desirable insert. Thus, sequence-verified CAT4-Avi/pJet1.2 was used as amplicon template with the HR primers and the amplicon reinserted into the HR pipeline. However, after restriction digestion the

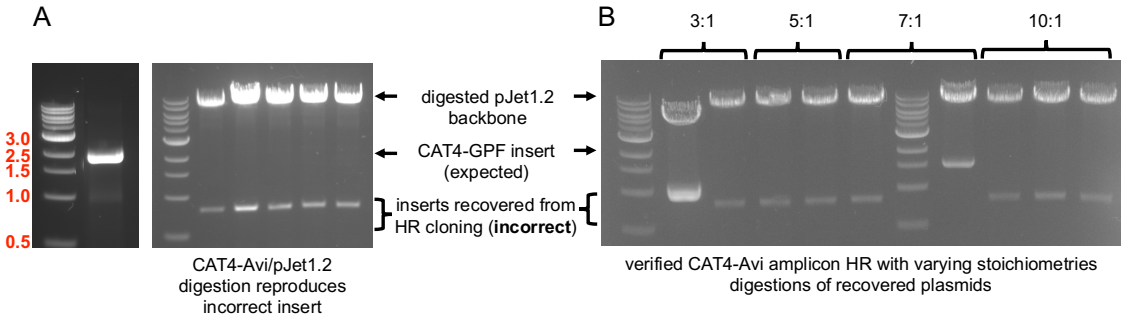


Figure 0.6 (A) Repetition of PCR amplification of CAT4-Avi with HR primers using CAT4-Avi/pJet1.2 as template resulting in a 2.6 kb primary product (left) and restriction digests from repetition of yeast HR cloning using this amplicon. (B) Attempt to optimise the recombination efficiency by screening different insert:vector stoichiometries ranging from 3:1-10:1. Bands at 2.6 kb were not detected.

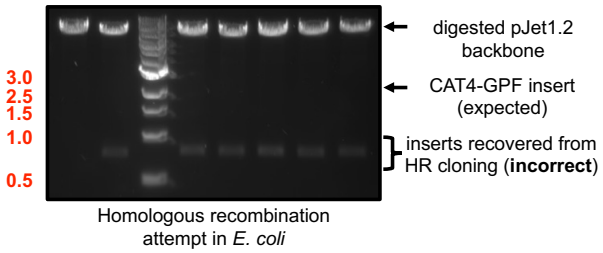


Figure 0.5 Repetition of in vivo homologous recombination in *E. coli* host instead of *S. cerevisiae*. The recovered insert is the same size as the one recovered from *S. cerevisiae* cloning.

CAT4-Avi-GFP band was absent and only the persistent 0.8 kb insert could be recovered (**Figure 0.6-A**). Having cross-examined the amplicon from two plasmid templates (pLeu2D and pJet1.2) we deciphered that although the CAT4-Avi amplicon is correctly generated by PCR it is incorrectly recombined *in vivo with* linearised Leu2D. A new *S. cerevisiae* strain BJ5460 was grown and used from a glycerol stock to assess whether the yeast stock used was problematic. This did not result in different observations and the 0.8 kb band was still recovered. Perhaps the usual protocol used in the group of mixing 5:1 insert:linearised-pLeu2D stoichiometry is not optimal for CAT4-Avi. To this end, different ratios of mixing were used ranging from 3:1 to 10:1 insert:vector molar ratio to optimise HR efficiency in yeast. However, this too did not result in the correct digestion band

(Figure 0.6-B). Since the inserted gel-extracted PCR amplicon was sequence-verified, different insert:vector ratios were attempted and the yeast stock was not problematic, it was hypothesised that the increased length of the CAT4 gene may be leading to recombination of a portion of the total CAT4-Avi insert, explaining the 0.8 kb band. As an alternative to HR in *S. cerevisiae* we tested HR in *E. coli*, as reported here (Jacobus and Gross, 2015), at the reported optimal 2:1 insert:vector ratio. It was observed that both the +insert and –insert (negative control) resulted in the same number of transformed colonies. Nevertheless, the inserts were analysed by colony PCR and restriction digestion which still proved to lack the desirable insert (Figure 0.5).

An orthogonal strategy not involving *in vivo* HR was traditional restriction enzyme subcloning methods (Figure 0.7). As was explained above, the *At*CAT4-avi could be produced by PCR and inserted into pJet1.2. The WT-*At*CAT4 gene-string has a BglII restriction site towards the C-terminal end of its sequence. There is also a HindIII site downstream of yeGFP in pLeu2D. The BglII/HindIII WT fragment in pLeu2D was exchanged via ligation subcloning by the equivalent avi-tagged fragment (Figure 0.7-C). While in theory the method should have worked, no colonies upon ligation were observed. The supercoiled WT-*At*CAT4/pLeu2D plasmid is known to be resistant to sticky-end ligation cloning within the Newstead Group. The supercoiled population of pLeu2D can be observed by DNA electrophoresis of the undigested plasmid (Figure 0.7-A).

With traditional subcloning proving resistant in pLeu2D, the HR methodology

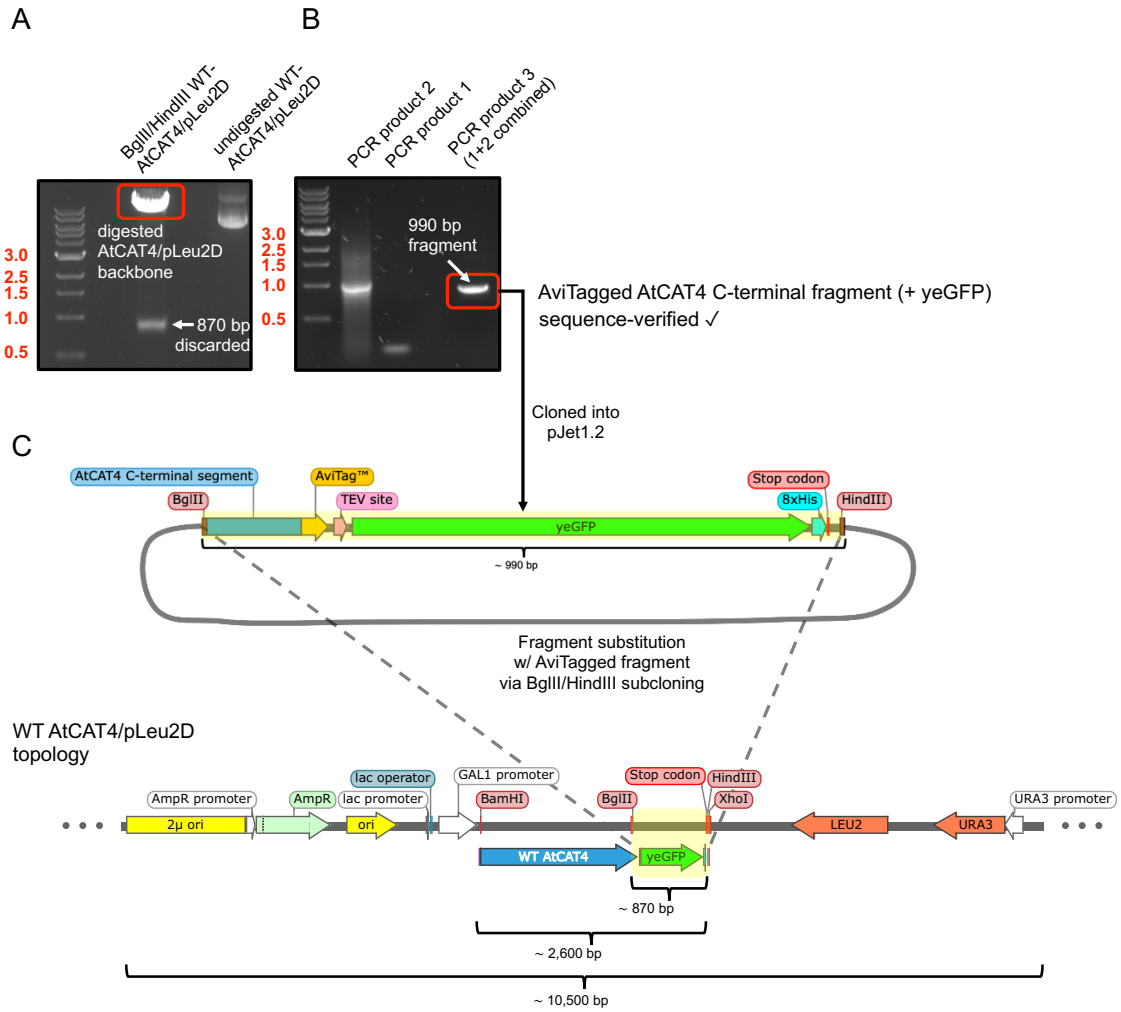


Figure 0.7 Procedure for subcloning of a C-terminal AtCAT4-Avi-tagged fragment from pJet1.2 into pLeu2D containing the remaining AtCAT4 sequence. (A) Restriction digest of WT AtCAT4/pLeu2D plasmid with BglII/HindIII produces two bands; an 870 bp fragment highlighted in yellow in (C) which was discarded and the backbone purified. (B) PCR modification of WT AtCAT4 insert by overlap PCR (described in **chapter 2**); Product 1 corresponds to the BglII-site→AviTag; Product 2 corresponds to AviTag→HindIII-site; Product 3 combines Products 1 and 2 by overlap extension. Product 3 was subcloned into pJet1.2 to be verified by sanger sequencing (as shown in C). (C) Schematic outline of final cloning step involving BglII/HindIII digestion of AtCAT4 C-terminal-AviTag-TEV_{site}-yeGFP-His_8 fragment and subcloning into BglII/HindIII digested AtCAT4/pLeu2D by sticky-end ligation.

was revisited upon to the observation that a persistent 0.8 kb insert was being cloned into the target vector. Why is this insert contaminating cloning even though PCR products are being gel extracted? To understand this discrepancy

the 0.8 kb insert was sanger sequenced. We hypothesised that it may originate from a non-specific PCR reaction. To test this, the PCR amplification of the WT-*At*CAT4-yeGFP insert from pLeu2D was repeated but at a lowered annealing temperature to boost production of non-specific products. Indeed, a prominent band was recovered at 0.8 kb (**Figure 4.5-A**). This surprising result was explained since the HR-specific sequence at the 5'-end of *CAT4* and the sequence directly upstream of the last 18 residues had a sequence identity >85%. As a result, the PCR HR forward primer was priming this non-specific site within the *CAT4* sequence and generating a truncated CAT4 avi-tagged PCR product (albeit in smaller yields; faint bands in agarose gel). Alas, HR cloning methods were failing previously due to a low-abundance PCR by-product that was more efficiently being recombined with the target vector due to its smaller size.

Chapter 5 Supplementary Information

Other Sybodies as CryoEM SPA Candidates

As discussed in **chapter 4**, the sybodies are from the concave library, having similar geometry, CDR length and residue composition to SybB5. These Syb should possess similar selectivity as SybB5, to CAT4, however, the CDR per residue identity, was found to be low in **chapter 4**, suggesting diversity amongst binders. Mild sequence differences in the CDRs are likely to modulate the side-chain atomistic interactions. Computational residue scanning analysis (**Figure 0.8**), showed that the affinity could be improved by exchanging M31 to either N/H/Y (other Sybs), or W57Y (SybA6) or Q56K (SybA6). The predicted affinity is however lowered when mutating SybB5 interacting residues, i.e., W33E (SybA6),

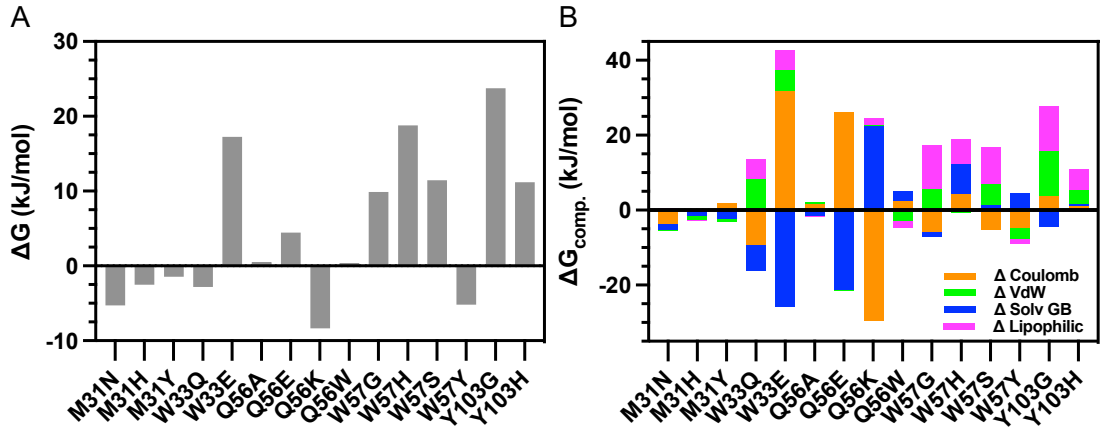


Figure 0.8 (A) Computational residue scanning analysis of SybB5 residues involved in atomistic interactions. Residues were mutated to residues of the other Syb. The y-axis shows change in binding affinity, in kJ/mol, calculated with molecular mechanics generalised-Born surface area (MMGBSA) method. Mutated residues and those within 4.0 Å where energy minimised with the amber14sb force field prior to MMGBSA. The more negative the ΔG value is the higher the CAT4-Syb_{mut} predicted affinity. (B) Decomposition of the free energy values reported in (A) to the main contributing energetic components.

W57H/S/Y (all other Sybs) and Y103H (SybA6) and Y103G (SybH7). These predictions might include bias towards the experimental binding pose of SybB5 but with this limitation considered, the data suggest that although SybA6 would improve some interactions, it would negate others making it an ambiguous second choice to SybB5. SybH7 has the W33E and Y103G mutations (ΔG>0) , making it also a suboptimal choice.

SybD8, conserves Y103, has an aromatic substitution W57H and a W33Q mutation, the latter giving negative MMGBSA ΔG. SybD8 was also ranked 2nd to SybB5 in BLI off-rate providing confidence that it is a good alternative Syb to test in SPA. A protein sample bound with SybD8, was prepared in an identical manner to SybB5. The SybD8 sample showed comparable particle distribution to SybB5 in Cu grids and data collection with TITAN KRIOS was done. However, vibration in the microscope hardware due to the column valve not isolating

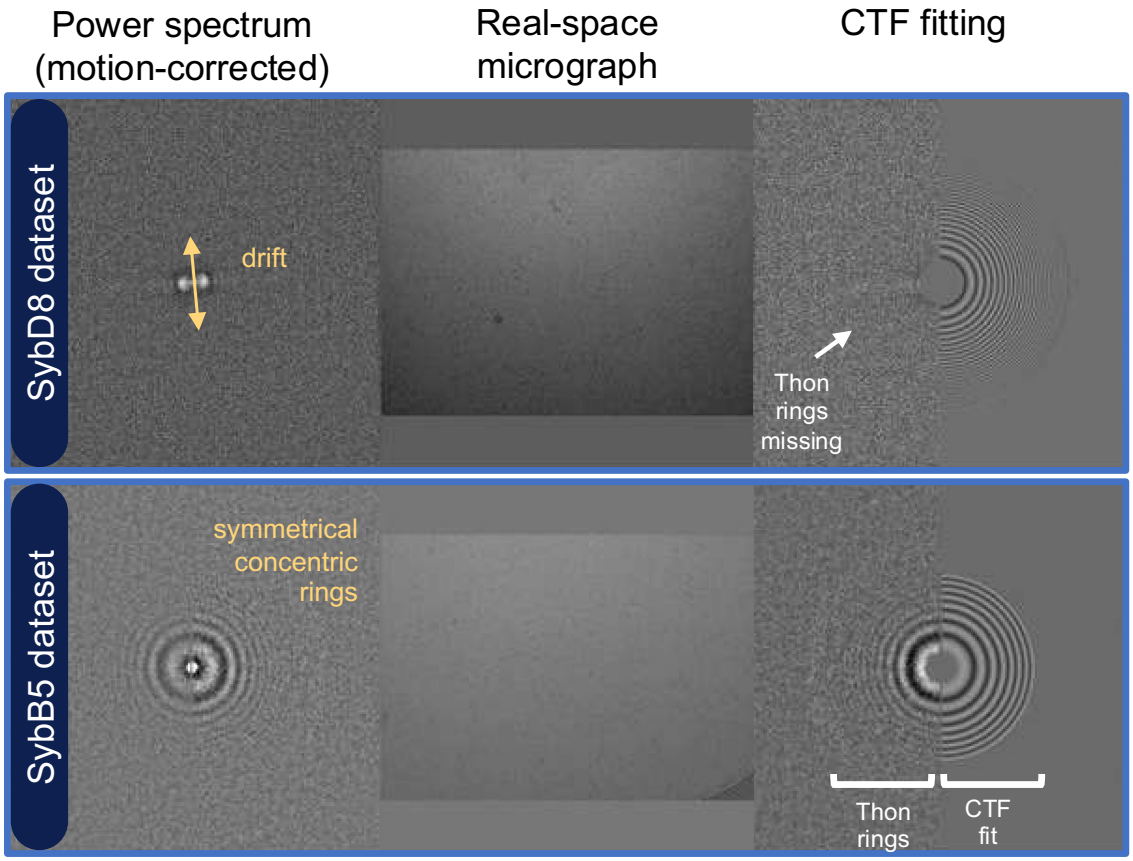


Figure 0.9 Comparison of micrograph quality between then SybB5 and SybD8 datasets. SybD8 dataset has problematic for SPA motion drift.

properly induced noticeable sample motion by the directional loss of Thon rings in the power spectrum of the motion-corrected micrographs (**Figure 0.9**). Thon rings report the information content of the image as a function of spatial frequency, i.e., resolution. The white rings represent spatial frequencies that contain information, while the gaps between rings represent spatial frequencies with low or zero information. The spatial information in the case of SybD8 is non-symmetrical and dampens faster than SybB5, i.e., the vibration masked the high-resolution signals. Only low resolution information (> 8 Å) – i.e., "blobs" – was available for SybD8, but showing Syb binding.

The data from the motion-drifted SybD8 dataset was 2D-classified and *ab*-

initio models (N=2) were generated, with lowpass filtering at 12 Å. 53.5% of the particles corresponded to a Syb-bound low resolution map with the remaining likely consisting of a combined unbound CAT4 and junk particle set. For comparison, this value ranged from 74.6%-52.1% for SybB5 measured in the initial "test" (P3-J194) and "collection" (P8-J70) datasets. The similar % of bound CAT4 particles seems to suggest it would be similar to SybB5. With this preliminary information, could a structure with SybD8 be solved with cryoEM SPA? Possibly, but it is challenging to suggest this with confidence without full single particle analysis of a dataset without motion-induced drift. However, due to the high expense associated with each data collection and with the simultaneous successful reconstruction of the posterior SybB5 dataset (main dataset), another data collection with SybD8 was abandoned.

AtCAT4 unmodelled loops

EL2-EL4 Diffuse Electron Densities

The 294-305 EL4 loop prediction is confidently structured by Alphafold2 (90>pLDDT>70) and forms a small α-helical region. However, N141 is the predicted N-glycosylation site for *At*CAT4 (chapter 3), and it is not surprising that this loop would be hypermobile when glycosylated, a feature not accounted for by Alphafold2. Attempts to model the EL2 and EL4 loop residues in the diffuse densities (Figure 0.10) led to overall low local cross-correlation values prohibiting further attempts to bias towards the Alphafold2 prediction. Interestingly EL4 is in the direct vicinity of SybB5 and modelling an α-helical like density (as in Alphafold2) caused steric clashes with the SybB5 density. Therefore, Syb binding causes translation and disordering of said loop, hindering modelling.

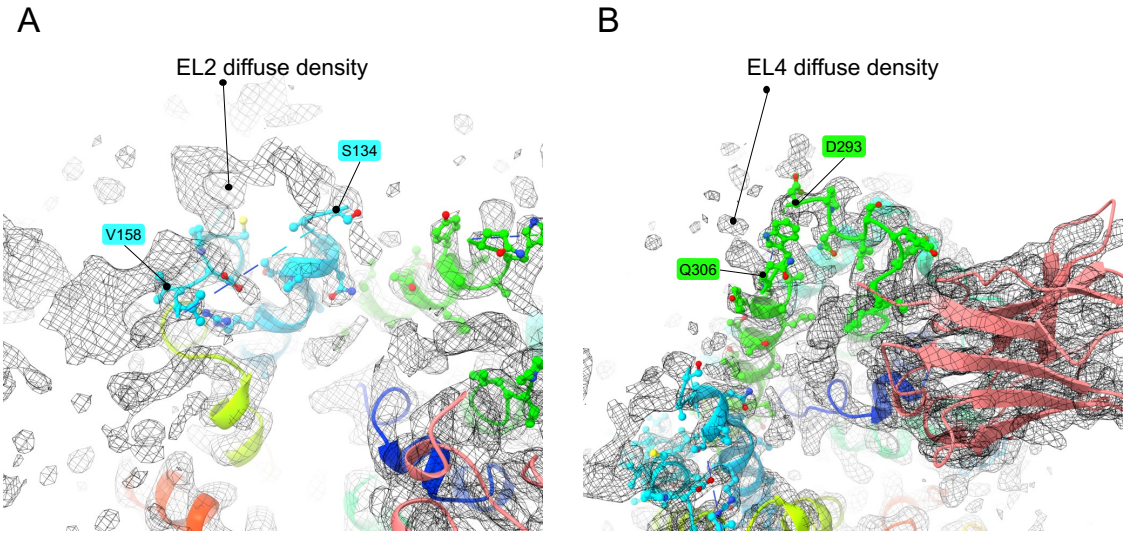


Figure 0.10 Diffuse extracellular loop densities corresponding to (A) EL2 and (B) EL4. The residues before the discontinuous, unmodelled regions are annotated. SybB5 is shown in salmon colouring. The electron density map is shown as a black mesh at step = 1 and level = 0.392 settings on ChimeraX-1.2.5.

IL5 Disordered Propensity

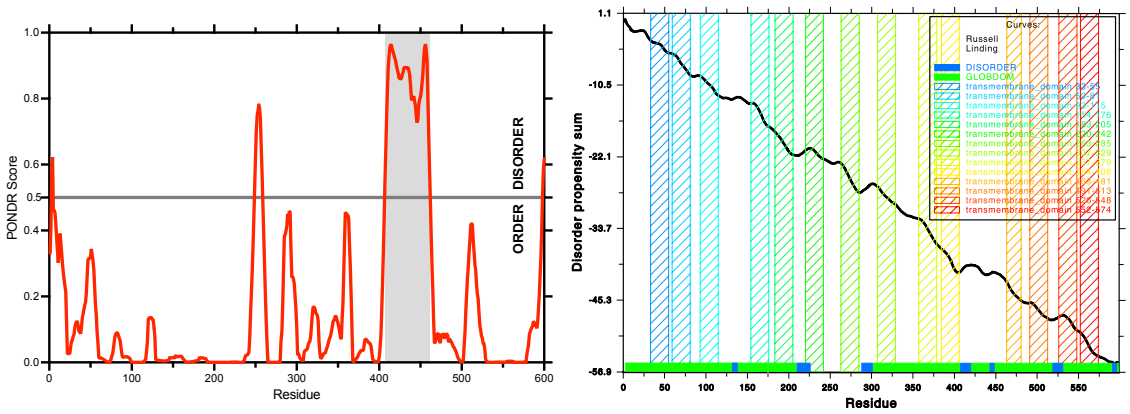


Figure 0.11 CAT4 sequence analysis outputs of two intrinsically disorder prediction algorithms, PONDR (left) and DisEMBL (right). Data plotted in Prism 9.

Diffuse densities are either statements of highly-mobile loops and/or intrinsically disordered regions (IDRs) a very common feature of signaling membrane proteins (Tusnády, Dobson and Tompa, 2015).. Indeed, the IL5 48-residue stretch (residues 409-457) are predicted disordered from the sequence; having a PONDR score $>$ 0.75, with 0.5 being the cutoff to assign a region as disordered;

and according to Russell-Linding disorder definition (Linding *et al.*, 2003) (Figure 0.11). The lack of density for IL5 in the cryoEM map for CAT4 further testifies to these predictions while the Alphafold2 structural prediction has very low confidence (pLDDT $<$ 50) for this region (Figure 5.8-A).

AtCAT4 Purification in GDN

The purification procedure in GDN is described in detail in chapter 2. The resulting gel-filtration traces and SDS-PAGE analysis of protein purity is shown below showing homogeneous *At*CAT4 in SEC and high purity in SDS-PAGE analysis of the SEC fractions.

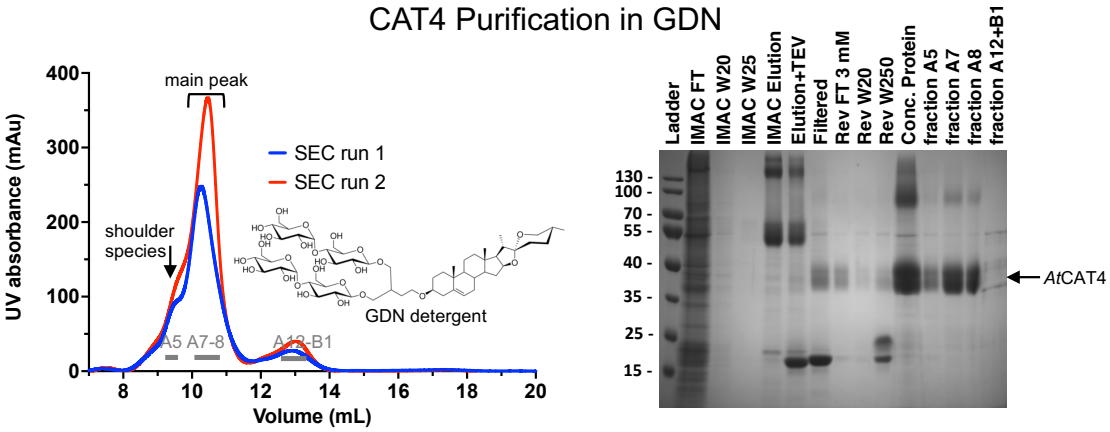


Figure 0.12 AtCAT4 purification in GDN detergent belt.

Electrophysiological Characterisation of AtCAT4 with TEVC

No measurable currents were generated by *At*CAT4 at –140 mV upon the addition of 10 mM lysine at 2-3 days post-injection (Figure 0.13-A). DSF thermal shift data suggested that purified *At*CAT-4 may bind to histidine at pH 7.5 (Figure 5.17). However, no measurable currents were obtained with L-histidine at pH 6 or 7 (Figure 0.13-B). None of the proteinogenic amino acids had a significant

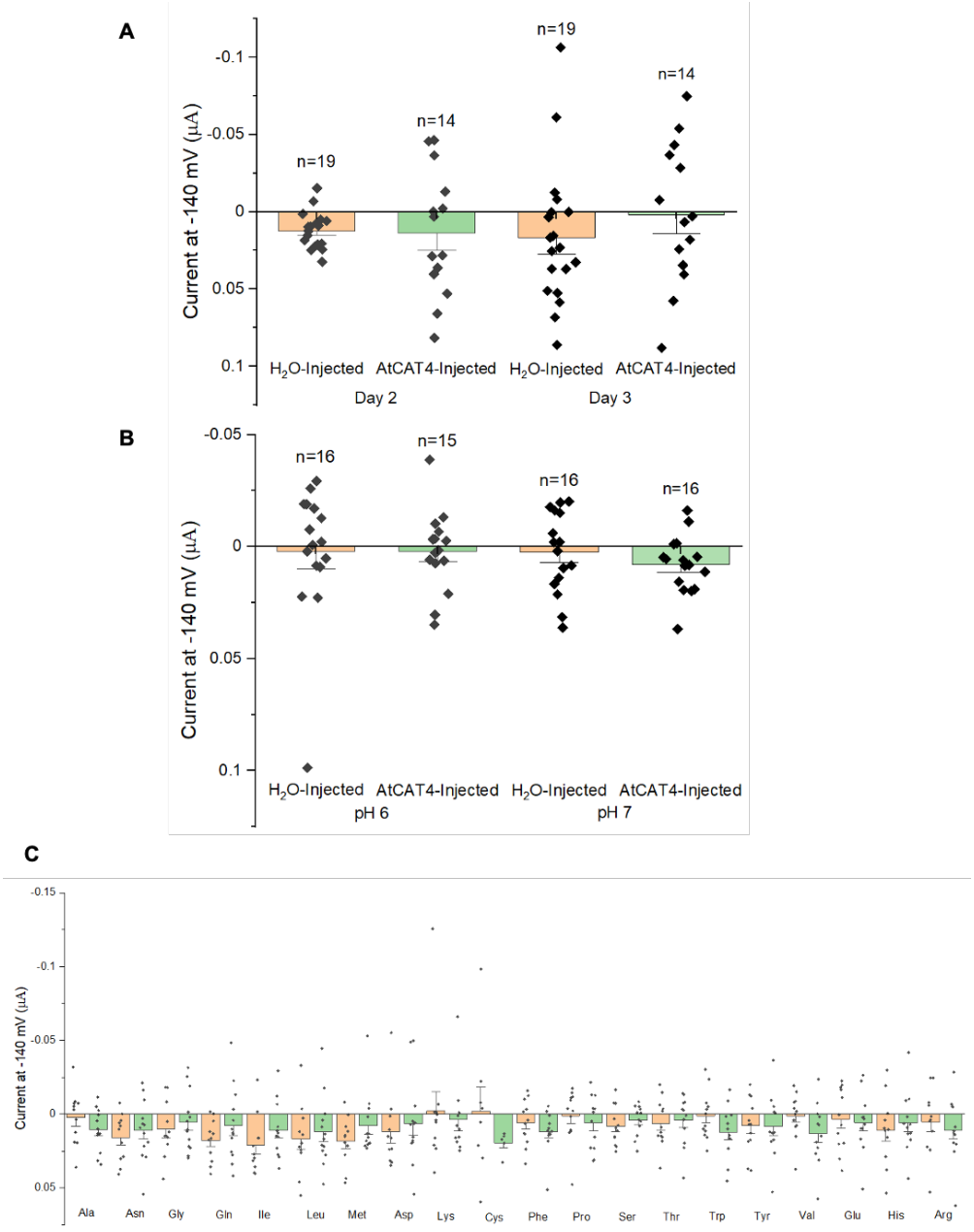


Figure 0.13 TEVC measurements on *At*CAT4-expressing and water-injected (negative control) *X.* oocytes. (A) Currents measured at -140 mV with 10 mM L-Lysine at pH 6.0, after 2 and 3 days. (B) Currents measured with 10 mM L-histidine at pH 6.0 and 7.0 at –140 mV. (C) Proteinogenic amino acid screening at –140 mV using –160 to +40 mV at 10 mV steps (charged: 5 mM, L-Cys: 1 mM, others: 10 mM). Biological repeats, *N*=2.

effect (**Figure 0.13-C**). Gly, Gln, Ile, Met, Trp and Val resulted in a slightly more positive or negative current in the *At*CAT4-injected oocytes compared to the

water-injected control. This results from the difference in the control measurement between the two groups (water vs *At*CAT4), shifting the *At*CAT4-injected group for the first 10 set of amino acids below the water-injected mean and the last 10 above the water-injected mean (different control for first and last 10 amino acids). Different control currents in water and cRNA-injected oocytes are expected and they are noticeable here because no significant currents are present. Leaky oocytes were excluded from the analysis for the charged amino acids. These oocytes were identified based on the presence of currents at the –40 mV resting potential.

Chapter 6 Supplementary Information

Benchmarking Cholesterol Binding

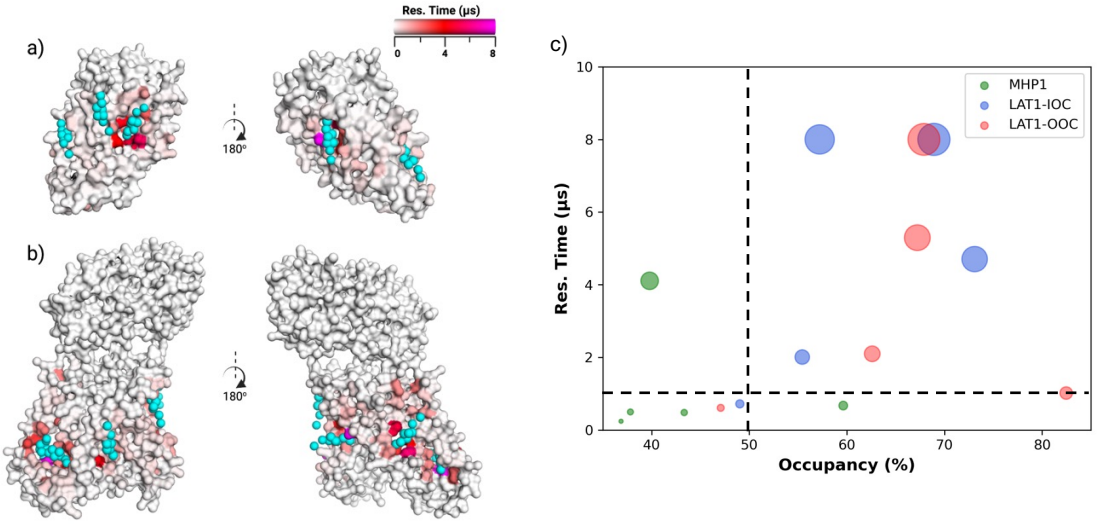


Figure 0.14 CG representations of (a) bacterial nucleobase-cation symporter, Mhp1, and (b) inward-open LAT1 (LAT1-IOC) with bound cholesterols, shown as cyan spheres, indicative of a top-ranking binding site location. CG residues are coloured according to residence time. (c) Correlation plot of the top 5 binding sites ranked by occupancy and residence time in each of the proteins: LAT1-IOC (blue), LAT1-OOC (red), and Mhp1 (green). Each point represents a cholesterol binding site with the point's size proportional to the correlation between binding site occupancy and RT. Dashed lines indicate the 50% occupancy and 2 μs RT cutoffs used to distinguish nonspecific/annular interactions from specific/strong binding interactions. Figure produced in collaboration with Biochemistry Part-II student, Daniel Standing.

HAT Binding Site Residue Composition

The binding sites calculated using PyLipID python library from 100 μs CG MD simulations of xCT, LAT1, and $b^{0,+}$AT were decomposed to their individual residues. The residues were clustered in classes, according to their side-chain chemistry, as (i) hydrophobic, (ii) polar, (iii) basic i.e., positive-charged, (iv) acidic i.e., negative-charged. Due to their non-polar character, Gly, Met and Cys were considered hydrophobic for this analysis.

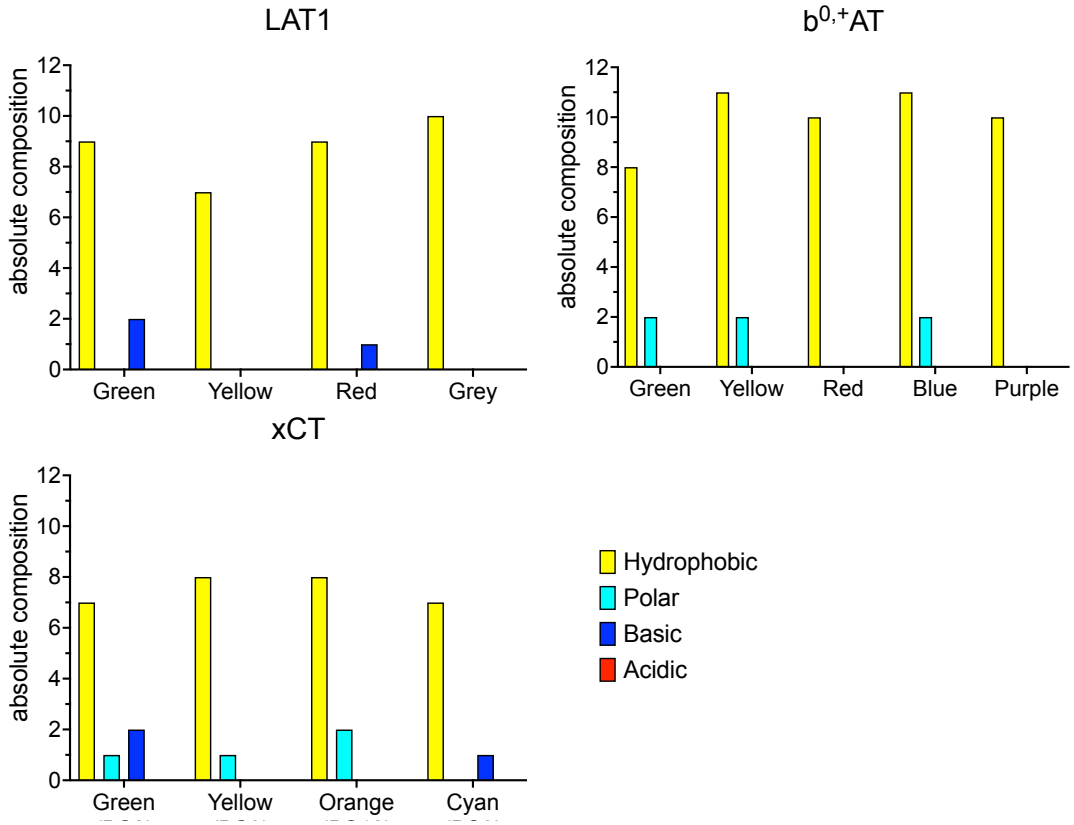


Figure 0.15 Residue decomposition analysis of primary binding sites identified in HATs by CG MD simulations. The presented data is from 100 μs simulations.

Supplementary Information

APPENDIX B

Molecular Dynamics Theory and Methods

This section provides a high-level overview of molecular dynamics (MD) simulation theory that is relevant to the computational methods used within this thesis, including atomistic and coarse-grained MD and enhanced sampling algorithms. Detailed review of MD and free energy theory is beyond the scope of this thesis and is available from multiple excellent textbooks (Frenkel and Smit, 1996; Leach, 2001; Porohille and Chipot, 2008; Mann, 2018)

Classical Molecular Dynamics

The most successful theory to describe the microscopic world is undoubtedly Quantum Mechanics (QM) and its time-evolution counterpart, Quantum Dynamics. The time-evolution of atoms and molecules is important in protein biochemistry since information extracted from structural biology methods offer a static, ensemble description of the underlying dynamics. QM offers detailed description of particle interactions by assigning the system a wavefunction, Ψ, which is described by the time-independent Schrödinger equation:

$$\widehat{H}\,|\Psi\rangle = E\,|\Psi\rangle$$

, where H is the Hamiltonian operator, $|\Psi\rangle$ is the system wavefunction and E is

the energy associated with the wavefunction state. For an arbitrary system composed of n electrons with N nuclei with Z_I charges the Hamiltonian can be written as:

$$\widehat{H}= \left[\frac{-\hbar^2}{2\mu}\nabla^2+V(\mathbf{r})\right]= \frac{-\hbar^2}{2\mu}\sum_{i=1}^{n}\nabla_{\mathrm{i}}^2-\sum_{i=1}^{n}\sum_{i=1}^{N}\frac{(Z_I e)e}{4\pi\epsilon_0\mathbf{r}_{iI}}+\sum_{i<j}^{n}\frac{e^2}{4\pi\epsilon_0\mathbf{r}_{iI}}$$

, where $\hbar$ is the reduced Planck's constant, μ is the reduced mass of the electron, ∇^2 is the Laplace operator, e is the charge of the electron, $\mathbf{r}_{iI}$ is the separation of the ith electron and I^{th} nucleus, ϵ_0 is the permeability of free space, and $\mathbf{r}_{ij}$ the separation of electrons i and j.

Although accurate, there is an associated computational cost with calculating the Schrödinger equation which limits its usability in cases where systems are composed of more than a few hundred atoms. However biological systems of a membrane protein in a solvated membrane typically consist of thousands of particles necessitating the use of an equation describing the system that is quickly solvable by modern computers such that many iterations can be achieved to simulate as large times of a given system as possible.

The standard method for systems of biological relevance is a classical method based on the Born-Oppenheimer approximation, which says that the position of an atom can be approximated by the position of its nucleus since electronic configuration can instantaneously adjust to the nuclear position. Classical molecular dynamics of a system uses Newton's second law of motion,

$$m_i\mathbf{a}_i=\mathbf{F}_i(t)=-\nabla_i U(\mathbf{r}_1,\ldots,\mathbf{r}_{\mathrm{i}},\ldots,\mathbf{r}_N)=\sum_{i\neq j}^{N}\mathbf{F}_j$$

, where $\mathbf{F}_i$ is the force exerted on an atom i as a resultant of the forces exerted

to atom i by all other atoms of the system j containing N atoms. Additionally, $\mathbf{a}_i$ is the atom i acceleration with mass m_i. A major limitation of the classical approach is that it ignores electronic effects such as polarisation. To address this an "effective" force field is developed for the atoms/particles that captures the QM effects by fitting parameters describing the classical potential energy function U.

For each time step in the simulation the forces must be calculated for the system atoms as the derivative of the potential energy. This is achieved by algorithms using finite difference methods, known as integrators, to integrate Newton's equation and generate updated particle coordinates every time step, Δt, subject to the potential energy U (or equivalently, the forces $\mathbf{F}$). An assumption of constant forces during Δt is made, which necessitates the choice of Δt to be as small as possible without loss of accuracy, commonly selected to match the vibrational frequency of lightweight bonds with hydrogen (approx. 1 fs). A common discretisation of the equations of motion used in this thesis is the leap-frog integrator, which calculates to position $\mathbf{r}_i$ and velocity $\dot{\mathbf{r}}_I$ of a particle in interleaved time points, according to,

$$\dot{\mathbf{r}}_i\left(t+\tfrac{\Delta t}{2}\right)=\dot{\mathbf{r}}_i\left(t-\tfrac{\Delta t}{2}\right)+\Delta t\,\ddot{\mathbf{r}}_i(t) \text{ and } \mathbf{r}_i(t+\Delta t)=\mathbf{r}_i(t)+\Delta t\,\dot{\mathbf{r}}_i\left(t+\tfrac{\Delta t}{2}\right).$$

The leap-frog integrator is time-reversible and symplectic – i.e. it preserves the system's total energy (Berendsen and Van Gunsteren, 1986). To increase the time step and speed up simulation, bonds with hydrogens are typically constrained to their equilibrium value allowing for a 2 fs time step, for example via the LINCS constraint solver (Hess *et al.*, 1997) – used in this thesis. Other methods such as SHAKE (Ryckaert, Ciccotti and Berendsen, 1977) – using Lagrange multipliers are also common, albeit slower.

The Force Field

In the classical approach instead of a QM $V(\mathbf{r})$ describing the systems' potential energy, a given MD force field refers to the functional of the system's potential energy $U(\mathbf{r})$, as the sum of two interactions: (a) bonded interactions between covalently bound atoms and (b) non-bonded interactions between atoms. The bonded interactions are commonly defined as the sum of three terms, namely, bond stretching, angle bending, torsions bending and improper torsion penalties.

The bonds between atoms at a distance r_{ij} are modelled as a classic harmonic oscillator of reference length r_{ij}^0, following Hooke's law, which dictates that a spring has potential energy U, quadratically related to its deviation from its equilibrium length, $\left(r_{ij} - r_{ij}^0\right)^2$, times a force constant, k_{ij}, related to the spring's resistance to deviation.

$$U_{bonds}(r_{ij}) = \frac{1}{2} k_{ij}\left(r_{ij} - r_{ij}^0\right)^2$$

Similarly, the angular bending between three atoms of angle θ_{ijk} follows the harmonic oscillator functional form around and equilibrium value θ_{ijk}^0 for the angle.

$$U_{angles}(\theta_{ijk}) = \frac{1}{2} k_{ijk}\left(\theta_{ijk} - \theta_{ijk}^0\right)^2$$

A different functional is used for dihedrals, a variant on the Ryckaert-Bellemans function, firstly used in the OPLS-AA and AMBER families of force fields, is and is described by equation 2.8.

$$U_{dihedrals}(\varphi_{ijkl}) = U_0 + \frac{1}{2}\sum_{n=1}^{4} U_n\left(1 + \cos\left(n\varphi_{ijkl}\right)\right)$$

Improper dihedrals are used to restrict off-plane motion of atoms belonging to an aromatic system or inversion of chirality.

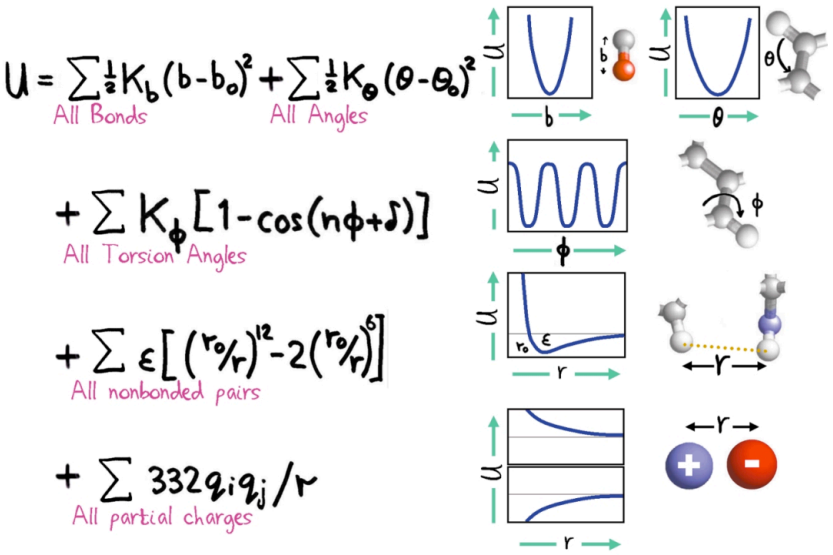


Figure 0.1 Schematic representation of a "modern" force field functional components. The total potential energy of a system is the sum of simple terms allowing for bond stretching, bond angle bending, bond twisting, van der Waals interactions and electrostatics. Figure reproduced from (Levitt, 2001).

The non-bonded interactions consist of coulombic interactions and soft repulsive and attractive interactions modelled by the Lennard-Jones potential.

$$U_{nonbonded}(\mathbf{r}_{ij}) = \sum_{\substack{non-bonded \\ atom\ pairs\ ij}} [U_{coulomb}(\mathbf{r}_{ij}) + U_{LJ}(\mathbf{r}_{ij})]$$

$$= \sum_{\substack{non-bonded \\ atom\ pairs\ ij}} \frac{q_i q_j}{4\pi\epsilon_0 \mathbf{r}_{ij}} + 4\epsilon_{ij}\left(\left[\frac{\sigma_{ij}}{\mathbf{r}_{ij}}\right]^{12} - \left[\frac{\sigma_{ij}^6}{\mathbf{r}_{ij}}\right]\right)$$

, where $\mathbf{r}_{ij}$ is the distance between particles i and j with point charges q_i and q_j, ϵ_0 is the permittivity of free space ϵ_{ij} is the well-depth and σ_{ij} is the distance at which the Lennard Jones potential vanishes. The latter two are calculated with "combination rules" or in the case of CHARMM and AMBER force fields, Lorentz-Berthelot rules, $\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj})$ and $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$.

Periodic Boundary Conditions & Long-Range Interactions

The non-bonded terms are more expensive to calculate than the bonded terms as they concern all pairs between N atoms, and as such scale N^2. For the Lennard-Jones potential a distance cutoff is applied as it scales to zero quickly as a function of $\mathbf{r}_{ij}$, while for electrostatics (which scale slowly as r^{-1}) the Particle Mesh Ewald (PME) summation method is used to split the interaction into two parts, the real-space component (described above) and a Fourier-space component. The latter "smears" particle charges on a predetermined mesh at longer ranges and using a Fast Fourier Transformation to obtain a potential that is applied to a specific particle associated with that particular point on the mesh.

For the PME Fourier space term to work, the simulated system must be encapsulated within a cell of x, y, z dimensions, the unit cell, which has a neutral net charge, such that it may form an infinite lattice by periodically staggering the system in x, y, and z dimensions. Periodic boundaries also alleviate the problem of simulating a molecular system with finite boundaries with a vacuum.

Thermodynamic Ensembles

Simulation with the above equations of motion leads to a system with constant particle number (N), internal energy (E) and volume (V), yielding the microcanonical ensemble (NVE) of system (coordinate, momentum) configurations ($\mathbf{q}$,$\mathbf{p}$), i.e., a collection of microstates whose Hamiltonian is constrained to the constant energy hypersurface of phase space.

However, as one typically wishes to simulate a system as realistically close to experiment as possible (i.e., using simulation as a "computational microscope"),

one must consider that a system may exchange energy with the environment but must do so under conditions of constant temperature (T) and pressure (P). Therefore, with a constant particle number, the desired thermodynamics ensemble is the isothermal-isobaric (NPT). In this ensemble the microstates are on the constant enthalpy hypersurface since the conserved quantity is $\mathcal{H}(\mathrm{x})+PV=H$. To achieve this a thermostat and barostat coupling is implemented by mathematical adjustment of coordinates and velocities. Various thermostats and barostats formalisms exist. In this thesis the Bussi, also known velocity-rescaling, and Nosé-Hoover thermostat, along with the Parrinello-Rahman barostat were employed for production simulations and are presented.

The Bussi thermostat scales particle velocities by a factor $\dot{\mathbf{r}}_i^{'}=\dot{\mathbf{r}}_i\left(\frac{\overline{K}}{K}\right)^{\frac{1}{2}}$ where K is the instantaneous kinetic energy and $\overline{K}=\frac{N_f k_\mathrm{B} T}{2}$ is the target kinetic energy of particles with N_f degrees of freedom drawn from the kinetic energy equilibrium distribution. The target is achieved by a stochastic differential equation with two terms, (a) one that exponentially relaxes K to its target value $\overline{K}$ with a coupling constant τ_T, and (b) one that ensures sampling from the correct NPT ensemble with a Wiener process (the stochastic term) dW. The thermostat is given by,

$$\mathrm{d}K=\left(\overline{K}-K\right)\frac{dt}{\tau_T}+2\sqrt{\frac{K\overline{K}}{N_f}}\frac{dW}{\sqrt{\tau_T}}$$

Another formalisation is that of the Nosé-Hoover thermostat, once a system has been equilibrated. It relies on extending the system's Hamiltonian with a fictitious particle, with an associated "mass" Q and velocity s, that acts as a coupling heat bath (Evans and Holian, 1985). Q has units energy×time^2, and is the parameter that determines the time scale on which the daemon acts. The

Nosé Hamiltonian takes the form:

$$\mathcal{H}(\mathbf{q}, \mathbf{p}, \mathrm{s}, \mathbf{p}_\mathrm{s}) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{q}) + \frac{\mathbf{p}_\mathrm{s}^2}{2Q} + gkT \ln s$$

To sample from NPT ensemble, a barostat is also needed. The Parrinello-Rahman barostatic algorithm (Parrinello and Rahman, 1981) is formulated in a similar fashion to that of the Nosé-Hoover thermostat to couple the system's pressure to a reference value by implementation of an extended Hamiltonian, which includes a volumetric contribution.

Coarse-Grain Modelling & the MARTINI Force Field

We have seen the basic equations governing a modern atomistic molecular dynamics simulation. Although parameterised to reproduce QM calculations, the pairwise additive force field terms described above necessitate the use of time steps in the order of 1-5 fs. While classical atomistic force fields lend themselves to the study of protein dynamics, biological phenomena such as protein-lipid interactions and multiple association/dissociation events routinely occur in the order of μs. One approach is to cluster groups of atoms into sites/beads, effectively making the system a coarser set of grains, instead of fine atoms. These CG sites are parameterised to capture "effective" interactions with their environment. This along with a reduction in the number of degrees of freedom (DoF) allows for a jump in spatiotemporal scales. However leads to reduction in chemical resolution (i.e. potentials, DoF) and a smoother energy landscape which on one hand allows for faster exploration of the landscape but one the other is escorted with reduced accuracy of the landscape due to loss of rugged features (Figure 0.2).

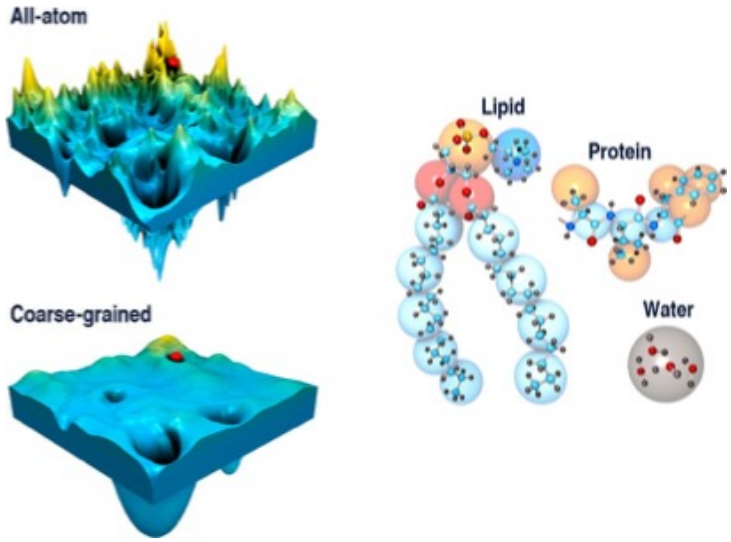


Figure 0.2 Schematic illustration of coarse-graining showing the comparison of all-atom versus coarse-grained energy landscape. In CG the energy landscape is smoothened/flattened enabling its exploration in search for the global minima, while avoiding traps in the local minima. The smoothing is achieved by grouping of atoms into CG sites. Figure adapted from (Kmiecik et al., 2016).

Careful selection of the grouping of atoms into CG sites and their parameterisation is key so that the underlying physics of the system is respected. Several CG methods exist and are reviewed comprehensively elsewhere (Voth, 2008). The MARTINI 2 force field for biomolecules (Monticelli *et al.*, 2008; de Jong *et al.*, 2013; Marrink and Tieleman, 2013; Melo, Ingólfsson and Marrink, 2015) is used in this thesis. In the MARTINI model there are four particle classes: polar (P), nonpolar (N), apolar (C), and charged (Q). For more complex chemical moieties, such as structures containing rings or small molecules, an S-type particle is used. Subtypes also exist denoting the H-bonding capabilities (d = donor, a = acceptor, da = both, 0 = none) or by a number indicating the degree of polarity (from 1 = low polarity to 5 = high polarity). There are 18 fundamental particles to accurately describe amino acid side chains and other molecules.

Connectivity and bond terms between particles are derived from a molecule's

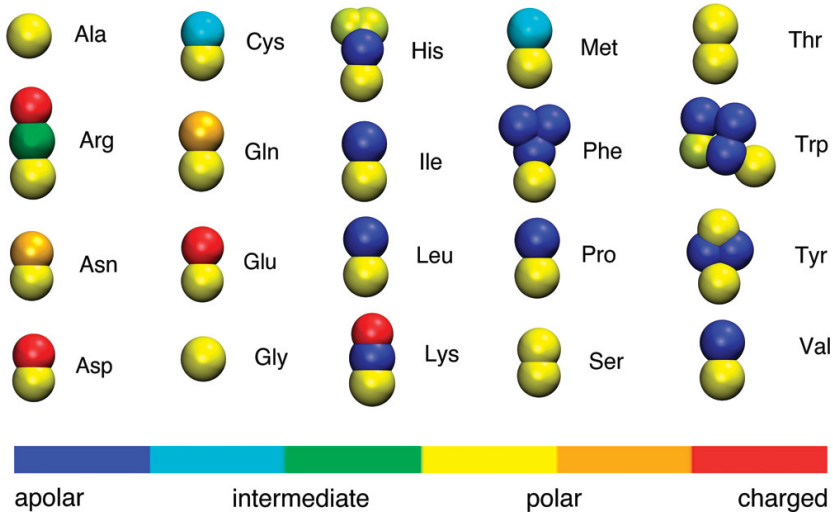


Figure 0.3 Representation of MARTINI 2 force field amino acid side chains as CG beads of different size and charge. Figure reproduced from (Monticelli et al., 2008)

atomistic representation while the conservation of protein conformation is assured by the means of an elastic network between residues while allowing internal dynamics. An example is the ElNeDyn network (Periole *et al.*, 2009), which applies harmonic potentials between C_α beads within a cut-off, R_c (0.8-1.0 nm) with a force constant, K_{spring} (500-1000 kJ mol^{-1} nm^{-2}). For non-bonded interactions, the 12-6 LJ potential functional and coulombic functional are used with the following modifications. The well depth is ε_{ij} = 2.0-5.6 kJ mol^{-1} for interactions between polar and apolar groups or strongly polar groups, respectively. All beads have a size σ = 0.47 nm, except S-type beads which have a reduced value 0.43 nm and ε'_{ij}=0.75 ε_{ij}. Charged interactions are explicitly calculated up to r_{cut} = 1.1 (with dielectric constant, ε_r = 15) by a modified coulombic potential containing an additional reaction-field term (Tironi, Sperb, Smith and Van Gunsteren, 1995), which ensures $\mathbf{F}_{coul,rf} = \frac{\partial U_{coul,rf}}{\partial \mathbf{r}}$ is zero at r_{cut}. In MARTINI simulations the reaction-field dielectric ε_{rf} beyond r_{cut} is set to ∞ (De Jong et al., 2016).

$$U_{coul,rf} = f\frac{q_i q_j}{\varepsilon_r r_{ij}}\left[1 + \frac{\varepsilon_{rf} - \varepsilon_r}{2\varepsilon_{rf} + \varepsilon_r}\frac{r_{ij}^3}{r_c^3}\right] - f\frac{q_i q_j}{\varepsilon_r r_c}\frac{3\varepsilon_{rf}}{2\varepsilon_{rf} + \varepsilon_r}$$

The water model used in this thesis is the regular MARTINI water which groups four atomistic water molecules into one large water bead. Although this speeds up calculations, it reduces the complexity of interactions with the solvent since the water model does not bear charges and is therefore blind to electrostatics and polarisation effects. A polarisable MARTINI water model has been developed by adding $+q$ and $-q$ drude charges (Yesylevskyy *et al.*, 2010), and although useful, the computational overhead associated with it was not justified in this thesis where solute-solvent interactions were not of critical importance. Resolution change to atomistic simulations addresses water interactions.

Free Energies, Umbrella Sampling & Potential of Mean Force

The free energy change from state A to state B is an informative thermodynamic quantity that has been used to describe protein-ligand interactions between a bound and unbound state. However, the free energy is a function of $3N$ coordinates, where N is the number of simulated particles and therefore knowledge of the complete function describing a free energy change process would include insignificant dimensions. The potential of mean force (PMF) is defined as a projection of the free energy landscape on a chosen reaction coordinate of interest. As it is derived from the free energy, the PMF, $W(\chi)$, will be defined from the average distribution function $\varrho(\chi)$ using the reaction coordinate and a chosen reference point along that reaction coordinate (χ^*).

$$\mathrm{W}(\chi)=\mathrm{W}(\chi^*)-k_\mathrm{B}T\ln\left[\frac{\langle\varrho(\chi)\rangle}{\langle\varrho(\chi^*)\rangle}\right]$$

MD simulations as we have described them above progress freely in phase

space. Given an ergodic trajectory generated by a Hamiltonian $\mathcal{H}(x)$, the ensemble average of a property, such as the free energy change, can be replaced by the time average of the property over the trajectory. Therefore, in principle, free energy changes from state A to state B can be obtained from an MD trajectory by simply calculating the ratio of probability of state A, $\varrho_A(\chi)$, over probability of state B, $\varrho_B(\chi)$. However unbiased MD tends to stay in the vicinity of microstates residing near local minima and may need to overcome high energy barriers to sample the available configurational space along the reaction coordinate of interest. Additionally, routine simulation times are not enough to allow for spontaneous jumps of barriers or to allow the exploration of all the configurations either. For these reasons the ratio of probabilities, $\varrho_A(\chi)/\varrho_B(\chi)$, does not converge and the free energy estimates from typical unbiased simulations are not accurate.

Enhanced sampling methods, such as umbrella sampling, metadynamics, steered MD, accelerated MD, adaptive biasing force, alchemical transformations permit efficient exploration of configurational space along the desired reaction coordinate by pushing the system towards the desired change (Porohille and Chipot, 2008). In this thesis umbrella sampling is used in conjunction with spatial stratification of the reaction coordinate to compute the PMF. The reaction coordinate is stratified in N_w windows, each with a reaction coordinate value χ_w. An additional harmonic potential, $\omega_w(\chi)$, is added to the system's potential energy to constrain the reaction coordinate within the window with a force constant F_c, $\omega_i(\chi) = 1/2\ F_c(\chi\text{-}\chi_i)^2$. Simulation of each window then produces a local, biased probability distribution. The distributions are recombined into one cumulative unbiased probability distribution along the reaction coordinate which is produced by removing the biasing work. The weighted histogram analysis method

(WHAM) was used to assemble the different window distributions to produce the unbiased profile (Kumar *et al.*, 1992; Hub, de Groot and van der Spoel, 2010).

Approximate Free Energies by the End-Point MMPBSA Method

MMPBSA approximately calculates the absolute binding free energy (ABFE) of host-guest systems. MMPBSA is an intuitive approach that combines molecular mechanics and continuum models to calculate end-point binding free energies. The polar part of the solvation free energy was calculated using Poisson-Boltzmann (PB) equations. In these calculations, a dielectric constant of $\varepsilon_{\text{solute}} = 4$ was assigned to the binding area owing to the highly charged residues present, and $\varepsilon_{\text{solute}} = 80$ for water. A variable dielectric was assigned along the membrane normal, as demonstrated here (Greene *et al.*, 2019). The binding free energy for each complex was calculated using equation (1)

$$\Delta G_{\text{eff}} = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - \text{T}\Delta S_{\text{MM}} \quad (1)$$

In equation (1) ΔG_{eff} is the binding free energy for each calculated complex. The entropic contribution, $\Delta S_{\text{MM,}}$ to binding free energy was calculated with the interaction-entropy method, which has sound roots in statistical mechanics, shown here (Duan, Liu and Zhang, 2016). Explicit entropic contributions to protein-ligand binding free energies have been shown to improve correlation with experimental measurements (Aldeghi *et al.*, 2017). ΔE_{MM} defines the interaction energy between the protein and the ligand as calculated by molecular mechanics in the gas phase. ΔG_{sol} is the desolvation free energy for transferring the ligand from water to the binding area calculated, using the PBSA model. The terms for each complex ΔE_{MM} and ΔG_{sol} are calculated using equations (2) and (3)

$$\Delta E_{\text{MM}} = \Delta E_{\text{elec}} + \Delta E_{\text{vdW}} \quad (2)$$

$$\Delta G_{\mathrm{sol}} = \Delta G_{\mathrm{P}} + \Delta G_{\mathrm{NP}} \quad (3)$$

In equation (2) ΔE_{elec} and ΔE_{vdW} are the electrostatic and the van der Waals interaction energies, respectively. In equation (3) ΔG_{P} is the electrostatic or polar contribution to the free energy of solvation and the term ΔG_{NP} is the non-polar or hydrophobic contribution to the solvation free energy.