2.24Å and 2.13Å structures of divalent anion sodium symporter (DASS) member LaINDY from *L. acidophilus* reveal highly-coordinated structural waters and a lipid at the dimer interface.

David Robert Benjamin Speedman

Centre for Medicines Discovery, Nuffield Department of Medicine

& Exeter College

A thesis submitted for the partial fulfilment of the degree of Doctor of Philosophy

October 2024

I. Abstract

The divalent anion sodium symporter (DASS) family are low affinity dicarboxylate phosphate transporters found across all domains of life. The DASS family contains two distinct clades: cotransporters, which utilise a sodium gradient to drive transport across the plasma membrane; and exchangers which utilise dicaboxylate antiport. The cotransporter clade is most well characterised, and contains the human SLC13 family which are involved in the disease pathologies of diabetes, cancer and epilepsy. In contrast, the exchanger clade DASS members are exclusively found exclusively in plants and prokaryotes. Better understanding of the DASS-E clade may have implications for bacterial infections, bioproduction and agriculture. Through the model DASS exchanger LaINDY from *Lactobacillus acidophilus* this project aims to characterise the substrate binding, and the transport mechanism of the DASS exchangers.

Here I present 2.13Å and 2.24Å structures Cryo-EM structures of LaINDY, with the improved resolution revealing highly coordinated waters within the elevator domain that I hypothesise may have functional roles, and a phosphatidyl glycerol molecule modelled at the dimer interface. While the Cryo-EM maps were determined in the presence of ligands, no density was observed in the LaINDY binding site. This is consistent with *in vitro* binding experiments by tryptophan fluorescence quenching, which revealed LaINDYs low affinity for substrate at pH7.5. Ultimately my results reveal further complexity LaINDYs, transport mechanism, and substrate binding which may apply to the DASS exchanger clade as a whole.

II. Acknowledgements

Contributions to Experimental Procedures

All work in this thesis are my own, however I would specifically like to mention:

- Dr Alvaro Vinals Guitart for: guidance while troubleshooting LaINDY expression.
- Dr Mike Fairhead for: guidance while troubleshooting LaINDY expression and purification
- Dr Gamma Chi for: training during the 5mM malate Cryo-EM sample preparation, and throughout the entire data-processing pipeline.
- Dr Ashley Pike for: microscope operation during the collection of both 5mM malate and α-ketoglutarate datasets, assistance during the entire data-processing pipeline, and execution of micrograph dose-fractionation from the 5mM malate dataset which ended up not being used.
- Dr David Staunton for: training and guidance on executing the tryptophan fluorescence quenching assay.
- Dr Matthew Hankins for: training and guidance while optimizing the thin layer chromatography experiment.

Non-experimental Contributions

I would like to give special thanks to:

Firstly, my past supervisor Professor Liz Carpenter for the hiring of me as a research assistant, and her support and encouragement when taking me on as a DPhil student.

Secondly, Dr Karin Rodström for teaching me basically everything I know about wet lab work and being the best postdoc to work for.

Finally, my primary supervisor Dr David Sauer, your enthusiasm and guidance throughout the project has been invaluable, a special shout out to insane rate at which you gave written feedback on chapters.

My supervisors both past and present: Dr David Sauer, Professor Liz Carpenter, Professor Wyatt Yue, and Professor Alex Bullock for their continued support, feedback and empathy when I was at my lowest.

Professors: Benedikt Kessler and Frank Von Delft for their guidance in focusing my work to achieve more realistic goals.

Professors: Rob Gilbert, Esther Becker, Ervin Fodor and Dr Lucy Taylor for their help and support with administration.

The IMP group members both past and present, but in particular: Dr Ashley Pike, Dr Gamma Chi, Dr Roslin Adamson, Dr Tomas Pascoa, Dr Matthew Hankins, Dr Jesper Hansen, Dr Mingda Ye, Dr Huanyu Li, Dr Michael Miller, Dr Sagar Raturi, Dr Viijay Kumar, Dr Chris Fage, Dr Didi He, Dr Adam Evans and Dr Amy Chu.

The Structural Genomics Consortium (SGC)/Centre for Medicines Discovery (CMD) staff both past and present, but in particular: Dr Ellie Williams, Dr Alvaro Vinals Guitart, Dr Mike Fairhead, Dr Mel Sweeney, Dr Will Bradshaw, Leela Shrestha, Dong Wang, Kory Wang and all other colleges who have provided me with training, assistance and guidance for the best part of the last decade. The people really made this place an encouraging and joyful place to work, and despite the prolonged difficulties of the restructuring and relocations it has remained that way.

Outside of the university I would like to thank family but in particular my parents: Mum & Al; and Dad & Kay for their support, both financially and emotionally, but mainly for their patience with me during the most difficult years. My very-extended family, for all the joy your bring to my life. In particular my sister Anna and the Price-cousins, it is a shame we will all perpetually be out-shone by our partners.

My best friends: Kirsten & Tom, Larry, Barry & Annie, Nat, Rosie, Ciara, Rebecca, Katy & James. This has been without hyperbole the most difficult five years of my life, and your continued effort with me throughout, despite at times it being both poorly received and unreturned has genuinely been one of the most important factors in me getting this far. Thank you.

Most importantly my partner Annie, your unwavering support throughout this journey has been astounding, and I really cannot thank you enough for all the slack you've taken up to help me achieve this goal. You have made me a better person.

Finally, the many artists that have sound-tracked my life. Not just propelling me through the long days in the lab, or while grinding my fingers into the keyboard, but also providing the motivation to finish in the final days: Avalon Emerson, Joy Orbison, Skee Mask, Polygonia, Mia Koden, Overmono, Surusinghe, Charli xcx, Taylor Swift and London Grammar.

III. For Caitlin

You absolutely would never have read this even if you could, but if we’re being mathematical about it the 7 people who will might as well be a rounding error. For your sake I hope there is something else out there, and that its better than here, but your decision was still really stupid. I wish I did more, and it has taken a long time, but can accept now that it probably would not have mattered.

I love you.

IV. General Disruption Statement

A month before a I started my DPhil my stepsister Caitlin ended her life. It took a while for the ramifications of this event to properly manifest, but the ensuing depression severely impacted my ability to work over the course of my studies, resulting in excess of a year in which I was incapable of working.

Additionally, 5 months into my DPhil the UK went into national lockdown due to the COVID-19 pandemic in March of 2020, and all non-COVID related scientific work was halted for a significant period of time. Project work was able to return under a “2 weeks on, 1 week off” reduced shift schedule from to reduce the spread of infection (28.07.2020-27.08.2021).

During this time the Oxford Structural Genomics Consortium that my lab had been a part of restructured into the Centre for Medicines Discovery, the immediate result of that being my initial supervisor Professor Liz Carpenter leaving the organisation, and as a combination of both that and the pandemic the company sponsoring my DPhil work on ABCA7: Eisai Co., Ltd. pulled out of the sponsorship. Liz’s interim replacement Professor Wyatt Yue also left within a year of the restructuring to set up his own lab within the University of Newcastle. It was not until Dr David Sauer took over as group leader of the Integral Membrane Protein group in the summer of 2021 that the group had a full-time principal investigator and my work on the LaINDY project began. The final consequence of the restructuring were two lab relocations, the first from the Old Road Campus Research Building (ORCRB) to the Dorothy Crowfoot Hodgkin Building (DCHB) from 23.08.2021-15.09.2021 and the second from the DCHB to the Nuffield Department of Medicine Research Building (NDMRB) from 05.03.2023-06.04.2023. Both relocations dramatically caused disruptions to my DPhil project due to the amount of time packing and unpacking, a -80°C freezer failure that thawed all samples, ultracentrifuge install failure, water supply loss, and as of 21.08.2024 access to our cold room and all purification systems within.

From 23.05.2023 onwards I have been seeing a therapist weekly, and through hers and my supervisors David’s support, along with a 12-month extension, I have completed this study.

V. Table of Contents

I. Abstract… 1
II. Acknowledgements 2
III. For Caitlin 5
IV. General Disruption Statement 6
V. Table of Contents 8
VI. List of Figures 12
VII. List of Tables 14
VIII. Abbreviations 15
1. Introduction 18
1.1 Dicarboxylates as Important Biomolecules 18
1.2 The DASS Family of Transport Proteins 19
1.3 DASS Cotransporter VcINDY as the Prototype DASS Member 21
1.4 The Elevator Domain and the Canonical SNT Motif 22
1.5 DASS Exchangers Substitute the Sodium Sites with Basic Residues 24
1.6 Transport Cycles of the DASS Family 25
1.6.1 ScPho90 Structure Captures the Transport Cycle in Multiple States 26
1.7 Human DASS-C Structures 27
1.7.1 NaCT and Competitive DASS Inhibition 27
1.7.2 NaS1, NaDC1 and Allosteric DASS Inhibition 28
1.8 The DASS-E Clade is Poorly Studied 29
1.8.1 DASS-Es of the Plantae Chloroplasts and the Two-Translocator Model 30
1.8.2 *E. coli* DASS-Es 34
1.8.3 DASS-Es in Industrial Prokaryotes 38
1.8.4 Chlamydial DASS-Es 39
1.8.5 LaINDY 40
1.9 Membrane Protein Environments 41
1.9.1 Lipids and Membrane Protein Regulation 41
1.9.1.1 Lipids and the DASS Family 41
1.9.1.4 Phosphatidyl Glycerol Stabilisation and Allosteric Regulation of AmtB 45
1.9.1.5 Interfacial Lipids Stabilising Multimeric Proteins 46

1.9.2 Ordered Water Molecules Within Membrane Protein Structures 47
1.10 Aims 55
2. Materials and Methods 56
2.1 Constructs and Vectors 56
2.2 *E. coli* Strains 56
2.3 Chemically Competent Cell Generation 57
2.4 LaINDY Expression 57
2.4.1 Heat-shock Transformation 57
2.4.2 Plasmid Amplification 58
2.4.3 Overexpression of LaINDY 58
2.4.4 Overexpression Test-Purification 59
2.5 LaINDY Purification 60
2.5.1 Membrane Isolation 60
2.5.2 IMAC 60
2.5.3 SEC 61
2.6 SDS-PAGE 62
2.7 X-ray Crystallography 62
2.7.1 Sample Preparation 62
2.7.2 Crystal Screening and X-ray Data Collection 63
2.8 Cryo-EM 63
2.8.1 Sample Preparation 63
2.8.2 Grid Screening and Data Collection 63
2.8.3 Data Processing and Model Building 65
2.9 Biophysical Assays 70
2.9.1 NanoDSF Thermal Shift Assay 70
2.9.2 Tryptophan Fluorescence Quenching Assay 71
2.10 TLC 71
2.10.1 Lipid Extraction 71
2.10.1 Lipid Separation 72
2.11 Sequence Alignments 72
3. Improving the Yield of LaINDY for Characterization Experiments 73
3.1 Aims 73
3.2 Expression 73

3.2.1 *E. coli* Strain Swap from LOBSTR-RIL to R3-pRARE2 73
3.2.2 Auto Induction Media: ZYM-5052 to Formedium AIM TB Base 76
3.3 Purification 77
3.3.1 Solving the Precipitation Issue Post TEV Addition 77
3.3.2 IMAC Bind 78
3.4 Discussion 79
4. High Resolution Structures of LaINDY with Poor Substrate Occupancy 82
4.1 Aims 82
4.2 Reported Crystal Reproduction 82
4.2.1 Optimization Screen Design 82
4.2.2 Sitting-Drop Vapor Diffusion Crystallisation 83
4.2.3 Crystal Screening at Diamond Light Source 84
4.2.4 X-ray to Cryo-EM 85
4.3 Revised Cryo-EM Purification 85
4.3.1 2.24Å Map of LaINDY with 5mM α-ketoglutarate 85
4.3.2 LaINDY in the Outward Open State 92
4.3.3 2.13Å Map of LaINDY with 5mM Malate 93
4.4 No Substrate in the Elevator Domain in Either Dataset 96
4.4.1 Determining Whether the Membrane Mimetic Impedes Substrate Binding 99
4.4.2 Additional Substrate Thermostability Screen 102
4.4.3 Establishing a Dissociation Constant to Explain Poor Substrate Density with Tryptophan Fluorescence Quenching. 105
4.5 Sample Preparation in 75mM Disodium Succinate 111
4.6 Discussion 112
4.6.1 LaINDY Reconstituted in PMAL-C8 Resolution may be Pushed Higher 112
4.6.2 LaINDY Has Low Affinity for Succinate, Malate and α-ketoglutarate at pH7.5 113
4.6.3 Physiological Conditions May be the Key to Resolving Substrate Bound Structures 114
4.6.4 Capturing LaINDY in Multiple States May require Alternative Mimetics 117
5. Structural Waters and a Lipid Within an Improved LaINDY Model 118
5.1 Aims 118
5.2 Cα Comparison with Previous Best LaINDY Structure and Other DASS Members 118
5.2.1 Presented Malate and α-ketoglutarate Structures are Structurally Identical 120
5.3 Isoleucine 187 is a Valid Ramachandran Outlier 120

5.4 Peptide Flips Within the Model 122
5.4.1 Proline 257 Provides an Additional H-bond donor to Lysine 185 124
5.4.2 Scaffold Domain Beta Hairpin Double Flip 124
5.4.3 A Serine 397 Flip Removes Electronegative Carbonyl from Binding Site........................ 124
5.5 Novel Waters Within the LaINDY Protomer 125
5.5.1 Water 581 Interacts with the Arginine 159 – Glutamate 146 Salt Bridge.......................... 128
5.5.2 Water 659 and the “Aromatic Cage” Network.. 128
5.5.3 Conformationally Sensitive Water 649 sits in a break in Helix H4 131
5.5.4 Poorly Coordinated Water 662 in a Solvent Inaccessible Elevator Cleft 131
5.6 Acyl Chain Densities Surrounding the LaINDY Protomer... 132
5.6.1 Dimer Interface Lipids.. 134
5.6.2 Perpendicular Inserts Between Scaffold and Elevator... 137
5.7 Discussion.. 140
5.7.1 Highly Coordinated Waters Within the DASS Elevator 141
5.7.2 Waters Within the Binding Domain of the LaINDY Elevator 142
5.7.3 A PG lipid at the Dimer Interface 142
5.7.4 Understanding LaINDYs Interaction with Membrane Lipids may Require a More Native Membrane Environment 143
6. Concluding Remarks 145
6.1 Discussion.. 145
6.2 Future Work on the Whole DASS Subfamily.. 147
6.2.1 Observations on the sequence conservation in the DASS-E transporters 149
6.2.2 The Relevance of DASS-Es in the Pathology of Bacterial Infections............................... 150
7. References 152

VI. List of Figures

Figure 1.1 Metabolites of the TCA cycle and their respective structures. 19
Figure 1.2 Phylogenetic tree of all Published DASS members discussed in the text. 21
Figure 1.3 The DASS fold through the structure of prototype DASS-C: VcINDY. 23
Figure 1.4 LaINDY-C_o and VcINDY-C_i and their respective binding sites. 24
Figure 1.5 Contrasting the proposed transport cycles of DASS-Cs and DASS-Es. 26
Figure 1.6 The asymmetric structure of DASS-C ScPho90 and the novel Na3 site. 27
Figure 1.7 Competetive inhibition of NaCT by compound PF2. 28
Figure 1.8 Allosteric inhibition of NaDC1 by compound ACA. 29
Figure 1.9 DASS-E functionality in chloroplasts. 31
Figure 1.10 Sequence alignment of the canonical DASS SNT motifs. 36
Figure 1.11 Modelled lipids in DASS structures. 43
Figure 1.12 AmtB PG binding site. 46
Figure 1.13 LeuT-SSRI complex binding site waters. 49
Figure 1.14 Water molecules compensating for amine groups in SiaP mutants. 51
Figure 1.15 Water molecule hydrogen bonding networks throughout the bacteriorhodopsin photocycle. 53
Figure 1.16 Waters mediating dimer stabilisation in AdiC. 55
Figure 2.1 Simplified data processing pipeline for the 2.24Å α-ketoglutarate dataset. 65
Figure 2.2 Simplified data processing pipeline for the 2.13Å malate dataset. 66
Figure 2.3 NanoDSF data processing example. 70
Figure 3.1 Expression yields of LOBSTR-RIL and R3-pRARE2 BL21 E. coli transformed with LaINDY. 75
Figure 3.2 SDS-PAGE analysis highlighting precipitation of TEV. 78
Figure 3.3 SDS-PAGE analysis of IMAC binding techniques. 79
Figure 3.4 SDS-PAGE analysis of final LaINDY purification. 80
Figure 4.1 Crystallisation of NG purified LaINDY via sitting drop vapour diffusion. 83
Figure 4.2 5mM α-ketoglutarate grid, SEC trace and SDS-PAGE analysis. 87
Figure 4.3 5mM α-ketoglutarate grid, microscope session to dataset. 89
Figure 4.4 5mM malate purification to dataset. 94
Figure 4.5 Binding site of the presented structures, highlighting the solvent densities. 98
Figure 4.6 Establishing parameters for NanoDSF thermostability assay. 100
Figure 4.7 Thermostability of LaINDY reconstituted in DDM or PMAL-C8 in the presence of 10mM substrate. 101
Figure 4.8 Probing additional chemical moieties impact on dicarboxylate stabilisation of LaINDY. 103
Figure 4.9 Tryptophan fluorescence quenching of 2μM LaINDY with disodium succinate at 280nm. 106
Figure 4.10 Tryptophan fluorescence quenching of 2 μM DDM-LaINDY with disodium succinate. 109

Figure 4.11 Cryo-EM Sample of amphipol purified LaINDY with 75mM disodium succinate. 112
Figure 4.12 Lipids of the L. acidophilus lipidome...116
Figure 5.1 Reported LaINDY and DASS homolog structures superimposed with presented malate and α ketoglutarate models. 119
Figure 5.2 Examination of the LaINDY break in helix H5a containing isoleucine 187, and the respective helix in DASS-C VcINDY. 121
Figure 5.3 Overview of the peptide flips between presented LaINDY-malate and the LaINDY-X-ray structure. 123
Figure 5.4 The Serine 397 flip within the canonical SNT motif viewed from the binding site.. 125
Figure 5.5 Cartoon representation of the LaINDY malate model highlighting all modelled waters and acyl chains. 126
Figure 5.6 Overview of the interesting water densities within the LaINDY protomer. 127
Figure 5.7 Conservation of the “aromatic cage” within the DASS-E clade. 130
Figure 5.8 Modelled acyl chains compared with previous LaINDY structure. 132
Figure 5.9 Overview of the Acyl chain densities surrounding the LaINDY protomer. 133
Figure 5.10 Thin layer chromatography of the LaINDY-PMAL-C8 malate Cryo-EM sample.. 135
Figure 5.11 Electrostatic surface of the LaINDY dimer interface and a tentatively modelled DPPG. 137
Figure 5.12 Comparison of the perpendicular lipid inserts within the scaffold-elevator interfaces of LaINDY and ScPho90. 139
Figure 6.1 Sequence alignment of the SNT motifs of all discussed DASS members in the text. 148

VII. List of Tables

Table 2.1 Data collection and refinement statistics for LaINDY Cryo-EM datasets.................... 67
Table 3.1 Purification yields for all LaINDY purifications. 80
Table 4.1 Dilution factors of preliminary tryptophan fluorescence quenching assay in Figure 4.9. 107
Table 4.2 Dilution factors of tryptophan fluorescence quenching assay in Figure 4.10..............110
Table 5.1 Cα RSMD comparison of superimposed DASS structures..119

VIII. Abbreviations

Linnaean Taxonomy:

Aa	*A. aeolicus*	*Aquifex aeolicus*
As	*A. succinogenes*	*Actinobacillus succinogenes*
At	*A. thaliana*	*Arabidopsis thaliana*
Bl	*B. licheniformis*	*Bacillus licheniformis*
Ce	*C. elegans*	*Ca enorhabditis elegans*
Cg	*C. glutamicum*	*Corynebacterium glutamicum*
CtL2	*C. trachomatis*	*Chlamydia trachomatis serovar L2*
Cp	*C. pneumoniae*	*Chlamydia pneumoniae*
Dm	*D. Melanogaster*	*Drosophila melanogaster*
Ec	*E. coli*	*Escherichia coli*
Ef	*E. faecalis*	*Enterococcus faecalis*
Hs	*H. sapiens*	*Homo sapiens*
Hsa	*H. salinarum*	*Halobacterium. salinarum*
Kp	*K. pneumoniae*	*Klebsiella pneumoniae*
La	*L. acidophilus*	*Lactobacillus acidophilus*
Mt	*M. tuberculosis*	*Mycobacterium tuberculosis*
Ph	*P. horikoshii*	*Pyrococcus horikoshii*
Ps	*P. sativum*	*Pisum Sativum*
Sb	*S. bicolor*	*Sorghum bicolor*
Sc	*S. cerevisiae*	*Saccharomyces cerevisiae*
Sf	*S. frugiperda*	*Spodoptera frugiperda*
So	*S. oleracea*	*Spinacia oleracea*
Ss	*S. saprophyticus*	*Staphylococcus saprophyticus*
Vc	*V. cholerae*	*Vibrio cholerae*
Zm	*Z. mays*	*Zea mays*

General:

5052	0.5% glycerol, 0.05% glucose, 0.2% α-lactose
A$_{280}$	Absorbance at 280nm wavelength
ACA	N-(p-amylcinnamoyl) anthranilic acid
AIM	Auto induction media
APBS	Adaptive Poisson-Boltzman Solver
ArnA	Bifunctional polymyxin resistance protein ArnA
Bis-Tris	2,2-Bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol
C2	Axis with 180° rotational symmetry
Cα	Amino acid alpha carbon
CA	Cardiolipin
(win)COOT	Crystallographic object-oriented tool-kit for windows
COSMIC	Central Oxford structural molecular imaging centre
CHS	Cholesterol hemisuccinate tris salt
CPM	N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide

Cryo-EM	Cryogenic transmission electron microscopy
CryoSPARC	Cryo-EM single particle ab-initio reconstruction and classification
CTF	Contrast transfer function
CV	Column volume
DASS	Divalent anion sodium symporter
DDM	n-Dodecyl-Beta-maltoside
DNase	Deoxyribonuclease
DPPG	Dipalmitoylphosphatidylglycerol
DTT	Dithiothreitol
EAAT	Excitatory amino acid transporters
eLBOW	Electronic ligand builder and optimisation workbench
ENA	European nucleotide archive
ESI	Electrospray ionisation
FSC	Fourier shell correlation
FTMS	Fourier transform mass spectrometry
GDN	Glyco-diosgenin
Gui	Graphical user interface
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethane-sulfonic acid
HP	Hairpin
IMAC	Immobilized metal affinity chromatography
INDY	I’m not dead yet
IPTG	Isopropyl β-d-1-thiogalactopyranoside
K_A	Association Constant
K_D	Dissociation Constant
LB	Luria-Bertani medium
LDS	Lithium dodecyl-sulphate
LGV	Lymphogranuloma venereum
LMNG	Lauryl Maltose Neopentyl Glycol
ME	Malic Enzyme
MES	(2-(N-morpholino)ethanesulfonic acid)
MLCL	Monolysocardiolipin
MOPS	(3-(N-morpholino)propanesulfonic acid)
MSP2N2	Membrane scaffold protein (2N2)
MWCO	Molecular weight cut-off
NaCT	Sodium dependant citrate transporter
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NanoDSF	Nano differential scanning fluorimetry
NG	n-Nonyl-Beta-D-Glucopyranoside
NTA	Nitrilotriacetic acid
OD_{600}	Optical Density at 600nm wavelength
OPIC	Oxford Particle Imaging Centre
PA	Phosphatidic acid
PAGE	Polyacrylamide gel electrophoresis
PC	Phosphatidyl choline
PE	Phosphatidylethanolamine

PEPCK	Phosphoenolpyruvate carboxykinase
PES	Polyethersulfone
PF2	PF-06649298
PG	Phosphatidylglycerol
PHENIX	Python-based hierarchical environment for integrated crystallography
PI	Phosphatidyl Inositol
PMA	Phosphomolybdic acid
PMAL-C8	Poly (Maleic Anhydride-alt-1-Decene) substituted with 3-(Dimethylamino) Propylamine
PROMALS3D	Profile multiple alignment with predicted local structures and 3D constraints
PS	Phosphatidyl Serine
QGDG	Quadroglyco-diacylglycerol
RELION	Regularised likelihood optimisation
RMSD	Root mean squared distance
ROS	Reactive oxygen species
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SDS	Sodium dodecyl-sulphate
SEC	Size Exclusion Chromatography
SlyD	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD
SMILES	Simplified molecular input line entry system
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
SPR	Surface plasmon resonance
SSME	Solid supported membrane electrophysiology
TB	Terrific broth
TCA	Tricarboxylic acid
TCEP	(tris(2-carboxyethyl)phosphine)
TEM	Transmission electron microscopy
TEV	Tobacco etch virus protease
TGDG	Triglyco-diacylglycerol
TLC	Thin layer chromatography
T_m	Melting Temperature
Tris	Tris(hydroxymethyl)aminomethane
T-DNA	Transfer-Deoxyribonucleic acod
V_E	Coloumb potential
Xtal	Crystallisation
ZYM	1% N-Z-amine, 0.5% yeast extract, 25mM sodium phosphate monobasic, 25mM potassium phosphate dibasic, 50mM ammonium chloride, 5mM ammonium sulphate

1. Introduction

1.1 Dicarboxylates as Important Biomolecules

Small dicarboxylates such as α-ketoglutarate, succinate, fumarate, and malate are important molecules for all organisms **(Fig. 1.1)**. Not only are they intermediates of the Tri Carboxylic Acid (TCA) cycle, essential for the generation of electron carriers for oxidative phosphorylation; but are also precursors for many metabolites required for cell growth; and important signalling molecules [2-4]. More recently we are beginning to understand the way small dicarboxylates are utilised by pathogenic Enterobacteriacea and Pasteurellaceae bacteria in anaerobic conditions during gastrointestinal tract colonization [5, 6]. Under anaerobic conditions the TCA cycle is inactive, as such carbon sources are converted to fumarate which is reduced to succinate as the terminal step of fumarate respiration [7]. Succinate antiport is utilised to import dicarboxylates which can be converted to fumarate, and as such targeting transporters of small dicarboxylates is a novel and exciting target hypothesis for inhibiting these pathogens.

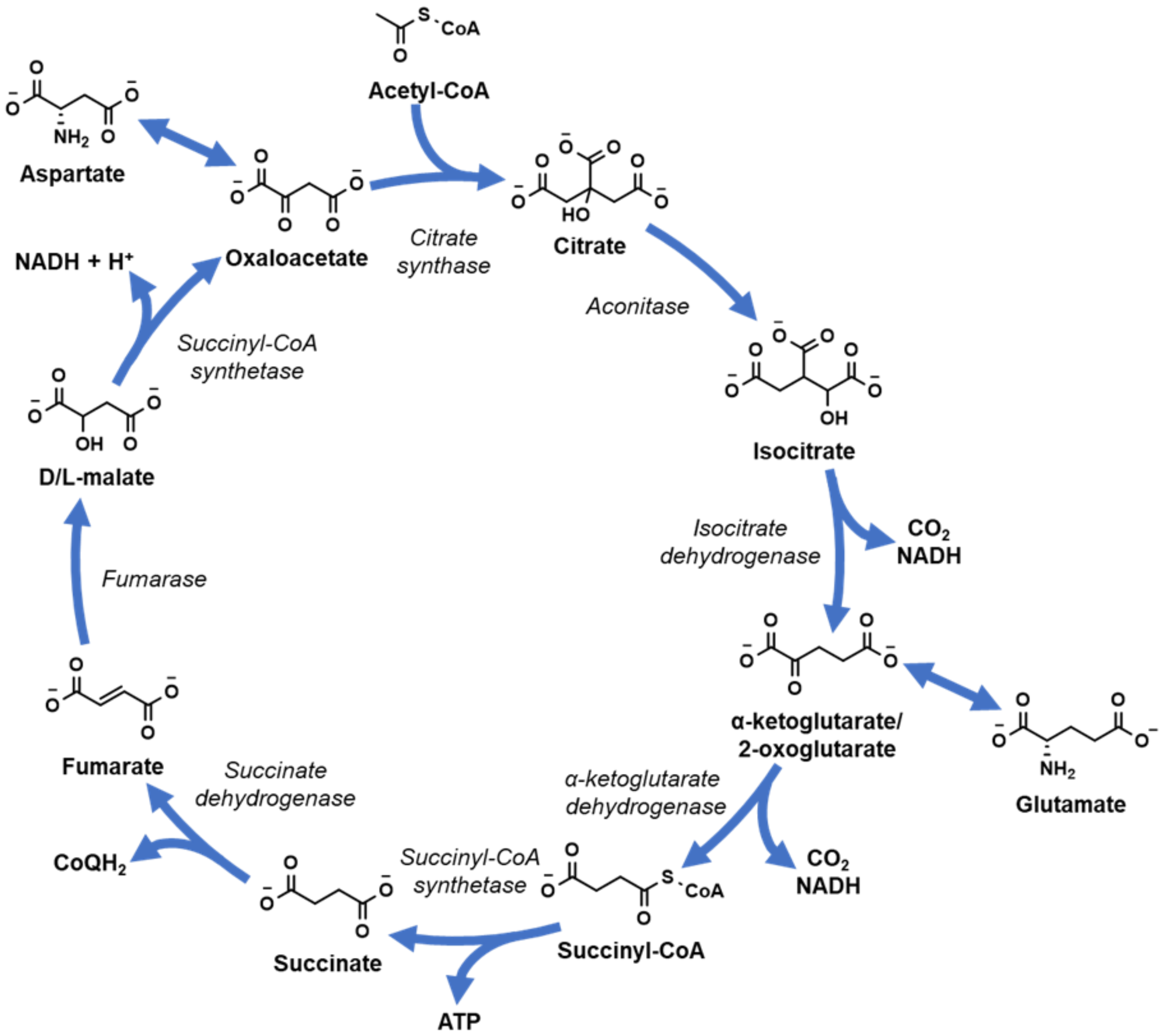


Figure 1.1 Metabolites of the TCA cycle and their respective structures.

Enzymes responsible for catalytic steps are labelled in italic. Steps producing reduced cofactors or ATP are highlighted with branched arrows. Figure adapted from Cannon 2014 [1].

1.2 The DASS Family of Transport Proteins

The Divalent Anion/Na+ Symporter (DASS) family is the largest within the Ion Transport (IT) superfamily of integral membrane proteins that transport ionic molecules across membranes [8]. DASS transporters are ubiquitous across all domains of life, with notable members: I’m Not Dead Yet (INDY) found in model organisms *D. melanogaster*, *C. elegans*, and in many prokaryotes; Na^+ dependent Citrate Transporter (NaCT) the mammalian INDY ortholog; and Dicarboxylate Transporters 1 and 2 (DIT1/2) from *A. thaliana* [9-12].

There are two distinct clades within the DASS family: Cotransporters (herein DASS-C) which utilize a favourable inward Na^+ gradient to concentrate sulphate or (TCA) cycle intermediates such as citrate, succinate, α-ketoglutarate or into cells; and Exchangers (herein DASS-E) which exchange one substrate for another in a Na^+ independent manner **(Fig. 1.2)** [13, 14]. Interestingly DASS-Es have been exclusively found in plants and prokaryotes, while DASS-Cs are ubiquitous.

The DASS family represent interesting therapeutic targets. INDY knockdown has been shown to increase the lifespan of *D. melanogaster* and *C. elegans* by mimicking calorie restriction [9, 11, 15]. NaCT upregulation has in turn been implicated in non-alcoholic fatty liver disease, with RNAi knockdown preventing the phenotype in mice [16]; and over 40 loss-of function mutations within NaCT are responsible for SLC13A5-epilepsy [17, 18].

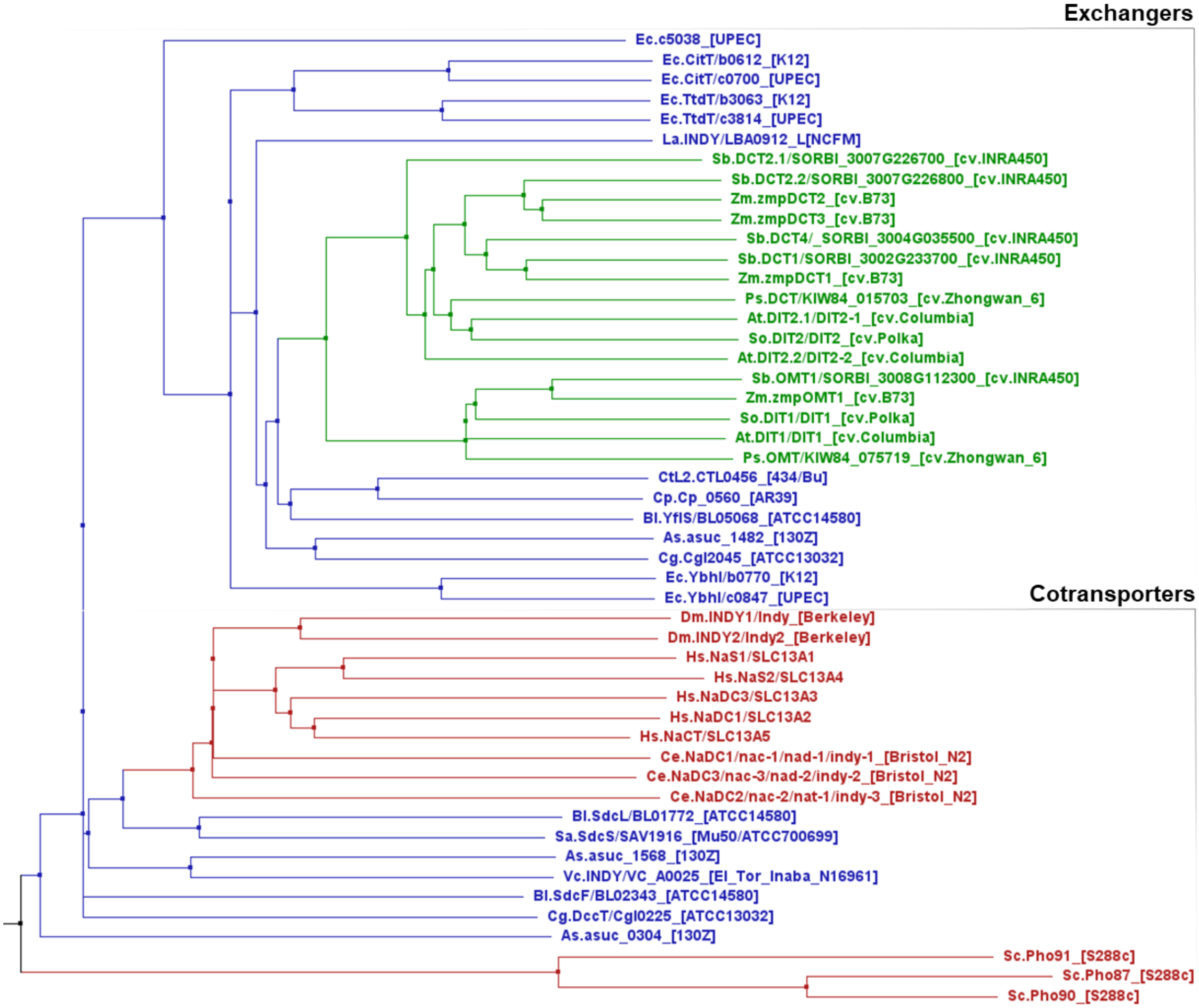


Figure 1.2 Phylogenetic tree of all Published DASS members discussed in the text.

Nomenclature for entries is presented as follows: "organism acronym.protein name(when available)/gene name_[uniport strain]" for example Ec.CitT/c0700_[UPEC] representing the DASS member CitT, gene c0700, from the O6:H1 UPEC strain of E. coli. DASS members of the prokaryotic, eukaryotic and plantae kingdoms are coloured blue, red, and green respectively.

1.3 DASS Cotransporter VcINDY as the Prototype DASS Member

The DASS-C member from *Vibrio cholerae* VcINDY is the best characterised of the of the DASS family with the first high-resolution structure solved in 2012 and 9 subsequent structures to date [14, 19-21]. VcINDY is a prototype for the DASS family, and its structures showed this family of proteins are typically between 90-130kDa homodimers with two discrete domains per protomer:

the scaffold domain, which facilitates protomer dimerization and supports the movement of the elevator domain; and the elevator domain which binds and translocates TCA substrates across the membrane **(Fig. 1.3A)**. A single protomer of VcINDY consists of 11 transmembrane helices, in which the helices H2-H6 and H7-H11 are inverted repeats of each other, displaying a 26% sequence similarity [19] **(Fig. 1.3C)**. Helices H1-4 and H7-H9 comprise the two halves of the scaffold domain, with helices 4c and, 9c acting as arm helices with flexible "wrists" that cradle the elevator domain. The elevator domain however is comprised of helices H5-H6, H10-H11 and two pairs of hairpin helices: HP_{in} and HP_{out} respectively which contain the SNT motif canonical to the DASS in their connecting turns.

1.4 The Elevator Domain and the Canonical SNT Motif

The elevator domain of each VcINDY protomer contains two highly conserved Ser-Asn-Thr (SNT) motifs. Residues within the SNT motif partially form the sodium ion binding sites: Na1, and Na2, with additional coordination provided by residues from L5ab, and L10ab respectively [19] **(Fig. 1.3B)**. The sodium ions balance the anionic charges of the dicarboxylate substrates and are essential for substrate binding, as the binding of these ions significantly increases stability within the substrate clamshell, creating a proper binding site for the substrate [21].

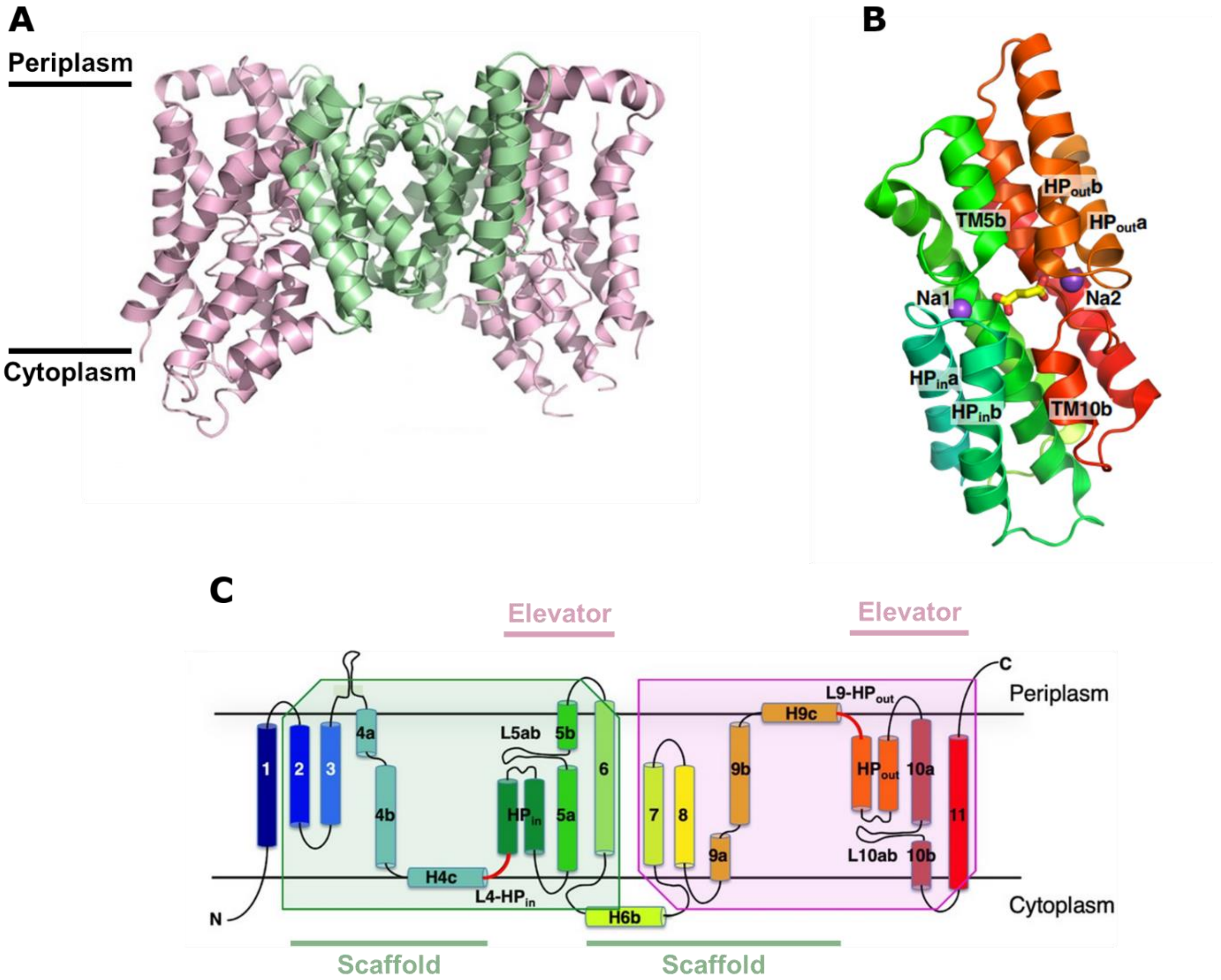


Figure 1.3 The DASS fold through the structure of prototype DASS-C: VcINDY.

A. *The VcINDY homodimer, with the scaffold domain coloured green and the elevator domain coloured pink.* ***B.*** *The elevator domain with helices coloured in accordance to the topology diagram in C. The Na1 and Na2 sodium ions are shown as purple spheres and succinate is shown in yellow.* ***C.*** *Topology diagram for the helices of a single VcINDY protomer. The inverted repeat symmetry is of the N and C termini is highlighted by the green and pink angled boxes. The helices which comprise the scaffold and elevator domains are denoted by green and pink bars. Figures A. and C. used with permission from Sauer et al., 2020. Figure B. used with permission from Sauer et al., 2022.*

The translocation of substrates across the membrane occurs via an elevator-type mechanism, in which the transport domain rotates as a rigid body 34.7°, moving the substrate binding site 13.0Å across the membrane [14]. This mechanism was originally discovered in GltPh of the EAAT family [22, 23], but demonstrated in VcINDY by Mulligan *et al.*, through non-native cysteine crosslinking and molecular dynamics (MD) simulations [24]. Interestingly, to date all structures of VcINDY are in an inward-facing conformation (C_i) and it required switching to the homologue LaINDY from *Lactobacillus acidophilus* for the first outward facing structure (C_o) to provide structural validation for the mechanism [14] **(Fig. 1.4A/B)**.

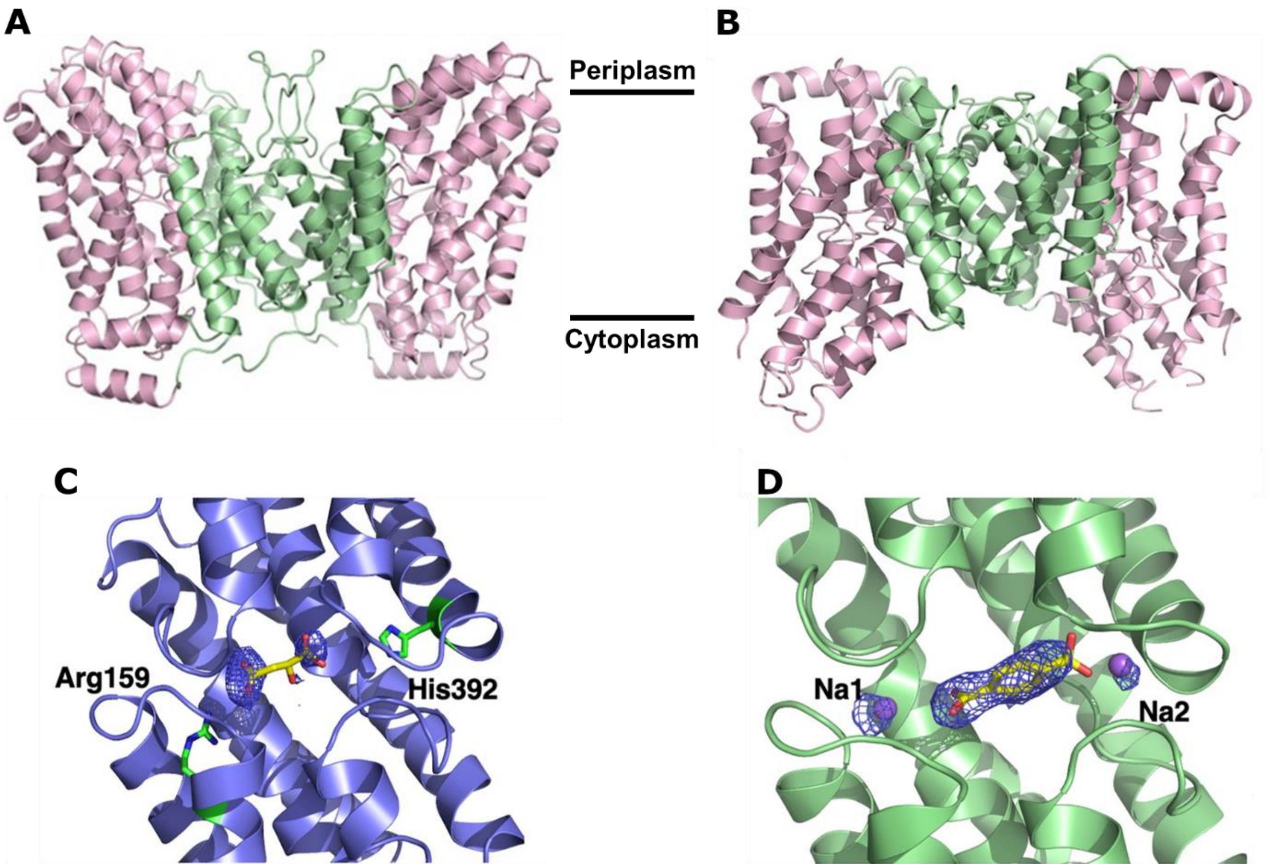


Figure 1.4 LaINDY-C_o and VcINDY-C_i and their respective binding sites.

A. *The LaINDY dimer, with the scaffold domain coloured green and the elevator domain coloured pink.* ***B.*** *The VcINDY dimer, coloured as in A.* ***C.*** *The LaINDY elevator coloured blue. The basic residues within the binding site that substitute the Na1 and Na2 sites in DASS-Cs are coloured green. An α-ketoglutarate molecule coloured yellow has been fit between the basic residues: Arg159 and His392. Otherwise atoms are coloured by heteroatom. Density for the substrate is shown as a purple mesh. D. The VcINDY binding site with sodium ions in the Na1/Na2 sites shown as purple spheres. A yellow terephthalate molecule has been fit into the binding site. Density for the substrates in shown as a blue mesh. Figures used with permission from Sauer et al., 2020.*

1.5 DASS Exchangers Substitute the Sodium Sites with Basic Residues

The structures of LaINDY highlighted key differences between the two clades of the DASS subfamily. LaINDY and other DASS-E members have basic residues that extend into what would be the Na1/Na2 sites in DASS-Cs, acting as surrogate cations to compensate for the negatively charged substrates **(Fig. 1.4C/D)**. There is also no significant difference within the LaINDY substrate binding site in the presence or absence of substrate, unlike the cooperative nature of Na^+ binding in VcINDY [21].

1.6 Transport Cycles of the DASS Family

The diversity within the binding site also translates to mechanistic differences between the transporters, although charge neutralization still permits transport domain translocation between C_i and C_o. The field's current hypothesis is elevator domain translocation is gated by its charge balance, such that when the binding site is electroneutral VcINDY can freely switch between C_i and C_o **(Fig. 1.5A)**. This can occur in two states: when VcINDY is sodium free, and when the both Na1/2 sites and the substrate binding site are occupied. As sodium binding increases the charge of the binding site, thus preventing transport domain translocation until a substrate has bound, the sodium gradient determines the direction of DASS-C substrate transport [21]. As the basic residues Arg159 and His392 within the LaINDY binding site provide a constant positive charge, the only way to neutralize that charge at physiological pH is for a substrate to bind, meaning a substrate-less transition between C_i and C_o is not possible **(Fig. 1.5B)**.

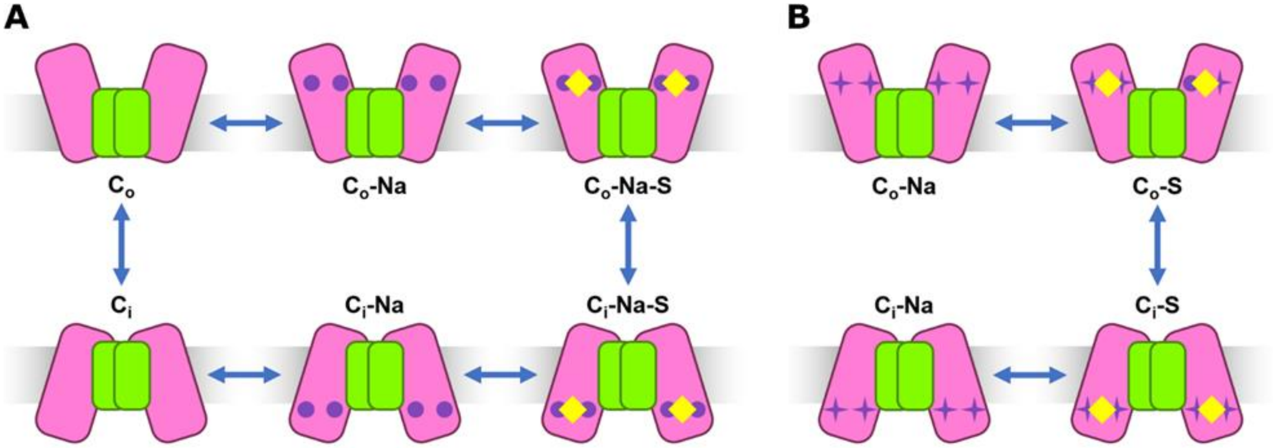


Figure 1.5 Contrasting the proposed transport cycles of DASS-Cs and DASS-Es.

***A**. The proposed reaction cycle of VcINDY and all DASS cotransporters. Scaffold domains are coloured green while the elevator domains are pink. The lipid bilayer is shown as grey. Sodium ions are represented as purple circles. The dicarboxylate substrate is a yellow diamond. Permitted state transitions are linked by bidirectional arrows. **B**. The proposed reaction cycle of LaINDY and all DASS exchangers. Colouring is identical to A, except the basic residues substituting the sodium ions, which are represented as a purple four-point star.*

1.6.1 ScPho90 Structure Captures the Transport Cycle in Multiple States

During the writing of this thesis the Cryo-EM structures of DASS-C member Pho90 a low affinity sodium-phosphate cotransporter from *S. cerevisiae* were published [25]. Of the three captured states presented, on displayed the two elevator domains in an asymmetric ‘one in one out’ state. This not only made it the first DASS member to have both states captured, but also indicated the two protomers can act independently of one another **(Fig. 1.6A)**.

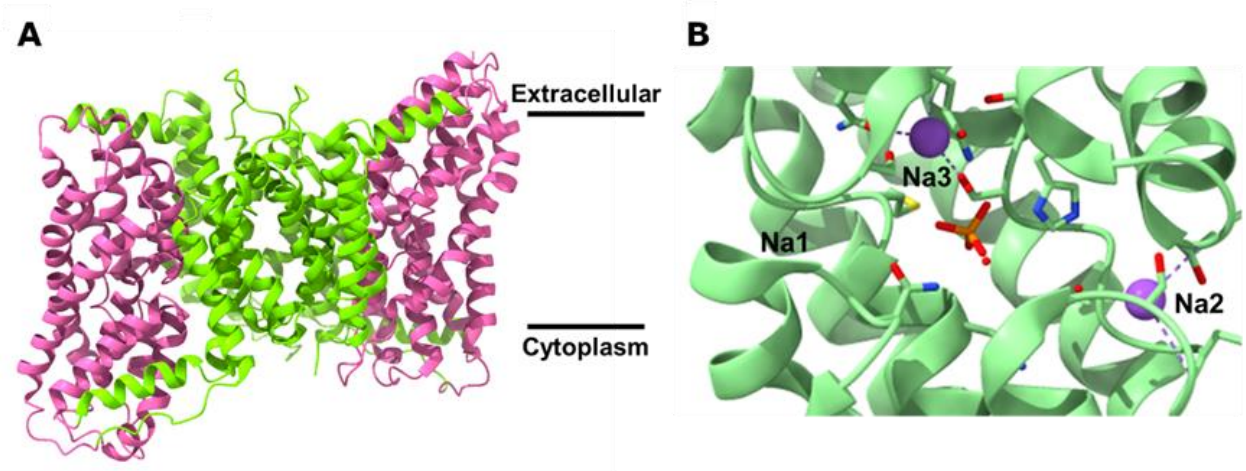


Figure 1.6 The asymmetric structure of DASS-C ScPho90 and the novel Na3 site.

A. *The Schneider et al., (2024) 3.12Å asymmetric-apo structure of ScPho90 (PDB 8R35). The scaffold domain is coloured green and the elevator domain is coloured pink.* ***B.*** *Elevator domain of the 2.62Å outward facing sodium and phosphate bound structure (PDB 8R34). The polypeptide is coloured green. The sodium ions are represented as purple spheres. Coordinating residues are shown, with coordination shown as dashed lines. Otherwise atoms are coloured by heteroatom.*

1.6.1.1 A Novel Sodium Site

Additionally, ScPho90 revealed a new sodium site for the DASS family, as the HPin SNT motif appears to not bind sodium despite high sequence conservation. Instead, a site I have designated Na3, is created by residues from H5b and H10a **(Fig. 1.6B)**. The Na2 site in ScPho90 is also much more poorly defined than other DASS-Cs, with only two residues with carbonyl moieties within coordination distance of the sodium ion. Finally, a histidine mutation within the SNT motif instead contributes hydrogen bond donors to the phosphate to be transported.

1.7 Human DASS-C Structures

1.7.1 NaCT and Competitive DASS Inhibition

In 2021 Sauer *et al.,* solved the first structures of a human DASS cotransporter via Cryo-EM, NaCT (SLC13A5) in an inward facing citrate bound, and inward-facing inhibitor-bound states [18]. The structure of NaCT bound to the inhibitor PF2 provided structural evidence for the molecule acting as a competitive inhibitor **(Fig. 1.7A)**, binding in the same site as the substrate would in the elevator binding pocket, which was at the time contested [26-29]. The PF2 molecule has a tert-butyl group extending from a benzene ring that contributes to its high selectivity for NaCT over other human DASS members NaDC1 and NaDC3, by its interaction with residues from helix H9b in the scaffold domain. Additionally, this benzene-tert-butyl extension also contributes to the hypothesis for its inhibition mechanism: which is that steric hindrance of the extension prevents elevator translocation effectively locking the transporter in the C_i-Na-S state **(Fig. 1.7B)**.

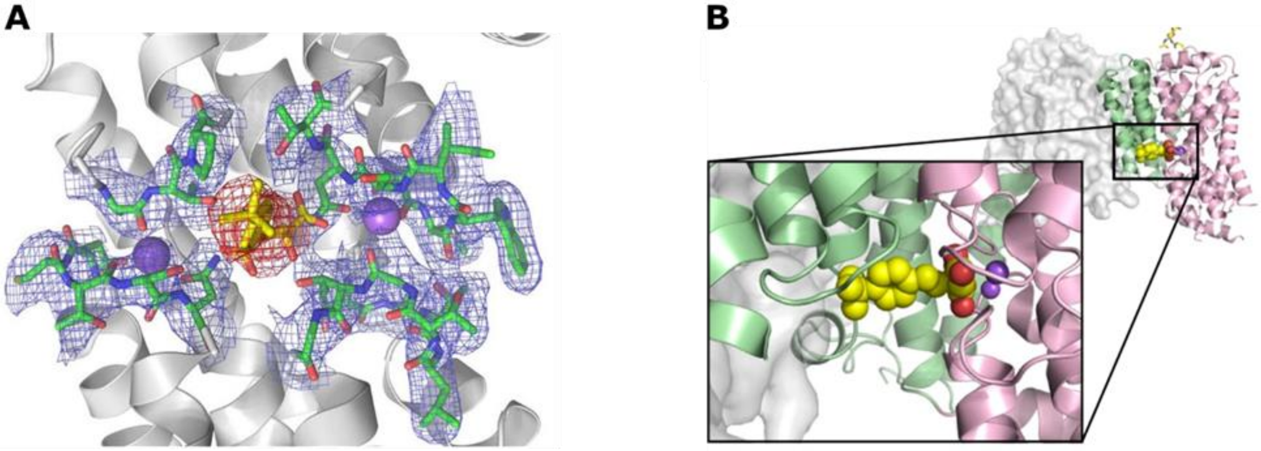


Figure 1.7 Competetive inhibition of NaCT by compound PF2.

A. *The Sauer et al., (2021) 3.12Å inward sodium-PF2 bound structure (PDB 7JSJ). The elevator domain is coloured grey. Sodium ions are coloured purple. LaINDY residues are coloured green, while the PF2 molecule is coloured yellow. Otherwise atoms are coloured by heteroatom. Density for LaINDY is shown as a blue mesh, while the PF2 density is a red mesh.* ***B.*** *View showing the PF2 benzene-tert-butyl extension indicating how it may lock the protein in the inward facing state by sterically hindering translocation. Figures used with permission from Sauer et al., 2021.*

1.7.2 NaS1, NaDC1 and Allosteric DASS Inhibition

Most recently Chi *et al.*, have solved the structures of sodium sulphate cotransporter NaS1 (SLC13A1) and sodium dependant citrate transporter NaDC1 (SLC13A2) via Cryo-EM [30]. Multiple structures were reported for both proteins, most notably an asymmetric state substrate bound state for NaS1, and an allosteric inhibitor ACA bound outward open state for NaDC1 **(Fig. 1.8A)**. ACA selectively inhibits NaDC1 and NaCT but not NaS1 [31], binding into a hydrophobic pocket between helices H6 and H11 with a salt bridge pairing while its tail interacts with hydrophobic residues in helix H2 of the scaffold domain. The exact mechanism of the inhibition is unknown, but the authors speculate that ACA molecule may rigidly couple the elevator and scaffold domain preventing elevator transition **(Fig. 1.8B)**.

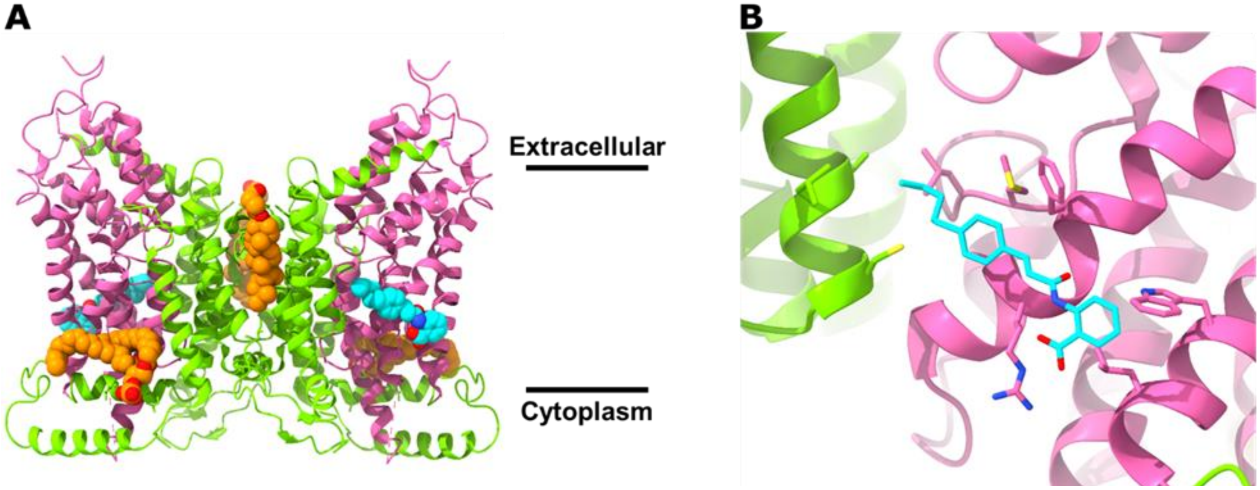


Figure 1.8 Allosteric inhibition of NaDC1 by compound ACA.

A. *The Chi et al., (2024) 3.30Å outward facing apo ACA bound structure (PDB 8W6G). The elevator domain is coloured green, while the elevator is coloured pink. The ACA inhibitor is coloured blue, and the modelled phospholipid and CHS molecules are coloured orange. Otherwise atoms are coloured by heteroatom.* ***B.*** *The hydrophobic binding pocket of ACA with interacting residues shown. The colours are as in A.*

1.8 The DASS-E Clade is Poorly Studied

The rapid improvements of Cryo-EM hardware [32, 33], and software [34-36], over the past decade has accelerated our understanding of the DASS family. In the last four years alone an additional six members have had structures solved compared to the previous eight years where VcINDY stood as the only experimentally determined atomic model. Despite this the DASS-E clade of the family has been underrepresented, particularly the bacterial members. As the DASS-C clade contains all the mammalian homologs this is perhaps to be expected. Illustrating this point, early work on DASS-Cs: SdcL, SdcF from *Bacillus licheniformis* [37, 38]; SdcS from *Staphylococcus aureus* [39]; DccT from *Corynebacterium glutamicum* [40]; and VcINDY highlights the fact that sodium dependence is shared with the mammalian homologs. Further, both

B. licheniformis and *C. glutamicum* both have predicted DASS-E genes with no direct literature on them to date. However, as DASS-E genes are specific to bacteria and plants, and their substrates are commodity chemicals, understanding the DASS-E clade may have impacts on bioproduction [41], agriculture [42], and bacterial infection [43].

1.8.1 DASS-Es of the Plantae Chloroplasts and the Two-Translocator Model

The DASS-E members of the plantae kingdom are the most functionally characterized of the exchangers, with experiments reporting dicarboxylate transport in chloroplasts dating as early as the 1960s [44-47]. Most of the work on was carried out between the C_3 model plants: *Arabidopsis thaliana*, *Spinacia oleracea*, and *Pisum Sativum* DASS-Es. The 2-oxoglutarate/malate antiporter or OMT (also referred to as DIT1) and the malate/glutamate antiporter or DCT (also referred to as DIT2) are highly expressed in chloroplasts, and are involved in nitrogen assimilation during normal photorespiratory conditions [48, 49]. During photorespiration, DIT1 imports α-ketoglutarate into the stroma of the plastid using malate as a counter anion **(Fig. 1.9A)**. α-ketoglutarate is then transaminated by Fd-GOGAT to glutamate [50]. DIT2 then exports glutamate out of the stroma to the cytosol to be utilised by the cell for biogenesis, again using malate as the counterion, in a process that has been termed the two-translocator model [51, 52].

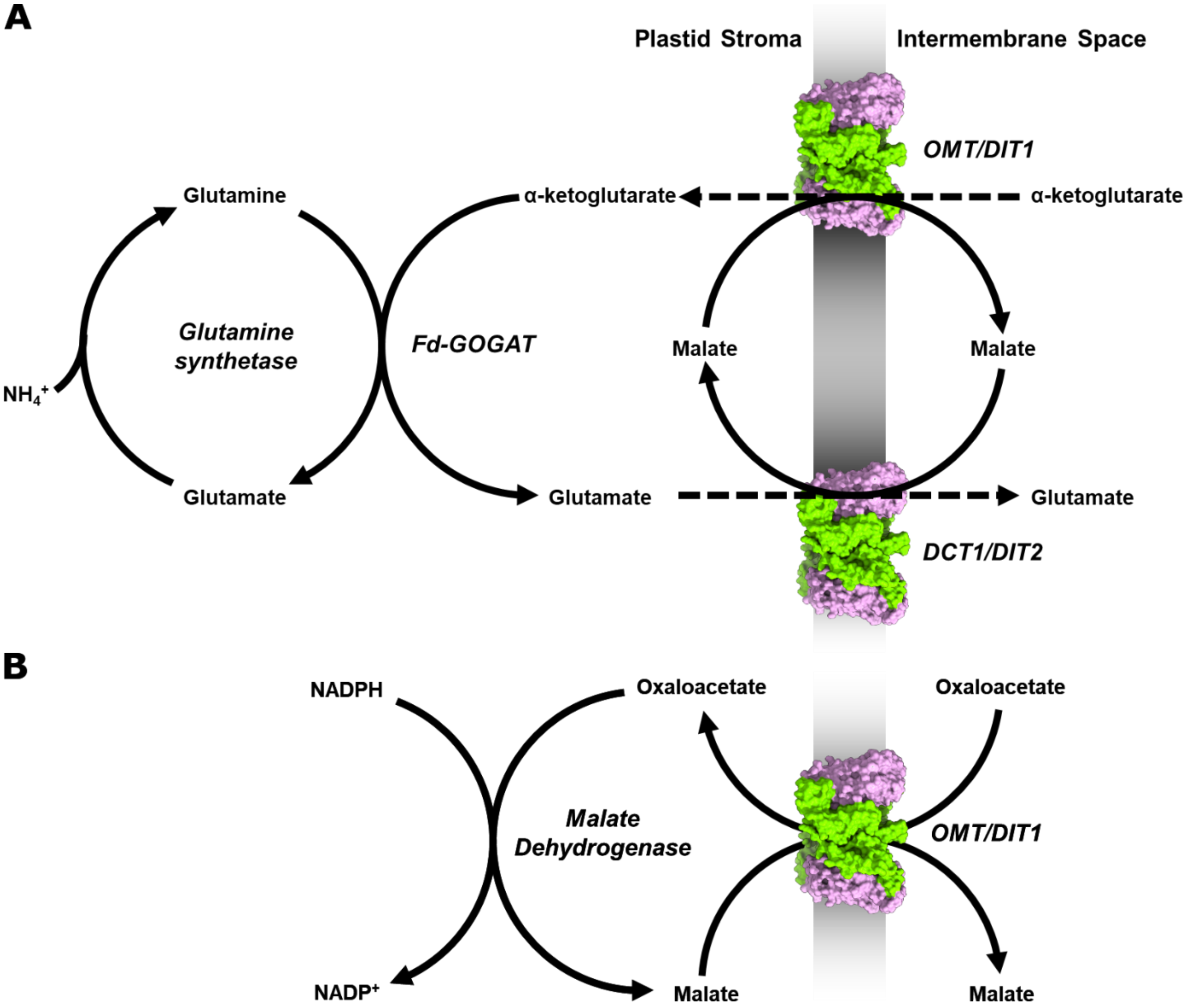


Figure 1.9 DASS-E functionality in chloroplasts.

A. *The two-translocator model nitrogen assimilation in chloroplasts. The Sauer et al. (2020) 2.86Å outward facing LaINDY structure (PDB: 6WTW) is used to represent both DIT1 and DIT2. The plastid inner membrane is represented as a dark grey line. Enzymes and Transporters are labelled in italics.* ***B.*** *The malate valve model for DIT1, removing redox potential from the stroma to prevent photoinhibition and ROS generation.*

1.8.1.1 An *Arabidopsis thaliana* conditional mutant reveals the DASS family

DIT2 was the first DASS member identified, revealed through the *A. thaliana* 'dct mutant' line [53]. The ‘dct mutant’ line in was non-viable under normal photorespiratory conditions due to defective dicarboxylate transport [53]. The cause of the phenotype was later identified as the loss of a 42kDa polypeptide from the chloroplast envelope membrane [54]. In 2003, following the sequencing of the *A. thaliana* genome [55], the 'dct mutant line' was molecularly characterised as a single point mutation in AtDIT2.1 resulting in a single non-neutral change from the highly conserved glycine 206 to glutamate [56].

The two-translocator model presented above was proposed following silicone oil centrifugation filtration experiments on intact *S. oleracea* and *P. sativum* chloroplasts, and radiolabelled transport assay on *S. oleracea* plastid envelopes reconstituted into liposomes [48, 49, 57]. All experiments indicated that two antiporters were involved, distinguishable by a difference in apparent affinity for amino acids aspartate and glutamate. Despite the two-translocator model, Renne *et al.* were able to use T-DNA knockouts to demonstrate that only loss of AtDIT2.1 caused the photorespiratory phenotype, not AtDIT1 despite the latter's overall higher expression levels in leaves [56]. This indicating that DIT2 can compensate for the lack of DIT1 functionality, acting as both the OMT and DCT.

1.8.1.2 DIT1 Oxaloacetate/Malate Antiport and Photoinhibition

An additional role of DIT1 as a malate valve was hinted at through reports revealing DIT1 accepts oxaloacetate as a substrate with an apparent affinity of 0.04mM (SoDIT1) and 0.042mM (AtDIT1) [10, 56], which would permit oxaloacetate import at cytosolic concentrations in illuminated leaves [58]. The malate valve is critical for maintaining the redox potential of the stroma, by exporting malate and importing oxaloacetate, the latter of which can be reduced back to malate, preventing the accumulation of ROS under high-light stress **(Fig 1.9B)**. This was confirmed in 2011 by Kinoshita *et al.*, generating two T-DNA knockouts of AtDIT1 that reduced oxaloacetate uptake into chloroplasts [59]. Additionally, both mutants showed a photoinhibition phenotype in high-light conditions attributed to a high concentration of reducing equivalents in the stroma.

1.8.1.3 Additional DASS-E Genes in C_4 Plants

In C4 plants photosynthetic activity is partitioned between bundle sheath and mesophyll cells, ultimately concentrating CO_2 around RuBisCO and resulting in increased photosynthetic efficiency relative to C3 plants that do not compartmentalise this reaction [60].

In C_4 photosynthesis, bicarbonate is fixed by PEPC in the mesophyll cell forming oxaloacetate, which depending on the C_4 biochemical subtype is reduced to either malate or aspartate. The malate or aspartate then freely diffuse across the plasmodesmata into bundle sheath cells, where one of three malic enzymes: plastidic NADP-ME, mitochondrial NAD-ME, or cytoplasmic PEPCK perform the final decarboxylation step [61]. The flux of dicarboxylates between the two cell types and their respective subcellular compartments is distinct between species of C4 plants, and as such additional DASS-E genes have been discovered in *Zea mays* (NADP-ME and PEPCK pathways) and *Sorghum bicolor* (NADP-ME) [62, 63].

Taniguchi *et al.*, found four DASS-E transcripts in *Z. mays*; OMT1, DCT1, DCT2, and DCT3 and demonstrated mesophyll expression of OMT1 and DCT1, and bundle sheath expression of DCT2/3 [62]. ZmDCT2/3 are nearly identical in both nucleotide and protein sequence, and are presumed to be alleles from the parental lines used to generate the *Z. mays* hybrid strain for the study. In a subsequent study, Weissmann *et al.* (2016) made mutants of ZmDCT2 confirming its role as the primary malate transporter in bundle sheath chloroplasts [64]. Though not lethal, The ZmDCT2 mutant resulted in 3% of photosynthetic activity of the wild type, despite upregulation of the PEPCK pathway from 25% to 55% of total carbon assimilation.

Most recently, in a study of the C4 grass *S. bicolor*, Weissmann *et al.* (2021) found an additional DCT4 gene not syntenic to any of the DASS genes in *Z. mays* [63]. Genomic analysis of the C_4 grasses revealed homologous DCT4 genes in a number of other C_4 grasses with NADP-

ME as the primary decarboxylase. Expression of OMT1, DCT1, and DCT2.1 was demonstrated in mesophyll cells, with expression of DCT2.2 and DCT4 in bundle sheath cells. Both the Taniguchi *et al.* and Weissmann *et al.* (2021) studies indicate that DASS-Es may have evolved multiple roles in C_4 photosynthesis, and understanding them may be key to our understanding of the subtypes of C_4 photosynthesis.

1.8.1.4 Plantae DASS-Es are of Chlamydial Origin

Interestingly the DASS-Es of the plantae kingdom are one of four transporter families thought to be of Chlamydial origin [65]. Chlamydia are obligate intracellular pathogens [66], and as such the proposed mechanism is that a chlamydia-like ancestor was also endocytosed during the cyanobacterial endosymbiosis of the common ancestral algae. Indeed, the phylogenetic tree of the DASS family including predicted DASS members CTL0456 from *Chlamydia trachomatis serovar L2* and Cp_0560 from *Chlamydia pneumoniae* shows the Chlamydial members at the root of the plantae DASS-Es **(Fig. 1.2)**.

1.8.2 *E. coli* DASS-Es

1.8.2.1 CitT is an Anaerobic Citrate/Succinate Antiporter

As the most widely studied prokaryotic model organism, it is perhaps unsurprising that multiple DASS-Es have been identified and characterised in *E. coli*. CitT was discovered following the sequencing of the E. coli K12 genome when Pos *et al.* compared the *E. coli* citAB operon with that of *K. pneumoniae* [67-69]. It was noted that the *E. coli* citAB operon contained an additional ORF encoding a protein with 34% sequence identity to the plant DASS-E SoDiT1 [70], and hence renamed CitT.

Pos *et al.* demonstrated elevated efflux of radiolabelled citrate in exchange for succinate, fumarate and tartrate in *E. coli* BL21 cells transformed with a pET24-CitT expressing plasmid [68].

Combined with the anaerobic expression of CitT and succinate being the final product of citrate fermentation, this observation led to the prediction of CitT as a citrate/succinate antiporter.

As it is well documented that most known *E. coli* strains cannot utilise citrate as the sole carbon source under aerobic conditions, Pos *et al.* demonstrated *E. coli* DH5α cells transformed with a lac-inducible CitT plasmid were able utilize citrate as a carbon source aerobically. Multiple evolution experiments have demonstrated *E. coli* aerobic utilisation of citrate requires gain of function through gene dupilication of CitT [71, 72]. However, although CitT gain of function is essential, a strong aerobic citrate utilization phenotype is only seen following additional upregulation of the non-DASS dicarboxylate transporter dctA, which recaptures the succinate exported during CitT citrate uptake [73].

1.8.2.2 TdtT is an Anaerobic L-Tartrate/Succinate Antiporter

TtdT was identified in the Pos *et al.* CitT report [68]. TtdT has 44% sequence identity with CitT and is found downstream of the gene for L-tartrate dehydratase, so it was theorised that it functioned as a tartrate/succinate antiporter. This however, wasn't confirmed until 2007, when Kim *et al.* used knockdown mutants to demonstrate that TtdT was essential for anaerobic growth of *E. coli* on L-tartrate [74]. Additionally, TtdT antiport was shown to be directional, with 13-fold greater transport efficiency for L-tartrate import and succinate export over the inverse. Finally, TdtT was shown to transport succinate and fumarate, however is stereospecific for L-tartrate, while D-tartrate is imported via the non-specific C4 dicarboxylate transporter DcuB for anaerobic growth [75]. Additionally, TdtT is the most characterised of the DASS-Es that possess a sequence difference within the SNT2 motif, substituting the histidine for substrate recognition with an arginine **(Fig. 1.10)**. This change may have implications for substrate recognition, but has not been experimentally explored.

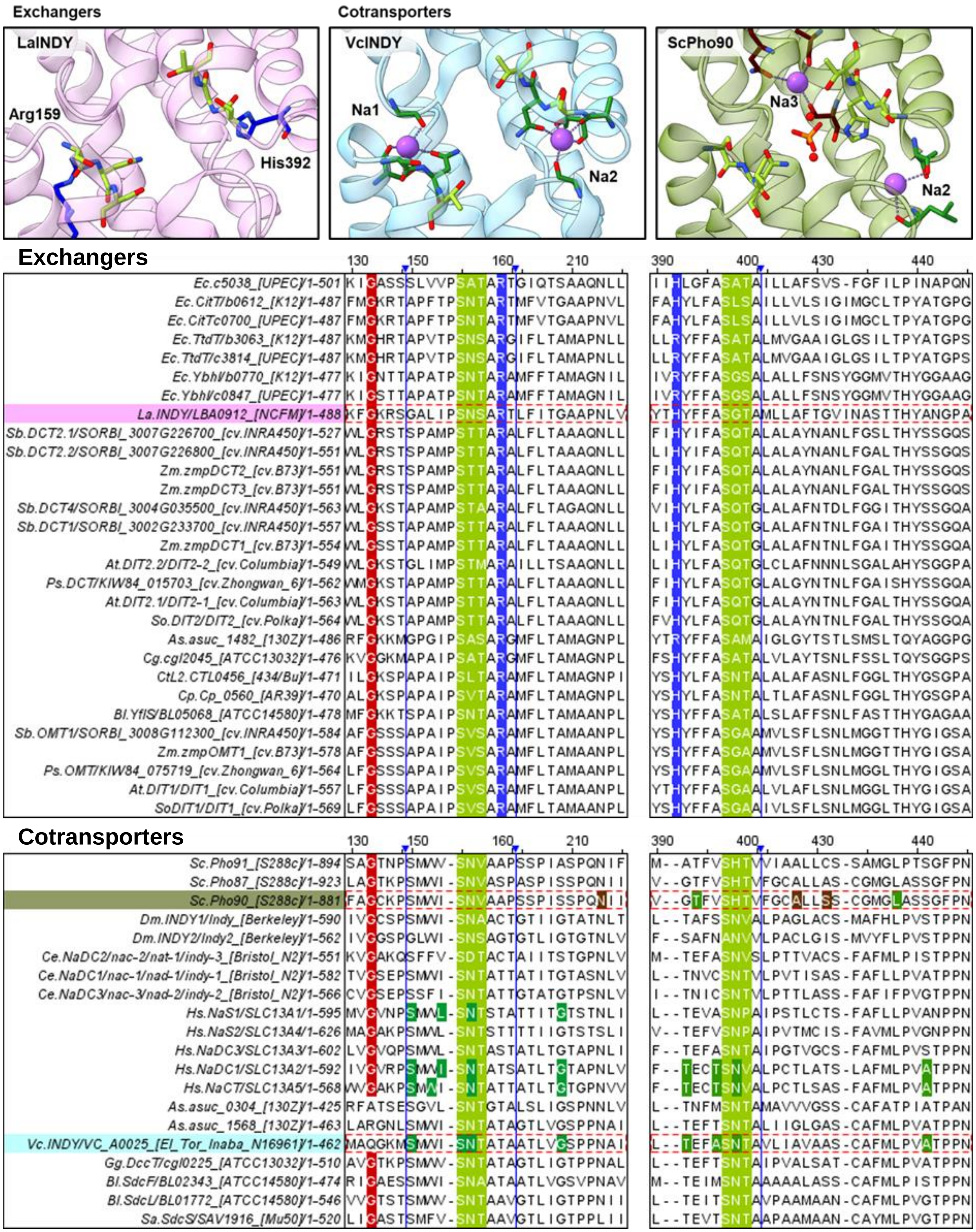


Figure 1.10 Sequence alignment of the canonical DASS SNT motifs.

continued on next page

Figure 1.10 continued.

The binding sites of DASS exchanger LaINDY, and DASS cotransporters VcINDY (PDB 7T9G) and ScPho90 (PDB 8R34) are shown as pink, blue and dark-green respectively. The SNT motif residues are coloured green. For LaINDY: The sodium-substituting basic residues Arg159 and His392 are coloured blue. For VcINDY and ScPho90: residues coordinating the Sodium ions in Na1/Na2 are coloured dark green, while the sodium ions are purple. For ScPho90 residues within the Na3 site are coloured brown. Otherwise atoms are coloured by heteroatom. In the sequence alignment DASS exchangers are in the top panels while the cotransporters are in the bottom panels. Breaks in the alignment are indicated by blue vertical lines. LaINDY, VcINDY, and ScPho90's positions are highlighted in the same colours are the peptide chain from the cartoons. Residues displayed in the cartoon are highlighted with identical colours. Additionally, the G206E mutation from the AtDIT2.1 'dct' mutant is highlighted in red. The alignment is numbered with respect to the amino acid positions of LaINDY. Sequence alignment was done in Clustal Omega v1.2.2.

1.8.2.3 YbhI is a DASS-E yet to be Characterised

YbhI was also identified by Pos *et al.* in their CitT report, having 35% sequence identity to the CitT peptide [68]. The exact function of YbhI remains unknown, however it is sandwiched between the two genes YbhH and YbhJ in the same orientation [76], indicating a common functional pathway. A structure of YbhH has been deposited with no accompanying literature (PDB 6OTV), and is predicted to be an isomerase in the PrpF family. YbhJ has 22% sequence identity to the aconitase AcnA, including 15 of 20 active-site residues, however no aconitase activity was for reported for recombinant YbhJ in vivo [77]. Assuming YbhH/I/J are all functional proteins, they may share a common substrate.

1.8.2.4 Uropathogenic *E. coli* Possess an Additional DASS Gene

c5038 is a DASS member found on a genomic island that is unique to uropathogenic *E. coli*, hence is absent in both diarrheagenic and laboratory strains [78]. Genes within this island encode subunits of the α-ketoglutarate dehydrogenase complex, and Cai *et al.* demonstrated α-ketoglutarate dependant expression of c5038 via the KguS/KguR two component system. In a 2017 report Cai *et al.* demonstrated the α-ketoglutarate sensitivity of c5038 occurs via RNA polymerase sigma factor σ54, and expression is also upregulated by FNR and ArcA, which are master regulators for mediating adaptation to oxygen availability [43]. Additionally, c5038 knockout mutants showed significant growth lag when grown in minimal media with α-ketoglutarate as the sole carbon source. However, a double knockout of c5038 and KgtP, an aerobically expressed α-ketoglutarate transporter from the major facilitator subfamily was required to fully prevent growth. Finally, Cai *et al.* showed loss of c5038 reduced the fitness of UPEC strain CFT073 during colonization of murine urinary tracts, although the competition factor was less than twofold.

1.8.3 DASS-Es in Industrial Prokaryotes

The unique metabolisms of many prokaryotes have been utilised to produce chemicals for over half a century. Of particular interest is *Corynebacterium glutamicum* which has been used to produce the dicarboxylate glutamate and lysine [79, 80], and *Actinobacillus succinogenes* which is used to produce succinate [81, 82].

C. glutamicum (originally designated *Micrococcus Glutamicus*) was isolated in 1957 by Kinoshita *et al.* and was demonstrated to produce high yields of L-glutamate from glucose [80, 83]. Following the sequencing of the *C. glutamicum* genome one DASS-C gene (DccT) and one suspected DASS-E gene have been found [40, 79]. DccT was characterised as sodium dependent importer of L-malate, succinate, fumarate, and oxaloacetate. However native expression levels of DccT were deemed too low to permit growth using its substrates as the sole carbon source, so the physiological role is currently unknown [40]. Nothing has been published on the DASS-E gene cgl2045, however it clusters with the plantae and chlamydial DASS genes in the sequence alignment **(Fig. 1.10)**. If cgl2045 is a transporter of glutamate, like its plant DCT homologues, then understanding the physiological role of the gene may further our understanding of glutamate production.

A. succinogenes was originally isolated from the bovine rumen, has an incomplete TCA cycle due to lacking citrate synthase, isocitrate dehydrogenase and succinate dehydrogenase [81]. As such a its metabolism is adapted to sugar fermentation, producing succinate, acetate and formate [84]. *A. succinogenes* expresses two DASS-C genes: asuc_0304, asuc_1568 and a single DASS-E gene asuc_1482 [85]. Only the asuc_0304 DASS-C gene has been characterised as a constitutively expressed sodium dependant importer, with a role in fumarate import under aerobic conditions [86]. This finding was supported by a transcriptomics study on dicarboxylate transport, noting asuc_0304 as constitutively expressed, while finding asuc_1568 and DASS-E asuc_1482 had differential expression during growth on fumarate [85]. As with the *C. glutamicum* DASS members, characterisation of these transporters will improve our understanding of the organism's metabolism and may provide insights into yield improvement of succinate.

1.8.4 Chlamydial DASS-Es

As mentioned in **Section (1.8.1.4)** the plantae DASS-Es are believed to have originated from a chlamydial ancestor during the original plastid endosymbiosis [65]. Modern day chlamydia species are the causes of many human diseases depending on the tissues they infect: *C. trachomatis* serovars A/B/C causing ocular trachoma, serovars L1/2/3 causing lymphogranuloma venereum sexually transmitted disease; and *C. pneumoniae* causing pneumonia [87-89]. Chlamydia species express a single DASS-E gene, with high sequence identity within the clade, however to date no literature has been directly reported on the DASS gene. This is understandable, as the two-phase lifecycle: reticulate bodies embedded within the host cell, and infectious elementary bodies with a dense protective cell wall has presented significant challenges with developing genetic tools for chlamydia [90]. However, since the first successful transformation of recombinant DNA into *C.*

psittaci via electroporation by Binet & Maurelli in 2009 the last decade has seen great strides in Chlamydial molecular biology [91].

In 2015 Brothwell *et al.* generated temperature sensitive mutants of *C. trachomatis* L2 in order to assess essential genes for the Chlamydia lifecycle, and three of thirteen genotyped mutants contained mutations in the DASS-E gene CTL0456 [92]. A later study by Kokes *et al.* in 2015 using chemical mutagens to generate single nucleotide variants of *C. trachomatis* L2 showed that of the 8,205 SNV generated the DASS-E gene CTL0456 produced no viable mutants with nonsense mutations [93]. In comparison the non-DASS dicarboxylate transporter GltT/CTL0658 produced 2 viable nonsense mutants, indicating that of the two, GltT is the non-essential transporter.

While neither of these studies were extensive, both implicate DASS-Es as an important part of the Chlamydial lifecycle, however a study focussed on Chlamydial transporter genes in general would be needed to probe this further.

1.8.5 LaINDY

L. acidophilus is a gram positive, acid tolerant bacteria initially isolated from infant faeces in 1900 [94]. *L.* acidophilus are probiotic bacteria, and many strains have demonstrated benefits to humans when colonising the digestive tract. More specifically, *L. acidophilus* has been shown to improve lactose digestion due to it high lactase activity [95], reduce cholesterol levels via both incorporation into its own membrane and conversion to the more poorly absorbed coprostanol [96, 97], and reduced the instance of colon cancer in rats [98]. *L. acidophilus* have a single DASS-E gene LaINDY but little is known about it and it has not been linked to any of the *L. acidophilus* health benefits. Rather LaINDY was used by Sauer *et al.* to solve the first structure of a DASS member in the outward facing state when this could not be achieved with their other target VcINDY [14]. As such little is known about the physiological role of LaINDY other than it is predicted as a α-ketoglutarate/succinate antiporter based on their *in vivo* transport data. However, as the only DASS-E member with a reported structure LaINDY is the model transporter for the DASS-E clade.

1.9 Membrane Protein Environments

By carrying out biological functions at or across a lipid bilayer, integral membrane proteins are unique among the proteome with the need to fold and function in both aqueous and hydrophobic environments. Furthermore, this lipid bilayer's composition is defined by the host organism and organelle, and its chemical makeup and biophysical properties can be changed with the cell's response to signals, metabolic state or environment. Accordingly, lipids can exert forces and functional effects on membrane proteins. However, studying a protein's interactions with both aqueous and lipidic solvents requires focused and specialised experiments, such as studies of the protein reconstituted into nanodiscs, proteoliposomes, or molecular dynamics. Alternatively, serendipitous discoveries of co-purifying lipids and highly-ordered waters have been found in structures of membrane proteins, providing unexpected insights into the tight binding and functional roles of these molecules.

1.9.1 Lipids and Membrane Protein Regulation

1.9.1.1 Lipids and the DASS Family

Exploratory studies of lipids' effects on the function of DASS family members have been limited. Experimentally phosphatidyl glycerol (PG) and cardiolipin (CA) have been shown to increase the thermostability of LaINDY homolog VcINDY and may synergistically effect substrate binding [99], and within the 3.15Å VcINDY-MSP-Fab84 structure (PDB: 6WW5) an unidentified lipid density was seen periplasmic to the H9c helix between the scaffold and elevator domain [14] **(Fig. 1.11)**. LaINDY was simulated within a lipid bilayer with molecular dynamics in the Sauer *et* *al.* report, but lipid protein interactions were not reported [14]. Furthermore, the Cryo-EM map of human NaS1 had a modelled cholesterol, while human NaDC1 had a tightly bound but unidentified phospholipid [30]. The unidentified phospholipid bound to NaDC1 is of particular interest due to its' phosphate moiety appearing to interact with Arg108, as the mutation of this residue retards transport. Additionally, densities for several lipids were modelled at the dimer interface of the recent ScPho90 structure paper [25]. However, no attempt was made to elucidate their importance to protein function. Consequently, while several specific interactions between lipids and DASS transporters have been reported, their importance to protein folding and transporter function remains unknown.

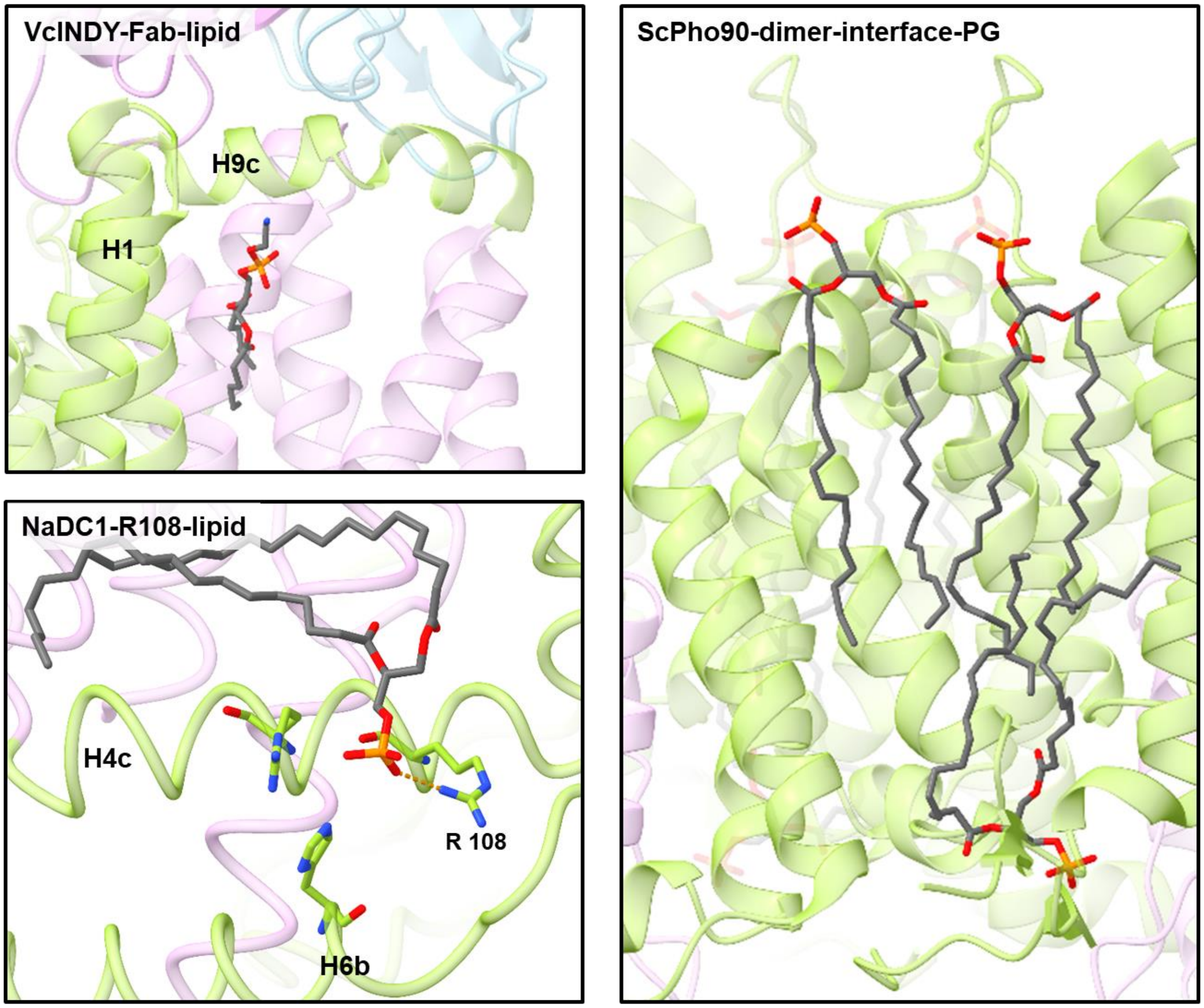


Figure 1.11 Modelled lipids in DASS structures.

The structures of VcINDY-Fab (PDB 6WW2), ScPho90 (PDB8R33) and NaDC1 (PDB 8W6G) are shown, with the elevator domains coloured pink and the scaffold domains coloured green. The Fab heavy and light chains in the VcINDY panel are coloured red and blue respectively. Lipid acyl chains are coloured grey. Otherwise atoms are coloured by heteroatom.

1.9.1.2 Phosphatidyl Inositol Sensitises *Mycobacterium tuberculosis* MscL to Membrane Mechanical Stress

MscL is a mechanosensitive ion channel highly conserved throughout the bacterial kingdom that responds to osmotic pressure by detecting tension in the inner membrane [100]. Although first discovered in *E. coli*, MscL from *Mycobacterium tuberculosis* was more amiable to crystallographic studies, and a 3.5A homopentameric structure was reported in 2006 [101]. Moe *et al.* discovered that recombinant Mt.MscL expressed in MscS/MscL double knockout *E. coli* strain MJF455, couldn't rescue cells from lysis due to osmotic shock, because it required over twice the membrane tension than when natively expressed in *M. tuberculosis* [102]. In a following report it was determined that the difference in activity was due to the lack of phosphatidylinositol (PI), an *M. tuberculosis* native lipid, in *E. coli* membranes [103]. This finding was further supported by gas phase unfolding experiments by Laganowsky et al, which showed that although all tested lipid stabilised Mt.MscL to a similar degree, PI was bound more strongly, with Mt.MscL stability increasing linearly upon binding of 4 PI molecules [104].

1.9.1.3 Aquaporin Passive Diffusion Modulated by Lipids

Aquaporins are 6 transmembrane helical membrane proteins found across all domains of life [105, 106]. The 6 helices form a pore, which facilitate the diffusion of small polar solutes such as water, glycerol, and urea across plasma membranes [107]. All aquaporins form native homotetramers, and homotetrameric aquaporins have been demonstrated to be most active oligomeric states [108].

GlpF is an aquaglyceroporin from *E. coli* facilitating the passive transport of water and glycerol across the plasma membrane [109]. A study by Klein *et al.,* in 2015 on reported that physiological concentrations of anionic lipids CA and phosphatidyl glycerol PG reduce GlpF activity, while not impacting formation of the functional tetramer [110]. More specifically, they were able to replicate the result with non-native anionic lipid phosphatidyl serine (PS) indicating the lipids electrochemistry is responsible.

Aquaporin Z (AqpZ) from *E. coli* is a "classical" aquaporin, in that it is highly selective for water only, with a 2.5Å structure reported in 2003 [111]. The gas phase unfolding mass spectrometry experiments by Laganowsky *et al.* discussed in **Section 1.9.1.2** also demonstrated CA significantly stabilising Aquaporin Z (AqpZ) with a single binding event compared to other lipids: phosphatidic acid (PA), phosphatidyl choline (PC), PS, PG, and phosphatidyl ethanolamine (PE); which instead linearly increased stability in successive binding events. Using *E.coli* Strain BKT22 [112], deficient in cardiolipin synthases A and B, Laganowsky *et al.* were able show a 2.6x reduction in AqpZ water transport, which could be recovered to near wild type levels when transformed with an additional plasmid restoring cardiolipin synthesis [104].

1.9.1.4 Phosphatidyl Glycerol Stabilisation and Allosteric Regulation of AmtB

AmtB from *E. coli* is a homotrimeric ammonium transporter from the Amt/Mep/Rh family, with homologs across all domains of life [113]. The first crystal structures solved by Khademi *et al.,* in 2004 [114], and to date there are 61 structures of AmtB homologs in the PDB but it wasn't until the previously mentioned Laganowsky *et al.* study that any lipid moieties were modelled [104] **(Fig. 1.12)**. After demonstrating significant stabilisation when binding CA and multiple PG lipids with stabilisation scaling linearly compared with PA, PS, PC, and PE, Laganowsky *et al.* solved an AmtB structure with a tenfold molar excess of PG to 2.3Å revealing eight lipids bound to the protein. The most interesting of these sites revealed the PG forming hydrogen bonds with residue Asn72 and a water bridge between the phosphate group and Asn79. An alanine double mutant of both asparagine residues was able to nullify the PG induced gas phase stabilisation. A subsequent study using solid supported membrane electrophysiology (SSME) and MD simulations demonstrated that without POPG present, AmtB is correctly folded, yet cannot complete the full substrate translocation cycle [115]. However, the Asn72/Asn79 alanine double mutant was not tested in the Mirandela *et al.* study, and there are other sites the POPG could bind from the Laganowsky *et al.* structure.

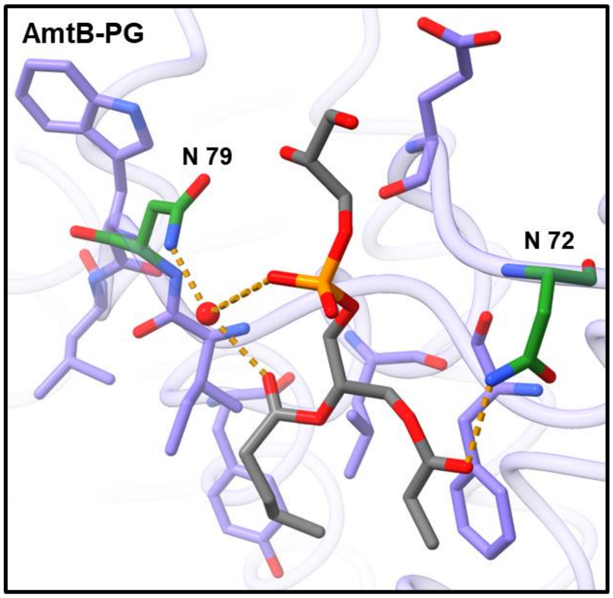


Figure 1.12 AmtB PG binding site.

A monomer of the 2.3Å AmtB structure (PDB 4NHZ) is shown coloured purple. The asparagine residues that mediate the PG binding are coloured green. The PG lipid is coloured grey. Hydrogen bonds are shown as orange dashed lines. Otherwise atoms are coloured by heteroatom.

Finally, a study determined the binding of POPG (16:0/18:1 PG) and POPE to AmtB positively allosterically regulates the formation of the complex between AmtB and inhibitor GlnK [116]. Additionally, lipid binding to AmtB was determined to be selective for both chain length and stereochemistry of the acyl tails.

1.9.1.5 Interfacial Lipids Stabilising Multimeric Proteins

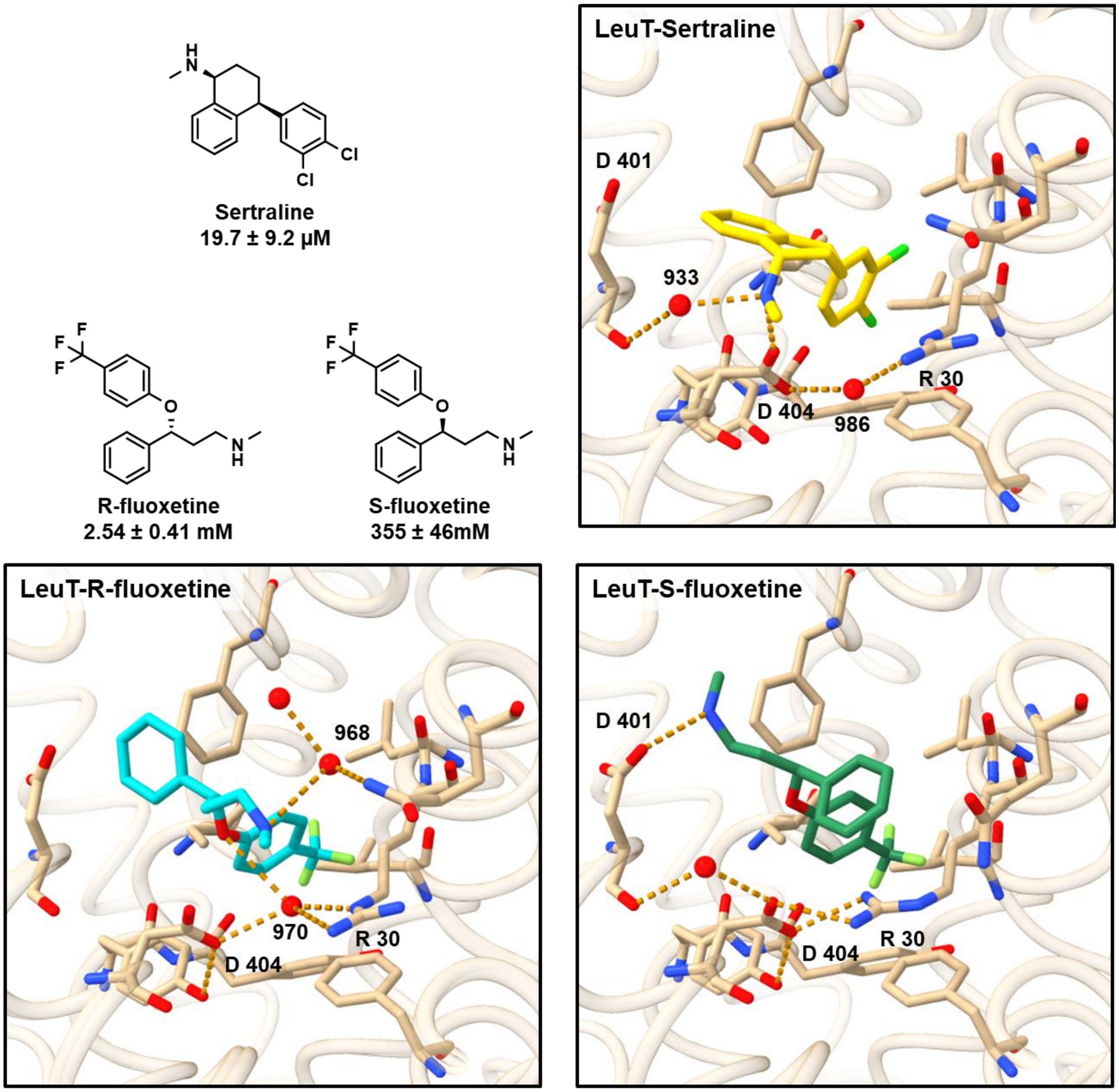
Lipids have also been shown to play a role in the multimerization of proteins. The seminal study by Gupta *et al.* in 2017 combined mass spectrometry, bioinformatics and molecular dynamics to show that altering plasma membrane lipid composition can propagate changes in oligomeric state [117]. For dimeric proteins: LeuT from *A. aeolicus* and NhaA from *E. coli*, Gupta *et al.* were able to show that the delipidation of the proteins, specifically the removal of cardiolipin, shifted the oligomeric state in favour of the monomeric form, completely abrogating the dimer in the case of LeuT. Further, using molecular dynamics they were able to show cardiolipin interacting with both monomers of LeuT at the binding interface as a bidentate ligand, with each phosphate moiety binding to basic residues K376, H377 and R506 on either side of the monomer. However, the Mancusso *et al.* VcINDY structure (PDB 4F35) was one of the 125 membrane proteins analysed in the Gupta *et al.* report [19]. Although, VcINDY was not discussed in the text, NapA from *Thermus thermophilus* has similar buried surface area and number of salt bridges, and thus similar dimer stability. Mass spectrometry of NapA revealed the dimer to be completely lipid free, indicating a highly stable dimer, which can be implied for VcINDY and possibly the whole DASS family. With the phospholipids modelled at the dimer interface of DASS-C member ScPho90**(Fig. 1.11)**, it is initially tempting to hypothesise that lipids could perform a similar function in the DASS family to that of cardiolipin in LeuT [25]. However, given the VcINDY/NapA comparison indicating a strong dimeric interface, and the fact that to my knowledge there have been no reports of DASS members in the monomeric state, lipids conferring dimer stabilisation does seem unlikely. Nevertheless, this is an interesting hypothesis to test in the futures.

1.9.2 Ordered Water Molecules Within Membrane Protein Structures

Transmembrane enzymes that require water as a substrate, such as V-type ATPases that hydrolyses ATP to translocate protons across a lipid bilayer [118], or the photosystem II complex that harnesses photons of light to oxidise water to oxygen, have either soluble subunits, or channels for water to enter the active site [119]. There are some proteins in which their primary function revolves around unique interaction with water molecules, such as the antifreeze proteins found in bacteria, fish, fungi, insects, plants, and yeast inhabiting low-temperature environments [120]. For example, the crystal structure of TmAFP, an antifreeze protein from *Tenebrio molitor*, revealed water molecules tightly bound to the structure into a polypentagonal arrangement allowing the protein to bind to unorganised ice, restricting the planes that ice crystals can nucleate, limiting the growth of ice for prolonged periods [121]. As the number of structures with sufficient resolution to resolve water molecules has increased we are finding more complex interactions between proteins and water, likely common across many protein families.

1.9.2.1 Waters as Part of a Ligand Binding Interface in LeuT and SiaP

LeuT is a leucine transporter from thermophile *Aquifex aeolicus,* and homolog of the mammalian neurotransmitter sodium symporter (NSS) family, many of which are targets for antidepressants. As LeuT is highly stable and amenable to crystallization many early structural studies on antidepressants used LeuT to elucidate information on how potential inhibitors would interact with the NSS family [122-124]. The Zhou *et al.* 2009 study, presented three high-resolution structures of LeuT with sertraline, R-fluoxetine, and S-fluoxetine non-competitively bound. Notably all three inhibitors bound in the same site, with water mediating stabilising hydrogen bonds between key binding site residues and the inhibitors **(Fig. 1.13)**. Notably, differences in water-mediated hydrogen bonds between LeuT and the compounds directly correlates with their measured inhibitory activity. Whilst there are many chemical principles that will contribute to the much higher IC_{50}, the lack of hydrogen bonding mediated both directly and indirectly by water molecules within the binding interface is clearly important.



LeuT-Sertraline

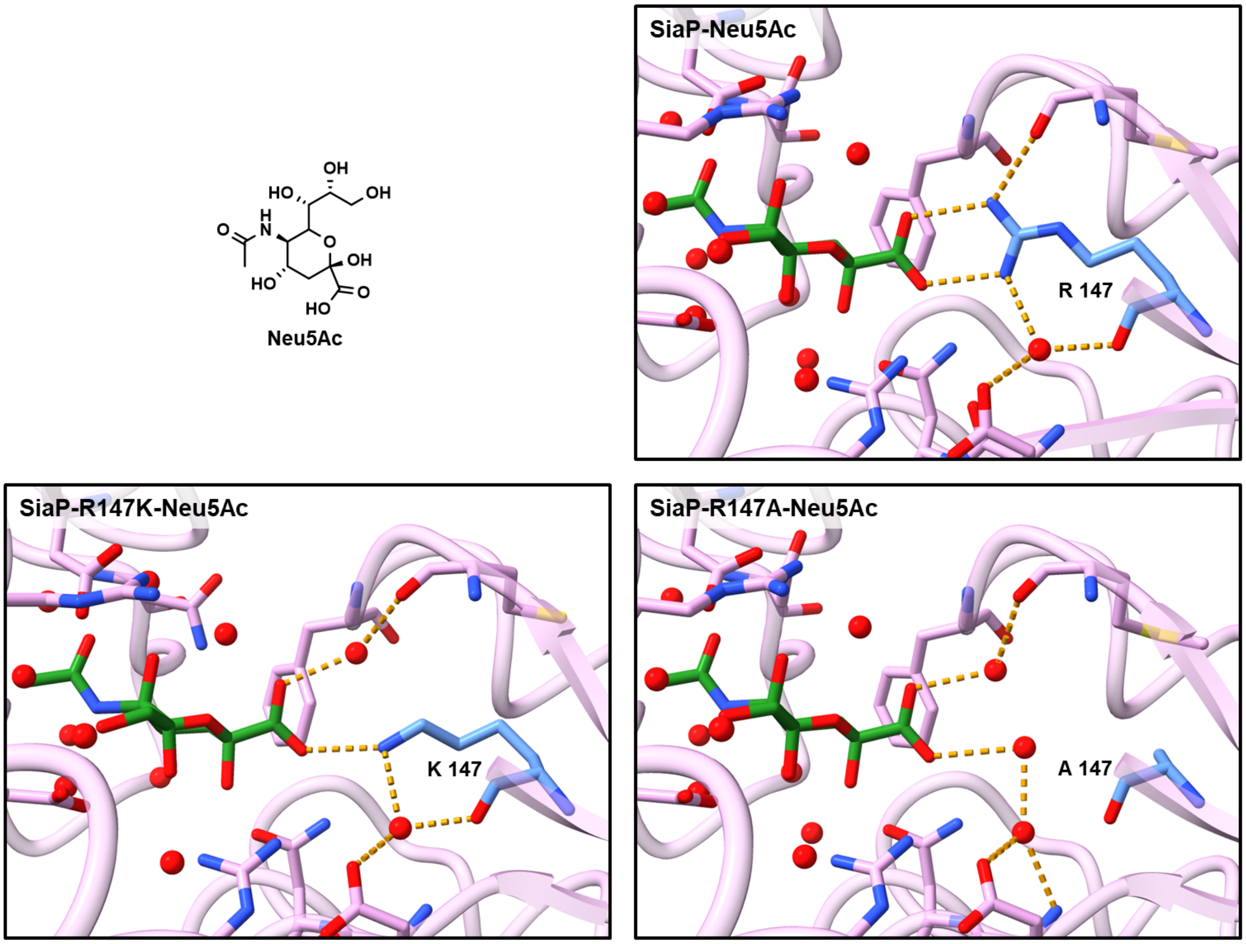
LeuT-R-fluoxetine

LeuT-S-fluoxetine

Figure 1.13 LeuT-SSRI complex binding site waters.

The structures of the SSRI's sertraline, R-fluoxetine and S-fluoxetine are shown, along with the calculated IC_{50}*s by Zhou et al. The LeuT-sertraline (PDB 36WU), LeuT-R-fluoxetine (PDB 36WV), and LeuT-S-fluoxetine are shown coloured brown. Sertraline, R-fluoxetine and S-fluoxetine are coloured yellow, cyan, and green respectively. Water molecules are shown as red spheres. Hydrogen bonds are shown as orange dashed lines. Otherwise atoms are coloured by heteroatom.*

The SiaPQM sialylic acid (Neu5Ac) transporter from *Haemophilus influenzae* is one of the most well characterised members of the Tripartite ATP-independent Periplasmic (TRAP) family that are widespread amongst bacteria and archaea [125]. TRAP transporters are comprised of three subunits: the small 4 transmembrane helix Q subunit, the large 12 transmembrane helix M subunit, and the soluble P subunit. The QM transmembrane subunits together have the same IT superfamily fold as the DASS family, unlike DASS members, the Q domains do not homodimerize like the scaffold domains of the DASS family, so functionally SiaPQM has a single elevator domain, whereas the LaINDY protomer has two elevators [126]. The soluble P domain is either free floating in the periplasm or anchored to the plasma membrane in Gram-positive bacteria, and binds the substrate with nanomolar affinity in the case of SiaP and Neu5Ac [127]. A highly conserved Arg147 residue within the SiaP binding is essential for high affinity Neu5Ac binding via the substrates dicarboxylate moiety was demonstrated by Fischer *et al.* to be essential for high affinity substrate binding [128] **(Fig. 1.14)**. In the study an *E. coli* strain with the Neu5Ac transporter deleted was cultured in minimal media with 5μM radiolabelled Neu5Ac. Rapid uptake of Neu5Ac was observed for *E. coli* recombinantly expressing wildtype SiaPQM, while the SiaP R147K mutant showed minimal uptake and R147A mutants did not. Interestingly, although transport was retarded, crystal structures of both mutants in millimolar concentrations of Neu5Ac revealed the substrate bound to SiaP with water molecules coordinating the substrate in analogous positions to the arginine amine groups **(Fig. 1.14)**.


Neu5Ac

SiaP-Neu5Ac

SiaP-R147K-Neu5Ac

SiaP-R147A-Neu5Ac

Figure 1.14 Water molecules compensating for amine groups in SiaP mutants.

The structure of Neu5Ac is shown. The structures of Neu5Ac bound: wildtype SiaP (PDB 2XWV), R147K SiaP (PDB 2XWI), and R147A SiaP (PDB 2XWK) are shown coloured pink. Neu5Ac is coloured green. The arginine 147 residue, and its R147K, and A147K mutant residues are coloured blue. Waters are shown as red spheres. The hydrogen bonds are shown as dashed lines, and for simplicity only hydrogen bonds coordinating the Neu5Ac carboxylate moiety and SiaP are shown.

Both LeuT and SiaP examples show a clear role of waters to introduce plasticity within a ligand binding interface. With the binding site for substrates within the elevator domain of the DASS family is a large solvent accessible cavity, substrates will have to displace waters to enter. Coupled with the wide variety of di and tri-dicarboxylates Sampson *et al.* have shown to stabilise VcINDY it is highly plausible that waters may be involved in the substrate recognition mechanism.

1.9.2.2 Functional Waters During the Bacteriorhodopsin Photocycle

Bacteriorhodopsin (bR) is a 7 TM helix proton pump originally discovered in the purple membranes of *Halobacterium. salinarum* [129], and drives light stimulated extracellular transport of protons due to the presence of a retinalidene chromophore covalently bonded to Lys216 (herein 'retinal SB') [130, 131]. The photocycle of bacteriorhodopsin is extremely well characterized in part due to the wealth of crystal and cryo-EM structures, from Henderson *et al.*'s first high-resolution structure in 1990 to present [132-134]. More recently the use of XFELs on room temperature crystals has allowed for time-resolved structures capturing transition states that exist for femtoseconds [135].

Although there are seven commonly agreed intermediates (I, K, J, K, L, M, N, and O) and the resting state (bR), the mechanism can be broadly classified into three steps: 'Isomerization', from all-*trans* to the 13-*cis*-retinal SB; 'Proton Release', from the retinal SB via Asp85 to the extracellular space; and 'Proton Uptake', from the cytoplasm via Asp96 to recharge the retinal SB [136].

Water molecules have been shown to be important for progression through the photocycle. During the resting state, the hydrogen bonding between water 402 and the protonated retinal SB is critical for maintaining the retinal SB's unusually high pK_a of 13.3 [137] **(Fig. 1.15)**. Further water 402 is also involved in the hydrogen bonding network along with Thr89 that stabilizes the pK_a of 2.2 for Asp85 [138]. The pK_a difference of 11 orders of magnitude between Asp85 and the retinal SB keeps the latter protonated and prevents proton leakage in the resting state [139]. Utilizing TR-SFX Nango *et al.* showed that retinal isomerization dislodges water 402 triggering the collapse of a water mediate hydrogen-bond network [139]. Subsequent structural changes reposition the protonated retinal SB in a hydrogen bonding network with a new water 452 and Thr89 that allows for transient proton transfer to Asp85 via Thr89 [135] **(Fig. 1.15)**.

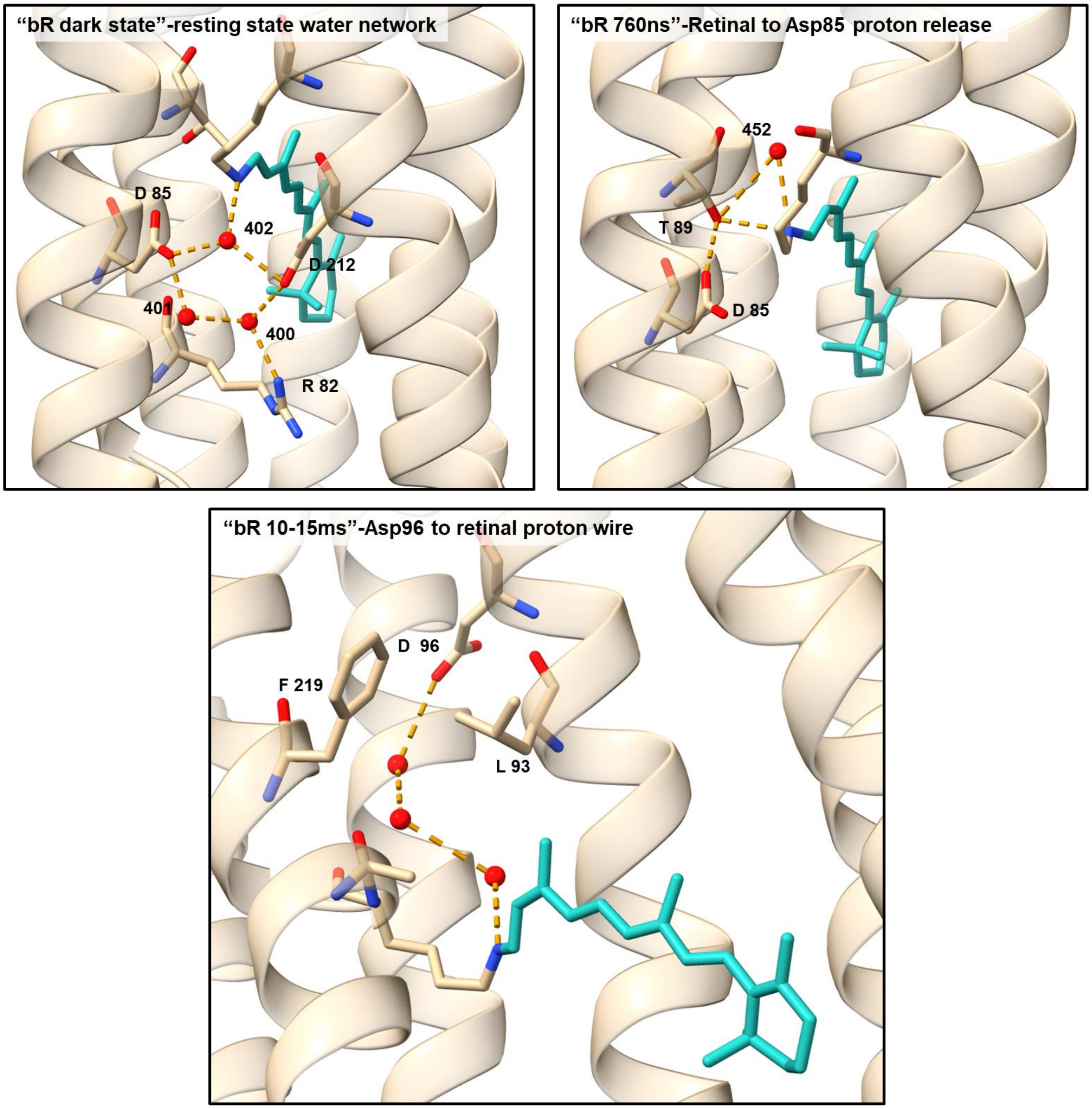


Figure 1.15 Water molecule hydrogen bonding networks throughout the bacteriorhodopsin photocycle.

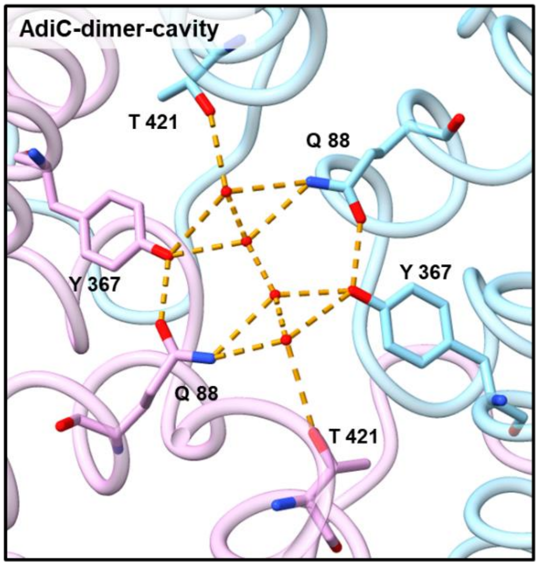
The structures of bacteriorhodopsin in: dark-state (PDB: 6RQP), and 10-15ms after photoexcitation (PDB 6RPH) from Wienert et al. 2019; and 760ns after photoexcitation (PDB 5B6X) from Nango et al. 2016) are shown coloured brown. The retinal Schiff base to lysine 216 is coloured cyan. Waters are shown as red spheres, otherwise atoms are coloured by heteroatom. In the "bR dark state", and "bR 760ns" panels helix 6 is removed for clarity, whereas in the "bR 10-15ns" panel helix 1 is removed.

Additionally, molecular dynamics simulations have implicated a transient water chain between Asp96 and the retinal SB to allow for the re-protonation via a Grotthuss proton wire [140, 141]. Weinert *et al.* demonstrated this with TR-SMX structures, showing the rotation of Leu93 and Phe219 creates space for the transient ordering of three water molecules, allowing the retinal SB to accept a proton from Asp96 over a 12Å distance [136] **(Fig. 1.15)**.

1.9.2.3 AdiC Structures Reveal Waters Within the Binding Site and Dimer Interface

A recent report on AdiC, an APC superfamily L-arginine/agmatine antiporter involved in decarboxylase-dependent acid-resistance in pathogenic *E. coli* [142], presented a 1.7Å outward-open structure with 357 modelled water molecules [143]. The structure showed some of these waters shaping and stabilizing the substrate binding site by linking different transmembrane helices via hydrogen bonding. Further, utilizing MD simulations, the authors demonstrate a specific water molecule helps to constrain the rotamer of tryptophan 293, the totally-conserved middle-gate residue that is important for substrate recognition via cation-π interaction [144, 145].

Finally, the AdiC structure also revealed a water-filled cavity at the dimer interface. These waters mediating intra-dimer hydrogen bonds between a threonine, glutamine and phenylalanine residues between the dimers **(Fig. 1.16)**. Mutating these residues to disrupt or change the hydrogen bonding capacity (Q88E, removing the amide donor; Y367F and T421V, removing hydroxyl donor/acceptors) all significantly reduced the thermostability of the AdiC dimer, indicating the waters contribute to dimer stabilization. This latter finding is particularly interesting because Sauer *et al.* found a similar cavity in the dimer interface of VcINDY with waters ordered in a square pyramidal confirmation and coordinated by a phenylalanine [14]. However, the significance of the waters within the cavity was not further analyzed in that report.



AdiC	Average Melting Temperature (°C)
Wild Type	61.1
Q88E	55.0
Y367F	51.8
T421V	57.9

Figure 1.16 Waters mediating dimer stabilisation in AdiC.

The dimer interface of the 1.7Å AdiC structure (PDB 7O82) is shown, with the A chain monomer shown as blue and the B chain monomer shown as pink. Waters are shown as red spheres. Hydrogen bonds are shown are orange dashed lines. Atoms are coloured by heteroatom. Average melting temperature of AdiC water cavity mutants calculated by Ilgü et al. are shown in the table.

1.10 Aims

This thesis aims to further understanding of the DASS family's interactions with substrates, the plasma membrane, and solvent water, through the model DASS-E member LaINDY from *L. acidophilus*. In particular I would like to push forward knowledge of the prokaryotic DASS-E's, for which research has been neglected in comparison to the DASS-C clade which contains the mammalian homologs. This will be achieved through high-resolution structural data of LaINDY in complex with substrates via x-ray crystallography and cryo-EM, and supported by substrate screening and biophysical methods.

2. Materials and Methods

2.1 Constructs and Vectors

The full-length wild type INDY construct from *L. acidophilus* (herein 'LaINDY') (ENA, accession AAV2769.1) had an N-terminal 10 Histidine purification tag with an upstream TEV site and was a gift from Professor Da-Neng Wang (Department of Cell Biology, New York University School of Medicine, New York, NY) [14].

For overexpression the LaINDY construct was cloned into a modified pET-28 vector, containing an ampicillin resistance cassette, downstream of an IPTG inducible T7lac promotor (herein 'overexpression plasmid') [146].

For growth assays the LaINDY construct was cloned into a pGEM-5Zf(+) vector (Promega #P2241), containing an ampicillin resistance cassette, downstream of a constitutively active SP6 promotor (herein 'native-expression plasmid') [147].

2.2 *E. coli* Strains

One Shot Mach1-T1R Chemically Competent *E. coli* #C862003 (herein 'Mach1') (Invitrogen) were used to for amplification and storage (as glycerol stock) of the LaINDY construct in overexpression and native plasmids.

The BL21(DE3)-LOBSTR-RIL #EC1002 (herein 'LOBSTR') strain (Kerafest), from the laboratory of Thomas U. Schwartz, PhD, Massachusetts Institute of Technology, have a chloramphenicol resistance plasmid [148].

The BL21(DE3)-R3-pRARE2 (herein 'R3-pRARE2') strain, a gift from Professor Opher Gileadi (Addgene plasmid #26242), have a chloramphenicol resistance plasmid [149].

The LOBSTR strain was used for growth assay experiments and initial overexpression for crystallization and cryo-EM. During optimization of LaINDY expression it was replaced by the R3-pRARE2 strain, which is used in the following protocol.

2.3 Chemically Competent Cell Generation

A LOBSTR or R3-pRARE2 glycerol stock was streaked on a 25ng/µL chloramphenicol LB-Agar plate and incubated 37°C, overnight [149]. A single colony was added to 10mL SOB medium (20g/L tryptone, 0.5g/L yeast extract, 0.5g/L NaCl, 0.186g/L KCl) which was split into 2x50mL falcon tubes and incubated overnight at 37°C with shaking at 250rpm. The preculture was added to 2x500mL SOB in autoclaved Ultra Yield 2.5L Flasks (Thomson), and incubated 250rpm, 37°C, until the OD_{600} was between 0.45-0.55. All subsequent steps were performed at 4°C unless stated otherwise. SOB culture medium was transferred to sterile, pre-chilled Corning 500mL PP Centrifuge tube (Corning) and centrifuged (3500xg, 10min). The supernatant was poured off and the cells were resuspended in 160mL ice-cold TFB1 (30mM Potassium acetate-acetic acid pH 5.85, 100mM RbCl, 10mM $CaCl_2$, 50mM $MgCl_2$ and 15% glycerol w/v), recombined and centrifuged (3500xg, 10min). The supernatant was poured off and the cells were resuspended in 35mL ice cold TFB2 (10mM MOPS-KOH pH 6.5, 75mM $CaCl_2$, 10mM RbCl and 15% glycerol w/v) and 500µL aliquots were flash frozen in liquid nitrogen and stored at -80°C.

2.4 LaINDY Expression

2.4.1 Heat-shock Transformation

Chemically competent *E. coli* cells were thawed on ice, while LaINDY overexpression/native plasmid was thawed from -20°C storage. 1.5µL of 100ng/µL LaINDY plasmid was pre-chilled on ice, before adding 40µL of chemically competent cells with gentle pipetting to mix, and incubating on ice for 20min. The cell-plasmid mixture was incubated in a water bath at 42°C for 45s to heat-shock the cells, before returning to ice for 2min. Under sterile conditions 160µL of SOC medium (SOB supplemented with 10mM $MgCl_2$ and 20mM glucose after autoclaving) was added to the cells, which were left rotating 250rpm, 37°C, 2h. 40µL of cells were streaked on a LB-Agar supplemented with 200ng/µL ampicillin and incubated 37°C overnight.

2.4.2 Plasmid Amplification

A single colony of Mach1 cells transformed with LaINDY was added to 10mL 2xLB broth supplemented with 200ng/µL ampicillin and left incubating at 250rpm, 37°C, overnight. The cells were harvested via centrifugation (4000xg, 4°C, 20min), and the supernatant poured off. The plasmids were then purified using the QIAprep Spin Miniprep Kit (Qiagen) following the supplied protocol, but eluted in Nuclease-free Water #BS100S (New England Biolabs) and stored -20°C.

2.4.3 Overexpression of LaINDY

A fresh transformation was performed for each expression. The LaINDY overexpression plasmid was transformed into R3-pRARE2 E. coli as previously described. All media herein was supplemented with 100ng/µL ampicillin. A single colony was added to 10mL 2xLB (20g/L tryptone, 10g/L yeast extract, 20g/L NaCl) and incubated 250rpm, 37°C, overnight. The 10mL starter culture was added to 120mL 2xLB and incubated 250rpm, 37°C, 4h. 10mL of preculture was then added to 1L AIM TB Base including trace elements (Formedium) supplemented with 1mL 10% antifoam-204 in ethanol and 1mL 1M $MgSO_4$ in autoclaved Ultra Yield 2.5L Flasks (Thomson) 250rpm. The incubator-shaker temperature was initially held at 37°C for 2h, then dropped to 25°C for 12h, and then to 18°C for 24h. Cells were harvested by centrifugation (4000xg, 4°C, 20min) and cell pellets from 2L of culture were flash frozen as aliquots in liquid nitrogen and stored -20°C.

2.4.4 Overexpression Test-Purification

A 5mL sample from each overexpression flask was taken and added to a UNIPLATE Collection and Analysis Microplate, 24-well, 10ml, natural polypropylene, round well bottom, irradiated with lid (Cytiva), which was spun 4000xg, 4°C, 20min. The supernatant was poured off and the cells resuspended in 2mL Lysis Buffer (50mM Tris-NaOH pH 8.0, 400mM NaCl, 10mM Imidazole and 5mM DL-Malic acid), supplemented with cOmplete protease inhibitor cocktail (herein 'protease inhibitors') (Roche). The resuspension was subjected to sonication on ice using a Vibra-Cell VC750 (Sonics & Materials, Inc.) equipped with a Multi-Element 24 Tip Mother Horn Assembly #4579 (Qsonica) at 35% amplitude, 5s pulse, 10s rest, for 4min. The lysed cells were transferred to a UNIPLATE Collection and Analysis Microplate, 96-well, 2ml, natural polypropylene, round well bottom (Cytiva) (herein '96-well block') and DDM >99% highly purified #NB-39-00015 (Neo-Biotech) was added to the lysed cells to a final concentration of 1%. The block was sealed with a Silicone Sealing Mat for 96 square well plates MAT-WP-96SQ (Elkay Manufacturing Company) and rotated, 4°C, 1h. The lysate was clarified via centrifugation (4000xg, 4°C, 20min) and transferred to a clean well with 100µL 50% Ni-NTA Agarose in Lysis Buffer. The '96-well block' was sealed and rotated 4°C, 1h. The resin was pelleted via centrifugation (500xg, 4°C, 2min) and transferred to a UNIFILTER Microplate, 96-well, 2 mL, 0.45 µm hydrophilic PVDF, filter bottom with long drip director (Cytiva) (herein 'filter block). The resin was washed with 40 CV Dialysis Buffer (50mM Tris-NaOH pH 8.0, 200mM NaCl, and 0.0261% w/v DDM) supplemented with 60mM Imidazole via centrifugation (500xg, 4°C, 3min) to remove excess buffer. The resin was incubated with 50µL Dialysis Buffer supplemented with 600mM Imidazole 4°C, 10min and eluted via centrifugation (500xg, 4°C, 3min) with a 15µL sample taken for SDS-PAGE to confirm LaINDY expression.

2.5 LaINDY Purification

Purification protocol adapted from Sauer *et al.*, (2020) [14]. All purification steps were performed at 4°C unless otherwise stated.

2.5.1 Membrane Isolation

R3-pRARE2 cell pellets overexpressing LaINDY were thawed in 5mL/g Lysis Buffer supplemented with protease inhibitors, and DNase using a magnetic stir bar at room temperature for 20min, then at 4°C for 20min until fully resuspended. The cell suspension was lysed by 4 passes at 1000bar through an EmulsiFlex-C5 homogenizer (Avestin, Inc.), and the lysate clarified by centrifugation (25,000xg, 20min). The clarified lysate was then ultracentrifuged (175,000xg, 3h) to pellet the membranes, which were flash-frozen in liquid nitrogen and stored at -80°C.

2.5.2 IMAC

Frozen membranes were resuspended in 10mL/g Solubilisation Buffer (50mM Tris-NaOH pH 8.0, 200mM NaCl, 10mM Imidazole, 5mM Malate) supplemented with protease inhibitors using a 40mL WHEATON® Dounce Tissue Grinder #357546 and the ‘Tight’ pestle (DWK Life Sciences). DDM was added to a final concentration of 1% and the lysate was stirred with a magnetic stirrer for 40min to maximize solubilisation. The insoluble material was removed via ultracentrifugation (175,000xg, 30min) and the resulting lysate was passed through 1mL His GraviTrap prepacked with Ni-Sepharose 6 Fast Flow #11003399 (Cytiva), 1 column/50mL lysate. The column was washed with 20CV Dialysis Buffer supplemented with 50mM Imidazole, then 20CV Dialysis Buffer supplemented with 60mM Imidazole. LaINDY was eluted incubating the resin for 5min with 2CV Dialysis Buffer supplemented with 300mM Imidazole, then a second elution after 5min incubation with 2CV Dialysis Buffer supplemented with 600mM Imidazole.

2.5.3 SEC

2.5.3.1 SEC for Crystallography

Elution fractions were desalted into Dialysis Buffer with PD-10 desalting columns packed with Sephadex G-25 resin #17085101 (Cytiva). TEV was added in a 5:1 (LaINDY:TEV) molar ratio and slowly rotated overnight. Cleaved LaINDY was passed through 1mL His GraviTrap prepacked with Ni-Sepharose 6 Fast Flow #11003399 (Cytiva) twice. The flow through was concentrated in a Vivaspin 20, 100kDa MWCO PES #28932363 (Cytiva) to approximately 500µL and centrifuged (20,000xg, 20min) to pellet aggregates. The supernatant was loaded onto a Superose 6 Increase 10/300 GL #29091596 (Cytiva) SEC column via an ÄKTA pure 25L #29018224 (Cytiva) chromatography system equilibrated with Xtal-SEC Buffer (25mM Tris-NaOH pH8.0, 150mM NaCl, 10% glycerol w/v, 1mM DTT) supplemented with 5mM substrate and NG, Anagrade N324 (Anatrace Products, LLC) to a final concentration of 0.4% w/v.

2.5.3.2 SEC for Cryo-EM

Elution fractions were combined and PMAL-C8 #P5008 (Anatrace Products, LLC) was added in a 1:5 (LaINDY:PMAL-C8) molar ratio and left rotating slowly overnight. Excess detergent was removed by adding Bio-Beads SM-2 Adsorbents #1528920 (Bio-Rad Laboratories) pre-equilibrated in Dialysis Buffer in a 1:100 detergent:BioBeads weight ratio, and left rotating slowly for 2 hours. The protein was then concentrated in a Vivaspin 20, 50kDa MWCO PES #28932362 (Cytiva) to approximately 500µL and centrifuged (20,000xg, 20min) to pellet aggregates. The supernatant was loaded onto a Superose 6 Increase 10/300 GL #29091596 (Cytiva) SEC column via an ÄKTA pure 25L #29018224 (Cytiva) chromatography system equilibrated with Amphipol-SEC Buffer (20mM HEPES-NaOH pH7.5, 150mM NaCl, 0.5mM TCEP) supplemented with 5mM substrate.

2.5.3.3 SEC for Biophysical Assays

Elution fractions were desalted into Assay Buffer (25mM HEPES-CholineOH pH 7.5, 100mM Choline Chloride, 0.0261% w/v DDM) with PD-10 desalting columns packed with Sephadex G-25 resin #17085101 (Cytiva). TEV was added in a 5:1 (LaINDY:TEV) molar ratio and slowly rotated overnight. Cleaved LaINDY was passed through 1mL His GraviTrap prepacked with Ni-Sepharose 6 Fast Flow #11003399 (Cytiva) twice. The flow through was concentrated in a Vivaspin 20, 100kDa MWCO PES #28932363 (Cytiva) to approximately 500µL and centrifuged (20,000xg, 20min) to pellet aggregates. The supernatant was loaded onto a Superose 6 Increase 10/300 GL #29091596 (Cytiva) SEC column via an ÄKTA pure 25L #29018224 (Cytiva) chromatography system equilibrated with Assay Buffer.

2.6 SDS-PAGE

Precast NuPAGE 4-12% Bis-Tris midi gels (Invitrogen) were assembled in a gel tank with NuPAGE MES SDS Running Buffer (Invitrogen). 15µL protein samples were mixed with 5µL NuPAGE LDS Sample Buffer (4x) supplemented with 30mM TCEP. 20µL was then loaded onto the gel along with 5µL Precision Plus Protein All Blue Prestained Protein Standards #1610373 (Bio-Rad Laboratories) and run at 150V for 1h. Gels were extracted from the cassettes and soaked in pre-warmed InstantBlue Coomassie Protein Stain (Abcam) to visualize the protein bands.

2.7 X-ray Crystallography

2.7.1 Sample Preparation

Peak fractions from SEC were concentrated to 2mg/mL in a Vivaspin 2, 100kDa MWCO, PES (Cytiva) and ultracentrifuged (20,000xg, 4°C, 20min) to pellet any aggregates. Crystals were grown using the sitting drop vapour diffusion method. Three drops of 2:1, 1:1, and 1:2 v/v protein:'reservoir solution' ratio, were dispensed for a final volume of 150nL by a mosquito®

Xtal3 (SPT Labtech) onto a 3-position 96-well Crystallisation Plate UVXPO-3LENS (SWISSCI) and stored in a Rock Imager 1000 (Formulatrix, Inc.) at 20°C. The crystallisation screen was designed around the conditions: 100mM HEPES pH 7.0, 25% PEG-1500; and 100mM HEPES pH 7.0, 30% Jeffamine ED-2001 and dispensed with a Formulator (Formulatrix, Inc.) using chemicals from Molecular Dimensions Limited. Crystals were mounted at room temperature using a Mounted CryoLoop (Hampton Research). No additional cryoprotectant was added before the mounted crystals were vitrified in liquid nitrogen directly from the drop.

2.7.2 Crystal Screening and X-ray Data Collection

Crystal screening was performed on the I24 microfocus beamline (Diamond Light Source, UK) set to a 0.9116Å wavelength at 100K. The crystals were located by carrying out diffraction grid scans and evaluated by collecting three diffraction images 45° apart.

2.8 Cryo-EM

2.8.1 Sample Preparation

Peak fractions from SEC were concentrated to ~2.5mg/mL in a Vivaspin 2, 50kDa MWCO, PES #28932236 (Cytiva) and ultracentrifuged (20,000xg, 4°C, 20min) to pellet any aggregates. Immediately before use, several Quantifoil R1.2/1.3 grid: Au 300 grids (Quantifoil Micro Tools) were glow discharged in air for 2min. On a Vitrobot Mark IV (FEI/Thermo Fisher Scientific) a 3µL sample was applied to each of the grids. The grids were blotted for 2.0s to 5.5s in 0.5s steps, -10 blot force, 100% humidity, 4°C and immediately plunged into liquid ethane. Grids were stored in liquid nitrogen until screening or data collection.

2.8.2 Grid Screening and Data Collection

Grid screening and preliminary data collection was performed on a 200kV Glacios Cryo-TEM equipped with a Falcon 3EC Direct Electron Detector (FEI/Thermo Fisher Scientific) at the

OPIC facility, or in the case of the 5mM α-ketoglutarate sample on a 300kV Titan Krios (FEI/Thermo Fisher Scientific) equipped with a K3 Direct Electron Detector (Gatan, Inc.) at the COSMIC facility. Grid quality was assessed on: ice thinness, and type (i.e. vitreous preferred); density, and particle distribution (i.e. minimal aggregation). High-quality grids were given mini-collections of 100-200 movies to further assess particle orientation before being greenlit for a large collection.

2.8.2.1 5mM α-ketoglutarate

A total of 4,815 movies were collected in super-resolution mode with the 300kV Titan Krios equipped with a K3 Direct Electron Detector using the EPU software (Thermo Fisher) at the COSMIC facility. Data collection parameters were as follows: 0.650Å/pixel (nominal 130,000x magnification), slit width 20eV, defocus range -0.8 to -2.2μm in 0.2μm steps, dose rate 22.8e^-/Å^2/s, and exposure time 2.2s (total dose 50.16e^-/Å^2 over 50 fractions).

2.8.2.2 5mM Malate

A total of 12,668 movies were collected in super-resolution mode with the 300kV Titan Krios (FEI/Thermo Fisher Scientific) equipped with a K3 Direct Electron Detector (Gatan, Inc.) using the EPU software (Thermo Fisher Scientific) in bin2 mode at the COSMIC facility. Data collection parameters were as follows: 0.832Å/pixel (nominal 105,000x magnification), slit width 20eV, defocus range -0.6 to -2.0μm in 0.2μm steps, dose rate 22.8e^-/Å^2/s, and exposure time 2.2s (total dose 50.16e^-/Å^2 over 50 fractions).

2.8.3 Data Processing and Model Building

Figures 2.1/2 and **Table 2.1** contain: simplified data processing flowcharts, and data collection and model refinement statistics respectively.

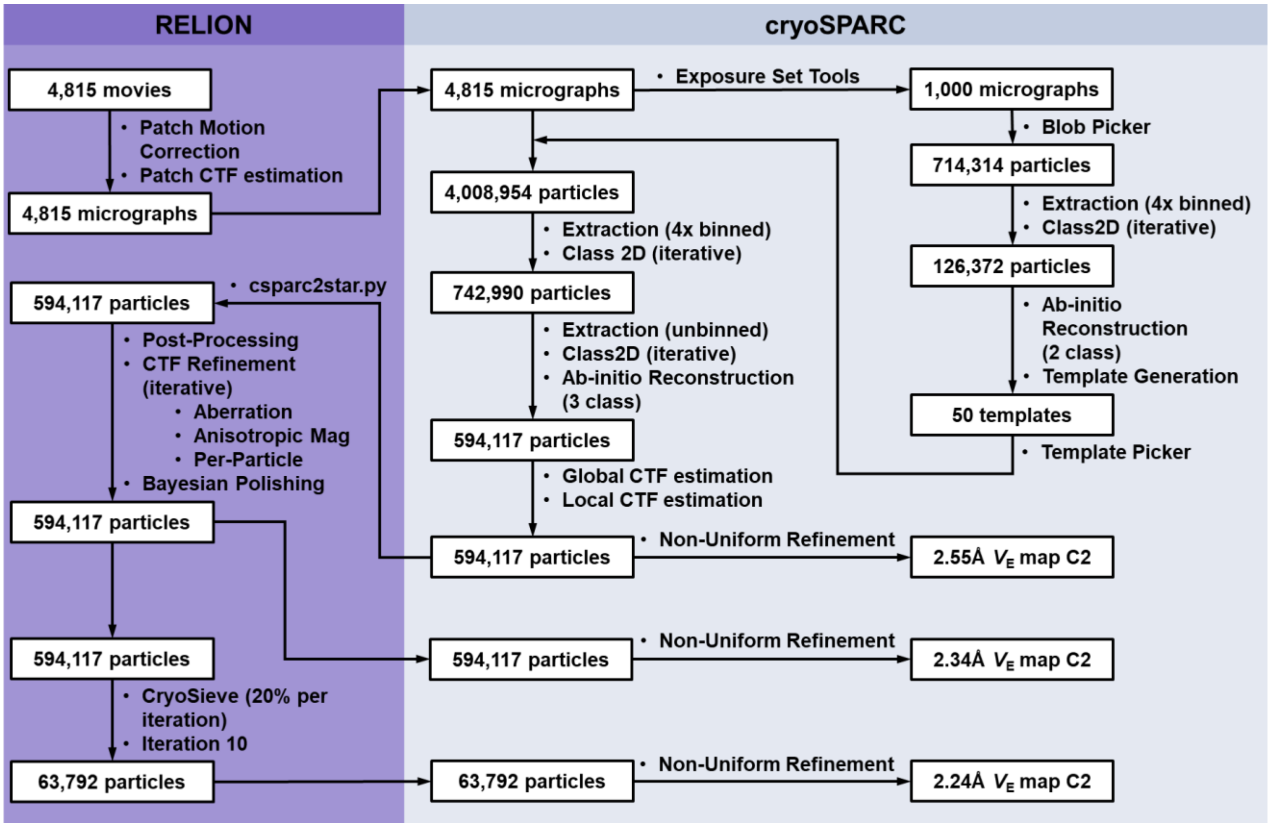


Figure 2.1 Simplified data processing pipeline for the 2.24Å α-ketoglutarate dataset.

Processes performed in RELION are within the purple box on the left, while those performed in CryoSPARC are in the grey box on the right.

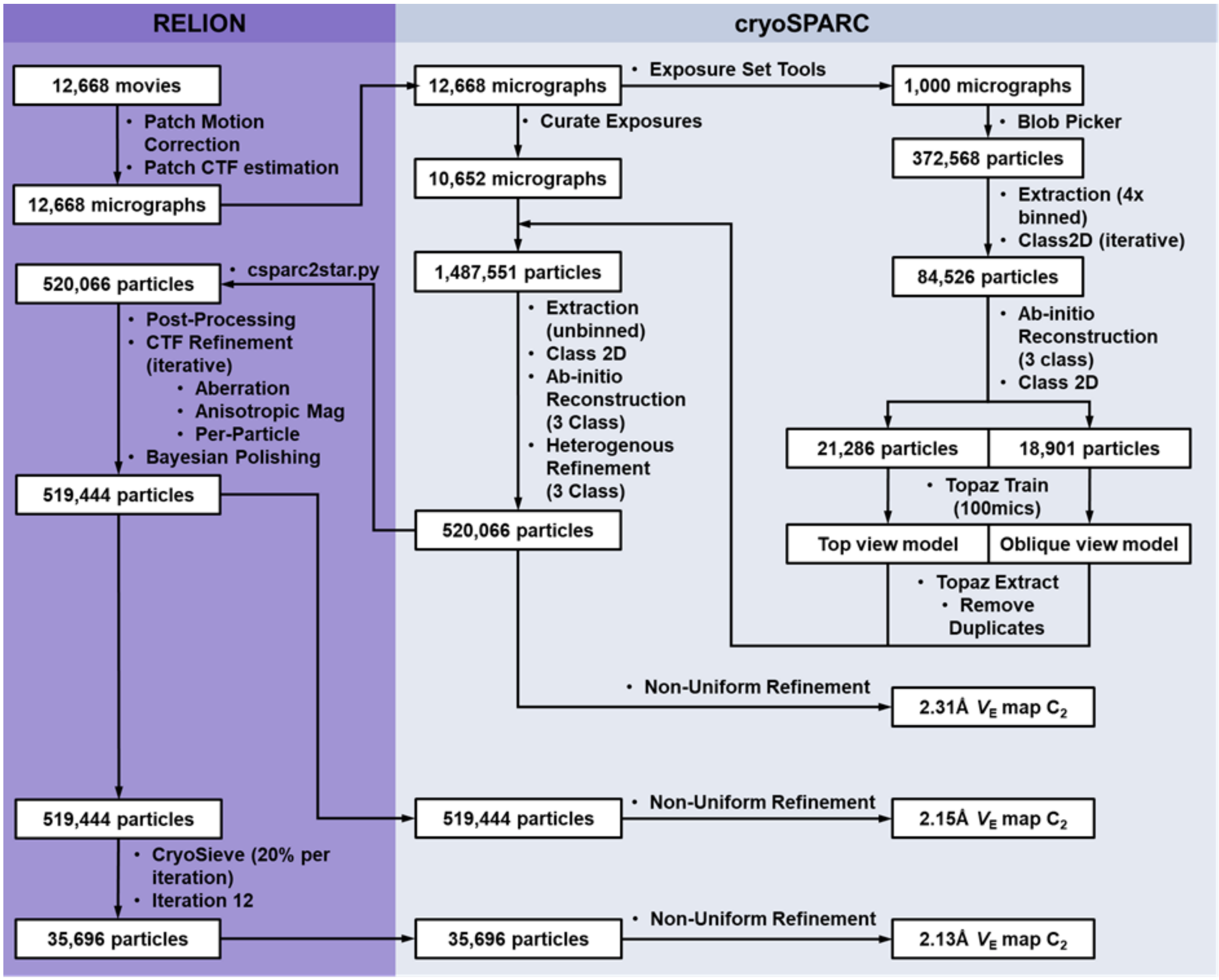
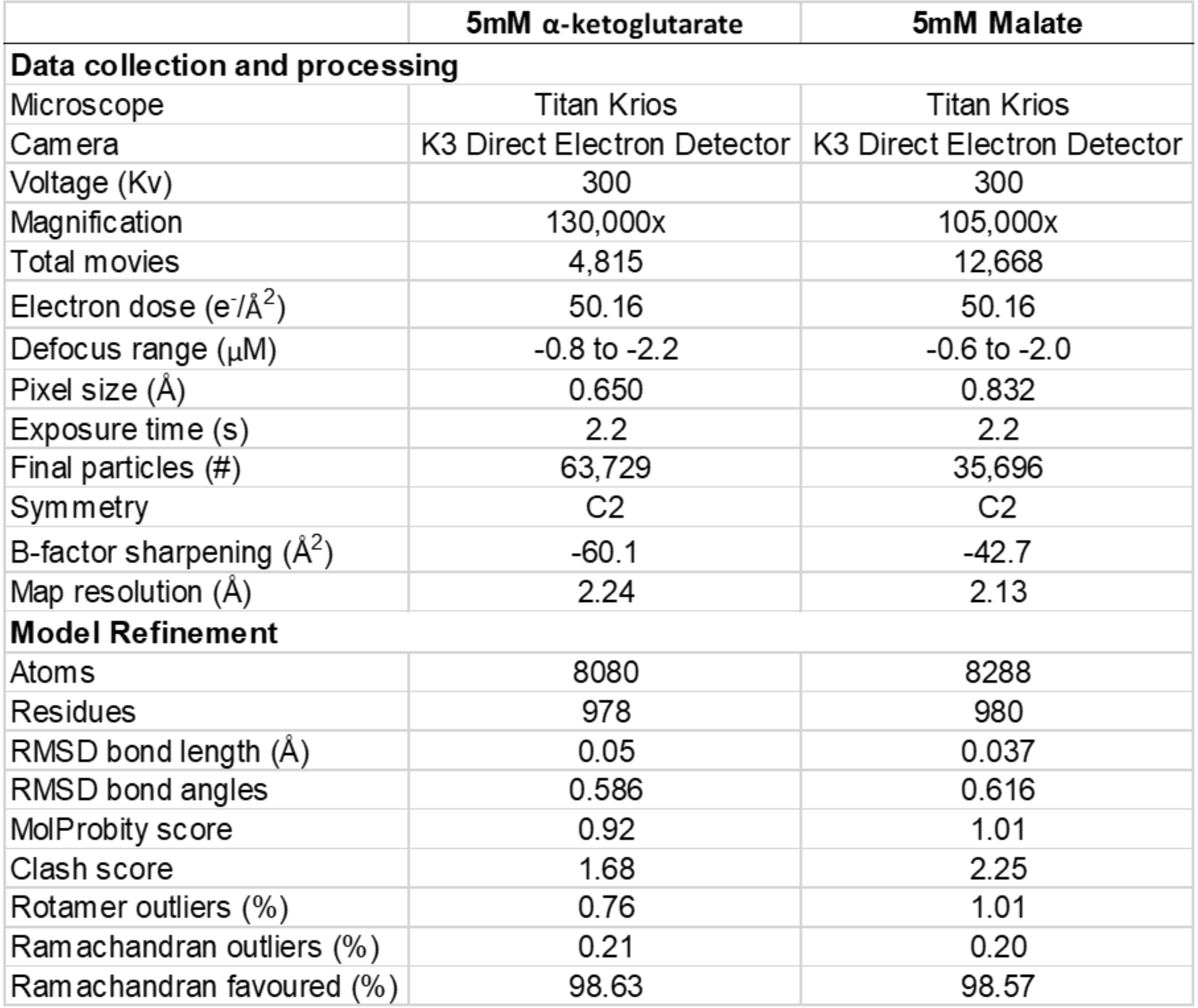


Figure 2.2 Simplified data processing pipeline for the 2.13Å malate dataset.

Processes performed in RELION are within the purple box on the left, while those performed in CryoSPARC are in the grey box on the right.



	5mM α-ketoglutarate	5mM Malate
Data collection and processing		
Microscope	Titan Krios	Titan Krios
Camera	K3 Direct Electron Detector	K3 Direct Electron Detector
Voltage (Kv)	300	300
Magnification	130,000x	105,000x
Total movies	4,815	12,668
Electron dose (e^-/Å^2)	50.16	50.16
Defocus range (μM)	-0.8 to -2.2	-0.6 to -2.0
Pixel size (Å)	0.650	0.832
Exposure time (s)	2.2	2.2
Final particles (#)	63,729	35,696
Symmetry	C2	C2
B-factor sharpening (Å^2)	-60.1	-42.7
Map resolution (Å)	2.24	2.13
Model Refinement		
Atoms	8080	8288
Residues	978	980
RMSD bond length (Å)	0.05	0.037
RMSD bond angles	0.586	0.616
MolProbity score	0.92	1.01
Clash score	1.68	2.25
Rotamer outliers (%)	0.76	1.01
Ramachandran outliers (%)	0.21	0.20
Ramachandran favoured (%)	98.63	98.57

Table 2.1 Data collection and refinement statistics for LaINDY Cryo-EM datasets.

2.8.3.1 Map Reconstruction of LaINDY in 5mM α-ketoglutarate

The 4,815 movies were imported into RELION v5.0.0 (beta) for motion correction and CTF parameter estimation, motion corrected exposures were then imported into cryoSPARC v3.3.1. A random subset of 1,000 exposures was split off and subject to blob picking, particles manually curated, and then selected 714,314 particles extracted into a 4x binned 400pixel box. The extracted particles were subjected to multiple rounds of 2D classification until top, side and oblique views were clearly visible in 17 “good” classes (126,372 particles), with obvious quaternary structure detail. An ab-initio reconstruction of the “good” classes was projected to generate 2D templates at 50 different orientations, which were then supplied to the template picker producing 4,008,954 particles from 4,815 exposures. The particles were then subjected to iterative rounds of 2D classification and class selection, until 742,990 particles in 53 classes remained. These particles were re-extracted into an unbinned 400pixel box and a multi-class ab-initio reconstruction was used to further remove "junk" particles. The 594,117 "good" particles were then iteratively globally and locally CTF-refined to correct for optical aberrations within the dataset [150]. The x and y coordinates of the particles were swapped using pyem v22.02.23 and the swapped particles were imported into RELION v4.0.0 for Bayesian particle polishing [150, 151]. The polished particles were fed into the CryoSieve 1.2.5 particle sorting algorithm which after 11 iterations yielded a final stack of 63,792 particles reporting a global map resolution of 2.24Å (Corrected Mask, FSC gold standard threshold y=0.143) from non-uniform refinement with C2-symmetry applied in cryoSPARC [152-154].

2.8.3.2 Map Reconstruction of LaINDY in 5mM Malate

The 12,668 movies were imported into RELION v5.0.0 (beta) for motion correction and CTF parameter estimation, motion corrected exposures were then imported into cryoSPARC v4.4.1 [36, 155]. A random subset of 1,000 exposures was split off and subject to blob picking and manual curation, leaving 372,568 particles which were extracted into a 4x binned 320pixel box. The extracted particles were subject to multiple rounds of 2D classification until top, side and oblique views were clearly visible in 12 classes (84,526 particles). Subsequent 3-class ab-initio reconstruction and 2D classification rendered 10 "good" classes, which were subdivided into 5 top views (21,286 particles); and 5 side or oblique views (18,901 particles). The subdivided classes were used to train two discrete Topaz neural network particle picker models on 100 exposures [34]. Topaz was then used to pick 1,485,551 particles from 10,652 exposures which were re-extracted into an unbinned 320pixel box and subjected to multiple rounds of sequential; 2D-classification, Ab-initio reconstruction and heterogenous refinement. The final heterogenous refinement had a class of 520,066 particles which reported a global map resolution of 2.31Å (FSC gold standard threshold y=0.143) from non-uniform refinement with C2 symmetry applied [154]. This class of particles had x and y coordinates swapped using pyem v22.02.23 and were imported back into RELION v5.0 (beta) for post processing [151]. The imported particles then had beam tilt, trefoil, 4th order aberrations, and anisotropic magnification refined before per-particle defocus, and astigmatism per-micrograph fitting. The particles were then subject to Bayesian particle polishing [150]. The polished particles were fed into the CryoSieve 1.2.5 particle sorting algorithm which after 12 iterations yielded a final stack of 35,696 particles reporting a global map resolution of 2.13Å (Corrected Mask, FSC gold standard threshold y=0.143) from non-uniform refinement with C2-symmetry applied in cryoSPARC [152-154].

2.8.3.3 Model Building

The Sauer *et al.* 2020 2.86Å LaINDY structure (PDB: 6WTW) was refined against the sharpened Cryo-EM maps in PHENIX 1.21.1-5306 using real-space refinement, and then iteratively rebuilt between WinCoot 0.9.8.93 and PHENIX using both the sharped and unsharpened maps [156-158]. Ligands had restraints generated using SMILES string inputs into PHENIX-eLBOW [159]. Once water molecules were built into the structure the Metric-Ion-Classification was run to probe potential ion sites and validate waters [160].

2.9 Biophysical Assays

2.9.1 NanoDSF Thermal Shift Assay

All thermal shift assays were performed on the Prometheus™ NT.48 (NanoTemper Technologies). LaINDY purified in either DDM or PMAL-C8 in Assay Buffer was concentrated to an absorbance measurement of approximately A_{280} 1.0. The LaINDY sample was then diluted 1:1 in buffer containing the potential interacting partner for a final A_{280} of 0.5 and incubated at room temperature for 10min. A 10µL sample was loaded into Prometheus NT.48 Capillaries, which were placed in the instrument. The intrinsic fluorescence of the samples at 330 and 350nm was recorded while heated from 20-95°C at a rate of 1°C/min. The T_m was determined by the inflection point from the first derivative of the 350:330nm ratio using the PR.ThermControl software (NanoTemper Technologies) **(Fig. 2.3)**.

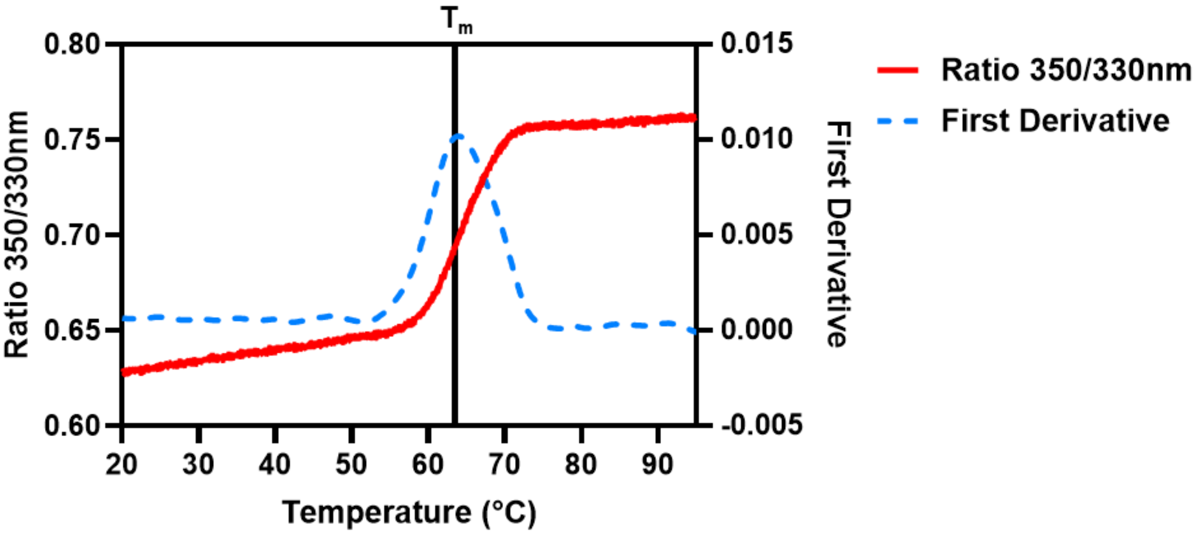


Figure 2.3 NanoDSF data processing example.

The 350/330nm ratio measurement is represented as a red line and uses the left y-axis. The first derivative of the 350/330nm ratio measurement is represented as a blue-dashed line and uses the right y-axis. The melting temperature or T_m *is determined by the peak of the first derivative and represents the 'inflection point' of the 350/330nm ratio curve, and is highlighted by a black vertical line.*

2.9.2 Tryptophan Fluorescence Quenching Assay

All measurements were made via a Horiba FluoroMax-4 (Horiba Scientific), using a Hellma-QS 10mm path length, 3500µL quartz cuvette #110-10-40 (Hellma GmbH & Co. KG) at 20°C. Excitation was set to 285nm and the emission window 310-400nm using slit widths of 2nm. The emission max of LaINDY was determined to be 330nm. LaINDY previously purified in Assay Buffer was diluted to 2µM into Assay Buffer to a working volume of 1900µL. After each addition of substrate, the solution was stirred for 3min, then allowed to settle for 30s before the measurement was taken. Substrate was added until a final concentration of 800mM was reached, and each experiment was repeated 3 times. Raw fluorescence data was corrected for assay volume change, then the corrected 330nm peak values were used to plot and fit binding curves in GraphPad Prism 10.3.0 **(Fig. 2.4)**.

2.10 TLC

All steps involving chloroform transfer or evaporation were performed under a fume hood.

2.10.1 Lipid Extraction

500µL of amphipol-purified LaINDY at 1mg/mL was thawed from -80°C storage. In a glass tube the LaINDY sample was combined with methanol and chloroform in a 1:2:1 v/v/v LaINDY:methanol:chloroform ratio, for a final volume of 2mL, and incubated at 60°C, 30min. 500µL of chloroform and 500µL of 0.3% w/v NaCl in water were added and the mixture was centrifuged 3000xg for 5min to induce phase separation. The upper aqueous layer and protein flocculent were carefully removed with a pipette, before using a nitrogen stream under a fume hood to evaporate the chloroform and methanol to leave a dried lipid film on the glass tube.

2.10.1 Lipid Separation

A solvent mixture of a 65:25:10 v/v/v chloroform:methanol:“acetic acid” ratio was made in a large glass tank [161]. A sheet of blotting paper was placed in the tank to improve equilibration and separation consistency, and the tank was allowed to equilibrate for 30min. Dried lipids extracted from LaINDY were resuspended in 100µL choloroform and spotted on to a 20x20cm Silica gel on TLC Al foils 60805-25EA (herein TLC plate) (Sigma) along with lipid standards for: DDM (Neo-Biotech); PMAL-C8 (Anatrace Products, LLC); *E.coli* CA #841199, *E.coli* PE #840027, *E.coli* PG #841188, and *E.coli* Polar Lipid Extract #100600 (Avanti Polar Lipids, Inc). The TLC plate was cut to allow for at least 13mm of solvent front migration, and then run in the pre-equilibrated glass tank. The TLC plate was air-dried and stained with 10% w/v PMA in ethanol, which was then developed with a heat gun until black spots for the lipids were revealed.

2.11 Sequence Alignments

Sequence alignments were either performed in Clustal Omega v1.2.2 via the EMBL Job Dispatcher [162, 163] for whole family alignments, or PROMALS3D for the structure-guided alignments [164, 165]. All sequence alignments and phylogenetic trees were processed and displayed used Jalview v2.11.4.1 [166].

3. Improving the Yield of LaINDY for Characterization Experiments

3.1 Aims

In reproducing the Sauer *et al.,* expression protocol for LaINDY [14], originally adapted from the original VcINDY structure paper from Mancusso *et al.,* in 2012 [19], I encountered multiple issues. Most importantly, the yields were low for both the cell mass of LaINDY transformed LOBSTR-RIL cells and final protein quantity. As an example: the purification p005, chronologically the first purification used for crystallisation trials, yielded only enough LaINDY to set up a single 96-well plate at 2.5mg/mL from 60µg of LaINDY from 6 litres of cells. This level of productivity would not only slow down structure determination efforts, but also the type of assays we could develop.

This chapter will detail key changes made to the overall protocol to improve productivity, with the secondary goal of reducing the time burden of the purification. It is important to note that almost every step of the protocol was looked at, from the number of passes through the C5 homogeniser to the duration of centrifuge spins. Experiments were often done by splitting a single purification into several parallel purifications at the appropriate step, keeping the number of variables as low as possible.

3.2 Expression

3.2.1 *E. coli* Strain Swap from LOBSTR-RIL to R3-pRARE2

Initial expression of LaINDY was performed in the BL21(DE3)-LOBSTR-RIL strain of *E. coli* [148], the closest cell line to in our possession to the rare codon-containing strain from the protocol provided to us by *Sauer et al.* LOBSTRs contain not only extra copies of: argU, ileY and leuW rare codon tRNAs [167], but also genetically modified copies of ArnA and SlyD [168] [169] [170], two histidine rich native proteins that often co-purify through IMAC. A notable missing feature of the strain is inherent bacteriophage resistance, an issue which had impacted our lab a year prior to embarking on this study. Due to the fact that we had completed our first lab relocation, our LOBSTR cell stock had not been opened since purchase, and all other currently used bacteria strains had been phage-resistance for a year, it was decided that the LOBSTRs lack of phage resistance was low-risk.

During expression e009 however, 20L of LaINDY transformed LOBSTRs after appropriate growth for 2 hours (final OD_{600}=0.150 from a starting 0.020), reported an OD_{600}=0.008 at 15h indicating the cells were dead **(Fig. 3.1A)**. Suspecting a potential lytic phage contamination, a double-layer plaque overlay assay was performed according to the *Kropinski et al.,* protocol [171, 172]. The assay result was negative, with no plaques forming even in the undiluted drop and a full lawn of bacteria, and to date a plausible explanation for the growth failure has not been suggested.

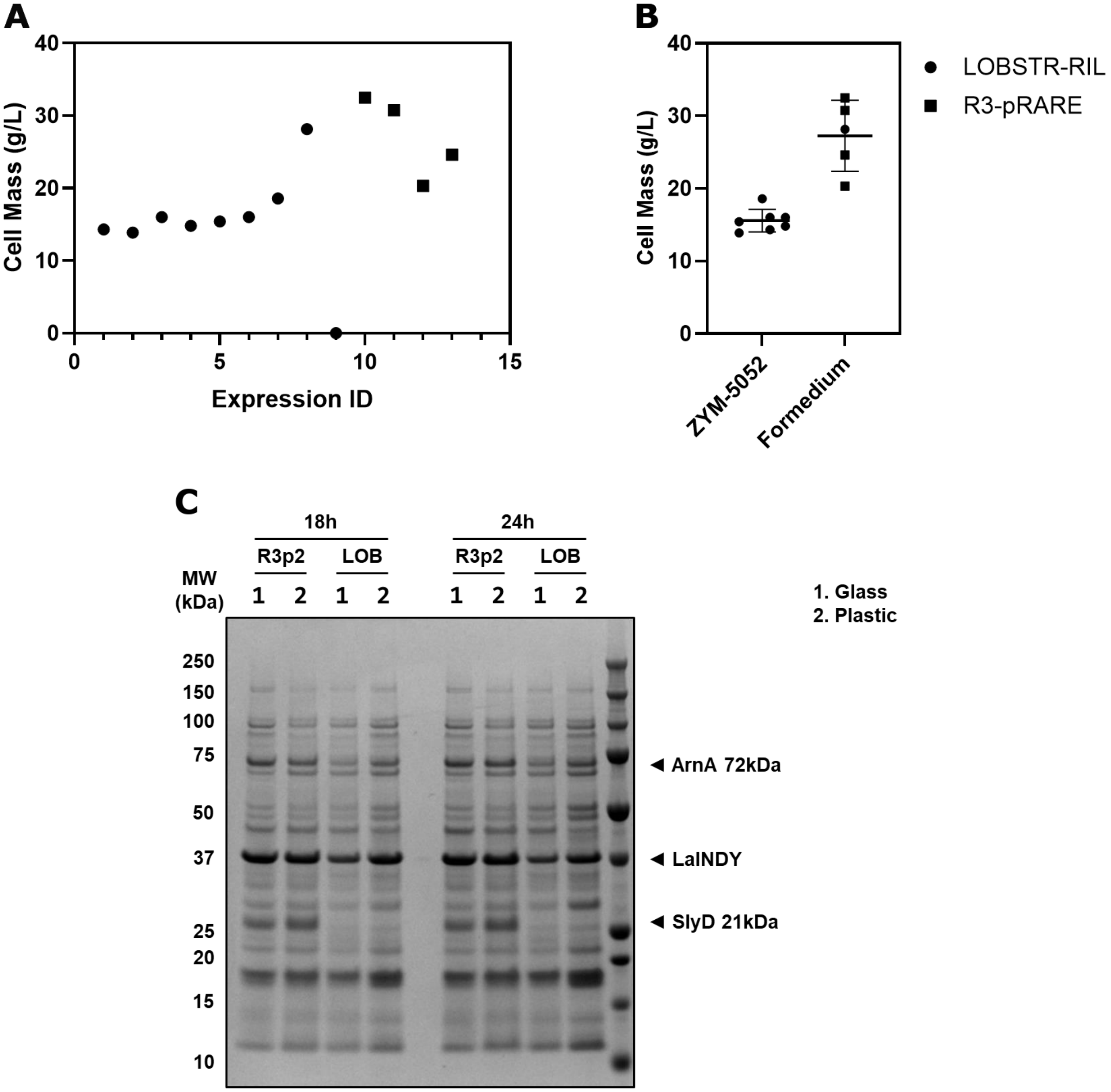


Figure 3.1 Expression yields of LOBSTR-RIL and R3-pRARE2 BL21 E. coli transformed with LaINDY.

A. *Mass of cells per litre of transformed cell culture. Purifications are in chronological order by expression ID. LaINDY expression in LOBSTR-RIL cells are represented by circles, while R3-pRARE expressions are squares.* ***B.*** *Scatterplot of A, grouping expressions by the media used. Failed expression e009 has been removed. Reported averages are 15.5g/L for ZYM-5052 and 27.26g/L for Formedium AIM TB. LOBSTR-RIL and R3-pRARE expressions are represented as in A.* ***C.*** *SDS-PAGE analysis of the expression test between the LOBSTR (LOB) and R3-pRARE2 (R3p2).*

I decided that rather than further probing the exact cause of the expression failure it would be quicker to circumvent future issues by reviving our phage-resistant BL21(DE3)-R3-

pRARE2 strain [149], removing a subset of labelled flasks from circulation for LaINDY-only work, and incorporating a rigorous cleaning procedure (Chemgene scrub, hot water soak, 2% glacial acetic acid soak, wet autoclave) between each expression.

During this time I implemented a test expression purification, adapted from my time as a technician for the Carpenter group working on the baculovirus-Sf9 expression trial pipeline [173, 174], so a failed or poor expression could be identified and discarded rather than wasting multiple days purifying low yielding cells. The procedure is described in full in **Section 2.4.4** and can be complete on the day of harvest in approximately 5 hours.

Comparison between the LOBSTR and R3-pRARE2 test expression SDS-PAGE gel appears to indicate that the R3-pRARE2 cell line produces LaINDY slightly better than the LOBSTR, based on the size of the gel bands in **Fig. 3.1C**, although no further effort was made to further quantify or explore this observation.

3.2.2 Auto Induction Media: ZYM-5052 to Formedium AIM TB Base

By far the biggest improvement to the cell mass per litre was the decision to switch from making ZYM-5052 autoinduction media in house [175], to using Formedium-manufactured AIM TB Base including trace elements supplemented with $MgSO_4$ and antifoam-204 **(Fig. 3.1B)** [176]. ZYM-5052 produced 15.58 ± 0.59 g/L while the Formedium-manufactured AIM TB produced 27.26 ± 2.19 g/L. A hypothesis for the greater variability of the Formedium data could be due to Maillard reaction by-products formed due to the reducing-sugars and amine sources being autoclaved together. The temperature ramp of the autoclave depends upon the quantity and type of material within, thereby creating an uncontrolled variable for the time the media is exposed to elevated temperatures. Therefore, introducing a variable time period in which the Maillard reaction can occur between reducing sugars and the amines in the media and hence a variable final concentration of sugar. This is not the case in the ZYM-5052, in which the tryptone and yeast extract are sterilised separately from the 5052 sugars. This hypothesis would be relatively simple to test, and could result in a more consistent cell mass.

3.3 Purification

From cell harvest to SEC purified protein our reproduction of the *Sauer et al.,* protocol took three days: one for membrane isolation, the second for solubilisation and IMAC purification, and the third for SEC. The second day of purification includes, a 3h batch bind and 1h dialysis which can result in a greater than 10h work day. As this lengthy process increases the likelihood of errors it was decided that streamlining the protocol, along with identifying points of failure, would improve the final protein yield of the purification procedure.

3.3.1 Solving the Precipitation Issue Post TEV Addition

After the completion of several purifications using the *Sauer et al.* protocol the steps that appeared to have the most negative effects on final yield, where the protein frequently precipitated, were the overnight TEV cleavage and preceding dialysis steps **(Fig. 3.2A)**. Analysing the precipitated protein, the majority of the protein within the lane is clearly the 27kDa TEV band, indicating that the TEV is precipitating at a greater rate than LaINDY. Standard protocols recommend a 3h dialysis for common salts [177], so we hypothesised that incomplete dialysis of the 450mM imidazole was causing the TEV to precipitate overnight. In support of this hypothesis, no precipitation was observed in the LaINDY amphipol purification, which omits both dialysis and TEV cleavage steps **(Fig 3.2B)**. In keeping with this hypothesis, the precipitation issue was entirely removed after switching to Cytiva PD10 columns for more rapid and complete buffer exchange. Additionally, as equilibration of the columns can be done during the IMAC elution phase, this reduces lab time on the second day by 50min.

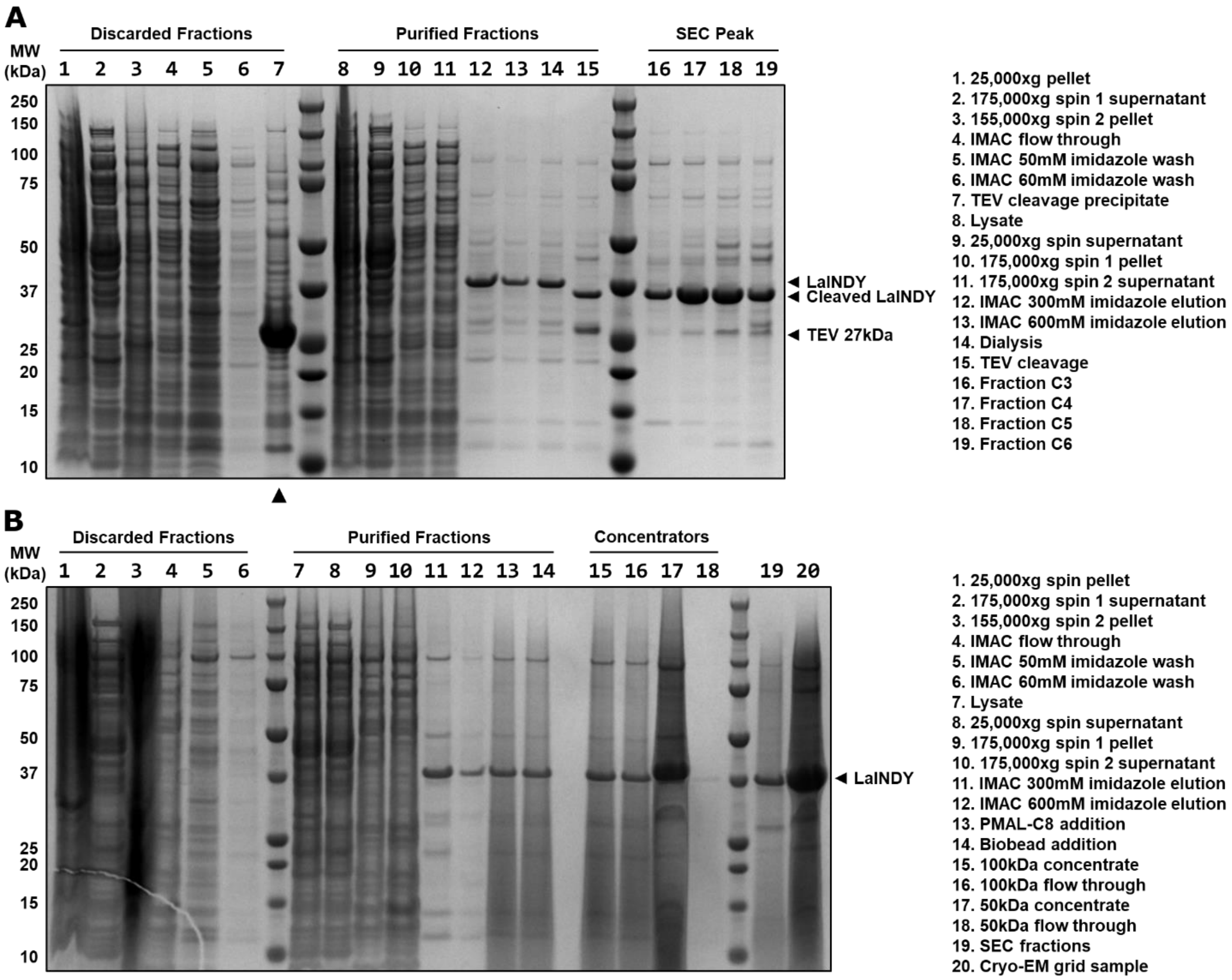


Figure 3.2 SDS-PAGE analysis highlighting precipitation of TEV.

A. *Purification p004, DDM/NG detergent reconstitution. The LaINDY, cleaved LaINDY, and TEV bands are highlighted with arrows. Lane 7, the pellet of the precipitate found after the addition of TEV, is highlighted with a vertical arrow.* ***B.*** *Purification p012, PMAL-C8 amphipol reconstitution.*

3.3.2 IMAC Bind

During the second day, the IMAC bind was a significant time expense, and point of protein loss due to the transferring of Nickel resin between multiple containers. Our crystallography method development group often performs up to 24 parallel 1L purifications utilising prepacked 1mL Nickel GraviTrap columns from Cytiva [178], and so switching to a batch bind would both drastically reduce the purification time, but also may improve the overall yield of the protein due to the removal of any resin transfer.

To test this, a single 200mL purification was split into four and either run through a 1mL prepacked gravity column 1, 2, or 3 times, or rotated with 1mL of resin for 3h. The resin was washed according to the standard protocol and then elution fractions were analysed via SDS-PAGE **(Fig. 3.3)**. Gravity flow was more productive than the batch bind, especially when running the lysate over the column twice, while also over 2h faster than the previous method.

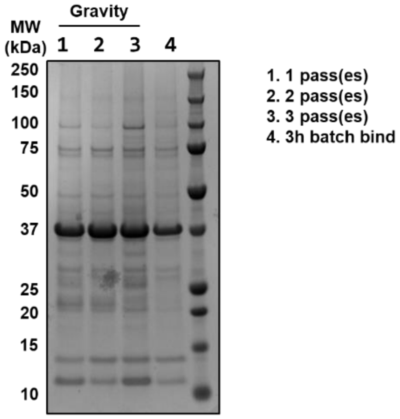


Figure 3.3 SDS-PAGE analysis of IMAC binding techniques.

Elution fractions of 50mL of resuspended LaINDY membranes either passed through a 1mL prepacked gravity column the specified number of times, or batch bound for 3h.

3.4 Discussion

The major changes mentioned have resulted in a huge improvement in LaINDY yield **(Table 3.1)**. Applying moving averages to the first eight non-zero purification yields (p001-p015) against the final seven (p016-p022) gives values of 1.26µg/g and 7.19µg/g of LaINDY per gram of cells respectively. In real terms, this means a 2L starting purification volume could produce enough protein for the comparatively demanding 1.9mL at 2µM for the tryptophan fluorescence requiring 203µg of protein. Additionally, the inclusion of a reverse purification step after TEV cleavage has significantly improved the purity, meaning the final samples should give more consistent results between purifications **(Fig. 3.4B)**.

					Membrane Mimetic		SEC Peak Fractions		
ID	Volume	Cell Mass	Substrate	Concentration	Solublisation	SEC	Volume	Measured A280	Yield per Cell Mass
	(L)	(g)		(mM)			(mL)		(µg/mg)
p001	2	28.64	malate	5	DDM	NG	n/a	n/a	n/a
p002	4	57.28	malate	5	DDM	NG	n/a	n/a	n/a
p003	4	57.28	malate	5	DDM	NG	n/a	n/a	n/a
p004	8	111.04	malate	5	DDM	NG	0.050	11.870	2.36
p005	6	96.00	α-ketoglutarate	5	DDM	NG	0.030	4.530	0.62
p006	6	88.98	α-ketoglutarate	5	DDM	NG	0.030	4.530	0.67
p007	8	123.20	n/a	n/a	DDM	NG	0.200	2.268	1.62
p008	8	128.24	malate	5	DDM	NG	n/a	n/a	n/a
p009	8	148.56	malate	5	DDM	NG	n/a	n/a	n/a
p010	6	61.30	malate	5	DDM	NG	n/a	n/a	n/a
p011	8	148.56	malate	5	DDM	NG	n/a	n/a	n/a
p012	3	84.48	malate	5	DDM	PMAL-C8	0.050	5.870	1.53
p013	3	84.48	α-ketoglutarate	5	DDM	PMAL-C8	0.020	2.210	0.23
p014	6	168.96	α-ketoglutarate	5	DDM	PMAL-C8	1.000	0.602	1.57
p015	8	259.84	α-ketoglutarate	5	DDM	PMAL-C8	1.000	0.859	1.46
p016	4	129.92	n/a	n/a	DDM	DDM	1.000	2.030	6.89
p017	4	129.92	n/a	n/a	DDM	DDM	1.000	2.227	7.56
p018	2	61.48	succinate	50	DDM	PMAL-C8	1.000	0.898	6.44
p019	2	61.48	n/a	n/a	DDM	DDM	1.000	0.823	5.90
p020	6	121.80	n/a	n/a	DDM	DDM	0.275	6.633	6.60
p021	5	101.50	n/a	n/a	DDM	DDM	0.300	5.498	7.17
p022	5	101.50	n/a	n/a	DDM	LMNG	0.300	7.488	9.76

Table 3.1 Purification yields for all LaINDY purifications.

Purifications with no yield value either precipitated before SEC (p001-p003), or contained no LaINDY due to expression issues (p008-p011).

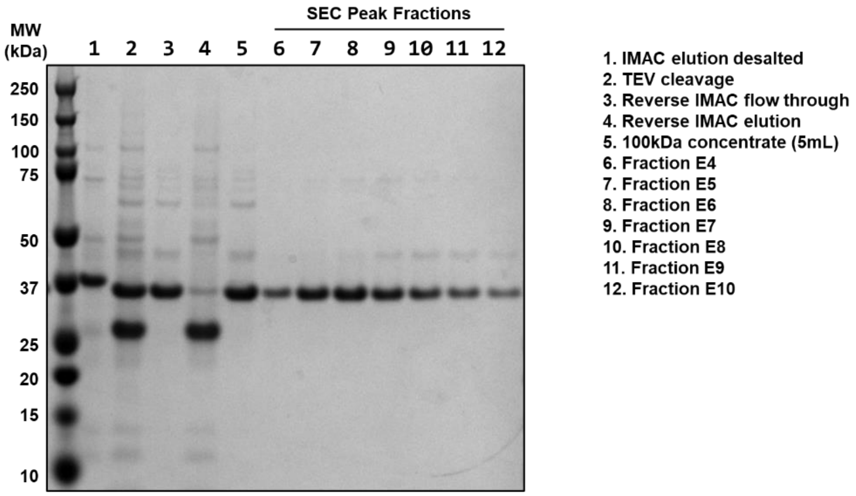


Figure 3.4 SDS-PAGE analysis of final LaINDY purification.

SDS-PAGE analysis of LaINDY purification p022, highlighting the purity of the final SEC sample.

Ultimately, my revisions to the *Sauer et al.* protocol have yielded a method for routinely and consistent LaINDY purification in sufficient quantities for all planned experiments including X-ray, Cryo-EM, and biochemical binding studies. I would also argue that this method could be used as a starting point for the purification of all bacterial DASS transporters, and my optimization work creates a template for improving those target-specific protocols. Should expression efficiency be comparable to that seen with LaINDY, then parallel purifications of multiple targets should also be possible, which could potentially greatly accelerate the understanding of the DASS family.

4. High Resolution Structures of LaINDY with Poor Substrate Occupancy

4.1 Aims

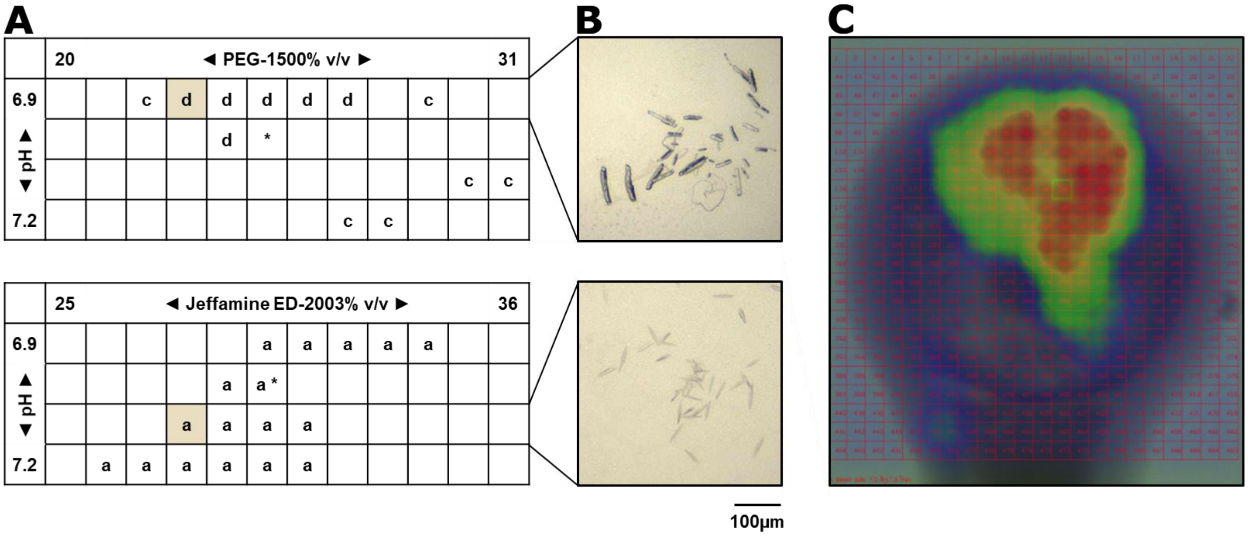
As introduced in **Section 1.8.5**, the previous structures of LaINDY has poorly resolved density for the substrate in the binding site. As such we still lack understanding of which, if any, key residues are involved in substrate recognition. I aim to solve substrate-bound structures of LaINDY to ~2.5Å via X-ray crystallography or Cryo-EM, to model sidechain and backbone interactions with a range of the confirmed substrates malate, α-ketoglutarate, and succinate. Understanding key interactions with substrates will not only aid in competitive inhibitor design but also increase our understanding of the DASS-E clade.

4.2 Reported Crystal Reproduction

Through private communication with Professor Da-Neng Wang we had information on two LaINDY crystallization conditions, (100mM HEPES pH 7.0, 25% PEG-1500; and 100mM HEPES pH 7.0, 30% Jeffamine ED-2001; both with 2mg/mL protein at 20°C) which gave incomplete datasets reaching 2.2Å. This resolution is well within the range to build small molecule ligands such as the TCA cycle intermediates transported by LaINDY. So, if these conditions could reliably produce these highly diffracting crystals we could rapidly produce high-resolution substrate-bound LaINDY datasets, skipping much of the laborious crystal condition screening. I therefore decided reproducing these crystals would be the first priority.

4.2.1 Optimization Screen Design

The expression and purification of LaINDY was reproduced as described in **Sections 2.4 & 2.5**. Notably despite the comparatively low target protein concentration of 2mg/mL, with 25µL of protein required by the mosquito Xtal3 for our 150nL 3-drop standard protocol, the yield was such that I only had enough protein for a single plate. With that in mind I designed an optimization screen 'LaINDYNYUhits' around the two conditions on a single 96 well plate **(Fig. 4.1A)**. At the time our lab was not in possession of the Jeffamine ED-2001 precipitant, but did have ED-2003 from Molecular Dimensions. The Hampton Research store page lists Jeffamine ED-2003 as a synonym of ED-2001, likely the two reagents are different nomenclatures for the same chemical.



	20	◄ PEG-1500% v/v ►										31
6.9			c	d	d	d	d	d		c		
▲					d	*						
pH ▼											c	c
7.2								c	c			

	25	◄ Jeffamine ED-2003% v/v ►										36
6.9						a	a	a	a	a		
▲					a	a *						
pH ▼				a	a	a	a					
7.2		a	a	a	a	a	a					

Figure 4.1 Crystallisation of NG purified LaINDY via sitting drop vapour diffusion.

A. *Optimisation screen 'LaINDYNYUhits' based around the two conditions that reported the highest diffraction reflections. The 96 well screen is split in with 48 wells per each of the original conditions (100mM HEPES pH 7.0, 25% PEG-1500 and 100mM HEPES pH 7.0, 30% Jeffamine ED-2003), which are denoted with an asterisk '*'. Letters indicate conditions in which crystals were detected, and further 'a', 'c', and 'd' represent the respective 1:2, 1:1, and 2:1 protein:'reservoir solution' ratio in the drops. Conditions in which example crystals are displayed in B are coloured tan.* ***B.*** *Crystals of LaINDY in 100mM HEPES pH 6.9, 23% PEG-1500 (top panel) and 100mM HEPES pH 7.1, 28% Jeffamine ED-2003 (bottom panel).* ***C.*** *Grid scan of the 100mM HEPES pH 6.9, 23% PEG-1500 in a loop conducted at the I24 beamline (Diamond Light Source, UK). The heatmap indicates the crystal within the vitreous well solution.*

4.2.2 Sitting-Drop Vapor Diffusion Crystallisation

A single sitting-drop vapor diffusion crystal tray was set up for LaINDY purified in Xtal SEC Buffer supplemented with 5mM α-ketoglutarate and 0.4% NG, using the LaINDYNYUhits screen and stored in the 20°C Rock Imager 1000 (Formulatrix, Inc.), and regularly imaged to monitor crystal growth. Crystals were visible around both target conditions at 7 days, and seemed to stop growing at 14 days. The PEG-1500 conditions grew crystals that were typically rods, and preferred drops with the higher 1:1 and 2:1 protein:’reservoir solution’ ratios whereas those found in the Jeffamine ED-2003 were small needles and only grew in the lower 1:2 protein:’reservoir solution’ ratio **(Fig. 4.1A/B)**. Both conditions gave small crystals, with only a few of the PEG-1500 rods exceeding 100µm and the all of the needles were closer to 50µm.

The crystals were harvested at room temperature with no additional cryo-protectant, as both PEG-1500 and Jeffamine ED-2003 precipitants are both medium molecular weight precipitants with cryoprotectant properties [179], and are present at concentrations sufficient for cryoprotection [180]. It has been previously noted that vapor diffusion membrane protein crystals are typically more fragile than those of soluble proteins, where detergent micelles reduce the number of protein-protein crystal contacts possible and resulting in loose-packing crystals with high solvent content [181]. Consistent with this, both the PEG-1500 rods and Jeffamine ED-2003 needles felt particularly fragile, with some of the larger rods starting to dissolve during the fishing process. Nevertheless, crystals were successfully mounted in cryo-loops and vitrified in liquid nitrogen.

4.2.3 Crystal Screening at Diamond Light Source

The vitrified crystals were dry-shipped to the microfocus I24 beamline (Diamond Light Source, UK) which has the capability of 7x6µm minimum focused beam size, a necessity for extracting the best diffraction from membrane protein crystals [182]. A diffraction grid scan was performed to locate the crystals within the loop **(Fig. 4.1C)**, and the crystals were evaluated from three diffraction images 45° apart. The highest resolution reflection reached 23Å and were from the 100mM HEPES pH 6.9, 24% PEG-1500 2:1 protein:’reservoir solution’ drop. This poor diffraction, significantly worse than privately communicated, could be due to a multitude of factors. Firstly, I noticed the cleavage of the 10xHis tag was incompletely removed during purification, and this heterogeneous protein would interfere with lattice formation. Secondly, having little experience with crystal mounting, slow crystal fishing and transfers will could have dehydrated the crystals. Finally, the 24% PEG-1500 condition is just below the aforementioned concentration for cryoprotectant properties, so it is entirely possible the decision to neglect additional cryoprotectant was a mistake.

4.2.4 X-ray to Cryo-EM

While I was not expecting to reach the 2.2Å on the first attempt, the large difference between reported and measured diffraction resolution suggested significant optimisation steps would be required, and potentially starting back at the core-screen stage. The former requiring a two-week period for the crystals to grow each iteration, and the latter an unknown number of iterations and growth periods. Due to this significant burden to improve diffraction, and the poor yield of the LaINDY purification at the time, I was decided to abandon Crystallography in favour of Cryo-EM. The Sauer *et al.*, 2020 paper also provided methodology for solving LaINDY structures via Cryo-EM utilising amphipol as the membrane mimetic, achieving a resolution of 3.09Å [14]. If reproducible, this offered a route to high-resolution LaINDY structures that would avoid the optimizations of both protein purification and grid quality. Furthermore, in the four years since the work of Sauer *et al.*, a new generation of electron detectors allows for the collection of larger datasets faster [33], novel software releases [152], and improvements to all major data processing software led us to believe that the jump in resolution would be achievable [35, 36].

4.3 Revised Cryo-EM Purification

4.3.1 2.24Å Map of LaINDY with 5mM α-ketoglutarate

LaINDY was purified according to the refined protocol adapted from Sauer *et al.* (2020) as described in **Sections 2.4 & 2.5**. In brief: LaINDY was extracted using DDM, and exchanging into

PMAL-C8 amphipol overnight following IMAC purification. After removing the excess detergent with activated Bio-Beads, the protein was concentrated for SEC. I noticed the difference in Stokes radius between a LaINDY-DDM micelle and the LaINDY-PMAL-C8 amphipol membrane mimetic was significant enough that the majority of the LaINDY in amphipol would pass through the 100kDa concentrator used for the crystallography purification. Therefore, all concentration steps with PMAL-C8 purified LaINDY utilised 50kDa concentrators from here onwards **(Fig. 3.2B/4.4B lanes 9-12)**. PMAL-C8 LaINDY was then run on a Superose 6 Increase 10/300 GL column in Amphipol-SEC Buffer supplemented with 5mM substrate **(Fig. 4.2A/B)** and further concentrated to 2.71mg/mL before centrifugation at 20,000xg for 20 minutes to pellet any aggregates.

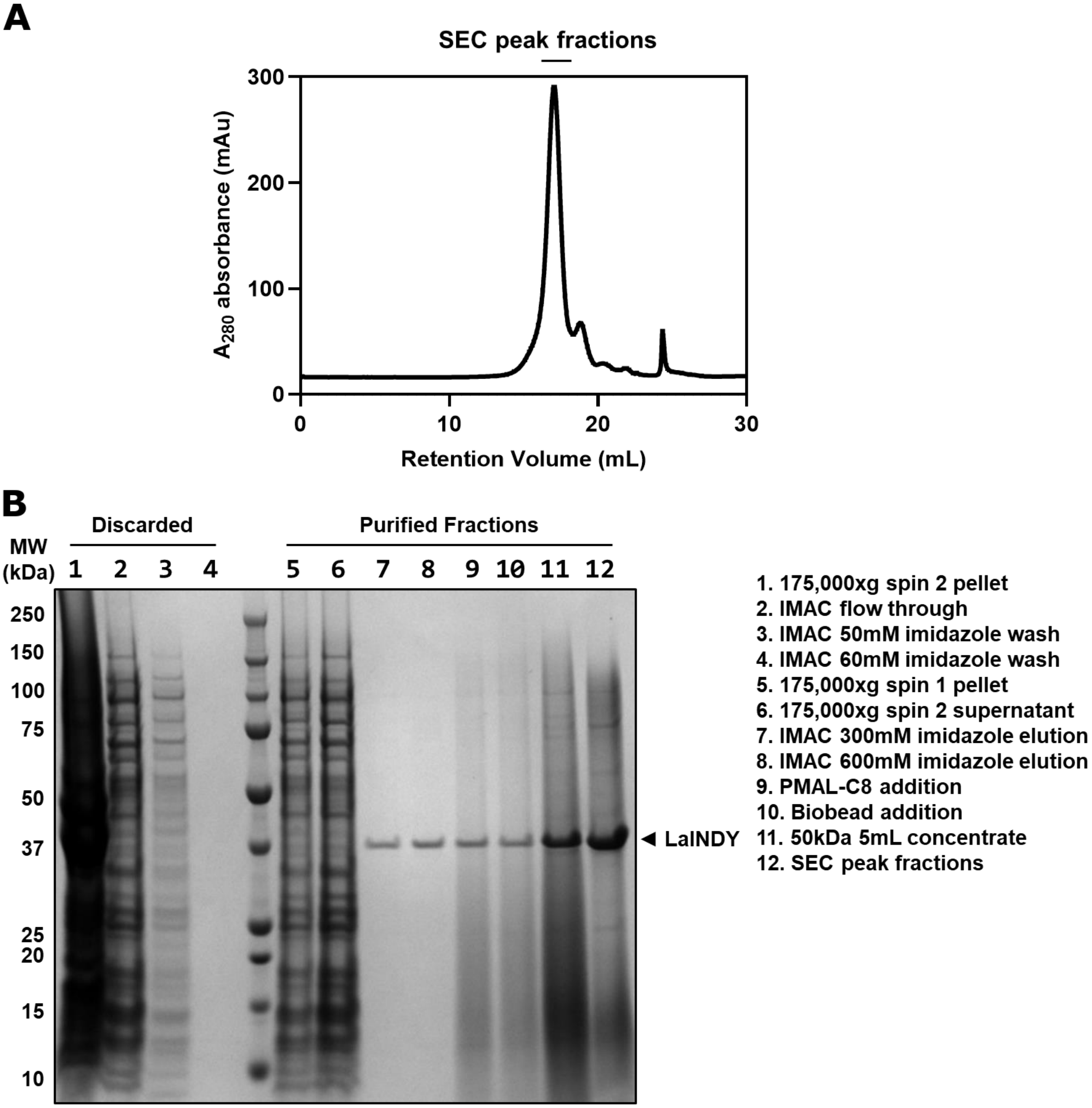


Figure 4.2 5mM α-ketoglutarate grid, SEC trace and SDS-PAGE analysis.

***A.** Elution profile for PMAL-C8 purified LaINDY from a Superose 6 Increase 10/300 GL column (purification p015). **B.** SDS-PAGR analysis of the purification, the corresponding LaINDY band is highlighted with an arrow.*

3µL of the purified protein was applied to a glow discharged Quantifoil R1.2/1.3 grid: Au 300 grid, and plunge frozen in liquid ethane using a Vitrobot Mark IV. The blot times sampled were between 2-5s in 0.5s increments. The screening of these grids took place on the OPIC Titan

Krios-Falcon 4i (Oxford, UK) during a collection trip for another project, in which the combination of a 5 image per hole collection method and the K3 Direct Electron Detector saw the microscope reporting 600+ movies per hour. This strategy resulted in running through others' collection quality grids quicker than anticipated and having time more hours of sufficient time to collect a modestly sized dataset for the LaINDY project. Unfortunately screening revealed that the blots were ineffective and inconsistent for the frozen grids, with the majority of nice hole too thick for collection. Only the 2.5s blotted grid had enough holes for collection **(Fig 4.3A)**. The protein in the usable grid squares was nicely distributed with minimal aggregation **(Fig 4.3B)**, so a collection route was planned via the few available holes and a small 4,815 movie dataset was collected at 0.650Å/pixel, slit width 20eV, defocus range -0.8-2.2μm in 0.2μm steps, dose rate 22.8e^-/$Å^2$/s, and exposure time 2.2s.

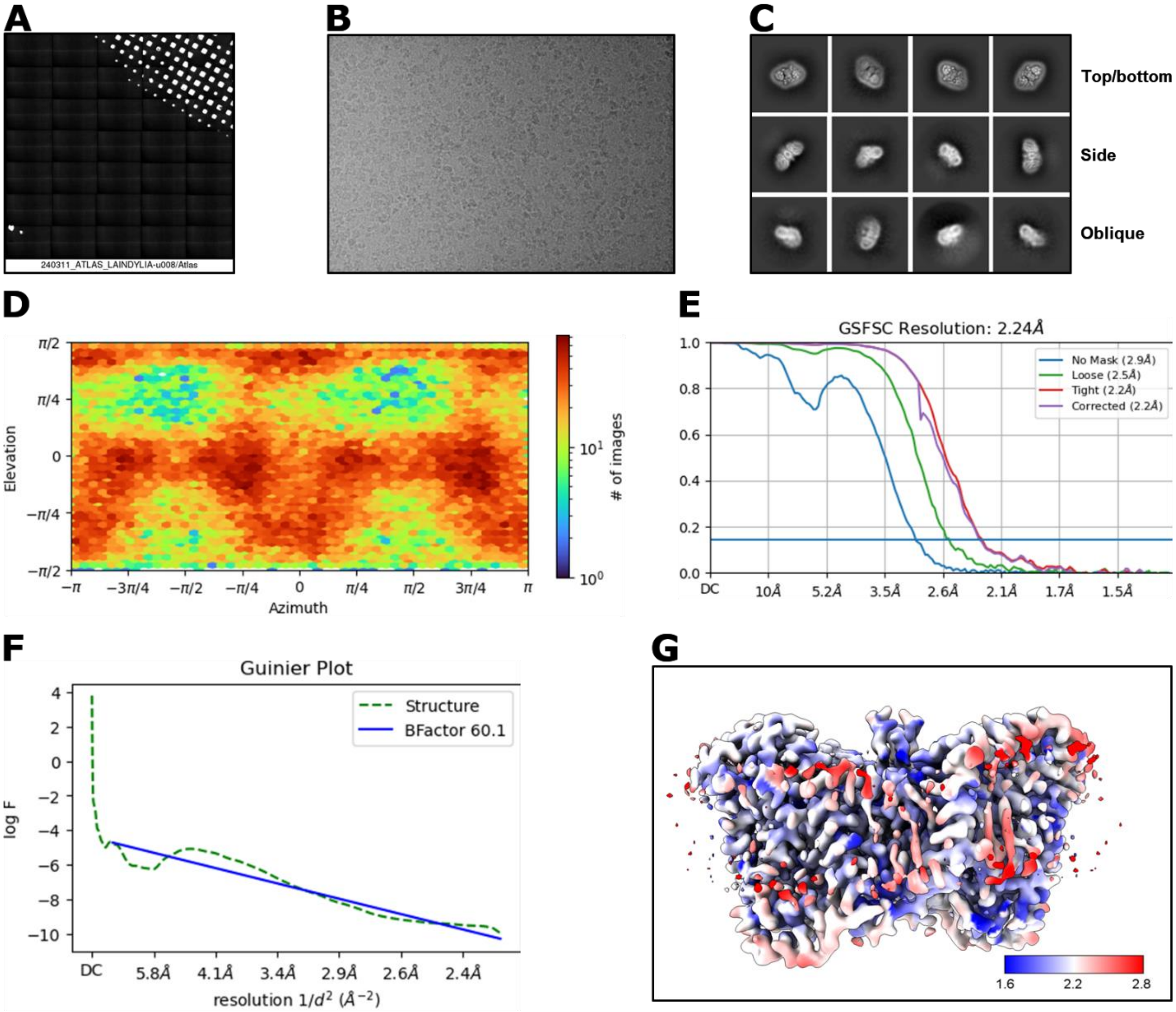


Figure 4.3 5mM α-ketoglutarate grid, microscope session to dataset.

***A.** ‘Atlas’ of the Cryo-EM grid the dataset was collected from, collectable holes are displayed as white. **B.** Representative micrograph acquired using a K3 Direct Electron Detector at 130,000x magnification. **C.** Representative 2D classes, grouped by their relative orientation. **D.** Orientation distribution heatmap for the final dataset (63,792 particles). **E.** Plot showing the FSC of the final dataset indicating a global map resolution of 2.24Å. **F.** Guinier plot showing the B-Factor used to generate the sharpened map with which the structure was built. **G** Unsharpened map coloured with a local resolution estimation done via cryoSPARC. The FSC threshold was set to 0.143.* Figures F-H are outputs from cryoSPARC,

The movies were motion corrected in RELION v5.0.0 (beta) before being imported into cryoSPARC v3.3.1 for the bulk of the data processing pipeline. The reasons for this were two-fold. Primarily; the final step of our standard processing pipeline employs a Bayesian Particle-Polishing algorithm which is not available in the version of cryoSPARC we possess, and would require the movies to be re-motion corrected by RELION if this step was carried out in cryoSPARC. Secondarily; it is well documented that radiation damage effects negatively charged residues within protein structures first [183], and reducing the electron dose of the final reconstruction by excluding later frames of each collected movie can reduce the effects of radiation damage with minimal loss of overall resolution [184]. We theorised that this technique could be employed to improve the density of LaINDY's dicarboxylate substrates, and restrictive dose reconstruction is not a feature in cryoSPARC v3.3.1.

After importing into cryoSPARC, I sub sectioned 1000 micrographs which were subjected to blob picking and yielded 714,314 particles. These particles were extracted, binned 4x, and subjected to multiple rounds of 2D classification, resulting in 126,372 particles in well-defined classes **(Fig 4.3C)**. An ab initio reconstruction of the best resolved particles, judged by its apparent likeness to previous LaINDY structures was then used to generate templates at 50 orientations, which were supplied to the template picker to pick particles from the whole dataset (4,815 micrographs).

The template picker found 4,008,954 particles, which were then extracted in an unbinned 400-pixel box, subjected to 2D Classification, then multiple rounds of 3 class ab-initio reconstruction with C1 symmetry (no symmetry) applied. One of the *ab initio* classes was intended to collect junk-particles and discarded after each iteration. The quality of the particle subset was evaluated empirically based on when the two non-junk classes converged, possessing similar percentage split of the input particles and identical orientation distributions.

The final subset of 594,117 particles then had global and local CTF parameters estimated, iteratively fitting the higher order; tilt, trefoil, spherical, tetrafoil, and defocus optical aberrations [185]. Ewald sphere curvature was ignored, despite the cryoSPARC guide recommending considering it for all sub 3Å datasets. This was decided based on DeRosier approximated resolution of phase errors due to the Ewald sphere, predicted from the protein diameter and electron wavelength [186]:

$$\frac{1}{R} = \sqrt{\frac{2 * 0.7}{t\lambda}}$$

Where R is spatial frequency, t is protein diameter (~98Å at longest axis), and λ is electron wavelength (0.0197Å at 300Kv), the mass of LaINDY gives an approximate resolution of 1.17Å where Ewald sphere curvature should be considered. However, the best LaINDY reconstructions were significantly above the DeRosier threshold at 2.64Å and 2.52Å with C1 and C2 symmetry applied respectively. Nevertheless the 2.64Å C1 symmetry map is already a 0.2Å improvement over the previous highest resolution LaINDY structure (PDB 6WTW) [14].

In order to proceed with the Bayesian polishing in RELION the '.cs' particles were converted to the RELION readable '.star' format via Asarnow's pyem suite [151], and as the patch motion correction was initially performed in RELION only the coordinates of the particles required adjustment. Following coordinate adjustment global and local and per particle CTF parameter aberrations were re-estimated and additionally anisotropic magnification using RELION v5.0.0 (beta)'s post-processing algorithms, and a subset of 10,000 particles was used to train the Bayesian polishing algorithm to output polishing parameters to be used on the whole dataset. The particles were then polished and re-imported back into cryoSPARC for a non-uniform refinement which reported an improved global map resolution of 2.34Å with C2 symmetry applied.

At this point we became aware of a recent publication by Zhu *et al.* (2023) reporting a novel particle sorting algorithm (CryoSieve) that iteratively reconstructs the map with reduced numbers of particles in an attempt to estimate the practical minimum number of particles required to achieve the given resolution of the dataset [152]. The authors demonstrate the "final stacks" of particles for three of the eight datasets analysed approach the theoretical particle minimum proposed by Henderson to allow for atomic reconstruction [187]. As a result, they propose that CryoSieve could therefore be used retroactively to quantitively evaluate sample preparation techniques. In five of the eight datasets the CryoSieveing procedure improved global map resolution by 0.16Å overall so we were interested as to whether it could be introduced into our data processing pipeline. As CryoSieve does not align particles in between each 3D reconstruction the mask created by the most recent cryoSPARC non-uniform refinement was used for the reconstructions. We used the default parameter of an 80% retention rate per iteration, and changed the high-pass cut-off frequency to 2Å. To determine the success of the algorithm each iteration subset of particles was imported back into cryoSPARC and subject to a non-uniform refinement with C2 symmetry applied to directly compare with the Bayesian polished set. The 10th iteration was the highest reporting global map resolution at 2.24Å with a B-Factor of -60.1 from 63,792 particles **(Fig 4.3E/F)**, the resolution a significant 0.1Å improvement over the pre-sieved subset of particles, and high enough resolution to see solvent, ligand and annular membrane lipid densities. The spread of particle orientations was similar to those of the previous structures **(Fig 4.3D)**, and interestingly consistent across the CryoSieve iterations, indicating that the high-resolution information within the dataset was not from a particular viewing angle.

4.3.2 LaINDY in the Outward Open State

The unsharpened map clearly shows LaINDY is in the outward-open state **(Fig 4.3G)**, as it has exclusively been reported in up to this point, and there was no hint of multiple conformations during the dataset processing. This is believed to be a function of the PMAL-C8 amphipol mimetic used during purification. In the last year DASS members Pho90 from *S. cerevisiae* and Human

NaS1 have been reported in multiple conformations within the same grid sample, with the proteins purified in the detergents LMNG/CHS and GDN respectively [25, 30]. The hydrodynamic radii of DDM, LMNG, and GDN micelles at 1.0% w/v are 3.5nm, 10.6nm and 4.2nm respectively [188]. These larger micelles will permit more movement of lipids within the micelle, a likely requirement of the 39° rigid body rotation the elevator domain undergoes during transport, while amphipols which preserve only the most auxiliary lipids and wrap tightly round the protein and will not permit such transitions [189]. This hypothesis is supported by my observation during purification procedure that clearly demonstrated LaINDY in a DMM micelle is large enough to concentrate utilising a 100kDa whilst PMAL-C8 purified LaINDY requires a 50kDa cut-off membrane. The fact that both VcINDY and LaINDY have only been resolved in amphipol in single states also provides evidence to this theory.

4.3.3 2.13Å Map of LaINDY with 5mM Malate

Determining the structure of LaINDY in the presence of malate followed a nearly-identical procedure to the structure of the protein with α-ketoglutarate described in the previous section, so the following will only highlight divergences.

LaINDY was purified in Amphipol SEC supplemented with 5mM Malate and concentrated to 2.5mg/mL post with a 50kDa concentrator.

Though the blotting parameters used when vitrifying the grids were identical, and the same Vitrobot Mark IV was used, blotting was much more consistent across all the grids. The collection grid, blotted for 4.5s was one of 4 grids deemed “collection quality” **(Fig 4.4C)**. This variability is expected between grid plunging, as blotting is the least reproducible part of the procedure and could be due to any combination of factors; such as the grid position within the tweezers, variations in grid curvature-out-of-plane, tweezer variability over time, blotting paper variability.

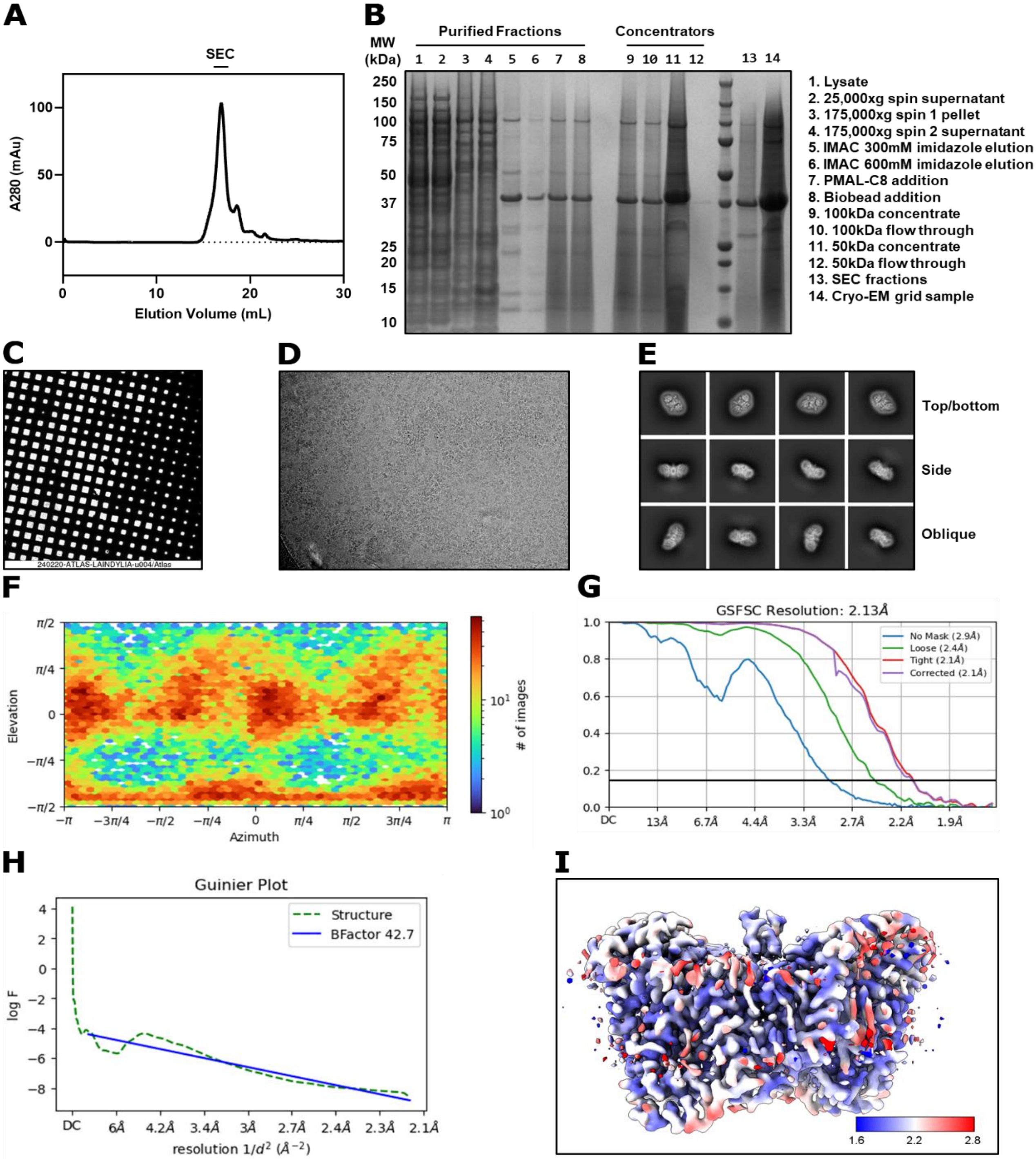


Figure 4.4 5mM malate purification to dataset.

A. *Elution profile for PMAL-C8 purified LaINDY from a Superose 6 Increase 10/300 GL column.* ***B.*** *SDS-PAGE analysis of the purification p012. A larger version of this SDS-PAGE gel can be found in* ***Figure 3.2***. ***C.*** *'Atlas' of the Cryo-EM grid the dataset was collected from, collectable holes are displayed as white.* ***D.*** *Representative micrograph acquired using a K3 Direct Electron Detector at 105,000x magnification.* ***E.*** *Representative 2D classes, grouped by their relative orientation.*

continued on next page

Figure 4.4 continued.

F. *Orientation distribution heatmap for the final dataset (35,696 particles).* ***G.*** *Plot showing the FSC of the final dataset indicating a global map resolution of 2.13Å.* ***H.*** *Guinier plot showing the B-Factor used to generate the sharpened map with which the structure was built. I. Unsharpened map coloured with a local resolution estimation done via CryoSPARC. The FSC threshold was set to 0.143. Figures F-H are outputs from CryoSPARC,*

The collection method reverted to our standard protocol: 0.832Å/pixel (105,000x mag) slit width 20eV, defocus range -0.8-2.2μm in 0.2μm steps, dose rate 22.8e^{-}/Å^{2}/s, and exposure time 2.2s. The reduced magnification permits 3 shots per hole, resulting in a 12,668-movie dataset.

Though the sample distribution and aggregation were acceptable throughout the grid **(Fig 4.4D)**, a significant number of grid holes had visible ice-deposits, or obvious changes of phase in the ice, that were clear on the micrographs during collection. Both of these sample contaminants can be partially explained by two factors: chronologically these were the first LaINDY prepared, and as such a sub-optimal preparation procedure may have been used. Secondly, this was the third microscope trip that this particular grid had been on, and the gradual accumulation of small ice particles during the many transfers between storage dewars and microscopes may also have been a factor. The result of this poor ice became obvious during the curation process in which 2,016 (15.9%) of the micrographs were discarded after using cutoffs for "Relative Ice Thickness" of 0.6<x<1.2, "CTF fit resolution (Å)" to 1.907<x<10 and manually rejecting any remaining micrographs with obvious ice defects.

After an initial 1,000 micrograph subset had been used to sort a "clean" subset of 40,187 particles via iterative 2D classification and selection, the Topaz neural network particle picker was used as opposed to the template picker in the previous dataset [34]. In our experience Topaz has performed better when trained separately to identify top and side/oblique views. One network was given 21,286 top views and another 18,901 side/oblique and both were trained on the same 100 micrograph subset. These trained topaz models then picked 1,681,043 particles from 10,652 micrographs. This was reduced to 1,487,551 particles after removing duplicates, meaning 11.5% of total particles picked were found by both models.

After picking particles, the data processing pipeline was performed near identically to the previous dataset, resulting in final Coloumb maps generated from 594,117 particles reporting 2.15Å, and a CryoSieved set of 35,696 particles reporting 2.13Å (B-Factor -42.7), both with C2 symmetry applied **(Fig 4.4G/H)**. The sieving algorithm did not greatly improve the resolution compared to the α-ketoglutarate dataset, the reason for this is unclear as there are so many divergent points between the two data processing procedures. Nevertheless, the modelling was done in the 2.13Å CryoSieved map.

4.4 No Substrate in the Elevator Domain in Either Dataset

After building the model as described in **Section 2.8.3.3** we first checked the binding site of the elevator domain to see if the improved resolution global map resolution had improved the density of the substrate from the previous structures **(Fig. 4.5A/B)** [14]. There are numerous densities between the Arg159 and His392 basic residues, corresponding to the canonical DASS substrate binding site, however in both structures these are clearly discrete water molecules **(Fig. 4.5C/D)**. There are ‘chains’ of densities present that we initially thought might correspond to a mixture of water and substrates. However, the distances between modeled solvents are 2.5-3.2Å as opposed to the 1.5Å distance expected for single carbon-carbon bonds, thereby precluding the presence of substrate in the binding site [190]. The initial explanation posited was low occupancy, and so we also looked at the larger particle subset maps pre-CryoSieve and the C1 symmetry maps in case the symmetry application was averaging out the substrate. However, all of these maps showed only water molecules within the binding site. We hypothesize the low occupancy of both α-ketoglutarate and malate is due to a discrepancy in the protocol vs. the reported protocol by *Sauer et al.,* resulting in 5mM substrate added rather than 10mM, which would also contribute to reduced occupancy.

Another theory that could quickly be verified was whether the amphipol was distorting the binding site of LaINDY in some way. Both of the Sauer *et al.*, reported amphipol structures with substrate with malate (PDB 6WU2) and α-ketoglutarate (PDB 6WU4) have poor densities that were presumed to be the substrate, but these molecules were not in the deposited models due to ambiguity [14].

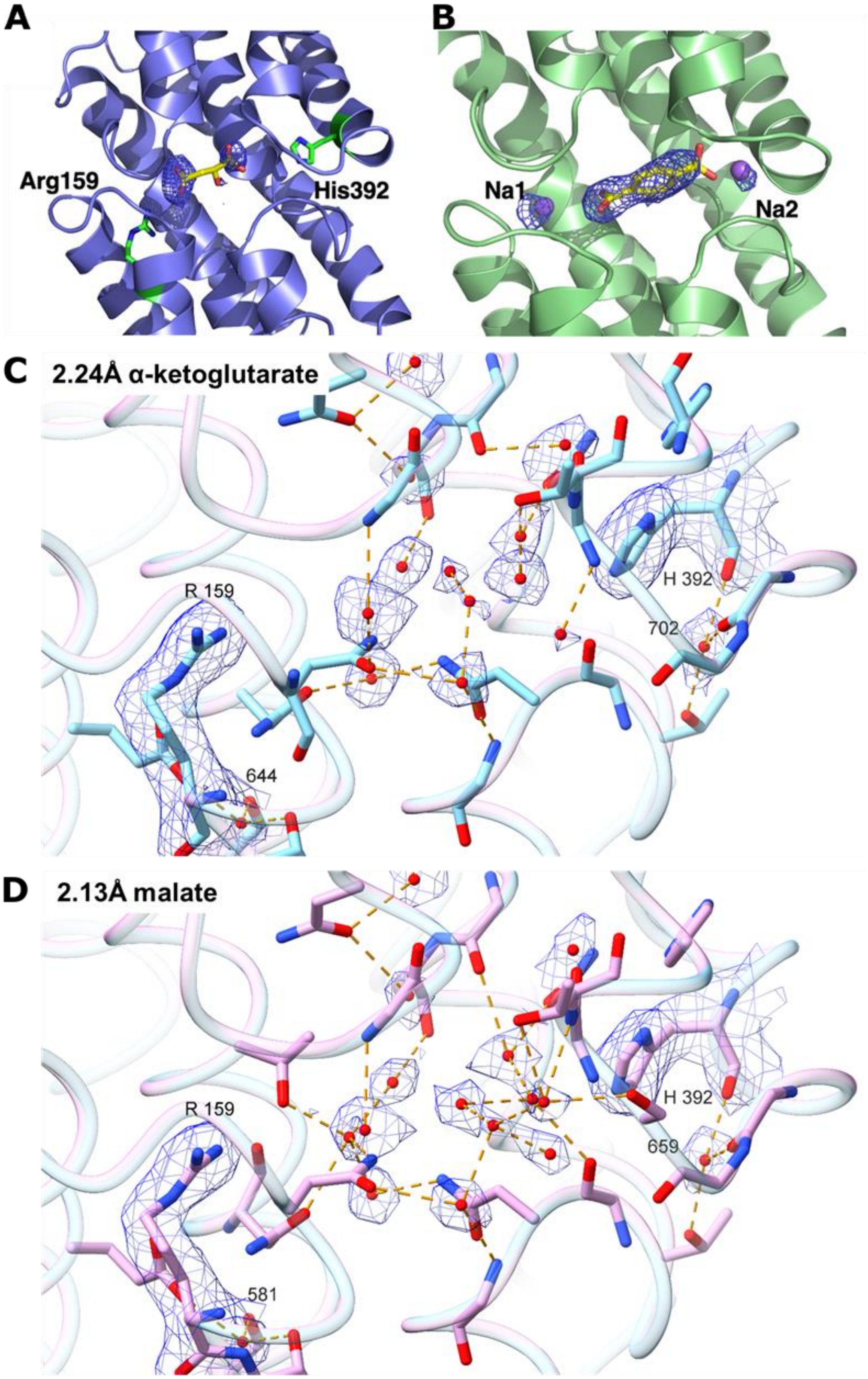


Figure 4.5 Binding site of the presented structures, highlighting the solvent densities.

continued on next page

Figure 4.5 continued.

***A.** Cartoon representation of the transport domain of the 2.86Å 6WTW x-ray structure of LaINDY with α-ketoglutarate fit into assumed density. The basic residues Arg159 and His392 within the binding site are highlighted in green, while α-ketoglutarate is yellow. **B.** Cartoon representation of the transport domain of LaINDY's DASS-C homolog VcINDY from the 3.92Å 6WTX x-ray structure with terephthalate fit into the assumed density. Sodium ion are highlighted as purple, while terephthalate is yellow. **C, D.** Binding sites of presented 2.24Å and 2.13Å LaINDY in the presence of 5mM α-ketoglutarate and malate respectively. Densities are shown as a blue mesh for the water molecules with basic residues Arg159 and His392 used to show the contouring. Only residues that are within hydrogen bonding distance (<3.2Å) are shown, and the respective hydrogen bonds are shown as orange dashed lines. Figures A and B are taken from the supplement of Sauer et. al., 2020.*

4.4.1 Determining Whether the Membrane Mimetic Impedes Substrate Binding

In order to biochemically probe LaINDY's binding affinity, and its possible dependence on lipid mimetic, we next examined the ability of various substrates to thermostabilize the solubilized LaINDY. A single purification was split after IMAC; half reconstituted into amphipol, while half was kept in DDM, both purifications were run on SEC in Assay Buffer, and the peak fractions were centrifuged 20,000xg to pellet aggregates. A sample from the DDM purified LaINDY was titrated from A_{280} 1.0-0.1 to determine which concentration and which intensity setting on the Prometheus would be optimal for the assay, settling on A_{280} 0.5 equivalent to 0.221 mg/mL or 4.1µM with 70% beam intensity **(Fig. 4.6B)**. Detection range of the instrument is between 0-20,000 fluorescence counts, and we have found a concentration and intensity that results in both 330 and 350nm peaks between 5,000 and 15,000 fluorescence counts gives the most consistent results.

The PMAL-C8 amphipol significantly absorbs UV light within the 280nm range used by our nanodrop to assess protein concentration, so ensuring the DDM and amphipol samples were equivalent by this UV-Vis spectroscopy was not possible. The LaINDY band intensity within the SDS-PAGE gel of the peak fractions in both conditions **Figure 4.6A columns 3 & 3** clearly indicates equivalent concentration, despite reading the drastically different A_{280} values of 1.253

for DDM and 1.862 (+49%) for the PMAL-C8 sample. Therefore, the dilution used to get the DDM-LaINDY sample to assay concentration was also applied to PMAL-C8 LaINDY and their equivalence for the protein concentration of both samples was confirmed by the 330 and 350nm peaks in the Prometheus.

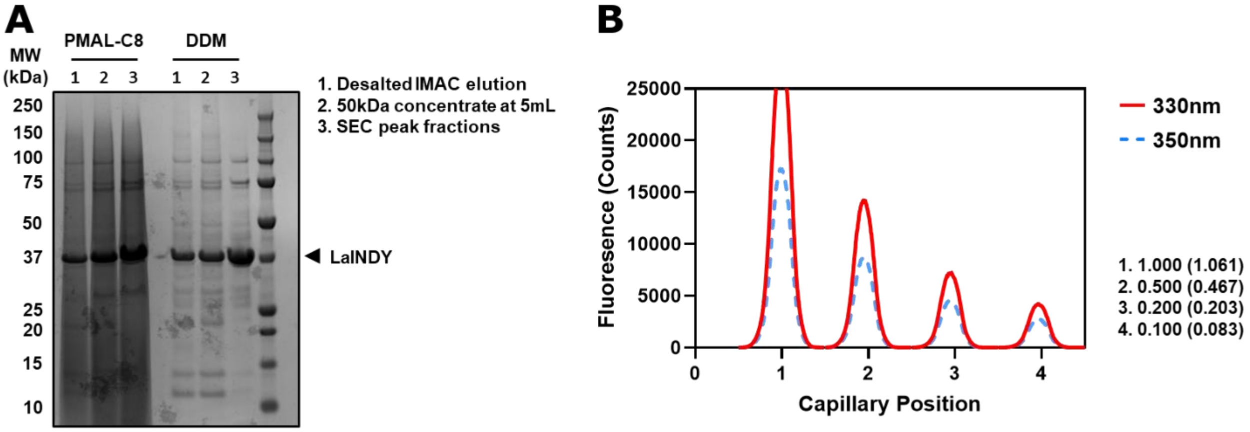


Figure 4.6 Establishing parameters for NanoDSF thermostability assay.

A. *SDS-PAGE analysis of the purifications for the thermostability assay, the corresponding LaINDY band is highlighted with an arrow.* ***B.*** *The titration of DDM-LaINDY to determine the appropriate concentration for use in the thermostability assay. The bracketed values are the actual concentrations as measured by nanodrop. The Prometheus intensity is at 70%.*

Test molecules were chosen based on previous experimental results and potential new substrates of interest: α-ketoglutarate, malate and succinate have all been confirmed transported by LaINDY [14] and private communication with Dr Chris Mulligan; terephthalate was shown to thermostabilise VcINDY and was modelled in the 3.92Å structure (PDB 6WTX) [14] **(Fig. 4.7A)**. Dimethyl-succinate was selected as a negative control, as the esterified carboxylates should prevent the electrostatic interaction with the basic residues in the LaINDY binding site. The substrates were made to a 20mM concentration in Assay Buffer and were incubated in a 1:1 ratio with LaINDY at 8.2µM (A_{280} 1.0) for a final LaINDY concentration of 4.1µM 10min at room temperature prior to loading the Prometheus NT.48 capillaries. A melting scan was then run from

20-95°C at 1°C/min with the ratio metric fluorescence at 330/350nm taken in triplicate **(Fig. 4.7C/D)**.

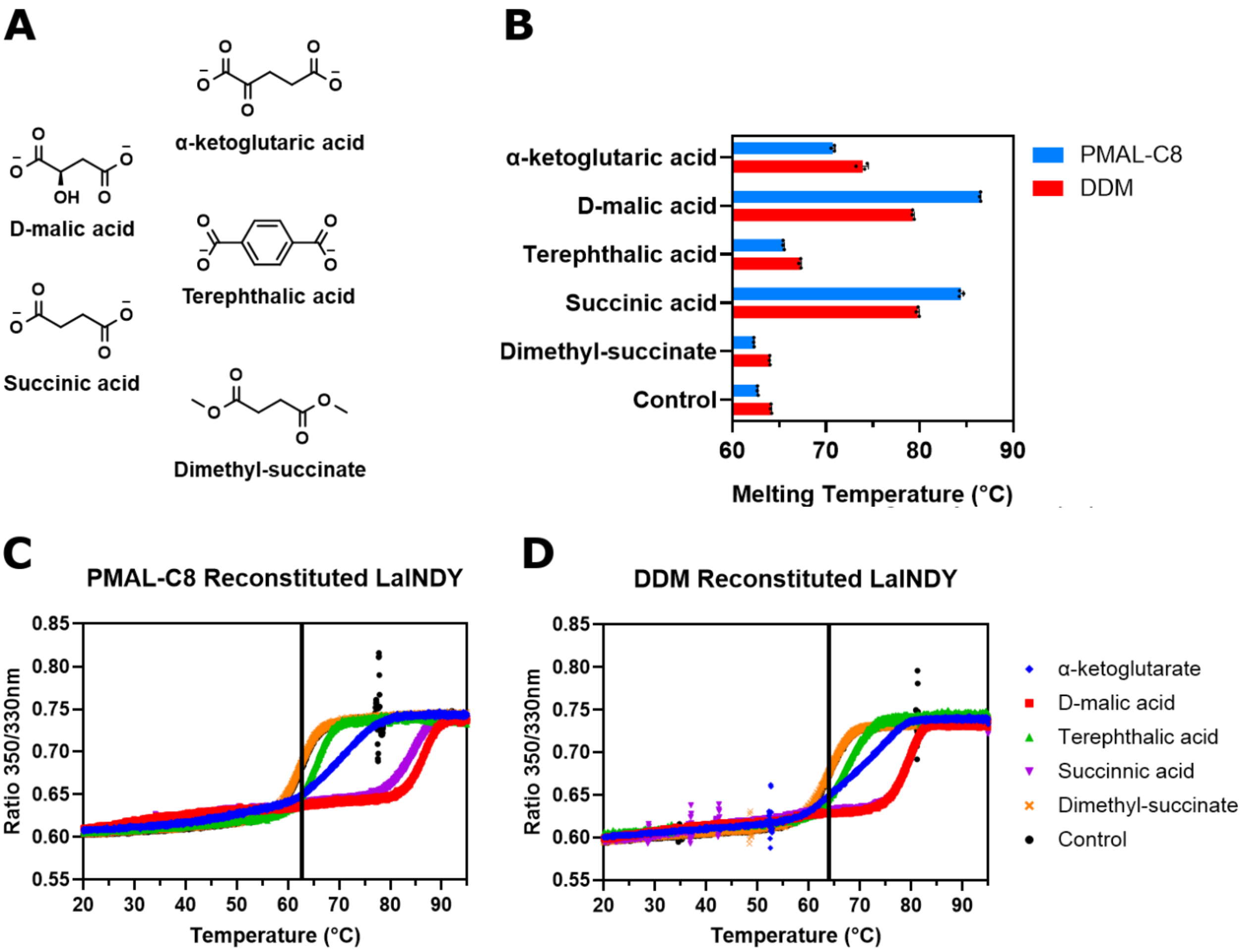


Figure 4.7 Thermostability of LaINDY reconstituted in DDM or PMAL-C8 in the presence of 10mM substrate.

A. *The structures of the substrates used in the thermostability assay.* ***B.*** *Average melting temperature results from a 20-95°C, 1°C/min melting scan, with 4μM DDM/PMAL-C8 LaINDY labelled red and blue respectively. n=3, repeats are technical rather than biological.* ***C.*** *Raw melting curves for the PMAL-C8 samples in B. A black vertical line indicates the average melting temperature of the 3 repeats of the control (LaINDY without substrate) at 62.7°C.* ***D.*** *Raw melting curves for the DDM samples in B. A black vertical line indicates the average melting temperature of the 3 repeats of the control (LaINDY without substrate) at 64.1°C. Both C and D use the same legend.*

The results clearly show that PMAL-C8 purified LaINDY is able to interact with all the selected substrates, showing clear stabilisation over the samples with dimethyl-succinate or without compound **(Fig 4.7B)**. The weakest stabilisation was seen with terephthalate in both DDM and PMAL-C8 was with +3.1°C and +2.8°C of stabilisation respectively. α-ketoglutarate followed (+9.8°C and +8.1°C), then malate (+15.2°C and +23.8°C) and succinate (+15.7°C and +21.7°C). The discrepancy between the DDM and PMAL-C8 results for succinate and malate in particular are interesting however, and may indicate some differential impacts of the two membrane mimetics on LaINDY's interaction with substrate. Regardless it is clear the LaINDY in amphipol is able to interact with substrates, suggesting the PMAL-C8 membrane mimetic is not responsible for the missing substrate density within the maps.

4.4.2 Additional Substrate Thermostability Screen

I performed a further ligand binding thermostability screen, picking substrates with the intention of probing the structure-activity relationship (SAR) of small molecules interacting with LaINDY **(Fig. 4.8A)**. The assay was set up as described in **Section 4.4.1**, with DDM reconstituted LaINDY incubated with 10mM substrate. However, the scope for SAR interpretation is limited without high resolution structural information to support the results.

For the four carbon dicarboxylates succinate, fumarate, and maleate isomerisation state appears to have a minimal impact on the substrate's ability to stabilise LaINDY, with +16.1°C, +15.2°C, and +14.0°C compared with the control **(Fig. 4.8A/B/C)**. This was not the case for DASS-C VcINDY, as in the report by Sampson *et al.* a CPM-based thermostability assay determined succinate, fumarate and maleate at 10 mM to stabilize VcINDY approximately +12°C, +7°C, and +0°C compared to the control [99]. The fumarate stabilisation contradicts the result from the Sauer at al. LaINDY thermostabilisation, which reported a SEC peak for LaINDY after 2h incubation with 100mM fumarate at 42°C 50% smaller than the no compound control [14]. The differences in methodology most likely explain this though, and highlights the need to measure binding by multiple methods.

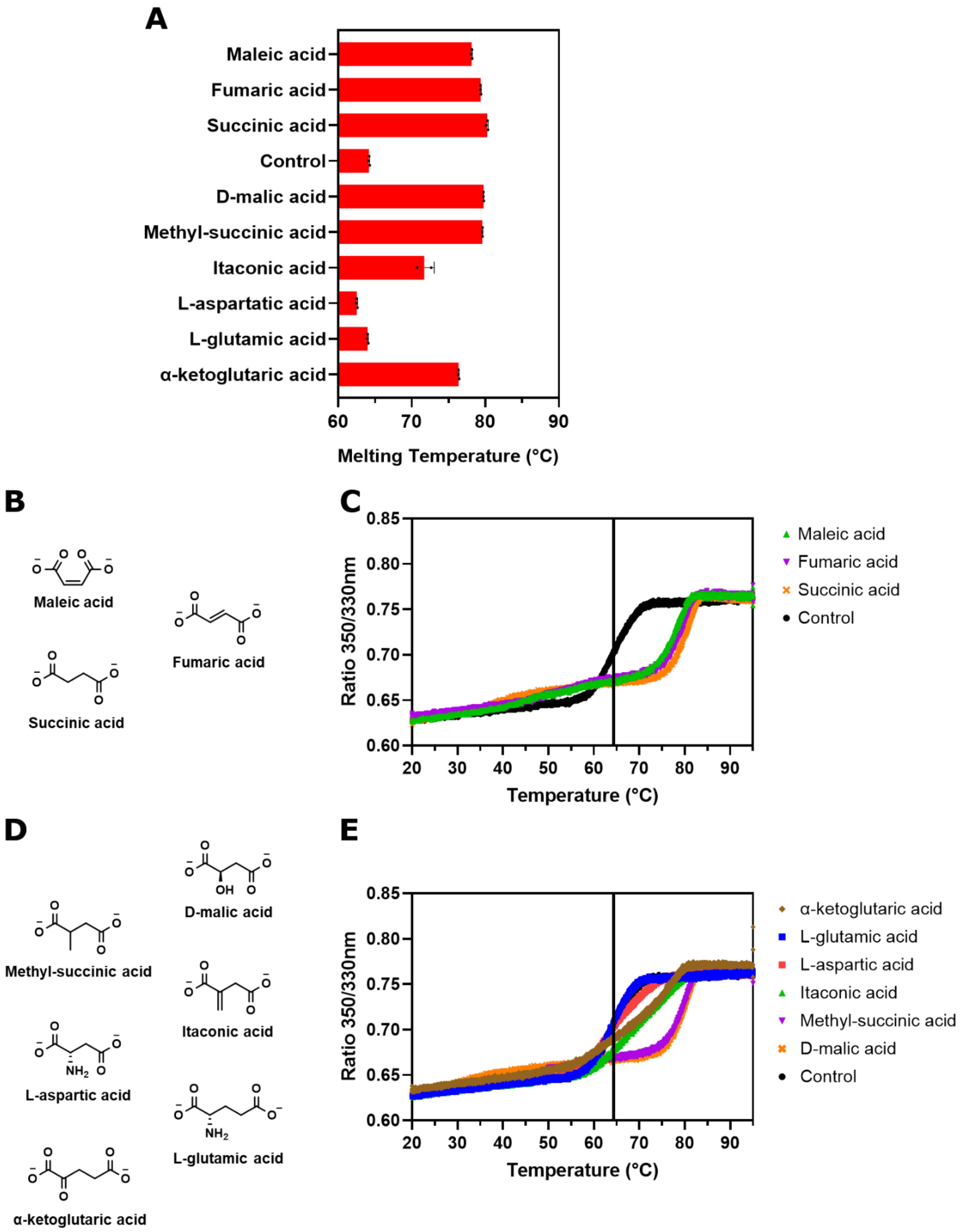


Figure 4.8 Probing additional chemical moieties impact on dicarboxylate stabilisation of LaINDY.

** continued on next page**

Figure 4.8 continued.

A. *Average melting temperature results from a 20-95°C, 1°C/min melting scan, with 4µM DDM LaINDY incubated with 10mM substrate. n=3 with the exception of Itaconate where n=2, because multiple peaks in the first derivate plot prevented assignment of a melting temperature. Repeats are technical rather than biological.* ***B.*** *The structures of the substrates used in the thermostability assay.* ***C.*** *Raw melting curves for the molecules in B. A black vertical line indicates the average melting temperature of the 3 repeats of the control (LaINDY without substrate) at 64.2°C.* ***D.*** *The structures of the substrates used in the thermostability assay. The structures of the substrates used in the thermostability assay.* ***E.*** *Raw melting curves for the molecules in D. A black vertical line indicates the average melting temperature of the 3 repeats of the control (LaINDY without substrate) at 64.2°C.*

In contrast, the functional group at the Cα may have more of an impact on substrate recognition (Fig 4.8A/D/E). A Cα hydrogen (succinate), hydroxyl (malate), and methyl (methyl-succinic acid) were able to stabilise LaINDY similarly, with +16.1°C, 15.6°C, and 15.4°C of stabilisation compared to the control. However, an alkene (itaconate) group resulted in half the stabilisation of succinate at +7.5°C, and an amino group (L-aspartate) actually destabilised -1.7°C when compared to the control. The amino group destabilisation is also seen when comparing the five carbon decarboxylates, with a carbonyl (α-ketoglutarate) stabilising +12.2°C and the amino group (L-glutamate) destabilising -0.2°C compared to the control. A charged amino group within the binding site could explain the destabilisation, however pK_a data for L-aspartate and L-glutamate has not been recorded at the >64°C that I recorded as a melting temperature for LaINDY, so this result alone cannot confirm that. I do not expect this destabilization difference to be steric, as the Cα amino group, and methyl groups of L-aspartate and methyl-succinic acid respectively would only differ by the mass of a single hydrogen atom.

Interestingly, considering the raw melting curves in **Figure 4.7C/D** and **Figure 4.8E**, α-ketoglutarate and itaconate appear to be less cooperative with respect to LaINDY with much shallower gradients. Both molecules have a double bonded atom at the Cα position, but comparison is complicated by the 1 carbon difference in chain length between the two dicarboxylate moieties.

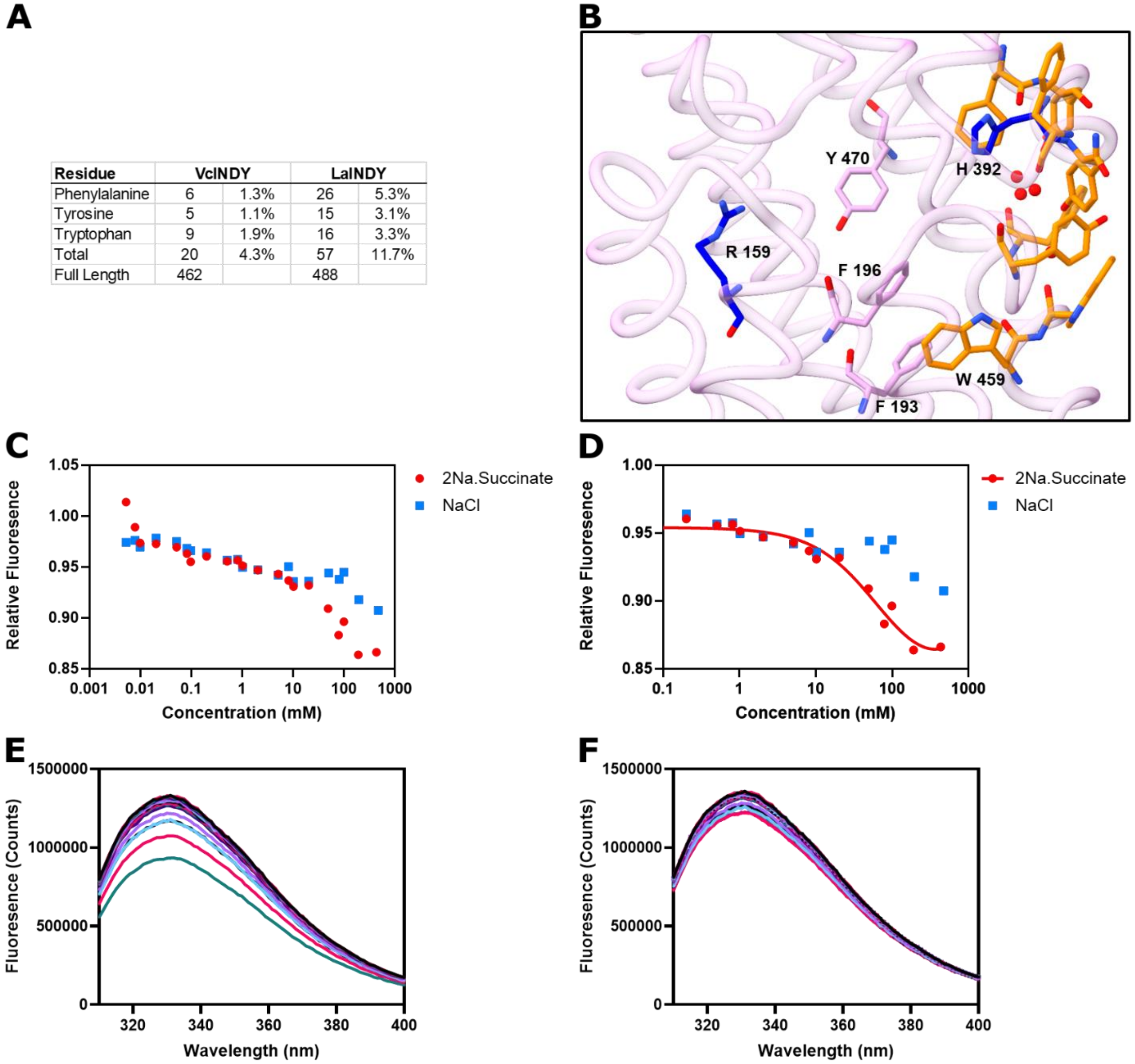
Repeating the experiment with four carbon oxaloacetate and five carbon α-methylene glutamate would provide more information.

Additionally, it also remains a possibility that these substrates adopt different orientations within the binding site, and as such the Cα groups interact with different residues within the binding site, but additional assays and structures would be needed to confirm which is correct.

4.4.3 Establishing a Dissociation Constant to Explain Poor Substrate Density with Tryptophan Fluorescence Quenching.

To give an estimate of the occupancy of substrate within the presented structures, the next step was to establish binding kinetics for the LaINDY-substrate interaction. Here I recognised that tryptophan fluorescence quenching had been successfully used to demonstrate the binding of Sodium ions to VcINDY [21], and LaINDY has significantly more intrinsically fluorescent residues than VcINDY **(Fig 4.9A)**. Although execution of the assay is time consuming, it is simple, requires minimal labor or modification to early sample prep methods, and would provide a highly sensitive readout of LaINDY-substrate interaction to yield a precise affinity. This assay would need to be performed on DDM-LaINDY rather than PMAL-C8-LaINDY. As mentioned in **Section 4.4.1** PMAL-C8 absorbs light within the UV range thereby making UV absorbance measurements of protein concentration less precise, as such the PMAL-C8 membrane mimetic incompatible with the assay.

In optimizing the tryptophan fluorescence quenching assay, I noticed that there are additional intrinsically fluorescent residues: Phe193, Phe196, Trp459 and Tyr 470 that are within the binding cavity between basic residues Arg159 and His392 **(Fig 4.9B)**. There is also an "aromatic cage" of residues shielding the elevator binding cavity from the lipid membrane on the His392 side which we predicted may also experience quenching via the substrate. Given the number of weakly fluorescent amino acids in addition to the Trp449 in the binding site I choose the excitation wavelength of 290nm rather than the tryptophan selective 295nm for the experiment [191], with the idea that multiple weakly fluorescent residues may be quenched by substrate binding. Without time constraints, my initial experiment would have been performing the assay with multiple excitation wavelengths and proceeding with which ever gave the best signal to noise.



Residue	VcINDY		LaINDY	
Phenylalanine	6	1.3%	26	5.3%
Tyrosine	5	1.1%	15	3.1%
Tryptophan	9	1.9%	16	3.3%
Total	20	4.3%	57	11.7%
Full Length	462		488	

Figure 4.9 Tryptophan fluorescence quenching of 2µM LaINDY with disodium succinate at 280nm.

continued on next page

Figure 4.9 continued.

A. *Table highlighting the difference in intrinsically fluorescent residue differences between DASS exchanger LaINDY and DASS cotransporter VcINDY.* ***B.*** *Cartoon representation of the LaINDY elevator domain binding site coloured pink. Basic binding site residues Arg159 and His392 are highlighted, along with Phe193, Phe196, Trp459, and Tyr470 which I predict may be quenched by substrate binding. The aromatic cage and its coordinated waters are coloured orange but not labelled.* ***C.*** *Initial fluorescence quenching binding curves of Disodium Succinate and Sodium Chloride coloured red and blue respectively.* ***D.*** *Nonlinear regression fit for total and nonspecific binding for the disodium succinate curve. The dissociation constant from the fit is 74.35mM. E. The raw emission scan from 310-400nm for the disodium*

Ligand Target Concentration (mM)	Sodium Chloride				Disodium Succinate			
	Stock Concentration (mM)	**Stock Added (µL)**	**Assay Total Volume (µL)**	**Dilution Factor**	**Stock Concentration (mM)**	**Stock Added (µL)**	**Assay Total Volume (µL)**	**Dilution Factor**
0	0	0	1900	1.0000	0	0	1900	1.0000
0.005	1	10	1910	1.0053	1	10	1910	1.0053
0.008	1	5	1915	1.0079	1	5	1915	1.0079
0.01	1	4	1919	1.0100	1	4	1919	1.0100
0.02	10	2	1921	1.0111	10	2	1921	1.0111
0.05	10	6	1927	1.0142	10	6	1927	1.0142
0.08	10	6	1933	1.0174	10	6	1933	1.0174
0.1	10	3	1936	1.0189	10	3	1936	1.0189
0.2	100	2	1938	1.0200	100	2	1938	1.0200
0.5	100	6	1944	1.0232	100	6	1944	1.0232
0.8	100	6	1950	1.0263	100	6	1950	1.0263
1	100	4	1954	1.0284	100	4	1954	1.0284
2	2000	1	1955	1.0289	2000	1	1955	1.0289
5	2000	3	1958	1.0305	2000	3	1958	1.0305
8	2000	3	1961	1.0321	2000	3	1961	1.0321
10	2000	2	1963	1.0332	2000	2	1963	1.0332
20	2000	10	1973	1.0384	2000	10	1973	1.0384
50	5000	12	1985	1.0447	2000	29	2002	1.0537
80	5000	12	1997	1.0511	2000	31	2033	1.0700
100	5000	8	2005	1.0553	2000	22	2055	1.0816
200	5000	40	2045	1.0763	2000	104	2159	1.1363
500	5000	125	2170	1.1421	2000	334	2493	1.3121

Table 4.1 Dilution factors of preliminary tryptophan fluorescence quenching assay in Figure 4.9.

Following the guide from Yammine *et al.,* a scout experiment was carried out with 2µM LaINDY using Succinate as the ligand of choice across a broad concentration range (5µM-500mM) to get an approximation of the K_D before a more rigorous experiment with finer concentration steps for a more accurate value [192]. Succinate was chosen for several reasons. Firstly, this ligand demonstrating one of the biggest stabilizations' vs control in our thermostability assay (+21°C) which could indicate a stronger interaction and a lower K_D. Secondly recent radiolabeled assay transport confirmation through the aforementioned communication with Dr Chris Mulligan. Finally, as a proposed succinate antiporter utilized during anaerobic respiration, kinetic data on this substrate is directly informative to bacterial physiology. Most dicarboxylates come as di-sodium salts, so cognizant of the fact that for each application of substrate within the assay, double the concentration of sodium would be added, I also carried out a control experiment of an NaCl titration. This also presented a further issue, as tryptophan fluorescence requires high concentration stocks of substrates, and so the highest concentration of disodium-succinate possible to dissolve in Assay Buffer was 2M (4M sodium), which is only soluble when warmed to 90°C. Once dissolved it remained soluble at room temperature when kept in motion on a rotating wheel.

The preliminary quenching experiment showed an obvious drifting baseline **(Fig 4.9C)**, which I believe is to be expected with the change in the ionic strength of the solution. The disodium succinate curve diverges from the sodium chloride curve at two points: a low micromolar and low millimolar range, which could possibly indicate two sites for succinate binding of different affinities. When applying a nonlinear regression for total and nonspecific binding to the data a K_D of 74.4mM is given for the low affinity site **(Fig 4.9D)**.

Regarding the low micromolar divergence, this range was only sampled in a single experiment and with limited measurements in this succinate concentration regime. As the fluorescence quenching assay is a time-consuming procedure, I made the decision to pursue the low affinity succinate binding identified in my scout experiment which would explain my structural data.

Following the scout experiment the target concentration range of disodium succinate was adjusted to 1-800mM with finer measurements made at either end of the range to ensure the curve properly flattened. Three technical replicates of 2μM LaINDY quenching assays were performed and the average fitted with nonlinear regression for total and non-specific binding yielding a K_D of 395.1mM **(Fig 4.10)**. This K_D value is an order of magnitude higher than the scout experiment, though still within the 95% confidence limits of its estimation. Despite this, the key point illustrated is that the K_D at least an order of magnitude greater than the concentrations of substrate used thus far for Cryo-EM studies, and thereby explains the poor substrate occupancy seen in my structures. Although substrate binding kinetics have been measured for some DASS cotransporters [18, 21, 25], and there are many DASS family members with reported transport assays [14, 42] [41], I believe this to be the first measurement of binding kinetics for a bacterial DASS exchanger.

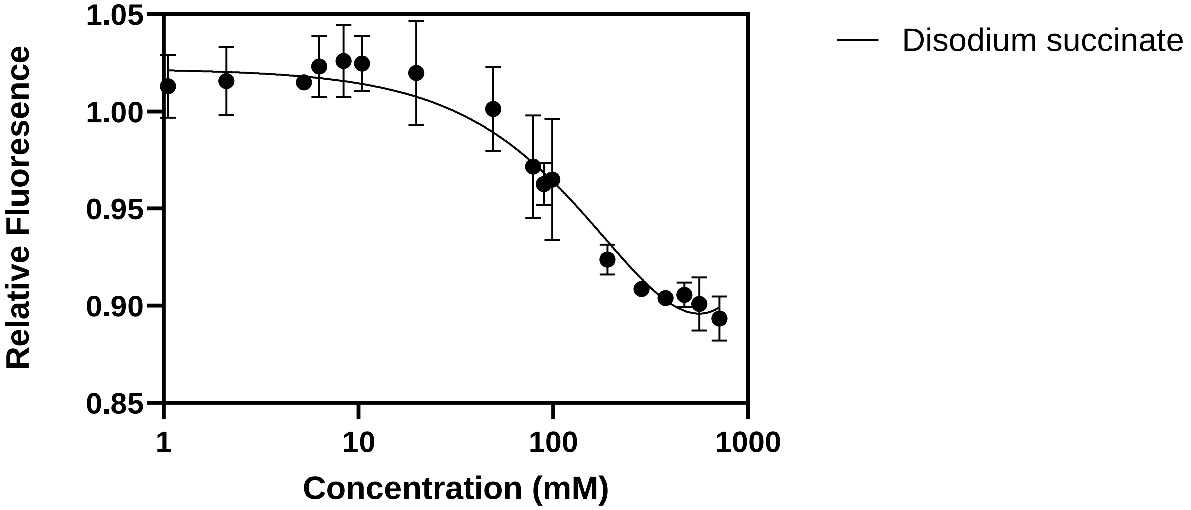
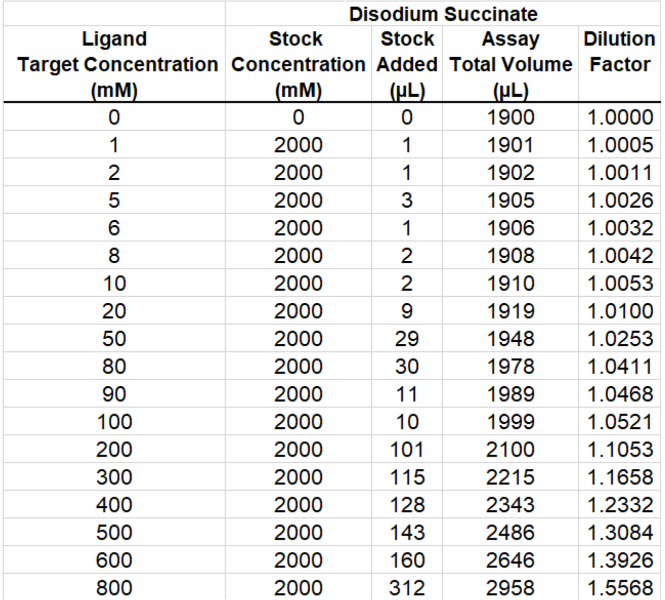


Figure 4.10 Tryptophan fluorescence quenching of 2 µM DDM-LaINDY with disodium succinate.

Binding saturation curve with concentration fitted with a $\log_{10}$ transformation from 1-800mM. The average of n=3 technical replicates are plotted with error bars showing standard deviation. Nonlinear regression fit for total and nonspecific binding of the disodium succinate to LaINDY reporting a dissociation constant of 395.1mM.



Ligand Target Concentration (mM)	Disodium Succinate Stock Concentration (mM)	Stock Added (µL)	Assay Total Volume (µL)	Dilution Factor
0	0	0	1900	1.0000
1	2000	1	1901	1.0005
2	2000	1	1902	1.0011
5	2000	3	1905	1.0026
6	2000	1	1906	1.0032
8	2000	2	1908	1.0042
10	2000	2	1910	1.0053
20	2000	9	1919	1.0100
50	2000	29	1948	1.0253
80	2000	30	1978	1.0411
90	2000	11	1989	1.0468
100	2000	10	1999	1.0521
200	2000	101	2100	1.1053
300	2000	115	2215	1.1658
400	2000	128	2343	1.2332
500	2000	143	2486	1.3084
600	2000	160	2646	1.3926
800	2000	312	2958	1.5568

Table 4.2 Dilution factors of tryptophan fluorescence quenching assay in Figure 4.10.

There are however, two major caveats in interpreting the apparent affinity of LaINDY for succinate. Firstly, the solubility of the substrate stock is most likely supersaturated as a heat block was required to dissolve the disodium succinate to 2M in the Assay buffer. During the second technical replicate, I noticed precipitate forming within the stock and so the stock was suspended over a heatblock for the remainder of the assay. This could introduce errors both in actual concentration of substrate added. Secondly, for accurate calculation of the association constant Yammaine *et al.,* recommend the final ligand concentration to be 5-10x greater than the measured K_D. With a final concentration of 715mM we have not achieved this. Adding additional measurements at higher substrate concentrations is not possible due to the solubility limit of the substrate.

While these issues might be resolved by performing the assay at a higher temperature, in tandem keeping the substrate stock in a water bath set to the assay temperature in an attempt to stabilize the supersaturated state of the stock. Nevertheless, the apparent K_D of LaINDY for succinate suggest the ligand concentration in my Cryo-EM samples is far below what is required for fully-occupied and modellable substrates in the density maps.

4.5 Sample Preparation in 75mM Disodium Succinate

Building from these observations that LaINDY has millimolar affinity for succinate, which may have caused partial occupancy in my previous LaINDY grids with 5mM succinate, I therefore pursued the obvious next step of drastically increasing the substrate concentration and remaking grids. The biggest issue is that substrate concentrations approaching concentrations in the high mM range may affect the contrast of the cryo-EM images, which is further amplified by the high sodium concentration brought with the ligand [193-195]. As the current optimized conditions for LaINDY grids require 150mM sodium chloride, I reasoned that this salt could be entirely replaced with 75mM disodium succinate while maintaining the ionic strength of the grid sample. Thus, I expected this new sample condition would not affect protein distribution on the grid, and also minimally impact image contrast, whilst increasing the substrate concentration 15-fold.

I therefore purified LaINDY in PMAL-C8 on a Superose 6 10/300 increase column in Assay Buffer with 75mM disodium succinate. Grids were made as described in previous sections, and screened on a 200kV Talos Arctica (FEI/Thermo Fischer Scientific) equipped with a Falcon 4i Direct Electron Detector (FEI/ Thermo Fischer Scientific) at the COSMIC facility. Atlasing of the grids revealed reasonable blotting, with ice of sufficient quality for data collection **(Fig 4.11A)**. Visual inspection of the grids revealed the particles were as visible as the previous samples and particle distribution was also comparable, indicating that substrate concentration could potentially be pushed even higher **(Fig 4.11B)**. Unfortunately, the all the grids were damaged during the vitrification process and as such the surfaces of many grid squares were uneven to the point of not being viable for large scale Krios collection. Nevertheless, data from this grid screening indicates that particles can be resolved within the ice at this high substrate concentration, and therefore a high-resolution substrate structure is possible. However, due to time constraints freezing of additional LaINDY grids further work could not be completed in this study.

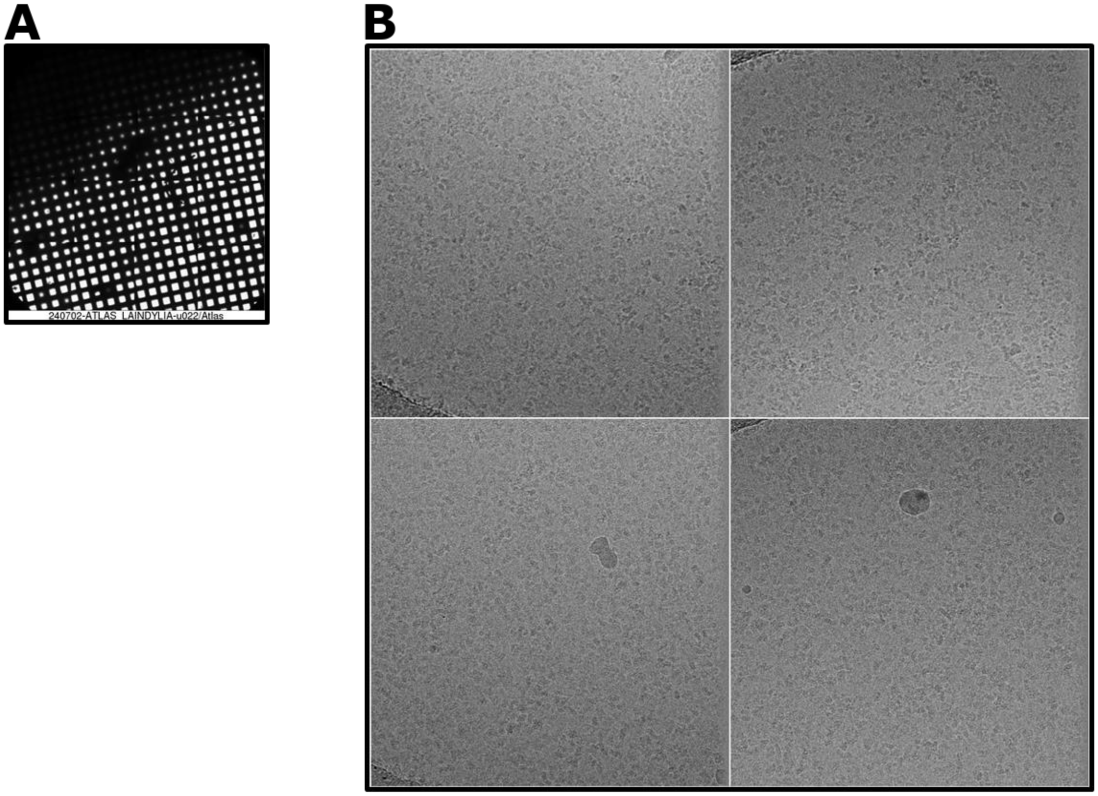


Figure 4.11 Cryo-EM Sample of amphipol purified LaINDY with 75mM disodium succinate.

A. *'Atlas' of the Cryo-EM screened, collectable holes are displayed as white.* ***B.*** *Four representative micrographs of the 75mM disodium succinate grid acquired using a Falcon 4i Direct Electron Detector.*

4.6 Discussion

4.6.1 LaINDY Reconstituted in PMAL-C8 Resolution may be Pushed Higher

In this chapter, two structures of LaINDY purified in the PMAL-C8 amphipol have been resolved to 2.24Å and 2.13Å, significantly improving the resolution over the previous best LaINDY structure at 2.86Å. This improvement is essentially all due to the improvements in cryo-EM hardware and software in the four years since the Sauer *et al.,* paper and the high resolution was achieved despite the suboptimal conditions: bad-blotting, and poor ice quality for the malate, and α-ketoglutarate samples respectively. Hence, I would suggest that purely through optimization of blotting conditions the current LaINDY PMAL-C8 sample could have the global map resolution pushed even higher.

4.6.2 LaINDY Has Low Affinity for Succinate, Malate and α-ketoglutarate at pH7.5

The tryptophan fluorescence quenching assay with succinate indicates that LaINDY's affinity for substrate may be significantly lower than anticipated at pH7.5 used in the Cryo-EM sample. Although issues with supersaturation of the substrate stock may have resulted in imprecision with the final calculated K_D. I have discussed adjustments to the protocol, such as raising the assay temperature and substrate stock to maintain its solubility. Despite the inaccuracies, a dissociation constant within the high millimolar range explains the poorly resolved substrate in previous LaINDY structures, and the lack of substrate in my structures with 5mM substrate.

However, assuming the worst-case scenario in which the measurable KD is 395.1mM and using the Hill equation to calculate the occupancy of succinate:

$$\theta = \frac{[L]^n}{K_D + [L]^n}$$

Where θ is the fraction of succinate bound LaINDY, [L] is succinate concentration, K_D is the measured dissociation constant (395.1mM), and n is the hill coefficient (1 for a single independent binding event). There is no literature supporting cooperativity in DASS transporters, and there are two basic residues per protomer for a single dicarboxylate, so a Hill coefficient of 1 is a reasonable assumption. At 5mM succinate concentration the LaINDY occupancy would be at 1.2%, and even at the elevated 75mM succinate of the optimized grid sample occupancy would only reach 16%, suggesting density for the substrate may still be poor. There are still multiple avenues to improve this however.

Firstly, increasing substrate concentration further is still a possibility. Sauer *et al.*, solved LaINDY homolog VcINDY with 300mM sodium chloride [21], which implies that a 150mM disodium succinate LaINDY structure would be possible. Using the above Hill calculation, this would yield an occupancy of 27.5%. Further, utilising substrates with chemical features that will give stronger density can be used to more clearly identify bound molecules, such as the aromatic ring of terephthalate.

Additionally, at the data processing stage there are methods to resolve partial occupancy, with a high enough map resolution. Generating a Columbic potential map without symmetry expansion applied, it may be possible to selectively refine elevators with occupied binding sites.

4.6.3 Physiological Conditions May be the Key to Resolving Substrate Bound Structures

Another approach would be to abandon the proven Cryo-EM sample condition, instead taking into consideration the physiological conditions that LaINDY would be in. I believe there are three major impacting factors.

Firstly, *L. acidophilus* is an acidophile, and the optimum pH range for its growth in culture is between pH5.5-6.0[94]. The pH of 7.5 I selected for Cryo-EM sample prep and subsequent thermostability and tryptophan fluorescence quenching to address my structures lack of substrate were based on the successful structures in the Sauer *et al.* report [14]. However, this is not the pH the protein will be exposed to physiologically, and it is within the realm of possibility that lowering the pH by two orders of magnitude (at the extreme) may impact the binding affinity of di-anionic substrates. Cytoplasmic pH for most organisms tends to be closer to neutral, however at present LaINDY has only been reported in the C_o state, and as such the binding domain of the elevator is periplasmic. This hypothesis is consistent the DASS-C member VcINDY, for which Sauer *et al.*

measured the K_D for succinate of 92.2 ± 5μM at pH7.5. The optimum pH range for culture of *V. cholerae* is between 6.5-9, and so the assay condition used by Sauer *et al.* are much more physiological for that protein's native environment [196].

Secondly, the impact of specific lipids on the function of membrane proteins has been widely demonstrated [103, 104, 115]. More specifically for the DASS family, GFP-based thermostability studies on VcINDY by Sampson *et al.* reported a synergistic stabilization when adding succinate and PE, while adding PE in the absence of substrate significantly destabilized the protein [99]. The inverse was true with CA, which significantly stabilized the protein alone, however that stability was abolished with the addition of succinate. The exact mechanism is not yet understood, but it suggests that if there is a similar trend for LaINDY, there may be a lipid and substrate combination that are most optimal for substrate resolving. Considering, the widely accepted fact that most of the host lipids are removed from membrane proteins during the purification process - the Laganowsky *et al.* work on AmtB required doping a 10-fold molar excess of PC prior to crystallisation form modellable lipid density to be apparent in their maps [104], this will require screening for the appropriate lipid first, and adjusting the Cryo-EM sample prep procedure. As such, understanding the differences between the native lipidome of the plasma membrane and the expression system will contribute to this. While there has been no comprehensive study on the *L. acidophilus* NCFM lipidome, Hansen *et al.* have reported that of *L. acidophilus* L-5α [197]. Assuming the composition of NCFM and L-5α is similar, we can estimate: ~60% CA, ~25% monolysylcardiolipin (MLCA), ~5% PA, ~5%PG, ~0.5% triglycodiacylglycerol (TGDG) [198, 199], and ~0.5% quadroglycodiacylglycerol (QDGQ), as the major contributing lipids; compared with *E. coli* 75% PE, 20% PG and 5% CA [200-202] **(Fig. 4.12).** Monolysylcardiolipin, phosphatidic acid, and the glycodiacylglycerols are not present in *E. coli*, and there is a significantly larger percentage of cardiolipin than in the *L. acidophilus* membrane. Additionally, many species of *Lactobacilli* including *L. acidophilus* ATC 314 and FTCC 0291 have been demonstrated to incorporate host cholesterol into their own membrane [97]. Given that the human DASS-Cs NaS1 and NADC1 have had cholesterol bound in recently reported structures. LaINDY binding to cholesterol should also be considered [30].

Finally, it is a possibility that we do not yet know the physiological ligands for LAINDY. The Sauer *et al.* report predicted LaINDY to be an α-ketoglutarate/succinate antiporter based on their thermostability results, and have demonstrate transporter for α-ketoglutarate, malate, and succinate using radiolabeled succinate in cell assays [14]. My nanoDSF thermostability assay gave

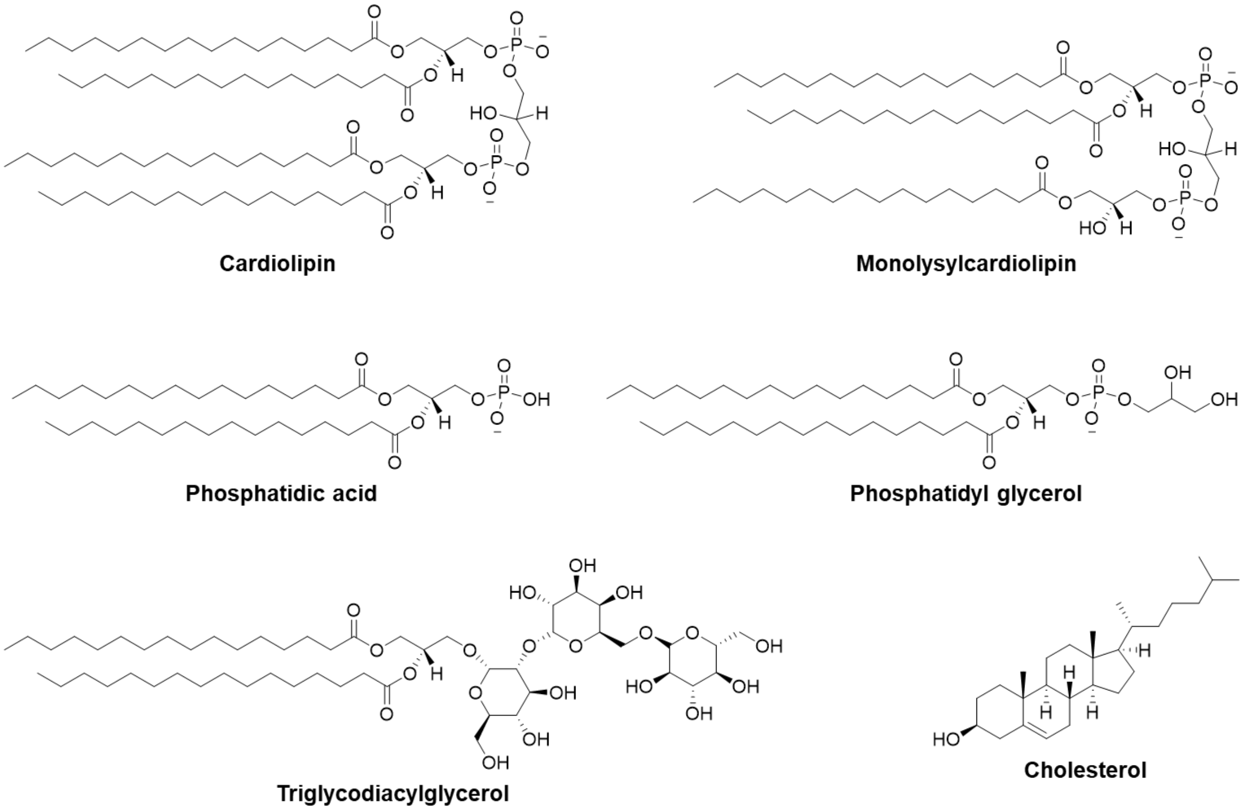


Figure 4.12 Lipids of the L. acidophilus lipidome.

Structures of the lipids from the Hansen et al. (2015) study on L. acidophilus L-5α, and cholesterol. The structure of TGDG as αGlc(1→6)αGal(1→2)αGlc(1→1)glycerol is assumed from Hölzl & Dormann (2007) and Shaw et al. (1968), in the absence of data on L. acidophilus glycolipids. QGDG is not shown for simplicity.

different results to their report, in particular for succinate and fumarate. The STRING database puts LaINDY in the same genomic neighborhood as citrate lyase, and as such their expression may be co-regulated and they may share the same substrates [203]. Citrate lyase catalyzes the cleavage of citrate to oxaloacetate and acetate [204]. Although although Sauer *et al.* reported no thermostabilisation for LaINDY following incubation with 100mM citrate for 2h at 42°C compared with the control, only ligands which thermostabilise the target give a signal in nanoDSF and therefore the technique is prone to false negatives. Consequently, LaINDY could function as a citrate or oxaloacetate transporter and may have higher affinities for these physiological ligands. It is also worth mentioning that in the characterization of DASS-E member TdtT transport of tartrate and succinate was determined to be directional, with transporter activity for tartrate import 13-fold greater than tartrate export [74]. The mechanism by which the transporter achieves this is not understood it implies differing affinities between C_i and C_o states, which if it is true for LaINDY will obviously be important.

4.6.4 Capturing LaINDY in Multiple States May require Alternative Mimetics

The amphipol purification process for LaINDY has also been contrasted with the recent work on its' homologs ScPho90, HsNaS1, and HsNaDC1 which were able to resolve multiple states of the protein in a single sample by utilizing larger micelle detergents such as LMNG or GDN [25, 30]. In the case of ScPho90, that target yielded comparable resolution in one state to my presented structures. Although the authors did not elaborate on how this was achieved, this demonstrates that with enough well-resolved particles it is possible to determine multiple states of LaINDY to high resolution in alternative membrane mimetics.

5. Structural Waters and a Lipid Within an Improved LaINDY Model

5.1 Aims

Despite the absence of density for substrate within the 2.24Å 5mM α-ketoglutarate and 2.13Å 5mM malate structures of LaINDY, the structures are the highest reported for the protein or any DASS transporter to date. I therefore examined the experimental maps and refined the models to determine what the increased resolution has revealed about LaINDY.

5.2 Cα Comparison with Previous Best LaINDY Structure and Other DASS Members

For an initial comparison the MatchMaker tool in UCSF ChimeraX v1.8 was used to superimpose the A chains of the structures and calculate the RMSD between the Cα's of each residue [205, 206]. The reported RMSD between the α-ketoglutarate and malate structures is 0.154Å between 489 atom pairs **(Table 5.1)**. For comparison the RMSD between the two structures and the previous highest resolution X-ray structure of LaINDY reported by Sauer *et al.,* (herein LaINDY-X-ray) (PDB 6WTW) are 0.348Å and 0.352Å for the α-ketoglutarate and malate structures respectively **(Fig. 5.1A)** [14]. The superimposition of DASS cotransporter VcINDY in the inward sodium bound conformation has significantly larger RMSD of >10Å due to the elevator domains inward conformation **(Fig. 5.1B)** [21]. In contrast superimposition of the LaINDY α-ketoglutarate and malate structures with the more distantly related DASS cotransporter ScPho90 yielded a RMSD of >5Å, with greater structural similarity due to the outward-facing conformation of the elevator domains for both proteins **(Fig. 5.1C)** [25].

	6WTW LaINDY		α-ketoglutarate		malate		7T9G VcINDY		8R33 ScPho90	
	Cα atoms	rmsd (Å)	Cα atoms	rmsd (Å)	Cα atoms	rmsd (Å)	Cα atoms	rmsd (Å)	Cα atoms	rmsd (Å)
6WTW LaINDY	489	n/a	489	0.348	489	0.353				
α-ketoglutarate	489	0.348	489	n/a	489	0.142	430	10.530	451	5.067
malate	489	0.353	489	0.142	490	n/a	433	10.439	451	5.478
7T9G VcINDY			430	10.530	433	10.439	444	n/a		
8R33 ScPho90			451	5.067	451	5.478			484	n/a

Table 5.1 Cα RSMD comparison of superimposed DASS structures.

The number of Cα atoms the Matchmaker algorithm used to calculate the RMSD values is also given for reference.

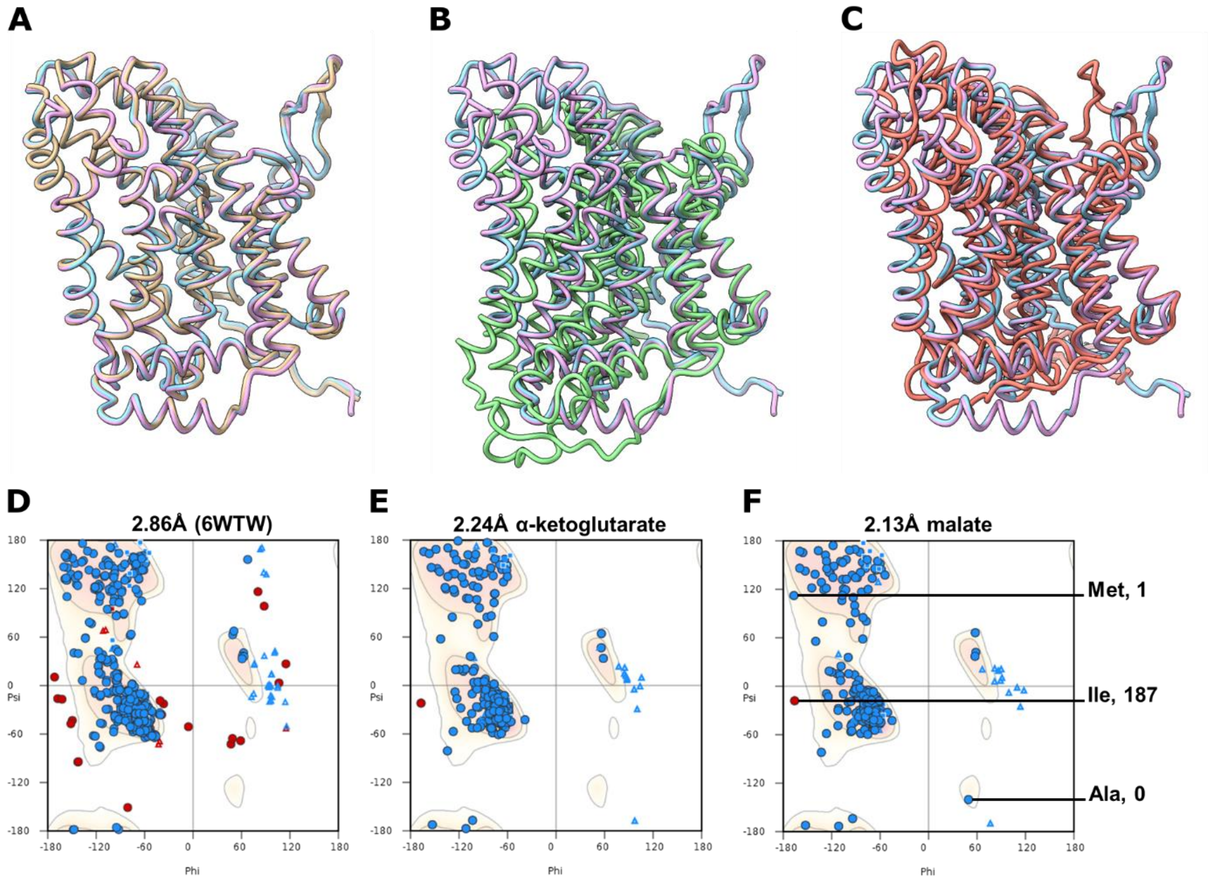


Figure 5.1 Reported LaINDY and DASS homolog structures superimposed with presented malate and α ketoglutarate models.

A, B, C. *Are single protomers of the structures of: LaINDY malate/α-ketoglutarate 2.86Å outward-open (PBD 6WTW) coloured brown, VcINDY 2.83Å inward-Na$^+$ bound (PDB 7T9G) coloured green, and ScPho90 2.29Å outward-open (PDB 8R33) coloured coral, superimposed with the 2.24Å α-ketoglutarate and 2.13Å malate LaINDY structures, coloured blue and pink respectively.* ***D, E, F.*** *Are the Ramachandran plots for the 6WTW, 5mM α-ketoglutarate and 5mM malate structures respectively. Triangles represent glycine, squares represent proline and circles represent all other amino acids. A red symbol indicates a residue with unfavorable bond angles, while a blue symbol indicates favoured bond angles. The altered N-terminal residues between E and F: alanine 0, and methionine 1 are highlighted in F. The single Ramachandran outlier common to both E and F: Isoluecine 187 is also highlighted in F. Ramachandran plots were generated with WinCoot.*

5.2.1 Presented Malate and α-ketoglutarate Structures are Structurally Identical

With Cα RMSD of 0.154Å between the presented α-ketoglutarate and malate LaINDY structures are structurally identical. An obvious minor difference is that in the higher resolution malate structure there was reliable density to model an N-terminal asparagine (N, -1). The modelling of this residue has altered the positions and hence the bond angles of alanine (A, 0) and methionine (M, 1) **(Fig. 5.1E/F)**. As the asparagine and alanine are both part of the flexible purification tag this should not be of any significance to the interpretation of the LaINDY structure. For the methionine the N-terminal phi angle difference is 39.33°, while the C-terminal psi angle is changed by only 3.95°. Combined with the fact the Cα for Met 1 has only shifted 0.39Å indicates the minor difference in methionine positioning is due to the modelling of the addition purification tag residue.

5.3 Isoleucine 187 is a Valid Ramachandran Outlier

The clearest improvement of the presented LaINDY structures to LaINDY-X-ray is the improvement of the bond angles between residues due to a number of peptide flips. Analysis of the LaINDY-X-ray map indicate this is due to the improved resolution, which for the following examples was ambiguous at the 2.86Å resolution of the LaINDY-X-ray structure. The Ramachandran plots in **Figure 5.1D/E/F** show the 14 unfavoured angles in LaINDY-X-ray reduced to a single residue (isoleucine 187) with unfavourable dihedral angles, and these changes are consistent for both malate and α-ketoglutarate structures.

Isoleucine 187 sits in a break in helix H5a within the elevator domain, and density for this outlier within the map is supported by the experimental map **(Fig. 5.2)**. There is also a salt-bridge pairing within 3 residues of the helix break between Arg184 (H5a) and Glu458 (H11). While these residues and the Arg-Glu salt bridge appear conserved within the VcINDY elevator, that protein's H5a helix does not have a break.

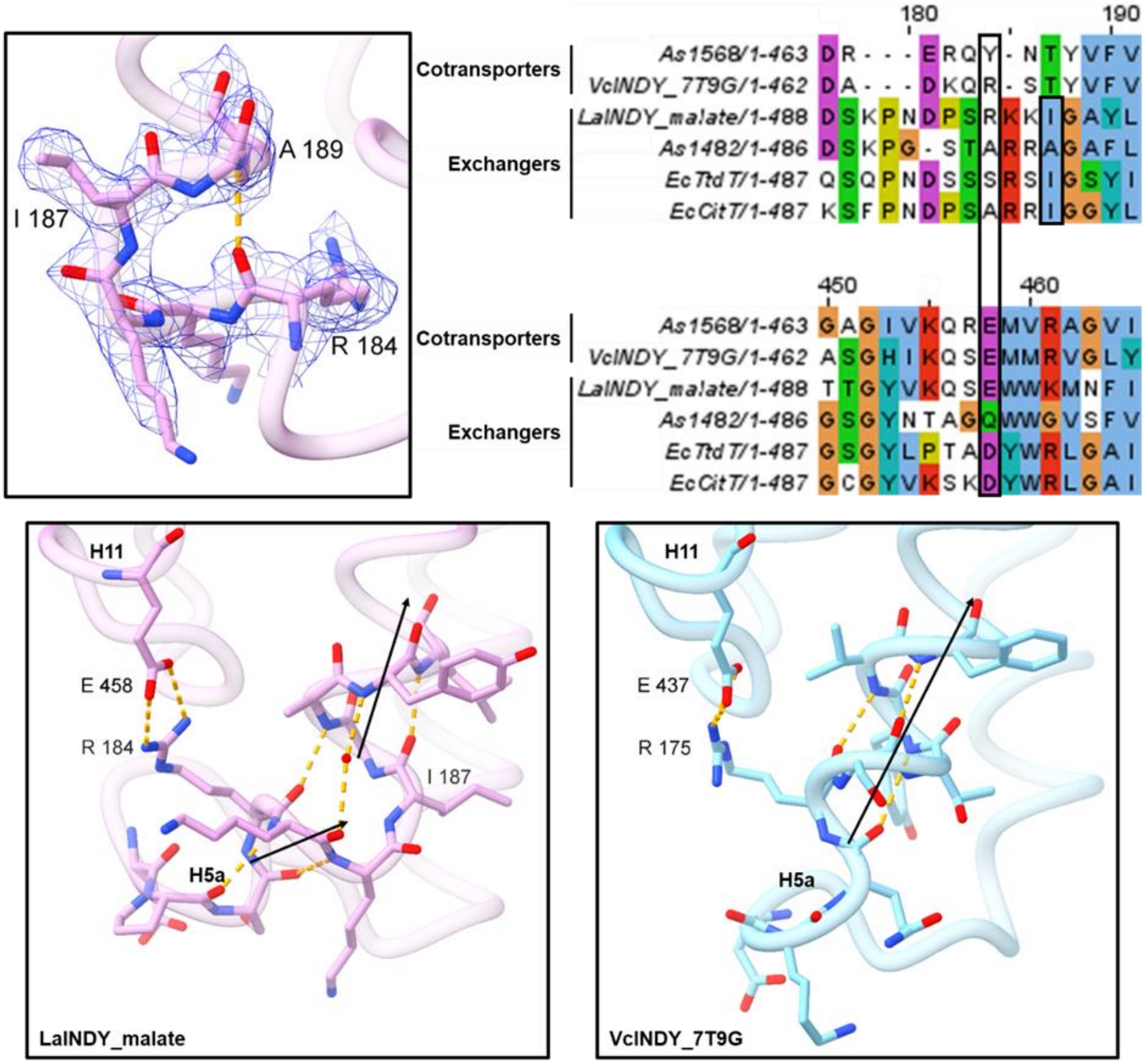


Figure 5.2 Examination of the LaINDY break in helix H5a containing isoleucine 187, and the respective helix in DASS-C VcINDY.

Helices H5a and H11 are shown for LaINDY malate and VcINDY (PDB: 7T9G) and are coloured pink and blue respectively, atoms are coloured by heteroatom. Hydrogen bonds were drawn to a cut-off of 3.2Å and are shown as orange dashed lines. Density for the H5a helix is shown as a blue mesh, to show the modelling is correct. A structural alignment of the LaINDY and VcINDY structures and the sequences of prokaryotic DASS members done via PROMALS3D using default settings is also shown. Alignment numbers are the respective positions in the LaINDY sequence. The positions of isoleucine 187 and the salt bridge pair are highlighted with a black box. Residues within the alignment are coloured according to the Clustal X colour scheme in Jalview v2.11.4.0.

Speculating a connection between these two features, a structural alignment of the elevator domains of LaINDY malate and VcINDY using the 2.83Å inward-sodium-bound structure (PDB 7T9G) [21], and the sequences of other prokaryotic DASS transporters from *E. coli* and *A. succinogenes* was done in PROMALS3D [207].

The sequence alignment indicates there is an insertion of an additional lysine (K, 185) residue in LaINDY, compared with VcINDY. The H11 acidic residue of the LaINDY salt bridge pair is conserved within all DASS members except for As1482, which coincidentally does not conserve isoleucine 187 either. The H5a arginine residue counterpart of the LaINDY salt bridge appears less conserved, but there are arginine residues within two residues of all the DASS members examined.

Considering all this information I propose that the isoleucine 187 Ramachandran outlier is caused by the combination of two factors: spatial constraints on local residues caused by maintaining the salt bridge, and the insertion of lysine (K, 185). I also proposed this may be a feature unique to LaINDY, as lacking the lysine insertion may effectively negate the +1 positioning of EcTtdT and EcCitT, allowing the helix H5 to fold as it does in VcINDY. While I do not believe there is any functional significance of the Isoleucine, it is certainly interesting and look forward to future DASS structures to test these hypotheses.

5.4 Peptide Flips Within the Model

The new high-resolution LaINDY malate structure was able to identify 10 regions of the peptide backbone that were mismodeled in the earlier LaINDY-X-ray structure. These have all been corrected in the revised model leaving the isoleucine 187 as the sole outlier as discussed in the previous section. I will highlight 3 regions which have important consequences for analysis and interpretation: Proline 257, the beta hairpin, and Serine 397 **(Fig. 5.3)**.

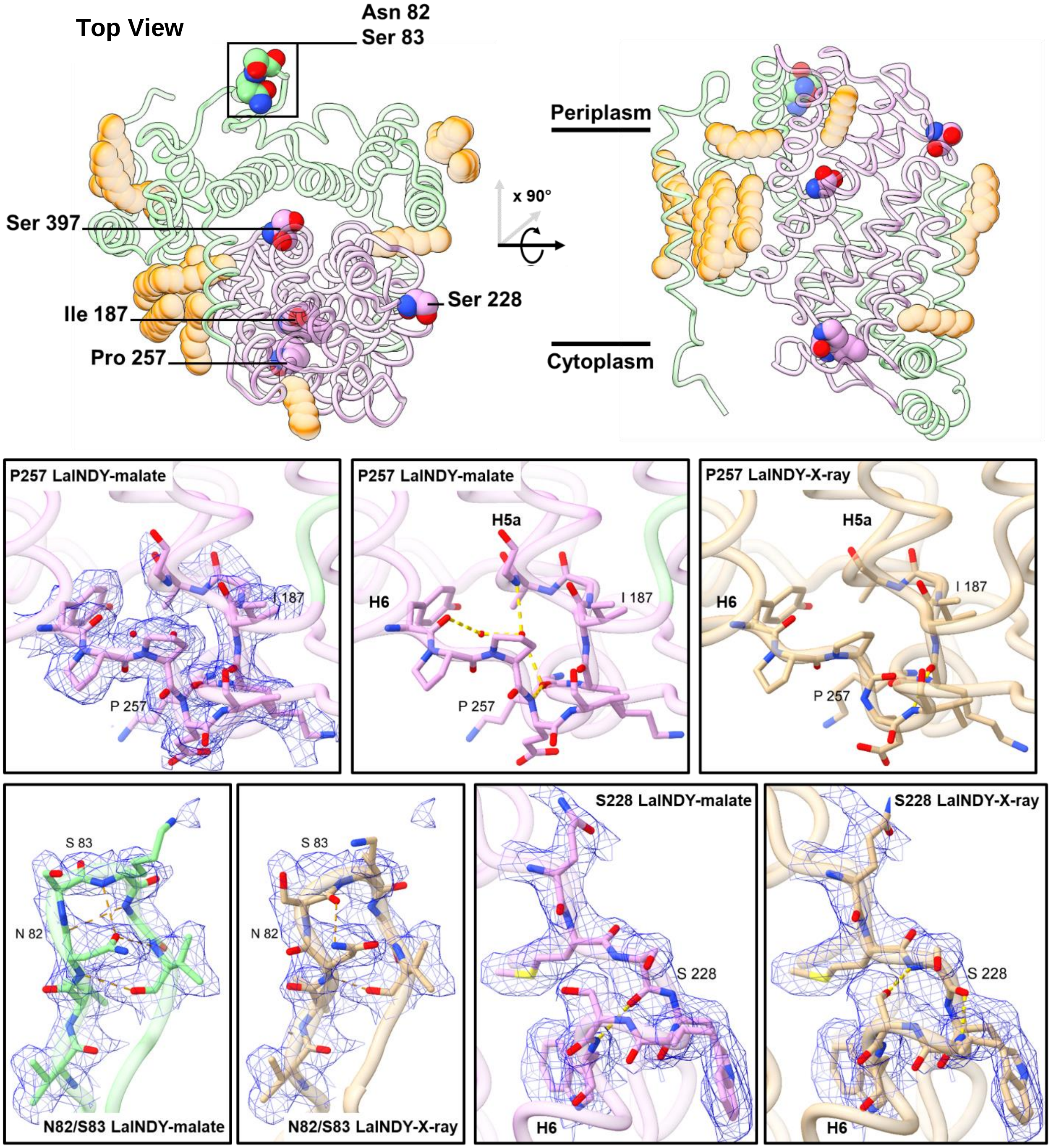


Figure 5.3 Overview of the peptide flips between presented LaINDY-malate and the LaINDY-X-ray structure.

In the overview, a single LaINDY protomer is viewed from the periplasmic side of the membrane, and within the membrane to aid with the positioning of the features. For the presented LaINDY structure the scaffold and elevator domains are coloured green, and pink respectively, while the whole of the LaINDY-X-ray peptide chain is coloured brown, otherwise atoms are coloured by heteroatom. Hydrogen bonds were drawn with a 3.2A cut off and are shown as orange dashed lines. The sharpened map is used to show the density for the features of the presented structure and is shown as a blue mesh.

5.4.1 Proline 257 Provides an Additional H-bond donor to Lysine 185

In immediate vicinity of the validated isoleucine 187 outlier is proline 257 which sits in the loop L6-6b. Correcting the peptide flip at proline 257 improves the dihedral angle metric for this region while also identifying an additional hydrogen bond donor for the backbone carbonyl of lysine 186. This additional hydrogen bond will definitely contribute to the stability of the helix break, and so shows a clear improvement of the model. Loop L6-6b in particular also highlights the other benefit of resolution this high: density for solvent molecules, two of which can be seen coordinated by residues within the H5a and H6-H6b loop. While this loop is solvent accessible, so the appearance of these densities is not surprising, all proteins will exist in contact with aqueous solvent and visualizing the interactions with solvent molecules will provide insights into protein-water interactions, which will be discussed in **Section 5.5**.

5.4.2 Scaffold Domain Beta Hairpin Double Flip

The short beta-hairpin within the scaffold domain was not well resolved in the LaINDY-X-ray structure, hence my new structure contains two sequential flips of asparagine 82 and serine 83. While obviously improving the dihedral angles of the hairpin residues, a side effect of this flip allows the carbonyl of the asparagine side chain to adopt a conformation in which it is an acceptor for hydrogen bonds from the amide groups for lysine 84 and valine 86. This small change should lower the energy of the model, therefore making it a more likely representation of LaINDY.

5.4.3 A Serine 397 Flip Removes Electronegative Carbonyl from Binding Site

Finally, the flipping of the first residue of the second SNT (LaINDY SGT) motif, the hallmark of the DASS subfamily, serine 397: which sits on the lip of the elevator binding domain in the HP_{out} loop **(Fig. 5.4)**. This replaces the electronegative carbonyl with the amide group which enables modelling of a hydrogen bond between the residue's backbone carbonyl and histidine 401.

As serine 397 is within the canonical binding site of DASS transporters this peptide flip will have significant effects on the modelled protein substrate interactions of LaINDY. In particular the MD simulations of LaINDY's interaction with succinate in the *Sauer et al.,* paper may have been impacted by the model error.

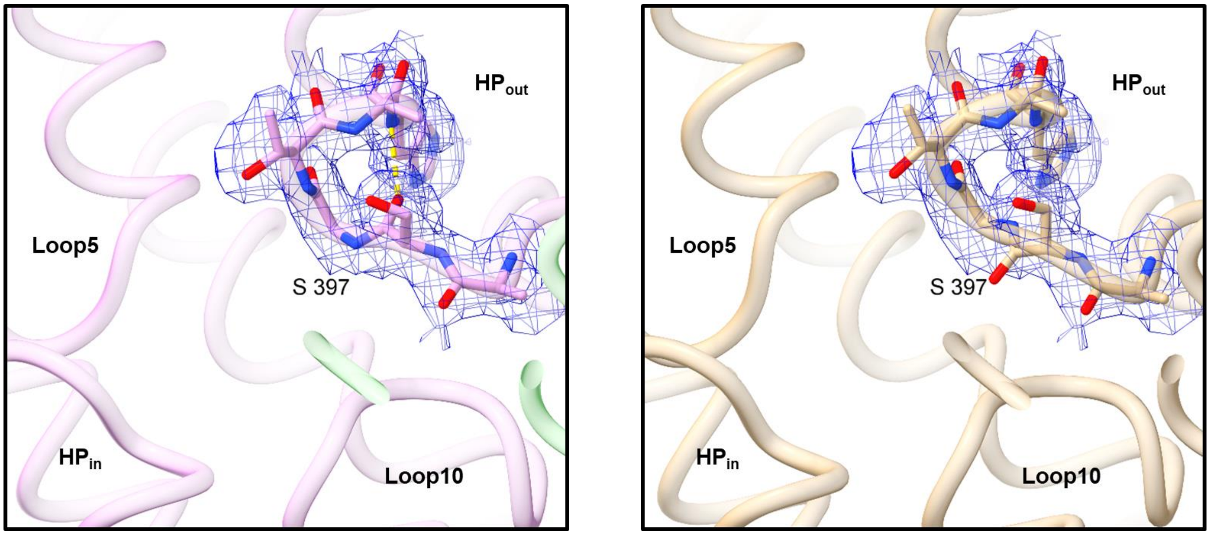


Figure 5.4 The Serine 397 flip within the canonical SNT motif viewed from the binding site.

For the presented LaINDY structure the scaffold and elevator domains are coloured green, and pink respectively, while the whole of the LaINDY-X-ray peptide chain is coloured brown, otherwise atoms are coloured by heteroatom. Hydrogen bonds were drawn with a 3.2A cut off and are shown as orange dashed lines. The sharpened map is used to show the density for the features of the presented structure and is shown as a blue mesh

A side effect of this serine 397 flip carbonyl flip is it that it becomes the first residue of the second HP$_{out}$ helix, rather than the preceding turn within the model. A similar flip occurs with serine 228, only becoming the first residue of the helix H6 **(Fig. 5.3)**.

5.5 Novel Waters Within the LaINDY Protomer

The higher resolution of my new structures yielded numerous non-protein densities that corresponded to solvent, lipids or the amphipol membrane mimetic. These were built as 496 waters and 22 acyl chains, and waters were validated using metric-ion-classification and CheckMyMetal v2.0 [160, 208] **(Fig. 5.5)**. While the waters in the binding site have already been discussed in

Section 4.4 numerous other waters occupy functionally important regions of the protein and are therefore discussed in greater detail below **(Fig. 5.6)**.

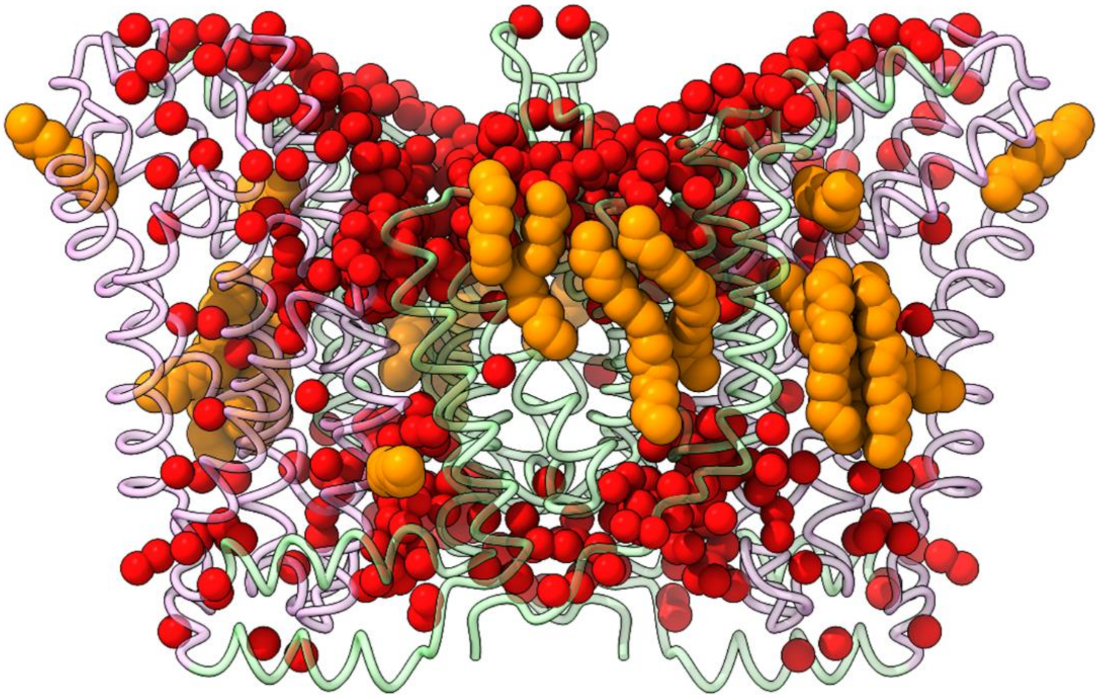


Figure 5.5 Cartoon representation of the LaINDY malate model highlighting all modelled waters and acyl chains.

The LaINDY homodimer is shown, with residues from the scaffold and elevator domains coloured green and pink respectively. The 496 modelled water molecules are shown as red spheres, while the 22 acyl chains are orange spheres.

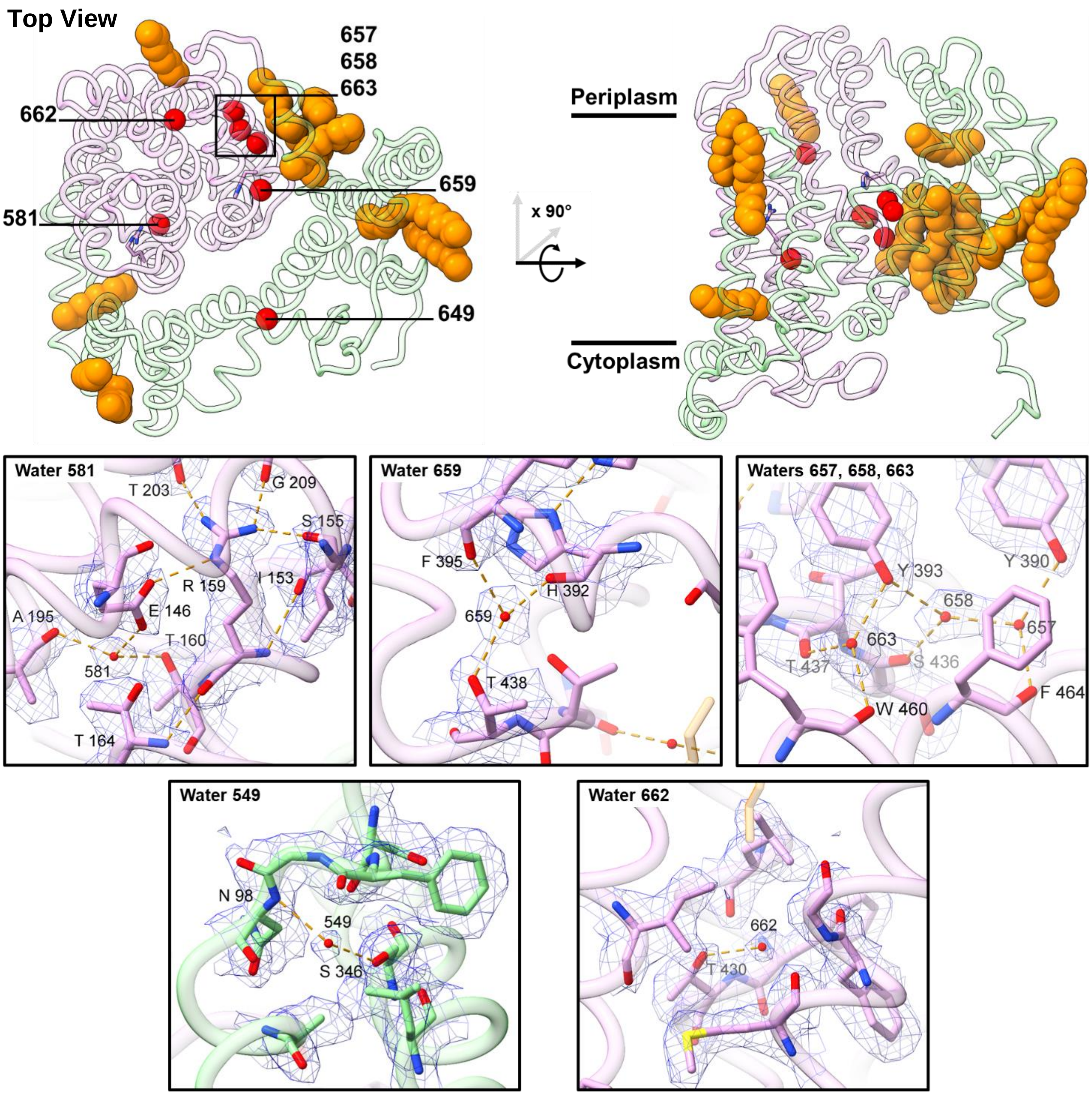


Figure 5.6 Overview of the interesting water densities within the LaINDY protomer.

In the overview, a single LaINDY protomer viewed from the periplasmic side of the membrane, and within the membrane to aid with the positioning of the features. The scaffold and elevator domain are coloured green, and pink respectively. Waters are shown as red spheres in the overview and in stick view in the highlight windows. Hydrogen bond pairs were drawn with a cutoff of 3.2Å and are shown as orange dashed lines. The sharpened map was used to model the waters, represented by a blue mesh. Densities of labelled interacting residues are shown to demonstrate appropriate contouring.

5.5.1 Water 581 Interacts with the Arginine 159 – Glutamate 146 Salt Bridge

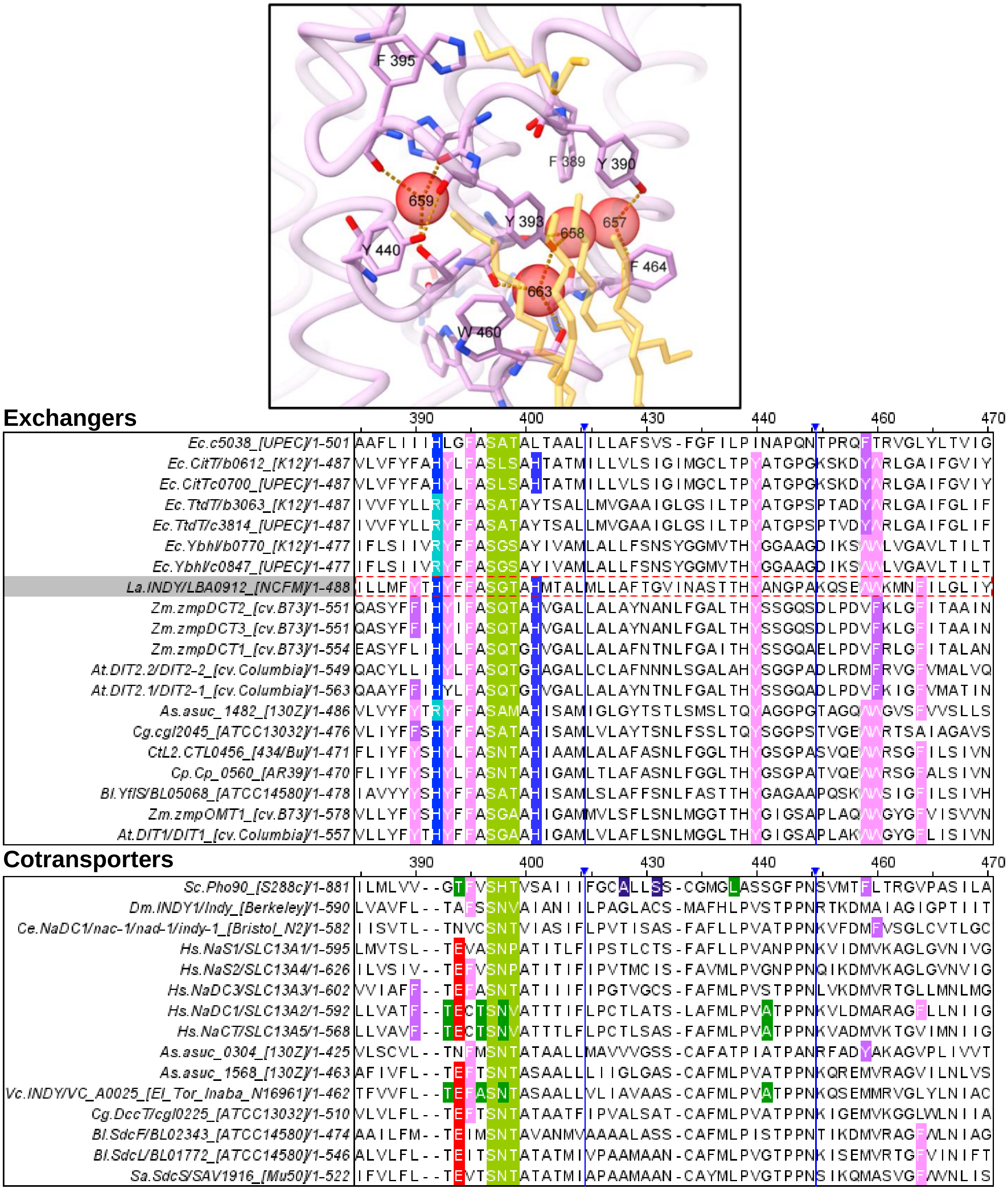
An obvious key solvent molecule revealed by the new structure is Water 581 located in between the HP_{in} helices within the LaINDY binding site and interacting with alanine 195, threonine 160 and glutamate 146 **(Fig. 5.6)**. This is notable, as an acidic residue in the position of glutamate 146 is highly conserved in the DASS exchangers, and is thought to be in a salt bridge pairing with arginine 159, the Na1 ion replacement residue [14]. It is worth noting that the rotamers of both glutamate 154 and arginine 159 are in alternate conformations compared to the previous LaINDY-X-ray structure. This results in more potential hydrogen bonding partners for both residues, lowering the energy of the new model. The short distances in apparent hydrogen bonds of 2.74Å, 2.79Å and 2.88Å for the glutamate, threonine and alanine, respectively, indicated tight coordination of the water. This strong coordination suggests the water is less dynamic that more peripherally bound waters elsewhere in the protein, but at present this is mere speculation and will require further experimentation.

5.5.2 Water 659 and the "Aromatic Cage" Network

On the opposing side of the elevator domain, water 659 is coordinated by the Na2 ion replacement residue histidine 392, phenylalanine 395 and threonine 438, with the respective hydrogen-bonding distances of 2.37Å, 2.98Å and 2.80Å. It is interesting that, either directly or indirectly, both sodium surrogate residues have strongly coordinated solvent molecules.

Immediately behind water 659, towards the elevator bundle of acyl chains, are water 657, 658 and 663, which are coordinated by a number of bulky aromatic residues which are highly conserved within the exchanger clade of the DASS subfamily **(Fig. 5.7)**. These four highly coordinated waters 657, 658, 659, and 663 could enter through the binding cavity, the bulky aromatics of helices HP_{out} and H11 conceivably creating the cavity through steric hindrance of each other. All atom RMSD is 0.203Å between the 58 atom pairs of the “aromatic cage” residues that coordinate the waters (Phe 390, Tyr 393, Phe 395, Tyr 440, Trp 440, Phe 464) in the LaINDY-X-ray and my presented malate structure. This indicates the solvent densities and tight coordination is not an artifact of the amphipol purification, and were probably present within the crystal structure without the resolution to resolve them.

This discovery raises and poses many questions about LaINDY and the DASS family. Now more than ever, a high-resolution structure of the LaINDY inward-open conformation is needed to see if these densities are consistent in both states. Do these waters become a de-facto part of the elevator domain? If so, are they co-transported, and if not, are they ejected prior to the conformational change?



Exchangers

Ec.c5038_[UPEC]/1-501	AAFLIIIHLGFASATALTAALILLAFSVS-FGFILPINAPQNTPRQFTRVGLYLTVIG
Ec.CitT/b0612_[K12]/1-487	VLVFYFAHYLFASLSAHTATMILLVLSIGIMGCLTPYATGPGKSKDYARLGAIFGVIY
Ec.CitTc0700_[UPEC]/1-487	VLVFYFAHYLFASLSAHTATMILLVLSIGIMGCLTPYATGPGKSKDYARLGAIFGVIY
Ec.TtdT/b3063_[K12]/1-487	IVVFYLLRYFFASATAYTSALLMVGAAIGLGSILTPYATGPSPTADYARLGAIFGLIF
Ec.TtdT/c3814_[UPEC]/1-487	IVVFYLLRYFFASATAYTSALLMVGAAIGLGSILTPYATGPSPTVDYARLGAIFGLIF
Ec.YbhI/b0770_[K12]/1-477	IFLSIIVRYFFASGSAYIVAMLALLFSNSYGGMVTHYGGAAGDIKSWWLVGAVLTILT
Ec.YbhI/c0847_[UPEC]/1-477	IFLSIIVRYFFASGSAYIVAMLALLFSNSYGGMVTHYGGAAGDIKSWWLVGAVLTILT
La.INDY/LBA0912_[NCFM]/1-488	ILLMFYTHYFFASGTAHMTALMLLAFTGVINASTTHYANGPAKQSEWWKMNFILGLIY
Zm.zmpDCT2_[cv.B73]/1-551	QASYFFIHYIFASQTAHVGALLALAYNANLFGALTHYSSGQSDLPDVFKLGFITAAIN
Zm.zmpDCT3_[cv.B73]/1-551	QASYFFIHYIFASQTAHVGALLALAYNANLFGALTHYSSGQSDLPDVFKLGFITAAIN
Zm.zmpDCT1_[cv.B73]/1-554	EASYFLIHYLFASQTGHVGALLALAFNTNLFGAITHYSSGQAELPDVFRLGFITALAN
At.DIT2.2/DIT2-2_[cv.Columbia]/1-549	QACYLLIHYLFASQTGHAGALLCLAFNNNLSGALAHYSGGPADLRDMFRVGFVMALVQ
At.DIT2.1/DIT2-1_[cv.Columbia]/1-563	QAAYFFIHYLFASQTGHVGALLALAYNTNLFGALTHYSSGQADLPDVFKIGFVMATIN
As.asuc_1482_[130Z]/1-486	VLVYFYTRYFFASAMAHISAMIGLGYTSTLSMSLTQYAGGPGTAGQWWGVSFVVSLLS
Cg.cgl2045_[ATCC13032]/1-476	VLIYFFSHYFFASATAHISAMLVLAYTSNLFSSLTQYSGGPSTVGEWWRTSAIAGAVS
CtL2.CTL0456_[434/Bu]/1-471	FLIYFYSHYLFASNTAHIAAMLALAFASNLFGGLTHYGSGPASVQEWWRSGFILSIVN
Cp.Cp_0560_[AR39]/1-470	FLIYFYSHYLFASNTAHIGAMLTLAFASNLFGGLTHYGSGPATVQEWWRSGFALSIVN
Bl.YflS/BL05068_[ATCC14580]/1-478	IAVYYYSHYFFASATAHISAMLSLAFFSNLFASTTHYGAGAAPQSKWWSIGFILSIVH
Zm.zmpOMT1_[cv.B73]/1-578	VLLYFYSHYFFASGAAHIGAMMVLSFLSNLMGGTTHYGIGSAPLAQWWGYGFVISVVN
At.DIT1/DIT1_[cv.Columbia]/1-557	VLLYFYTHYFFASGAAHIGAMLVLAFLSNLMGGLTHYGIGSAPLAKWWGYGFLISIVN

Cotransporters

Sc.Pho90_[S288c]/1-881	ILMLVV--GTFVSHTVSAIIIFGCALLSS-CGMGLASSGFPNSVMTFLTRGVPASILA
Dm.INDY1/Indy_[Berkeley]/1-590	LVAVFL--TAFSSNVAIANIILPAGLACS-MAFHLPVSTPPNRTKDMAIAGIGPTIIT
Ce.NaDC1/nac-1/nad-1/indy-1_[Bristol_N2]/1-582	IISVTL--TNVCSNTVIASIFLPVTISAS-FAFLLPVATPPNKVFDMFVSGLCVTLGC
Hs.NaS1/SLC13A1/1-595	LMVTSL--TEVASNPATITLFIPSTLCTS-FAFLLPVANPPNKVIDMVKAGLGVNIVG
Hs.NaS2/SLC13A4/1-626	ILVSIV--TEFVSNPATITIFIPVTMCIS-FAVMLPVGNPPNQIKDMVKAGLGVNVIG
Hs.NaDC3/SLC13A3/1-602	VVIAFF--TEFASNTATIIIFIPGTVGCS-FAFMLPVSTPPNLVKDMVRTGLLMNLMG
Hs.NaDC1/SLC13A2/1-592	LLVATF--TECTSNVATTTIFLPCTLATS-LAFMLPVATPPNKVLDMARAGFLLNIIG
Hs.NaCT/SLC13A5/1-568	LLVAVF--TECTSNVATTTLFLPCTLSAS-FAFMLPVATPPNKVADMVKTGVIMNIIG
As.asuc_0304_[130Z]/1-425	VLSCVL--TNFMSNTATAALLMAVVVGSS-CAFATPIATPANRFADYAKAGVPLIVVT
As.asuc_1568_[130Z]/1-463	AFIVFL--TEFTSNTASAALLLIIGLGAS-CAFMLPVATPPNKQREMVRAGVILNLVS
Vc.INDY/VC_A0025_[El_Tor_Inaba_N16961]/1-462	TFVVFL--TEFASNTASAALLVLIAVAAS-CAFMLPVATPPNKQSEMMRVGLYLNIAC
Cg.DccT/cgl0225_[ATCC13032]/1-510	VLVLFL--TEFTSNTATAATFIPVALSAT-CAFMLPVATPPNKIGEMVKGGLWLNIIA
Bl.SdcF/BL02343_[ATCC14580]/1-474	AAILFM--TEIMSNTAVANMVAAAALASS-CAFMLPISTPPNTIKDMVRAGFWLNIAG
Bl.SdcL/BL01772_[ATCC14580]/1-546	ALVLFL--TEITSNTATATMIVPAAMAAN-CAFMLPVGTPPNKISEMVRTGFVINIFT
Sa.SdcS/SAV1916_[Mu50]/1-522	IFVLFL--TEVTSNTATATMIAPAAMAAN-CAYMLPVGTPPNSIKQMASVGFWNLIS

Figure 5.7 Conservation of the "aromatic cage" within the DASS-E clade.

A cartoon representation of the LaINDY elevator domain, with the peptide coloured pink. Only aromatic residues or residues coordinating waters 657, 658, 659 and 663 are shown. Acyl chains are coloured orange, otherwise atoms are coloured by heteroatom.

continued on next page

Figure 5.7 continued.

Labeled residues are also labelled in the sequence alignment. A sequence alignment of residues surrounding the SNT2 motifs shown, done in Clustal Omega v1.2.2 via the EMBL Job Dispatcher. The top panels show the exchangers while the bottom panels show the cotransporters. Breaks in the alignment are indicated by blue vertical lines. LaINDY is highlighted with a grey box. SNT motif residues are green. The conserved aromatic cage is coloured light pink, while neutral mutations (aromatic preserving) residues are a darker pink. The DASS-E canonical histidine pairing are blue, while the arginine mutations are a light blue. DASS-C residues that coordinate sodium ions within the Na2 site based on the existing structures are coloured dark green. The DASS-C highly conserved glutamate is coloured red. Residues coordinating sodium within ScPho90s Na3 site are coloured purple.

5.5.3 Conformationally Sensitive Water 649 sits in a break in Helix H4

Water 649 sits at the dimer interface in the helix break between 4a and 4b, and hydrogen bonds with the amine of arginine 98 and hydroxyl of serine 346 of helix 9b at distances of 3.03Å and 2.74Å respectively. Directly above this helix break is the solvent proper, and it is not impossible to imagine that during the elevator transition subtle movements within helices 4a and 4b permit water access. Therefore, being located at this key dynamic region of the DASS transport fold, I hypothesize that changes in the conformation for the transporter regulate accessibility for a water at this site.

5.5.4 Poorly Coordinated Water 662 in a Solvent Inaccessible Elevator Cleft

Water 662 sits in a cleft of the elevator domain between helices HP_{out}, H10a and H11, which is a site also occupied by a single acyl chain. With only a single residue threonine 430, within the 3.2Å hydrogen bonding cutoff it has the least well-defined coordination site of all the mentioned waters. Furthermore, it is unclear how a water got to this location, 9.9Å away from the nearest water molecule, and 17.9Å away from the solvent accessible periplasmic surface of the protein. One possibility is the transition between inward and outward states could briefly allow solvent access, trapping a solvent molecule in the poorly defined site. However, leakage from the binding site during the natural flexing of random thermal motion cannot be excluded.

5.6 Acyl Chain Densities Surrounding the LaINDY Protomer

As mention in **Section 1.9.1.1** lipid densities have appeared in many of the recent structures of DASS members, and the relationship between lipid binding and DASS function has begun to explored with studies on VcINDY and NaDC1 [30, 99]. In the 3.36Å malate-amphipol LaINDY structure (herein LaINDY-malate-Sauer) (PDB 6WU2) Sauer *et al.,* modelled a number of short acyl chains surrounding the periplasmic side of their LaINDY structure: specifically, at the dimer interface of the scaffold domains; in an insert between H1 and H3; and in between the H9c arm and H11 of the elevator domain **(Fig. 5.8)**. While we do not believe the protein is surrounded by alkanes, their merit lies in modelling the visible portions of tightly bound lipids/amphipol to LaINDY. The increased resolution in the presented malate structure has not just improved and increased in length the densities for their modelled lipids but revealed novel densities as well **(Fig. 5.9)**. I have grouped the lipids into four categories based on their location within the protomer: The Dimer interface, the elevator bundle, the perpendicular arm helix inserts, and the elevator cleft. All four of the 'elevator bundle' acyl chains were present in the 3.36Å malate-amphipol LaINDY, and have been greatly extended to 3 dodecane chains and a single tetradecane.

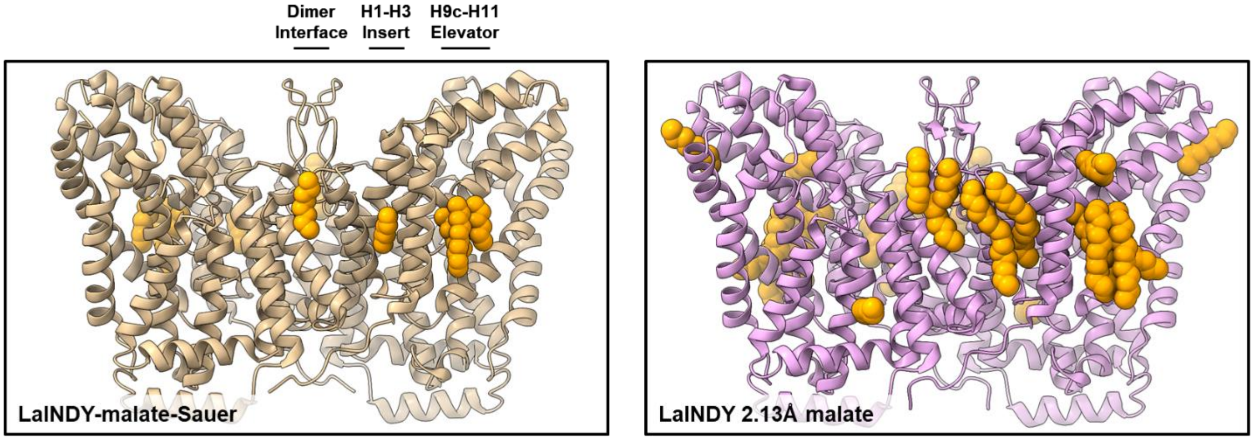


Figure 5.8 Modelled acyl chains compared with previous LaINDY structure.

The dimeric LaINDY-malate-Sauer (PDB 6WU2) and presented 2.13Å malate are shown, coloured brown and pink respectively. Modelled acyl chains are shown as orange spheres.

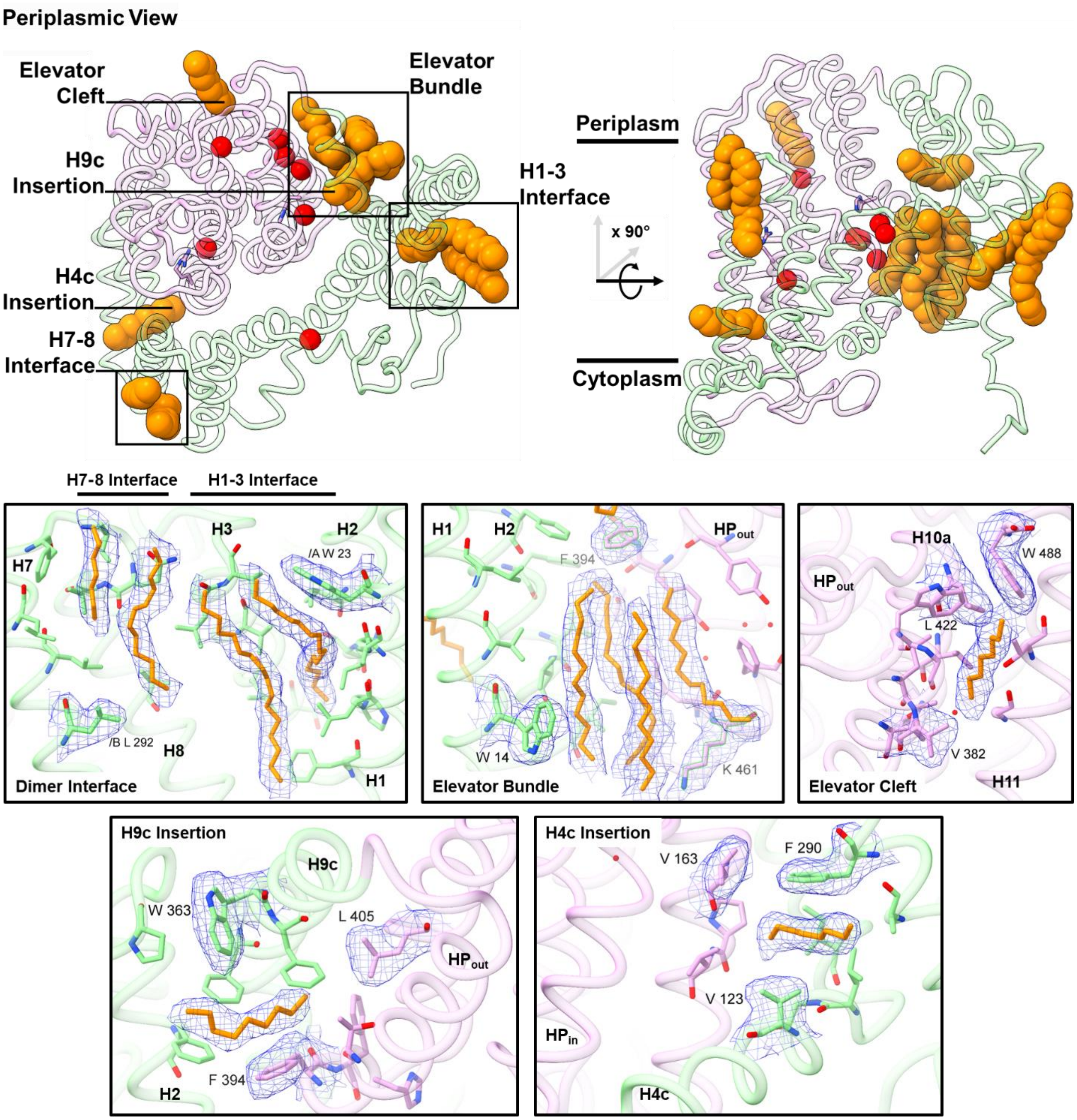


Figure 5.9 Overview of the Acyl chain densities surrounding the LaINDY protomer.

In the overview, a single LaINDY protomer viewed from the periplasmic side of the membrane and within the membrane, to aid with the positioning of the features. The scaffold and elevator domain are coloured green and pink respectively. Acyl chain densities are coloured orange, and the structural waters discussed in the previous section are coloured red. The unsharpened map was used to model the acyl chains as local densities were on the lower end of the resolution spread. Densities are shown along with labelled interacting residues to show appropriate contouring.

5.6.1 Dimer Interface Lipids

Within the LaINDY-malate-Sauer structure, an octane and hexane chain were modelled at the dimer interface. In their place my new malate structure has four densities, which have been tentatively modelled with a heptane, a dodecane, and two cetane acyl chains **(Fig. 5.9)**. While their exact molecular identity is unclear, intriguingly lipids with acyl chains between 16-18 carbon atoms are the most abundant in bacteria membranes [200], and the two cetane densities appear to connect at low map contours. However, as the model was built at much high contours to avoid fitting noise, I was reticent to build a phospholipid in which there was no obvious headgroup interaction with the LaINDY protomer without attempting to identify it experimentally. The obvious first step was to see if the density for the lipid could be improved by looking at the 2.15Å pre-CyroSieved map of 519,444 particles, to see if the headgroup density could be improved. The LaINDY-malate model was fit into the pre-CryoSieved map in ChimeraX, and indeed the density above where the two cetane chains joined appeared big enough to fit a phosphate moiety **(Fig. 5.11)**.

5.6.1.1 Identifying the Dimer Interface Phospholipid Density with TLC

As mentioned in **Section 2.4** *E. coli* was used as the system for LaINDY overexpression, and as such any potential lipids within the structure would be native to *E. coli*. *E. coli* possess a comparatively small pool of major membrane lipids, namely: ~75% PE, ~20% PG and ~5% CA [200-202]. Therefore, I decided to use TLC to help narrow the identity of the unknown lipid down, using the three major polar lipids, DDM and PMAL-C8 as standards **(Fig. 5.10A)**. Lipids from protein leftover from the LaINDY-malate-PMAL-C8 Cryo-EM sample were extracted and dried as a film onto a glass tube under nitrogen gas. Extracted lipids were resuspended in chloroform and titrated across a silica plate. A recipe for a solvent front of 65:25:10 v/v/v chloroform:methanol:’acetic acid’ from *Hughes et al.,* was used to improve the separation of PE and PG [161]. The plate was stained with PMA and developed with a heat gun **(Fig. 5.10B)**.

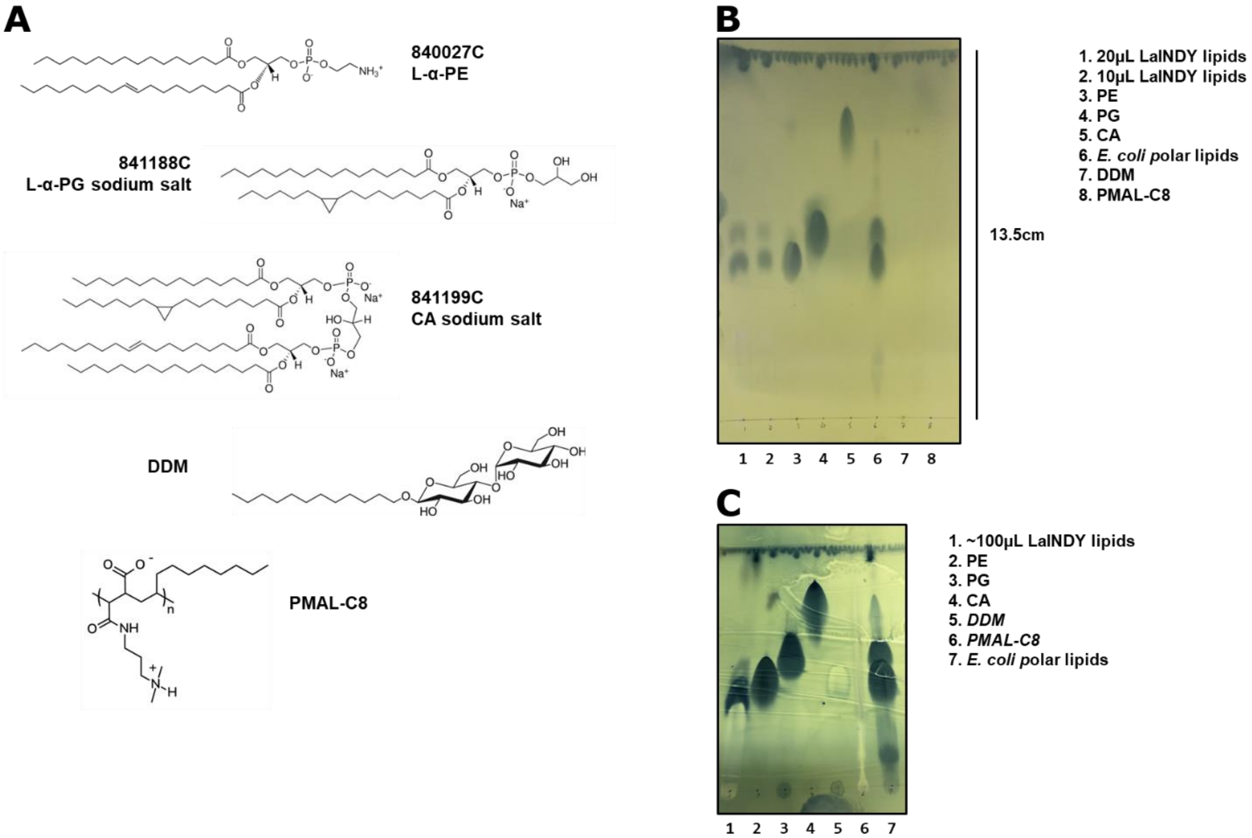


Figure 5.10 Thin layer chromatography of the LaINDY-PMAL-C8 malate Cryo-EM sample.

A. *Example structures of: the three major phospholipid components of the E.coli membrane PE, PG, and CA, with their Avanti Polar Lipids, Inc. product codes; and the purification additives DDM and PMAL-C8. Structures of the* E.coli lipids *were taken from their product pages on* www.sigmaaldrich.com*, while DDM and PMAL-C8 were taken from their product pages on* www.anatrace.com*.* ***B.*** *Silica plate from the TLC stained run for 13.5cm and stained with PMA.* ***C.*** *Silica plate from an overloaded, earlier TLC experiment stained with PMA. All spots are overloaded enough to visualize DDM and PMAL-C8 which do not react to the PMA stain.*

The TLC plate clearly shows that only PE and PG are present in the LaINDY-PMAL-C8 Cryo-EM sample, which is expected as the predominant lipids. As CA has the lowest abundance of the major lipids it could be argued that the titration of extracted lipids did not go high enough to rule show the CA. However, overloading (10x excess) all spots on the plate still reveals no CA spot **(Fig. 5.10C)**. Interestingly although DDM and PMAL-C8 are unreactive with the PMA stain, overloading the spots has revealed their migration pattern through the absence of the stain, and the PMAL-C8 smear can be clearly visualized in the LaINDY lipid lane.

5.6.1.2 Modelling PG as the "Dimer Interface" Phospholipid

Narrowing the lipid down to the either the zwitterionic PE and anionic PG, I used the surface chemistry of the dimer interface to decide which lipid to model at this site. Using PDB2PQR v3.6.1 and APBS v3.4.1 from the APBS suite of software tools [209], a potential map was generated. All parameters were kept as default except the pH which was adjusted from 7.0 to 7.5 to match the sample purification condition. The potential map revealed the surface of LaINDY to be largely electropositive, which was not surprising given its theoretical pI of 9.41, and function as a di-anion transporter. In particular at the dimer interface with the unidentified phospholipid density in question the local environment is positively charged **(Fig 5.11A)**.

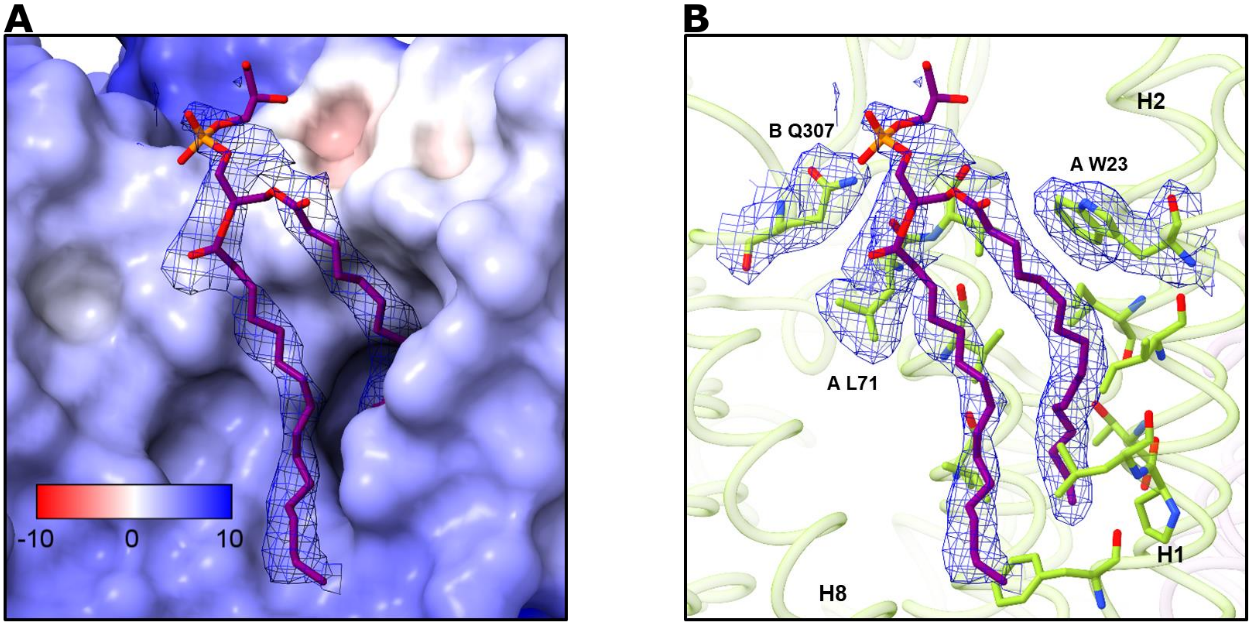


Figure 5.11 Electrostatic surface of the LaINDY dimer interface and a tentatively modelled DPPG.

A. *Electrostatic surface for the LaINDY dimer at pH 7.5 generated through the ABPS suite., focusing on the phospholipid density at the dimer interface. Red indicates a surface with negative electrostatic potential surface while blue positive. Density for the lipid is shown as a blue mesh of the 519,444-particle, 2.13Å C2 symmetry map. The modelled DPPG lipid is shown as purple, with non-carbons coloured by heteroatom.* ***B.*** *Cartoon representation of the potential DPPG lipid site showing the interacting residues. Density for the lipid is shown as a blue mesh of the 519,444-particle, 2.13Å C2 symmetry map, with density for chain A tryptophan 23 and chain B glutamine 307 to show appropriate contouring. The LaINDY scaffold domain is coloured green. The modelled DPPG lipid is shown as purple, with non-carbons coloured by heteroatom. Contouring between A and B has not been adjusted.*

5.6.2 Perpendicular Inserts Between Scaffold and Elevator

The most curious of the lipid densities present in the columbic map were two inserts, underneath the H9c and above the H4c arm helices respectively **(Fig. 5.9)**, near perpendicular compared to the usual orientation of lipids within a bilayer. On the His392 side of the elevator in an aromatic pocket in between H9c and the HP_{in} helices is a modelled nonane chain **(Fig. 5.12A/B)**. While in an equivalent hydrophobic pocket on the Arg159 side of the elevator in between the HP_{out} helices is a modeled octane chain. The first question we had was whether or not this was an artifact of the PMAL-C8 purification. We have demonstrated that amphipol reconstitution reduces the micelle size of LaINDY vs DDM and hypothesized that the tight polymer wrapped around the protein could force a lipid acyl chain into these pockets. Examining the densities for the previous

LaINDY structures, the local resolutions were not high enough to reliably confirm this theory. Local resolution of all the acyl chains is among the lowest found within both our maps, so it follows that these densities are the first to disappear at a lower global resolution.

5.6.2.1 Density for the H9c Insert of ScPho90 is Less Pronounced Than in the LaINDY Map

There is a small density seen between the H9c-HPout helices of ScPho90 **(Fig. 5.12C)**, although it is significantly smaller than that of the LaINDY equivalent. The helices in ScPho90 around the density are populated with smaller flexible hydrophobic valines and leucine residues, whereas the equivalent positions in LaINDY are comprised mainly of phenylalanines. I hypothesize that the high proportion of aromatic residues in LaINDY at these positions restrict motion, creating a well define pocket for the lipid tail to insert.

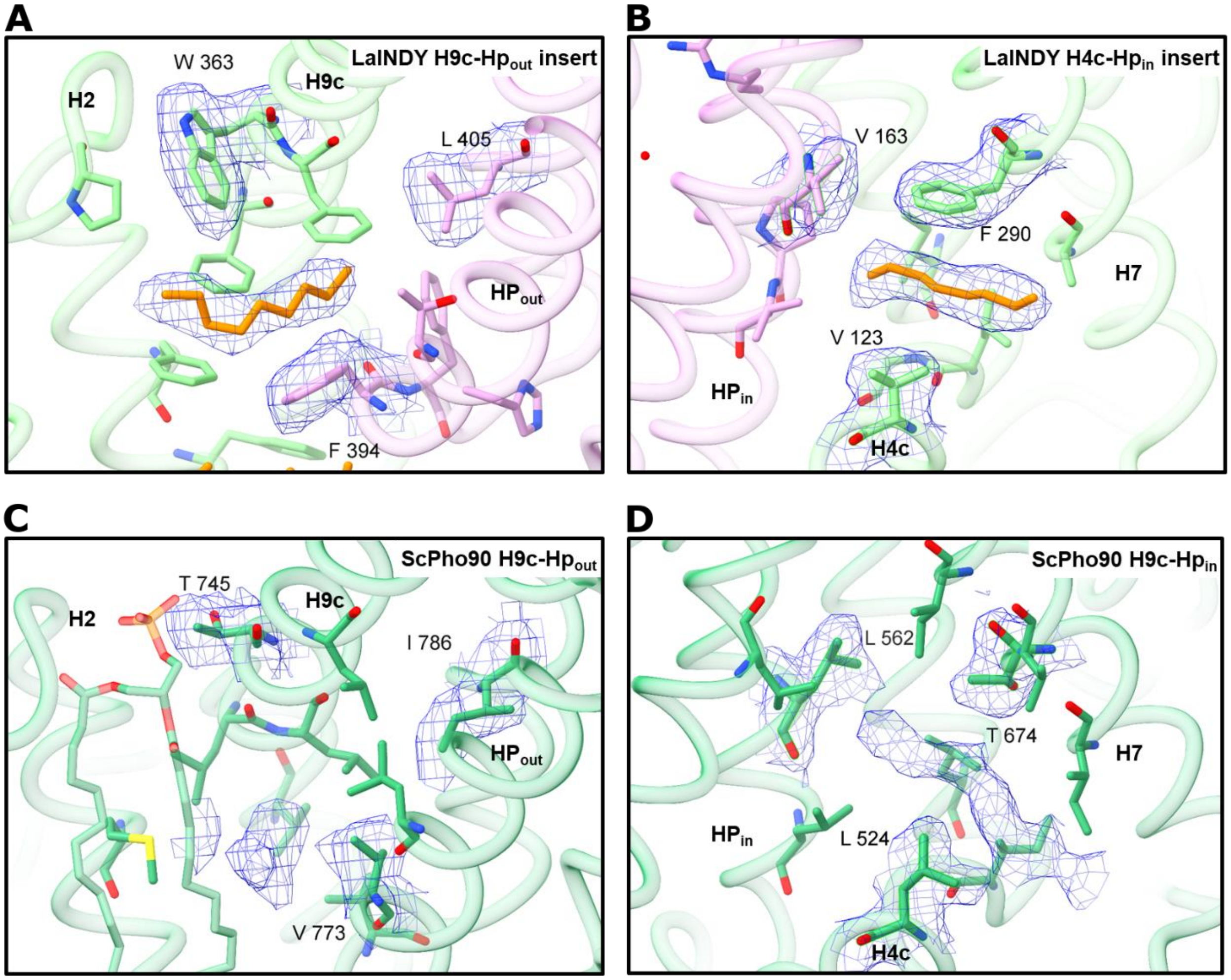


Figure 5.12 Comparison of the perpendicular lipid inserts within the scaffold-elevator interfaces of LaINDY and ScPho90.

A. *Cartoon representation of the H9c-HPout lipid insert of LaINDY. The scaffold domain is green while the elevator domain is pink. The nonane acyl chain is coloured orange, otherwise atoms are coloured by hetero atom, with nitrogen as blue, oxygen as red and sulphur as yellow. The unsharpened map is shown as a blue mesh, and labelled interacting residues density is shown to assure appropriate contouring.* ***B.*** *Cartoon representation of the H4c-HP$_{in}$ lipid insert of LaINDY. Colours are as in A.* ***C.*** *Cartoon representation of the H9c-HP$_{out}$ lipid insert of ScPho90. ScPho90 is coloured dark green.* ***D.*** *Cartoon representation of the H4c-HP$_{in}$ lipid insert of ScPho90. Colours are as in C. Deposited ScPho90 8R33 map densities are shown as a blue mesh, and labelled interacting residues densities are shown to assure appropriate contouring.*

5.6.2.2 Density for the H4c Insert of ScPho90 is More Pronounced Than in the LaINDY Map

The H4c-HPin insert comparison between LaINDY is ScPho90 is much clearer. In the ScPho90 map, there is an ~11 carbon density perpendicular to the membrane within the ScPho90 structure that was not modelled **(Fig. 5.11D)**. This density is greater in size compared to the corresponding LaINDY density in which an octane was modelled. This equivalency indicates these lipid insertions into the ‘arm helix’-‘hairpin helix’ cleft between the elevator and scaffold domains are likely also present for DASS transporters in their native membranes, rather than an artifact of the membrane mimetics necessary for protein purification and structure determination.

Whether there is a functional component to these inserts, or rather just a facet of acyl tails within the membrane filling space is unclear. It would be very interesting to see whether these insertions are present in the inward-open conformation with a high-resolution structure, or if the acyl chains provide some additional barrier to the transition between the outward-open – inward-open states with molecular dynamics.

5.7 Discussion

In this chapter the improvements to the LaINDY outward open model from the previous 2.86Å structure have been detailed through the lens of my presented 2.13Å LaINDY malate structure. The resulting peptide flips and rotamer alterations should result in a lower energy model that is truer to reality. In particular, the flipping of serine 397 within the canonical SNT motif od the DASS family alters the surface chemistry within the substrate binding domain of the model which would have impacted subsequent molecular dynamics simulations.

The higher resolution has led to the validation of a single Ramachandran outlier: isoleucine 187, residing in the elevator domain within a break in the H5a helix. Through structural comparisons with DASS cotransporter VcINDY, I have proposed that the combination of a conserved nearby salt bridge pairing (Arg184-Glu458), and a residue insertion (Lys185) to be the cause of this helix break. Further, I have hypothesized this may be a feature unique to LaINDY, based on the positioning of salt bridge pairing arginines in other DASS exchangers via structural alignment results. To test these hypotheses, solving the structures of EcTtdT or EcCitT which have arginine residues that are +1 positionally from that of LaINDY, and comparing their respective folds would be the obvious choice. Especially as to date LaINDY is still the only DASS exchanger with a structure.

5.7.1 Highly Coordinated Waters Within the DASS Elevator

My structures have revealed novel, tightly coordinated waters within the elevator domain, coordinated by the highly conserved “aromatic cage” residues within the DASS exchanger clade. Are these waters a *de facto* part of the elevator domain? If so are they co-transported, and if not, how are they ejected prior to the conformational change? I would suggest exploring these theories using molecular dynamics to simulate the water molecules as the elevator translocates between inward and outward states. The hydroxyl moieties of tyrosines 390 and 393 contribute to the hydrogen bond network of three of the highly coordinated waters. The AdiC study demonstrated that disrupting water mediated interactions between monomers at the dimer interface was able to significantly reduce the proteins thermostability [143]. As such performing similar experiments on LaINDY mutating both tyrosine residues to phenylalanine may yield similar results, revealing more information about nature of their role within LaINDY.

This discovery of these waters also raises questions about the whole DASS family. Whilst I would expect to see these highly coordinated waters in the DASS-E clade, could there be analogous sites in the other DASS-C members that would be revealed given high enough resolution? Only the ScPho90 outward-open structure has reached a comparable resolution, and it did not reveal any highly coordinated waters. It could be possible that this phenomenon is just in the prokaryote DASS members, however if highly ordered waters are revealed in subsequent DASS-C structures with appropriately high resolution then it could imply structural waters play an important role in the DASS translocation mechanism. It is clear we will have to wait for more structures to gain insight.

5.7.2 Waters Within the Binding Domain of the LaINDY Elevator

As the structures I solved had such low substrate occupancy they were effectively apo, it remains a possibility that some of the waters modelled in the binding site may form part of the binding interface when substrates are bound. The LeuT and SiaP studies mentioned demonstrate water molecules ability to mediate interactions between substrate and protein [124, 128]. Coupling my own thermostability results on LAINDY with the screen of DASS-C member VcINDY by Sampson *et al.* demonstrating the breadth of di- and tricarboxylates that can interact with the protein [99], it remains highly plausible that some of the waters modelled in the binding site may be involved in the substrate binding, and could potentially be different for each substrate. Exploring this will require multiple high-resolution structures of LaINDY with substrate bound.

5.7.3 A PG lipid at the Dimer Interface

Novel densities for acyl chains around the LaINDY protomer have been discovered. I have hypothesized these are from an anionic PG lipid that sits at the LaINDY dimer interface with one of its acyl chains tightly bound in-between helices H1, H2, and H3. Although there was no direct interaction of the modelled PG lipid head group and LaINDY, the question is raised as to whether or not the lipids in the membrane have an impact on LaINDY's function. This has been demonstrated to be the case for proteins such as: AmtB, AqpZ, GlpF, and MscL [103, 104, 110, 115]; and within the DASS family hinted at with reports on VcINDY and NaDC1 [30, 99]. As mentioned in **Section 1.9.1.5** lipids at the dimer interface have been shown to confer dimer stabilization. However, I do not believe that this is the function of this lipid, for multiple reasons. Namely the DASS family have a strong dimeric interface: indicated by the Gupta *et al.* comparison between VcINDY and NapA [117], and the fact that no DASS members have been reported as monomeric. Additionally, the lipid modelled only interacts with a single monomer at the dimeric interface. The most interesting aspect of the lipid is the acyl chain tail that is tightly buried between helices H1 and H3. Could this acyl confer some kind of rigidity within the scaffold domain during the elevator translocation, and as such would removal of the lipid retard transport? Generating mutants of these hydrophobic residues that interact with this acyl chain in H1 and H3 to large, more sterically hindering residues like phenylalanine and tryptophan should be able to force out the acyl chain. As a result, it may be possible to measure altered transport rates of mutants in live cells or proteoliposomes. The exclusion of this acyl chain from H1 and H3 would need confirmation, most likely through complementary structures.

5.7.4 Understanding LaINDYs Interaction with Membrane Lipids may Require a More Native Membrane Environment

In **Section 4.6.3** I highlighted the key differences in plasma membrane composition between *L. acidophilus* and *E. coli*, namely: CA and MLCA as the major membrane components, and the presence of PA, and the TGDG and QGDG species of glycolipids not being present in *E. coli*. While the PG lipid I modelled is common to both, the question remains whether one of the more abundant host lipids is a more likely candidate. More specifically, given the fact that approximately 85% of the lipids within the *L. acidophilus* membrane are CA and MLCA whilst only 5% of the *E. coli* membrane is CA, it is highly possible that in a native environment CA or MCLA may be bound. There are multiple approaches to take to address this.

Firstly, *in vitro* a thermostability screen of different lipids with LaINDY could be performed, with positive results then added to the Cryo-EM sample in large excess to overcome the loss of host lipids in the membrane solubilization stage, as this approach has been successful in structurally characterizing lipid-protein interactions for AmtB [104].

Additionally, *in silico* my presented structures of LaINDY could be run in molecular dynamics software such as MemProtMD [210], or the CHARMM-GUI used by Mirandela *et al.* in their AmtB study [115], to simulate a native *L. acidophilus* membrane to predict how the protein interacts with lipids in its native membrane. This approach may also be able to provide more information regarding the perpendicular acyl chain densities that I saw in my columbic potential map and were also found in the ScPho90 map [25].

Finally, now more than ever a high-resolution structure of LaINDY is required in the inward-facing state, which could provide additional information on the membrane and solvent interactions observed in this study. High-resolution structures would better inform molecular for the complete reaction cycle would better inform the structural changes underlying LaINDY's transport mechanism and provide greater insight into substrate binding and transport for the exchanger clade of the DASS family.

6. Concluding Remarks

6.1 Discussion

In this thesis I have presented the two highest resolution structures of the DASS family reported to date: 2.24Å and 2.13Å structures of LaINDY in the outward-open conformation. Although functionally identical due to low substrate occupancy, the improvement of the model through peptide flips and rotamer corrections is a clear benefit to understanding substrate binding. Additionally, the ability to resolve features of LaINDY's local environment in the form of solvent molecules and membrane lipid tails has raised many interesting questions about its mechanism. Our understanding of the role of lipids have on the DASS family is in its infancy but lipids have been modelled in the structures of VcINDY [21], ScPho90 [25], NaS1 and NaDC1 [30] and now with greater detail, LaINDY. Further understanding of the relationship between LaINDY and membrane lipids will require more careful consideration of the native lipids of *L. acidophilus*, as the lipidome is significantly different from that of the *E. coli* strain used to express the protein. The complex relationship between lipids and substrate binding within the DASS family has begun to be explored with VcINDY, for which both cardiolipin and phosphatidyl ethanolamine have shown synergistic and antagonistic relationships respectively with succinate [99]. Answering these questions with LaINDY will require further structural studies, molecular dynamic simulations, and biophysical assays.

I have addressed the lack of substrate within the binding site through a tryptophan fluorescence assay, indicating that at pH 7.5 LaINDY has a low affinity for succinate, reporting a 395.1mM dissociation constant. As a result, my structures in solved in 5mM substrate would have had 1.2% occupancy, and this explains the water observed in the binding site of my experimental maps. This is in conflict with my own thermostability results, which have reported substrate binding by at 10mM by an increased average melting temperature of LaINDY for many dicarboxylates at. 10mM at pH 7.5. This difference in binding between experiments could be explained by the differences in temperature of the tryptophan fluorescence binding, nanoDSF and Cryo-EM grid freezing. Given LaINDY's reported average melting temperature of64°C in DDM at pH 7.5 in the nanoDSF, meaning all the stabilising interactions with substrates are measure at significantly higher temperatures. Whereas the vitrobot used to vitrify the Cryo-EM samples was set to 4°C prior to the ethane plunge. Additionally, the dissociation constant I measured is significantly higher than dissociation constants measured for DASS-C members VcINDY and ScPho90, determining K_Ds for for succinate 92.2 ± 47.4µM and phosphate 1.9 ± 0.4mM respectively [21, 25]. However, as cotransporters, in which it has been demonstrate sodium ion binding induces structural reorganisation to form the substrate binding site [21], there a clear difference in substrate binding mechanism for LaINDY, and it may have a lower affinity. Further, *L. acidophilus* is an acidophile, therefore LaINDY may have evolved to optimize binding at low pH and therefore performing the assay accordingly may yield binding kinetics consistent with or explanatory of DASS-E's expected physiological function.

It is also possible that we do not have the correct substrate for LaINDY, and it will have a higher affinity for its physiological ligand. Given it is in the same genomic neighbourhood as citrate lyase they may share the same substrates, and as such oxaloacetate and citrate are good contenders for further binding studies.

Taking all this into consideration, revealing more about the LaINDYs relationship with lipids, and substrate dynamics will require careful consideration of the protein's native environment.

6.2 Future Work on the Whole DASS Subfamily

During the undertaking of this thesis I have worked exclusively on DASS-E LaINDY, through comparison between it and the solved structures of other DASS members there are a number of specific questions that I personally would like to see explored by the community, or have the opportunity to research myself **(Fig. 6.1)**.

```
Ec.c5038_[UPEC]              VTIVLSLVVPSATARTGIQTSAAQNLLIIHLGFASATALTAALILLAFSVS-FGFILPINAPQNTPRQFTRVGLY
Ec.CitT/b0612_[K12]          IDILLAPFTPSNTARTMFVTGAAPNVLFAHYLFASLSAHTATMILLVLSIGIMGCLTPYATGPGKSKDYWRLGAI
Ec.CitTc0700_[UPEC]          IDILLAPFTPSNTARTMFVTGAAPNVLFAHYLFASLSAHTATMILLVLSIGIMGCLTPYATGPGKSKDYWRLGAI
Ec.TtdT/b3063_[K12]          SELILAPVTPSNSARGIFLTAMAPNLLLLRYFFASATAYTSALLMVGAAIGLGSILTPYATGPSPTADYWRLGAI
Ec.TtdT/c3814_[UPEC]         SELILAPVTPSNSARGIFLTAMAPNLLLLRYFFASATAYTSALLMVGAAIGLGSILTPYATGPSPTVDYWRLGAI
Ec.YbhI/b0770_[K12]          LDLVLAPATPSNTARAMFFTAMAGNILIVRYFFASGSAYIVAMLALLFSNSYGGMVTHYGGAAGDIKSWWLVGAV
Ec.YbhI/c0847_[UPEC]         LDLVLAPATPSNTARAMFFTAMAGNILIVRYFFASGSAYIVAMLALLFSNSYGGMVTHYGGAAGDIKSWWLVGAV
La.INDY/LBA0912              LELMMGALIPSNSARTLFITGAAPNLVYTHYFFASGTAHMTALMLLAFTGVINASTTHYANGPAKQSEWWKMNFI
Sb.DCT2.1/SORBI_3007G22670   SEAVISPAMPSTTARALFLTAAAQNLLFIHYIFASQTAHVGALLALAYNANLFGSLTHYSSGQSGLGDVFKLGFI
Sb.DCT2.2/SORBI_3007G22680   SEAFISPAMPSTTARALFLTAAAQNLLLIHYIFASQTAHVGALLALAYNANLFGALTHYSSGQSDLPDVFKLGFI
Zm.zmpDCT2                   SEALISPAMPSTTARALFLTAAAQNLLFIHYIFASQTAHVGALLALAYNANLFGALTHYSSGQSDLPDVFKLGFI
Zm.zmpDCT3                   SEALISPAMPSTTARALFLTAAAQNLLFIHYIFASQTAHVGALLALAYNANLFGALTHYSSGQSDLPDVFKLGFI
Sb.DCT4/SORBI_3004G035500    GEACIAPAMPSTAARALFLTAGAQNLLVIHYLFASQTGHVGALLALAFNTDLFGGITHYSSGQADLPDVFRIGFI
Sb.DCT1/SORBI_3002G233700    SEACIAPAMPSTTARALFLTAAAQNLLLIHYLFASQTGHVGALLALAFNTNLFGAITHYSSGQAELPDVFRLGFI
Zm.zmpDCT1                   SEACIAPAMPSTTARALFLTAAAQNLLLIHYLFASQTGHVGALLALAFNTNLFGAITHYSSGQAELPDVFRLGFI
At.DIT2.2/DIT2-2             CETLMGLIMPSTMARAILLTSAAQNLLLIHYLFASQTGHAGALLCLAFNNNLSGALAHYSGGPADLRDMFRVGFV
Ps.DCT/KIW84_015703          SEVLIAPAMPSTTARALFLTAAAQNLLFIHYLFASQTGHVGALLALGYNTNLFGAISHYSSGQADLPDIFRTGFI
At.DIT2.1/DIT2-1             SEALIAPAMPSTTARALFLTAAAQNLLFIHYLFASQTGHVGALLALAYNTNLFGALTHYSSGQADLPDVFKIGFV
So.DIT2/DIT2                 SEALVAPAMPSTTARALFLTAAAQNLLFVHYLFASQTGHVGALLALAYNTNLFGALTHYSSGQADLPDVFKMGFI
As.asuc_1482                 ADVVIGPGIPSASARGMFLTAMAGNPLYTRYFFASAMAHISAMIGLGYTSTLSMSLTQYAGGPGTAGQWWGVSFV
Cg.cgl2045                   ADLVLAPAIPSATARGMFLTAMAGNPLFSHYFFASATAHISAMLVLAYTSNLFSSLTQYSGGPSTVGEWWRTSAI
CtL2.CTL0456                 TDFLLAPAIPSLTARAMFLTAMAGNPIYSHYLFASNTAHIAAMLALAFASNLFGGLTHYGSGPASVQEWWRSGFI
Cp.Cp_0560                   TDFFLAPAIPSVTARAMFLTAMAGNPLYSHYLFASNTAHIGAMLTLAFASNLFGGLTHYGSGPATVQEWWRSGFA
Bl.YflS/BL05068              SDLILSPAIPSNTARAMFLTAMAANPLYSHYFFASATAHISAMLSLAFFSNLFASTTHYGAGAAPQSKWWSIGFI
Sb.OMT1/SORBI_3008G112300    AEALLAPAIPSVSARAMFLTAMAANPLYSHYFFASGAAHIGAMMVLSFLSNLMGGTTHYGIGSAPLAQWWGYGFV
Zm.zmpOMT1                   AEALLAPAIPSVSARAMFLTAMAANPLYSHYFFASGAAHIGAMMVLSFLSNLMGGTTHYGIGSAPLAQWWGYGFV
Ps.OMT/KIW84_075719          SEALLAPAIPSVSARAMFLTAMAANPLYSHYFFASGAAHIGAMMVLSFLSNLMGGLTHYGIGSAPLAKWWGYGFV
At.DIT1/DIT1                 SEALLAPAIPSVSARAMFLTAMAANPLYTHYFFASGAAHIGAMLVLAFLSNLMGGLTHYGIGSAPLAKWWGYGFL
SoDIT1/DIT1                  SEALLAPAIPSVSARAMFLTAMAANPLYSHYFFASGAAHIGAMIVLSFLSNLMGGLTHYGIGSAPLPQWWGYGFL
Sc.Pho91                     VALFVSMWV-SNVAAPSSPIASPQNIFM--ATFVSHTVAAMIIVIAALLCS-SAMGLPTSGFPNTVGTFITRGVP
Sc.Pho87                     VVFFLSMWI-SNVASPASPISSPQNIIV--GTFVSHTVSAIIIFGCALLAS-CGMGLASSGFPNTVGAFISRGVP
Sc.Pho90                     VVFFLSMWI-SNVAAPSSPISSPQNIIV--GTFVSHTVSAIIIFGCALLSS-CGMGLASSGFPNSVMTFLTRGVP
Dm.INDY1/Indy                VTMFLSMWI-SNAACTGTIIGTATNLTL--TAFSSNVAIANIILPAGLACS-MAFHLPVSTPPNRTKDMAIAGIG
Dm.INDY2/Indy2               VSTFMGLWI-SNSAGTGTLIGTGTNLVF--SAFNANVVANIVLPACLGIS-MVYFLPVSTPPNKTKYFACCGIV
Ce.NaDC2/nac-2/nat-1/indy-3  ITSFISFFV-SDTACTAIITSTGPNLVM--TEFASNVSTGSIFLPTTVACS-FAFMLPISTPPNSMVEMIVCGFL
Ce.NaDC1/nac-1/nad-1/indy-1  VTGFLSMWI-SNTATTATITGTASNLVL--TNVCSNTVIASIFLPVTISAS-FAFLLPVATPPNKVFDMFVSGLC
Ce.NaDC3/nac-3/nad-2/indy-2  VTALLSSFI-SNTATTGTATGTPSNLVI--TNICSNTVTASIFLPTTLASS-FAFIFPVGTPPNKVSDMAFVGGI
Hs.NaS1/SLC13A1              STAFLSMWL-SNTSTATTITGTSTNLIL--TEVASNPATITLFIPSTLCTS-FAFLLPVANPPNKVIDMVKAGLG
Hs.NaS2/SLC13A4              CTTLLSMWL-SNTSTTTTIIGTSTSLIV--TEFVSNPATITIFIPVTMCIS-FAVMLPVGNPPNQIKDMVKAGLG
Hs.NaDC3/SLC13A3             TTSFLSMWL-SNTASTATLTGTAPNLIF--TEFASNTATIIIFIPGTVGCS-FAFMLPVSTPPNLVKDMVRTGLL
Hs.NaDC1/SLC13A2             VTAFLSMWI-SNTATSATLTGTAPNLVF--TECTSNVATTTIFLPCTLATS-LAFMLPVATPPNKVLDMARAGFL
Hs.NaCT/SLC13A5              VTALLSMWI-SNTATTATLTGTGPNVVF--TECTSNVATTTLFLPCTLSAS-FAFMLPVATPPNKVADMVKTGVI
As.asuc_0304                 ITGVMSGVL-SNTGTALSLIGSPNNLVL--TNFMSNTATAALLMAVVVGSS-CAFATPIATPANRFADYAKAGVP
As.asuc_1568                 ATAFLSMWI-SNTATAGTLVGSPPNAIL--TEFTSNTASAALLLIIGLGAS-CAFMLPVATPPNKQREMVRAGVI
Vc.INDY/VC_A0025             VTALLSMWI-SNTATAATLVGSPPNAIL--TEFASNTASAALLVLIAVAAS-CAFMLPVATPPNKQSEMMRVGLY
Gg.DccT/cgl0225              ATGFLSMWV-SNTATAGTLIGTPPNALL--TEFTSNTATAATFIPVALSAT-CAFMLPVATPPNKIGEMVKGGLW
Bl.SdcF/BL02343              ATAFLSMWI-SNAATAATLVGSVPNAVM--TEIMSNTAVANMVAAAALASS-CAFMLPISTPPNTIKDMVRAGFW
Bl.SdcL/BL01772              ATGFLSMWV-SNTAAVGTLIGTPPNIIL--TEITSNTATATMIVPAAMAAN-CAFMLPVGTPPNKISEMVRTGFV
Sa.SdcS/SAV1916              ATGFLSMFV-SNTAAVGTLIGTPPLIIL--TEVTSNTATATMIAPAAMAAN-CAYMLPVGTPPNSIKQMASVGFW
```

Figure 6.1 Sequence alignment of the SNT motifs of all discussed DASS members in the text.

The sequences are numbered with respect to the residue positions in LaINDY. Gaps in the sequence are highlighted with blue arrows and vertical lines. The SNT1 and SNT2 motifs are highlighted in light green. ***For the DASS-Es:*** *the SNT1 Arg159-Glu146, and SNT2 His392-His401 pairings are coloured blue, and silent mutations (Glu146Asp or His392Arg) are coloured light blue. The proline insert in SNT1 is coloured yellow. The residues of the "aromatic cage" are coloured orange, and silent mutations preserving the aromatic rings are coloured light orange.* ***For the DASS-Cs:*** *Residues coordinating the sodium ions in both SNT1 Na1 and SNT2 Na2 are coloured dark green, and taken from structures. The ScPho90 residues coordinating sodium in Na3 are coloured dark purple. The conserved glutamate residue in Na2 is coloured pink.*

6.2.1 Observations on the sequence conservation in the DASS transporters

6.2.1.1 The Canonical DASS-E Basic Residues and Their Bonding Partners

Arginine 159 is entirely conserved within the site equivalent to Na1 of the DASS-C of all known DASS-Es, and is predicted to recognise one of the substrates carboxylate moieties. Additionally, there is a salt-bridge between Arg159 and Glu148 that is mostly conserved, though in some instances the glutamate is replaced with an aspartate. Is this salt-bridge essential for substrate binding? Does an aspartate mutation alter transport kinetics? Notably, c5038 from uropathogenic *E. coli* is the only DASS-E member without this salt bridge, yet c5038 has been shown to transport α-ketoglutarate [43, 78]. Therefore, I would like to investigate the importance and role of this arginine and its interactions.

Similarly, the pairing between His392 and His401, which has been proposed as a proton donor to ensure the His392 is in a charged state for substrate recognition, is also highly conserved. Can LaINDY bind substrate in the absence of His401? Given the pK_a of histidine of 6.04 I would anticipate a His401 mutation having more of an impact that that of the previously mentioned Arg-Glu pairing. The *E. coli* transporters TdtT, and YbhI, give credence to this theory, as they do not conserve His392 and have an arginine at the position equivalent to His392. This arginine, which has a much higher pK_a and as such will be positively charged at most pHs. Further, *A. succinogenes* asuc1482 also contains an arginine at position His392, while retaining the histidine at position 401, hinting at further complexity. All three of these transporters would be interesting targets to work on, to probe how variation in the DASS-E SNT motif impacts the structure and function of the transporters.

6.2.1.2 Conserved Glutamate Within The DASS-C SNT2 Na2 Binding Site

There is a glutamate within the Na2 site, positioned just before the SNT2 motif (position 394 in LaINDY), that is highly conserved amongst DASS-C that have had sodium dependent transport verified [13, 19, 37-40, 86, 211]. The exceptions are ScPho90, which has a poorly defined Na2 site and the novel Na3 site, which may indicate different subtle differences in the mechanism of sodium binding [25], and asuc_0304. In the Sauer *et al.* 2022 paper in which a 3.23Å apo structure of VcINDY was solved in choline chloride (PDB 79TF), they reported significant structural changes around the SNT2 motif but made no mention of the glutamate residue [21]. Upon comparison between the apo and sodium bound structures the glutamate rotamer does appear to move significantly. As a conserved acidic residue in close proximity to the binding site on a sodium ion, it seems reasonable that this glutamate could be involved in sodium sensing. Accordingly, I would like to see if the mutation of this glutamate retards VcINDY transport.

6.2.2 The Relevance of DASS-Es in the Pathology of Bacterial Infections

Uropathogenic *E. coli* express the unique DASS member c5038, which is an α-ketoglutarate transporter upregulated under the α-ketoglutarate sensing KguS/KguR two-component system and FNR and ArcA under anaerobiosis [43, 78]. As such the idea that Ecc5038 could be a novel therapeutic target for uropathogenic infections is compelling. However, in the Cai *et al.* 2017 study found that a c5038 knockout strain only had a two-fold reduced fitness in murine urinary tract infections compared to the control. I believe a reasonable explanation for this is transporter redundancy. UPEC *E. coli* express four DASS-Es, one of which is uncharacterized, and we have seen through VcINDY; and LaINDY that this family of proteins can interact with many small dicarboxylates. Additionally, searching UniProt for UPEC strain CTF073 for dicarboxylate transporter protein families: DcuA/B, DcuC/D, CitM, 2-HCT, MHS, TRAP, DAACS, and DASS reveals 19 transporter genes which may be able to compensate for loss of c5038 [212]. Whilst I am certain not all of these transporters will bind α-ketoglutarate or be expressed anaerobically, if any do then this may explain the modest effect seen through c5038 knockout.

In addition to the *E. coli* gene, I would like to propose the chlamydial DASS-E member as a potential therapeutic target which have yet to be considered. CTL0456 in *C. trachomatis* L2 and Cp_0560 for *C. pneumoniae* are the genes responsible for the only chlamydial DASS member, referred to in the literature as sodTi [213, 214]. As mentioned in **Section 1.8.4** work on chlamydia has lagged behind that of other pathogens due to their obligate intracellular two-phase lifecycle, requiring culturing of a host cell to study. All species of chlamydia have an incomplete cell cycle due to the loss of citrate synthase, aconitase and isocitrate dehydrogenase [215], and as such must rely on import of host dicarboxylates through their two known dicarboxylate transporters: DASS-E member sodTi and DAACS member gltT. Further, two recent mutation studies on *C. trachomatis* L2 have indicated that SoDTi may be essential for survival [92, 93]. The Kokes *et al.* study being of particular note because *C. trachomatis* L2 mutant lines that could complete both stages of the lifecycle were generated in which the gltT gene contained nonsense mutations, indicating gltT is non-essential, while sodTi may be. Finally, a similar search of UniProt for dicarboxylate transporters in both *C. trachomatis* and *C. pneumoniae* reveals only two dicaboxylate transporter genes: gltT and sodTi. whilst this is very early stage evidence, I believe the chlamydial DASS members to be an exciting avenue worth exploring.

7. References

1. Cannon, W.R., *Concepts, challenges, and successes in modeling thermodynamics of metabolism.* Front Bioeng Biotechnol, 2014. **2**: p. 53.
2. Mills, E.L., K.A. Pierce, M.P. Jedrychowski, R. Garrity, S. Winther, S. Vidoni, T. Yoneshiro, J.B. Spinelli, G.Z. Lu, L. Kazak, A.S. Banks, M.C. Haigis, S. Kajimura, M.P. Murphy, S.P. Gygi, C.B. Clish, and E.T. Chouchani, *Accumulation of succinate controls activation of adipose tissue thermogenesis.* Nature, 2018. **560**(7716): p. 102-106.
3. Huergo, L.F. and R. Dixon, *The Emergence of 2-Oxoglutarate as a Master Regulator Metabolite.* Microbiol Mol Biol Rev, 2015. **79**(4): p. 419-35.
4. Arnold, P.K. and L.W.S. Finley, *Regulation and function of the mammalian tricarboxylic acid cycle.* J Biol Chem, 2023. **299**(2): p. 102838.
5. Cho, Y.B., J.W. Park, G. Unden, and O.B. Kim, *Asuc_0142 of Actinobacillus succinogenes 130Z is the l-aspartate/C4-dicarboxylate exchanger DcuA.* Microbiology (Reading), 2023. **169**(10).
6. Liu, B., L. Jiang, Y. Liu, H. Sun, J. Yan, C. Kang, and B. Yang, *Enterohaemorrhagic E. coli utilizes host- and microbiota-derived L-malate as a signaling molecule for intestinal colonization.* Nat Commun, 2023. **14**(1): p. 7227.
7. Unden, G., A. Strecker, A. Kleefeld, and O.B. Kim, *C4-Dicarboxylate Utilization in Aerobic and Anaerobic Growth.* EcoSal Plus, 2016. **7**(1).
8. Prakash, S., G. Cooper, S. Singhi, and M.H. Saier, Jr., *The ion transporter superfamily.* Biochim Biophys Acta, 2003. **1618**(1): p. 79-92.
9. Rogina, B., R.A. Reenan, S.P. Nilsen, and S.L. Helfand, *Extended life-span conferred by cotransporter gene mutations in Drosophila.* Science, 2000. **290**(5499): p. 2137-40.
10. Taniguchi, M., Y. Taniguchi, M. Kawasaki, S. Takeda, T. Kato, S. Sato, S. Tabata, H. Miyake, and T. Sugiyama, *Identifying and characterizing plastidic 2-oxoglutarate/malate and dicarboxylate transporters in Arabidopsis thaliana.* Plant Cell Physiol, 2002. **43**(7): p. 706-17.
11. Fei, Y.J., J.C. Liu, K. Inoue, L. Zhuang, K. Miyake, S. Miyauchi, and V. Ganapathy, *Relevance of NAC-2, an Na+-coupled citrate transporter, to life span, body size and fat content in Caenorhabditis elegans.* Biochem J, 2004. **379**(Pt 1): p. 191-8.
12. Inoue, K., L. Zhuang, D.M. Maddox, S.B. Smith, and V. Ganapathy, *Structure, function, and expression pattern of a novel sodium-coupled citrate transporter (NaCT) cloned from mammalian brain.* J Biol Chem, 2002. **277**(42): p. 39469-76.
13. Pajor, A.M., *Molecular properties of the SLC13 family of dicarboxylate and sulfate transporters.* Pflugers Arch, 2006. **451**(5): p. 597-605.
14. Sauer, D.B., N. Trebesch, J.J. Marden, N. Cocco, J. Song, A. Koide, S. Koide, E. Tajkhorshid, and D.N. Wang, *Structural basis for the reaction cycle of DASS dicarboxylate transporters.* Elife, 2020. **9**.
15. Knauf, F., N. Mohebbi, C. Teichert, D. Herold, B. Rogina, S. Helfand, M. Gollasch, F.C. Luft, and P.S. Aronson, *The life-extending gene Indy encodes an exchanger for Krebs-cycle intermediates.* Biochem J, 2006. **397**(1): p. 25-9.
16. Brachs, S., A.F. Winkel, H. Tang, A.L. Birkenfeld, B. Brunner, K. Jahn-Hofmann, D. Margerie, H. Ruetten, D. Schmoll, and J. Spranger, *Inhibition of citrate cotransporter Slc13a5/mINDY by RNAi improves hepatic insulin sensitivity and prevents diet-induced non-alcoholic fatty liver disease in mice.* Mol Metab, 2016. **5**(11): p. 1072-1082.

17. Thevenon, J., M. Milh, F. Feillet, J. St-Onge, Y. Duffourd, C. Juge, A. Roubertie, D. Heron, C. Mignot, E. Raffo, B. Isidor, S. Wahlen, D. Sanlaville, N. Villeneuve, V. Darmency-Stamboul, A. Toutain, M. Lefebvre, M. Chouchane, F. Huet, A. Lafon, A. de Saint Martin, G. Lesca, S. El Chehadeh, C. Thauvin-Robinet, A. Masurel-Paulet, S. Odent, L. Villard, C. Philippe, L. Faivre, and J.B. Riviere, *Mutations in SLC13A5 cause autosomal-recessive epileptic encephalopathy with seizure onset in the first days of life.* Am J Hum Genet, 2014. **95**(1): p. 113-20.
18. Sauer, D.B., J. Song, B. Wang, J.K. Hilton, N.K. Karpowich, J.A. Mindell, W.J. Rice, and D.N. Wang, *Structure and inhibition mechanism of the human citrate transporter NaCT.* Nature, 2021. **591**(7848): p. 157-161.
19. Mancusso, R., G.G. Gregorio, Q. Liu, and D.N. Wang, *Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter.* Nature, 2012. **491**(7425): p. 622-6.
20. Nie, R., S. Stark, J. Symersky, R.S. Kaplan, and M. Lu, *Structure and function of the divalent anion/Na(+) symporter from Vibrio cholerae and a humanized variant.* Nat Commun, 2017. **8**: p. 15009.
21. Sauer, D.B., J.J. Marden, J.C. Sudar, J. Song, C. Mulligan, and D.N. Wang, *Structural basis of ion - substrate coupling in the Na(+)-dependent dicarboxylate transporter VcINDY.* Nat Commun, 2022. **13**(1): p. 2644.
22. Crisman, T.J., S.G. Qu, B.I. Kanner, and L.R. Forrest, *Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(49): p. 20752-20757.
23. Reyes, N., C. Ginter, and O. Boudker, *Transport mechanism of a bacterial homologue of glutamate transporters.* Nature, 2009. **462**(7275): p. 880-885.
24. Mulligan, C., C. Fenollar-Ferrer, G.A. Fitzgerald, A. Vergara-Jaque, D. Kaufmann, Y. Li, L.R. Forrest, and J.A. Mindell, *The bacterial dicarboxylate transporter VcINDY uses a two-domain elevator-type mechanism.* Nat Struct Mol Biol, 2016. **23**(3): p. 256-63.
25. Schneider, S., W. Kühlbrandt, and Ö. Yildiz, *Complementary structures of the yeast phosphate transporter Pho90 provide insights into its transport mechanism.* Structure, 2024. **32**(7).
26. Huard, K., J. Brown, J.C. Jones, S. Cabral, K. Futatsugi, M. Gorgoglione, A. Lanba, N.B. Vera, Y.M. Zhu, Q.Y. Yan, Y.J. Zhou, C. Vernochet, K. Riccardi, A. Wolford, D. Pirman, M. Niosi, G. Aspnes, M. Herr, N.E. Genung, T.V. Magee, D.P. Uccello, P. Loria, L. Di, J.R. Gosset, D. Hepworth, T. Rolph, J.A. Pfefferkorn, and D.M. Erion, *Discovery and characterization of novel inhibitors of the sodium-coupled citrate transporter (NaCT or SLC13A5).* Scientific Reports, 2015. **5**.
27. Huard, K., J.R. Gosset, J.I. Montgomery, A. Gilbert, M.M. Hayward, T.V. Magee, S. Cabral, D.P. Uccello, K. Bahnck, J. Brown, J. Purkal, M. Gorgoglione, A. Lanba, K. Futatsugi, M. Herr, N.E. Genung, G. Aspnes, J. Polivkova, C.N. Garcia-Irizarry, Q. Li, D. Canterbury, M. Niosi, N.B. Vera, Z. Li, B. Khunte, J. Siderewicz, T. Rolph, and D.M. Erion, *Optimization of a Dicarboxylic Series for in Vivo Inhibition of Citrate Transport by the Solute Carrier 13 (SLC13) Family.* J Med Chem, 2016. **59**(3): p. 1165-75.
28. Pajor, A.M., C.A. de Oliveira, K. Song, K. Huard, V. Shanmugasundaram, and D.M. Erion, *Molecular Basis for Inhibition of the Na+/Citrate Transporter NaCT (SLC13A5) by Dicarboxylate Inhibitors.* Mol Pharmacol, 2016. **90**(6): p. 755-765.

29. Rives, M.L., M. Shaw, B. Zhu, S.A. Hinke, and A.D. Wickenden, *State-Dependent Allosteric Inhibition of the Human SLC13A5 Citrate Transporter by Hydroxysuccinic Acids, PF-06649298 and PF-06761281.* Mol Pharmacol, 2016. **90**(6): p. 766-774.
30. Chi, X.M., Y.M. Chen, Y.N. Li, L. Dai, Y.Y. Zhang, Y.P. Shen, Y. Chen, T.H. Shi, H.N. Yang, Z.L. Wang, and R.H. Yan, *Cryo-EM structures of the human NaS1 and NaDC1 transporters revealed the elevator transport and allosteric regulation mechanism.* Science Advances, 2024. **10**(13).
31. Pajor, A.M. and K.M. Randolph, *Inhibition of the Na+/dicarboxylate cotransporter by anthranilic acid derivatives.* Mol Pharmacol, 2007. **72**(5): p. 1330-6.
32. Kühlbrandt, W., *The Resolution Revolution.* Science, 2014. **343**(6178): p. 1443-1444.
33. Sun, M., C.M. Azumaya, E. Tse, D.P. Bulkley, M.B. Harrington, G. Gilbert, A. Frost, D. Southworth, K.A. Verba, Y. Cheng, and D.A. Agard, *Practical considerations for using K3 cameras in CDS mode for high-resolution and high-throughput single particle cryo-EM.* J Struct Biol, 2021. **213**(3): p. 107745.
34. Bepler, T., A. Morin, M. Rapp, J. Brasch, L. Shapiro, A.J. Noble, and B. Berger, *Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs.* Nat Methods, 2019. **16**(11): p. 1153-1160.
35. Zivanov, J., T. Nakane, and S.H.W. Scheres, *Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1.* IUCrJ, 2020. **7**(Pt 2): p. 253-267.
36. Kimanius, D., L. Dong, G. Sharov, T. Nakane, and S.H.W. Scheres, *New tools for automated cryo-EM single-particle analysis in RELION-4.0.* Biochem J, 2021. **478**(24): p. 4169-4185.
37. Strickler, M.A., J.A. Hall, O. Gaiko, and A.M. Pajor, *Functional characterization of a Na(+)-coupled dicarboxylate transporter from Bacillus licheniformis.* Biochim Biophys Acta, 2009. **1788**(12): p. 2489-96.
38. Pajor, A.M., N.N. Sun, and A. Leung, *Functional characterization of SdcF from Bacillus licheniformis, a homolog of the SLC13 Na(+)/dicarboxylate transporters.* J Membr Biol, 2013. **246**(9): p. 705-15.
39. Hall, J.A. and A.M. Pajor, *Functional characterization of a Na(+)-coupled dicarboxylate carrier protein from Staphylococcus aureus.* J Bacteriol, 2005. **187**(15): p. 5189-94.
40. Youn, J.W., E. Jolkver, R. Kramer, K. Marin, and V.F. Wendisch, *Identification and characterization of the dicarboxylate uptake system DccT in Corynebacterium glutamicum.* J Bacteriol, 2008. **190**(19): p. 6458-66.
41. Takahashi, S., M. Miyachi, H. Tamaki, and H. Suzuki, *The CitT transporter can be used as a succinate exporter for succinate production.* Bioscience Biotechnology and Biochemistry, 2021. **85**(4): p. 981-988.
42. Zamani-Nour, S., H.C. Lin, B.J. Walker, T. Mettler-Altmann, R. Khoshravesh, S. Karki, E. Bagunu, T.L. Sage, W.P. Quick, and A.P.M. Weber, *Overexpression of the chloroplastic 2-oxoglutarate/malate transporter disturbs carbon and nitrogen homeostasis in rice.* Journal of Experimental Botany, 2021. **72**(1): p. 137-152.
43. Cai, W., X. Cai, Y. Yang, S. Yan, and H. Zhang, *Transcriptional Control of Dual Transporters Involved in alpha-Ketoglutarate Utilization Reveals Their Distinct Roles in Uropathogenic Escherichia coli.* Front Microbiol, 2017. **8**: p. 275.
44. Heber, U., U.W. Hallier, and M.A. Hudson, *[Studies on the intracellular distribution of enzymes and substrates in leaf cells. II. Localization of enzymes of the reductive and*

oxidative pentosephosphate cycle in chloroplasts and permeability of the chloroplast membrane to metabolites]. Z Naturforsch B, 1967. **22**(11): p. 1200-15.

45. Heber, U., K.A. Santarius, M.A. Hudson, and U.W. Hallier, *[Studies on the intracellular distribution of enzymes and substrates in leaf cells. I. Intracellular transport of photosynthesis intermediates in steady-state photosynthesis and in the dark-light-dark cycle].* Z Naturforsch B, 1967. **22**(11): p. 1189-99.
46. Gimmler, H., G. Schafer, H. Kraminer, and U. Heber, *Amino acid permeability of the chloroplast envelope as measured by light scattering, volumetry and amino acid uptake.* Planta, 1974. **120**(1): p. 47-61.
47. Lehner, K. and H.W. Heldt, *Dicarboxylate transport across the inner membrane of the chloroplast envelope.* Biochim Biophys Acta, 1978. **501**(3): p. 531-44.
48. Woo, K.C., U.I. Flugge, and H.W. Heldt, *A Two-Translocator Model for the Transport of 2-Oxoglutarate and Glutamate in Chloroplasts during Ammonia Assimilation in the Light.* Plant Physiol, 1987. **84**(3): p. 624-32.
49. Flugge, I.U., K.C. Woo, and H.W. Heldt, *Characteristics of 2-oxoglutarate and glutamate transport in spinach chloroplasts : Studies with a double-silicone-layer centrifugation technique and in liposomes.* Planta, 1988. **174**(4): p. 534-41.
50. Lancien, M., M. Martin, M.H. Hsieh, T. Leustek, H. Goodman, and G.M. Coruzzi, *Arabidopsis glt1-T mutant defines a role for NADH-GOGAT in the non-photorespiratory ammonium assimilatory pathway.* Plant J, 2002. **29**(3): p. 347-58.
51. Linka, M. and A.P.M. Weber, *Shuffling ammonia between mitochondria and plastids during photorespiration.* Trends in Plant Science, 2005. **10**(10): p. 461-465.
52. Facchinelli, F. and A.P. Weber, *The metabolite transporters of the plastid envelope: an update.* Front Plant Sci, 2011. **2**: p. 50.
53. Somerville, S.C. and W.L. Ogren, *An Arabidopsis thaliana mutant defective in chloroplast dicarboxylate transport.* Proc Natl Acad Sci U S A, 1983. **80**(5): p. 1290-4.
54. Somerville, S.C. and C.R. Somerville, *A Mutant of Arabidopsis Deficient in Chloroplast Dicarboxylate Transport Is Missing an Envelope Protein.* Plant Science Letters, 1985. **37**(3): p. 217-220.
55. Arabidopsis Genome, I., *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.* Nature, 2000. **408**(6814): p. 796-815.
56. Renne, P., U. Dressen, U. Hebbeker, D. Hille, U.I. Flugge, P. Westhoff, and A.P. Weber, *The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2.* Plant J, 2003. **35**(3): p. 316-31.
57. Werner-Washburne, M. and K. Keegstra, *l-Aspartate Transport into Pea Chloroplasts : Kinetic and Inhibitor Evidence for Multiple Transport Systems.* Plant Physiol, 1985. **78**(2): p. 221-7.
58. Hatch, M.D., L. Droscher, U.I. Flugge, and H.W. Heldt, *A Specific Translocator for Oxaloacetate Transport in Chloroplasts.* Febs Letters, 1984. **178**(1): p. 15-19.
59. Kinoshita, H., J. Nagasaki, N. Yoshikawa, A. Yamamoto, S. Takito, M. Kawasaki, T. Sugiyama, H. Miyake, A.P.M. Weber, and M. Taniguchi, *The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism.* Plant Journal, 2011. **65**(1): p. 15-26.
60. Sage, R.F., T.L. Sage, and F. Kocacinar, *Photorespiration and the evolution of C4 photosynthesis.* Annu Rev Plant Biol, 2012. **63**: p. 19-47.

61. Hatch, M.D., *C4-Pathway photosynthesis in Portulaca oleracea and the significance of alanine labelling.* Planta, 1975. **125**(3): p. 273-9.
62. Taniguchi, Y., J. Nagasaki, M. Kawasaki, H. Miyake, T. Sugiyama, and M. Taniguchi, *Differentiation of dicarboxylate transporters in mesophyll and bundle sheath chloroplasts of maize.* Plant Cell Physiol, 2004. **45**(2): p. 187-200.
63. Weissmann, S., P. Huang, M.A. Wiechert, K. Furuyama, T.P. Brutnell, M. Taniguchi, J.C. Schnable, and T.C. Mockler, *DCT4-A New Member of the Dicarboxylate Transporter Family in C4 Grasses.* Genome Biol Evol, 2021. **13**(2).
64. Weissmann, S., F. Ma, K. Furuyama, J. Gierse, H. Berg, Y. Shao, M. Taniguchi, D.K. Allen, and T.P. Brutnell, *Interactions of C4 Subtype Metabolic Activities and Transport in Maize Are Revealed through the Characterization of DCT2 Mutants.* Plant Cell, 2016. **28**(2): p. 466-84.
65. Tyra, H.M., M. Linka, A.P. Weber, and D. Bhattacharya, *Host origin of plastid solute transporters in the first photosynthetic eukaryotes.* Genome Biol, 2007. **8**(10): p. R212.
66. Stelzner, K., N. Vollmuth, and T. Rudel, *Intracellular lifestyle of Chlamydia trachomatis and host-pathogen interactions.* Nat Rev Microbiol, 2023. **21**(7): p. 448-462.
67. Blattner, F.R., G. Plunkett, 3rd, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, and Y. Shao, *The complete genome sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-62.
68. Pos, K.M., P. Dimroth, and M. Bott, *The citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts.* Journal of Bacteriology, 1998. **180**(16): p. 4160-4165.
69. Bott, M., M. Meyer, and P. Dimroth, *Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae.* Mol Microbiol, 1995. **18**(3): p. 533-46.
70. Weber, A., E. Menzlaff, B. Arbinger, M. Gutensohn, C. Eckerskorn, and U.I. Flugge, *The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells.* Biochemistry, 1995. **34**(8): p. 2621-7.
71. Blount, Z.D., J.E. Barrick, C.J. Davidson, and R.E. Lenski, *Genomic analysis of a key innovation in an experimental Escherichia coli population.* Nature, 2012. **489**(7417): p. 513-8.
72. Van Hofwegen, D.J., C.J. Hovde, and S.A. Minnich, *Rapid Evolution of Citrate Utilization by Escherichia coli by Direct Selection Requires citT and dctA.* J Bacteriol, 2016. **198**(7): p. 1022-34.
73. Quandt, E.M., D.E. Deatherage, A.D. Ellington, G. Georgiou, and J.E. Barrick, *Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in Escherichia coli.* Proc Natl Acad Sci U S A, 2014. **111**(6): p. 2217-22.
74. Kim, O.B. and G. Unden, *The L-tartrate/succinate antiporter TtdT (YgjE) of L-tartrate fermentation in.* Journal of Bacteriology, 2007. **189**(5): p. 1597-1603.
75. Kim, O.B., S. Lux, and G. Unden, *Anaerobic growth of Escherichia coli on D-tartrate depends on the fumarate carrier DcuB and fumarase, rather than the L-tartrate carrier TtdT and L-tartrate dehydratase.* Arch Microbiol, 2007. **188**(6): p. 583-9.
76. Rudd, K.E., *Linkage map of Escherichia coli K-12, edition 10: The physical map.* Microbiology and Molecular Biology Reviews, 1998. **62**(3): p. 985-+.

77. Blank, L., J. Green, and J.R. Guest, *AcnC of Escherichia coli is a 2-methylcitrate dehydratase (PrpD) that can use citrate and isocitrate as substrates.* Microbiology-Sgm, 2002. **148**: p. 133-146.
78. Cai, W., Y. Wannemuehler, G. Dell'anna, B. Nicholson, N.L. Barbieri, S. Kariyawasam, Y. Feng, C.M. Logue, L.K. Nolan, and G. Li, *A novel two-component signaling system facilitates uropathogenic Escherichia coli's ability to exploit abundant host metabolites.* PLoS Pathog, 2013. **9**(6): p. e1003428.
79. Kalinowski, J., B. Bathe, D. Bartels, N. Bischoff, M. Bott, A. Burkovski, N. Dusch, L. Eggeling, B.J. Eikmanns, L. Gaigalat, A. Goesmann, M. Hartmann, K. Huthmacher, R. Kramer, B. Linke, A.C. McHardy, F. Meyer, B. Mockel, W. Pfefferle, A. Puhler, D.A. Rey, C. Ruckert, O. Rupp, H. Sahm, V.F. Wendisch, I. Wiegrabe, and A. Tauch, *The complete Corynebacterium glutamicum ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins.* J Biotechnol, 2003. **104**(1-3): p. 5-25.
80. Kinoshita, S., S. Udaka, and M. Shimono, *Studies on the amino acid fermentation. Part 1. Production of L-glutamic acid by various microorganisms.* J Gen Appl Microbiol, 2004. **50**(6): p. 331-43.
81. Guettler, M.V., D. Rumler, and M.K. Jain, *Actinobacillus succinogenes sp. nov., a novel succinic-acid-producing strain from the bovine rumen.* Int J Syst Bacteriol, 1999. **49 Pt 1**: p. 207-16.
82. McKinlay, J.B., M. Laivenieks, B.D. Schindler, A.A. McKinlay, S. Siddaramappa, J.F. Challacombe, S.R. Lowry, A. Clum, A.L. Lapidus, K.B. Burkhart, V. Harkins, and C. Vieille, *A genomic perspective on the potential of Actinobacillus succinogenes for industrial succinate production.* BMC Genomics, 2010. **11**: p. 680.
83. Udaka, S., *Screening method for microorganisms accumulating metabolites and its use in the isolation of Micrococcus glutamicus.* J Bacteriol, 1960. **79**(5): p. 754-5.
84. Van der Werf, M.J., M.V. Guettler, M.K. Jain, and J.G. Zeikus, *Environmental and physiological factors affecting the succinate product ratio during carbohydrate fermentation by Actinobacillus sp. 130Z.* Arch Microbiol, 1997. **167**(6): p. 332-42.
85. Rhie, M.N., B. Park, H.J. Ko, I.G. Choi, and O.B. Kim, *Transcriptome analysis and anaerobic C(4) -dicarboxylate transport in Actinobacillus succinogenes.* Microbiologyopen, 2018. **7**(3): p. e00565.
86. Rhie, M.N., H.E. Yoon, H.Y. Oh, S. Zedler, G. Unden, and O.B. Kim, *A Na+-coupled C4-dicarboxylate transporter (Asuc_0304) and aerobic growth of Actinobacillus succinogenes on C4-dicarboxylates.* Microbiology (Reading), 2014. **160**(Pt 7): p. 1533-1544.
87. Burton, M.J., *Trachoma: an overview.* Br Med Bull, 2007. **84**: p. 99-116.
88. Mabey, D. and R.W. Peeling, *Lymphogranuloma venereum.* Sex Transm Infect, 2002. **78**(2): p. 90-2.
89. Mabey, D. and N. Fraser-Hurt, *Trachoma.* BMJ, 2001. **323**(7306): p. 218-21.
90. Bastidas, R.J. and R.H. Valdivia, *Emancipating Chlamydia: Advances in the Genetic Manipulation of a Recalcitrant Intracellular Pathogen.* Microbiol Mol Biol Rev, 2016. **80**(2): p. 411-27.
91. Binet, R. and A.T. Maurelli, *Transformation and isolation of allelic exchange mutants of Chlamydia psittaci using recombinant DNA introduced by electroporation.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 292-7.

92. Brothwell, J.A., M.K. Muramatsu, E. Toh, D.D. Rockey, T.E. Putman, M.L. Barta, P.S. Hefty, R.J. Suchland, and D.E. Nelson, *Interrogating Genes That Mediate Chlamydia trachomatis Survival in Cell Culture Using Conditional Mutants and Recombination.* J Bacteriol, 2016. **198**(15): p. 2131-9.
93. Kokes, M., J.D. Dunn, J.A. Granek, B.D. Nguyen, J.R. Barker, R.H. Valdivia, and R.J. Bastidas, *Integrating chemical mutagenesis and whole-genome sequencing as a platform for forward and reverse genetic analysis of Chlamydia.* Cell Host Microbe, 2015. **17**(5): p. 716-25.
94. Gao, H., X. Li, X. Chen, D. Hai, C. Wei, L. Zhang, and P. Li, *The Functional Roles of Lactobacillus acidophilus in Different Physiological and Pathological Processes.* J Microbiol Biotechnol, 2022. **32**(10): p. 1-8.
95. Kim, H.S. and S.E. Gilliland, *Lactobacillus acidophilus as a dietary adjunct for milk to aid lactose digestion in humans.* J Dairy Sci, 1983. **66**(5): p. 959-66.
96. Liong, M.T. and N.P. Shah, *Acid and bile tolerance and cholesterol removal ability of lactobacilli strains.* J Dairy Sci, 2005. **88**(1): p. 55-66.
97. Lye, H.S., G. Rusul, and M.T. Liong, *Removal of cholesterol by lactobacilli via incorporation and conversion to coprostanol.* J Dairy Sci, 2010. **93**(4): p. 1383-92.
98. Goldin, B.R. and S.L. Gorbach, *Effect of Lactobacillus acidophilus dietary supplements on 1,2-dimethylhydrazine dihydrochloride-induced intestinal cancer in rats.* J Natl Cancer Inst, 1980. **64**(2): p. 263-5.
99. Sampson, C.D.D., C. Fabregas Bellavista, M.J. Stewart, and C. Mulligan, *Thermostability-based binding assays reveal complex interplay of cation, substrate and lipid binding in the bacterial DASS transporter, VcINDY.* Biochem J, 2021. **478**(21): p. 3847-3867.
100. Booth, I.R. and P. Blount, *The MscS and MscL families of mechanosensitive channels act as microbial emergency release valves.* J Bacteriol, 2012. **194**(18): p. 4802-9.
101. Chang, G., R.H. Spencer, A.T. Lee, M.T. Barclay, and D.C. Rees, *Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel.* Science, 1998. **282**(5397): p. 2220-6.
102. Moe, P.C., G. Levin, and P. Blount, *Correlating a protein structure with function of a bacterial mechanosensitive channel.* J Biol Chem, 2000. **275**(40): p. 31121-7.
103. Zhong, D.L. and P. Blount, *Phosphatidylinositol Is Crucial for the Mechanosensitivity of Mycobacterium tuberculosis MscL.* Biochemistry, 2013. **52**(32): p. 5415-5420.
104. Laganowsky, A., E. Reading, T.M. Allison, M.B. Ulmschneider, M.T. Degiacomi, A.J. Baldwin, and C.V. Robinson, *Membrane proteins bind lipids selectively to modulate their structure and function.* Nature, 2014. **510**(7503): p. 172-+.
105. Murata, K., K. Mitsuoka, T. Hirai, T. Walz, P. Agre, J.B. Heymann, A. Engel, and Y. Fujiyoshi, *Structural determinants of water permeation through aquaporin-1.* Nature, 2000. **407**(6804): p. 599-605.
106. Fujiyoshi, Y., K. Mitsuoka, B.L. de Groot, A. Philippsen, H. Grubmuller, P. Agre, and A. Engel, *Structure and function of water channels.* Curr Opin Struct Biol, 2002. **12**(4): p. 509-15.
107. Li, C. and W. Wang, *Urea transport mediated by aquaporin water channel proteins.* Subcell Biochem, 2014. **73**: p. 227-65.
108. Cymer, F. and D. Schneider, *A single glutamate residue controls the oligomerization, function, and stability of the aquaglyceroporin GlpF.* Biochemistry, 2010. **49**(2): p. 279-86.

109. Stroud, R.M., L.J. Miercke, J. O'Connell, S. Khademi, J.K. Lee, J. Remis, W. Harries, Y. Robles, and D. Akhavan, *Glycerol facilitator GlpF and the associated aquaporin family of channels.* Curr Opin Struct Biol, 2003. **13**(4): p. 424-31.
110. Klein, N., N. Hellmann, and D. Schneider, *Anionic Lipids Modulate the Activity of the Aquaglyceroporin GlpF.* Biophys J, 2015. **109**(4): p. 722-31.
111. Savage, D.F., P.F. Egea, Y. Robles-Colmenares, J.D. O'Connell, 3rd, and R.M. Stroud, *Architecture and selectivity in aquaporins: 2.5 a X-ray structure of aquaporin Z.* PLoS Biol, 2003. **1**(3): p. E72.
112. Tan, B.K., M. Bogdanov, J.S. Zhao, W. Dowhan, C.R.H. Raetz, and Z.Q. Guan, *Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(41): p. 16504-16509.
113. Andrade, S.L. and O. Einsle, *The Amt/Mep/Rh family of ammonium transport proteins.* Mol Membr Biol, 2007. **24**(5-6): p. 357-65.
114. Khademi, S., J. O'Connell, 3rd, J. Remis, Y. Robles-Colmenares, L.J. Miercke, and R.M. Stroud, *Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A.* Science, 2004. **305**(5690): p. 1587-94.
115. Mirandela, G.D., G. Tamburrino, P.A. Hoskisson, U. Zachariae, and A. Javelle, *The lipid environment determines the activity of the Escherichia coli ammonium transporter AmtB.* FASEB J, 2019. **33**(2): p. 1989-1999.
116. Cong, X., Y. Liu, W. Liu, X. Liang, and A. Laganowsky, *Allosteric modulation of protein-protein interactions by individual lipid binding events.* Nat Commun, 2017. **8**(1): p. 2203.
117. Gupta, K., J.A.C. Donlan, J.T.S. Hopper, P. Uzdavinys, M. Landreh, W.B. Struwe, D. Drew, A.J. Baldwin, P.J. Stansfeld, and C.V. Robinson, *The role of interfacial lipids in stabilizing membrane protein oligomers.* Nature, 2017. **541**(7637): p. 421-424.
118. Oot, R.A., P.M. Kane, E.A. Berry, and S. Wilkens, *Crystal structure of yeast V1-ATPase in the autoinhibited state.* EMBO J, 2016. **35**(15): p. 1694-706.
119. Linke, K. and F.M. Ho, *Water in Photosystem II: structural, functional and mechanistic considerations.* Biochim Biophys Acta, 2014. **1837**(1): p. 14-32.
120. Baskaran, A., M. Kaari, G. Venugopal, R. Manikkam, J. Joseph, and P.V. Bhaskar, *Anti freeze proteins (Afp): Properties, sources and applications - A review.* Int J Biol Macromol, 2021. **189**: p. 292-305.
121. Liou, Y.C., A. Tocilj, P.L. Davies, and Z. Jia, *Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein.* Nature, 2000. **406**(6793): p. 322-4.
122. Singh, S.K., A. Yamashita, and E. Gouaux, *Antidepressant binding site in a bacterial homologue of neurotransmitter transporters.* Nature, 2007. **448**(7156): p. 952-6.
123. Zhou, Z., J. Zhen, N.K. Karpowich, R.M. Goetz, C.J. Law, M.E. Reith, and D.N. Wang, *LeuT-desipramine structure reveals how antidepressants block neurotransmitter reuptake.* Science, 2007. **317**(5843): p. 1390-3.
124. Zhou, Z., J. Zhen, N.K. Karpowich, C.J. Law, M.E. Reith, and D.N. Wang, *Antidepressant specificity of serotonin transporter suggested by three LeuT-SSRI structures.* Nat Struct Mol Biol, 2009. **16**(6): p. 652-7.
125. Davies, J.S., M.J. Currie, R.C.J. Dobson, C.R. Horne, and R.A. North, *TRAPs: the 'elevator-with-an-operator' mechanism.* Trends Biochem Sci, 2024. **49**(2): p. 134-144.
126. Peter, M.F., J.A. Ruland, P. Depping, N. Schneberger, E. Severi, J. Moecking, K. Gatterdam, S. Tindall, A. Durand, V. Heinz, J.P. Siebrasse, P.A. Koenig, M. Geyer, C. Ziegler, U.

Kubitscheck, G.H. Thomas, and G. Hagelueken, *Structural and mechanistic analysis of a tripartite ATP-independent periplasmic TRAP transporter.* Nat Commun, 2022. **13**(1): p. 4471.

127. Muller, A., E. Severi, C. Mulligan, A.G. Watts, D.J. Kelly, K.S. Wilson, A.J. Wilkinson, and G.H. Thomas, *Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus influenzae.* J Biol Chem, 2006. **281**(31): p. 22212-22222.
128. Fischer, M., A.P. Hopkins, E. Severi, J. Hawkhead, D. Bawdon, A.G. Watts, R.E. Hubbard, and G.H. Thomas, *Tripartite ATP-independent Periplasmic (TRAP) Transporters Use an Arginine-mediated Selectivity Filter for High Affinity Substrate Binding.* J Biol Chem, 2015. **290**(45): p. 27113-27123.
129. Oesterhelt, D. and W. Stoeckenius, *Rhodopsin-like protein from the purple membrane of Halobacterium halobium.* Nat New Biol, 1971. **233**(39): p. 149-52.
130. Oesterhelt, D. and W. Stoeckenius, *Functions of a new photoreceptor membrane.* Proc Natl Acad Sci U S A, 1973. **70**(10): p. 2853-7.
131. Lewis, A., J. Spoonhower, R.A. Bogomolni, R.H. Lozier, and W. Stoeckenius, *Tunable laser resonance raman spectroscopy of bacteriorhodopsin.* Proc Natl Acad Sci U S A, 1974. **71**(11): p. 4462-6.
132. Henderson, R., J.M. Baldwin, T.A. Ceska, F. Zemlin, E. Beckmann, and K.H. Downing, *Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy.* J Mol Biol, 1990. **213**(4): p. 899-929.
133. Wickstrand, C., R. Dods, A. Royant, and R. Neutze, *Bacteriorhodopsin: Would the real structural intermediates please stand up?* Biochim Biophys Acta, 2015. **1850**(3): p. 536-53.
134. Taguchi, S., S. Niwa, H.A. Dao, Y. Tanaka, R. Takeda, S. Fukai, K. Hasegawa, and K. Takeda, *Detailed analysis of distorted retinal and its interaction with surrounding residues in the K intermediate of bacteriorhodopsin.* Commun Biol, 2023. **6**(1): p. 190.
135. Wickstrand, C., P. Nogly, E. Nango, S. Iwata, J. Standfuss, and R. Neutze, *Bacteriorhodopsin: Structural Insights Revealed Using X-Ray Lasers and Synchrotron Radiation.* Annu Rev Biochem, 2019. **88**: p. 59-83.
136. Weinert, T., P. Skopintsev, D. James, F. Dworkowski, E. Panepucci, D. Kekilli, A. Furrer, S. Brunle, S. Mous, D. Ozerov, P. Nogly, M. Wang, and J. Standfuss, *Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography.* Science, 2019. **365**(6448): p. 61-65.
137. Gat, Y. and M. Sheves, *A Mechanism for Controlling the Pk(a) of the Retinal Protonated Schiff-Base in Retinal Proteins - a Study with Model Compounds.* Journal of the American Chemical Society, 1993. **115**(9): p. 3772-3773.
138. Richter, H.T., R. Needleman, and J.K. Lanyi, *Perturbed interaction between residues 85 and 204 in Tyr-185->Phe and Asp-85->Glu bacteriorhodopsins.* Biophysical Journal, 1996. **71**(6): p. 3392-3398.
139. Nango, E., A. Royant, M. Kubo, T. Nakane, C. Wickstrand, T. Kimura, T. Tanaka, K. Tono, C. Song, R. Tanaka, T. Arima, A. Yamashita, J. Kobayashi, T. Hosaka, E. Mizohata, P. Nogly, M. Sugahara, D. Nam, T. Nomura, T. Shimamura, D. Im, T. Fujiwara, Y. Yamanaka, B. Jeon, T. Nishizawa, K. Oda, M. Fukuda, R. Andersson, P. Bath, R. Dods, J. Davidsson, S. Matsuoka, S. Kawatake, M. Murata, O. Nureki, S. Owada, T. Kameshima, T. Hatsui, Y. Joti, G. Schertler, M. Yabashi, A.N. Bondar, J. Standfuss, R. Neutze, and S. Iwata, *A three-*

dimensional movie of structural changes in bacteriorhodopsin. Science, 2016. **354**(6319): p. 1552-1557.

140. Freier, E., S. Wolf, and K. Gerwert, *Proton transfer via a transient linear water-molecule chain in a membrane protein.* Proc Natl Acad Sci U S A, 2011. **108**(28): p. 11435-9.

141. Agmon, N., *The Grotthuss Mechanism.* Chemical Physics Letters, 1995. **244**(5-6): p. 456-462.

142. Foster, J.W., *Escherichia coli acid resistance: tales of an amateur acidophile.* Nat Rev Microbiol, 2004. **2**(11): p. 898-907.

143. Ilgu, H., J.M. Jeckelmann, D. Kalbermatter, Z. Ucurum, T. Lemmin, and D. Fotiadis, *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 Biol, 2021. **19**(1): p. 179.

144. Kowalczyk, L., M. Ratera, A. Paladino, P. Bartoccioni, E. Errasti-Murugarren, E. Valencia, G. Portella, S. Bial, A. Zorzano, I. Fita, M. Orozco, X. Carpena, J.L. Vazquez-Ibar, and M. Palacin, *Molecular basis of substrate-induced permeation by an amino acid antiporter.* Proc Natl Acad Sci U S A, 2011. **108**(10): p. 3935-40.

145. Ilgu, H., J.M. Jeckelmann, V. Gapsys, Z. Ucurum, B.L. de Groot, and D. Fotiadis, *Insights into the molecular basis for substrate binding and specificity of the wild-type L-arginine/agmatine antiporter AdiC.* Proc Natl Acad Sci U S A, 2016. **113**(37): p. 10358-63.

146. Love, J., F. Mancia, L. Shapiro, M. Punta, B. Rost, M. Girvin, D.N. Wang, M. Zhou, J.F. Hunt, T. Szyperski, E. Gouaux, R. MacKinnon, A. McDermott, B. Honig, M. Inouye, G. Montelione, and W.A. Hendrickson, *The New York Consortium on Membrane Protein Structure (NYCOMPS): a high-throughput platform for structural genomics of integral membrane proteins.* J Struct Funct Genomics, 2010. **11**(3): p. 191-9.

147. Yanisch-Perron, C., J. Vieira, and J. Messing, *Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors.* Gene, 1985. **33**(1): p. 103-119.

148. Andersen, K.R., N.C. Leksa, and T.U. Schwartz, *Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification.* Proteins, 2013. **81**(11): p. 1857-61.

149. Savitsky, P., J. Bray, C.D. Cooper, B.D. Marsden, P. Mahajan, N.A. Burgess-Brown, and O. Gileadi, *High-throughput production of human proteins for crystallization: the SGC experience.* J Struct Biol, 2010. **172**(1): p. 3-13.

150. Zivanov, J., T. Nakane, and S.H.W. Scheres, *A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis.* IUCrJ, 2019. **6**(Pt 1): p. 5-17.

151. Asarnow, D., E. Palovcak, and Y. Cheng, *asarnow/pyem: UCSF pyem v0.5.* Zenodo, 2019.

152. Zhu, J., Q. Zhang, H. Zhang, Z. Shi, M. Hu, and C. Bao, *A minority of final stacks yields superior amplitude in single-particle cryo-EM.* Nat Commun, 2023. **14**(1): p. 7822.

153. Rosenthal, P.B. and R. Henderson, *Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy.* J Mol Biol, 2003. **333**(4): p. 721-45.

154. Punjani, A., H. Zhang, and D.J. Fleet, *Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction.* Nat Methods, 2020. **17**(12): p. 1214-1221.

155. Punjani, A., J.L. Rubinstein, D.J. Fleet, and M.A. Brubaker, *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nat Methods, 2017. **14**(3): p. 290-296.

156. Liebschner, D., P.V. Afonine, M.L. Baker, G. Bunkoczi, V.B. Chen, T.I. Croll, B. Hintze, L.W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams, and P.D. Adams, *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.* Acta Crystallogr D Struct Biol, 2019. **75**(Pt 10): p. 861-877.

157. Afonine, P.V., B.K. Poon, R.J. Read, O.V. Sobolev, T.C. Terwilliger, A. Urzhumtsev, and P.D. Adams, *Real-space refinement in PHENIX for cryo-EM and crystallography.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 6): p. 531-544.

158. Emsley, P., B. Lohkamp, W.G. Scott, and K. Cowtan, *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

159. Moriarty, N.W., R.W. Grosse-Kunstleve, and P.D. Adams, *electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1074-80.

160. Shub, L., W. Liu, G. Skiniotis, M.J. Keiser, and M.J. Robertson, *Metric Ion Classification (MIC): A deep learning tool for assigning ions and waters in cryo-EM and x-ray crystallography structures.* 2024.

161. Hughes, G.W., S.C.L. Hall, C.S. Laxton, P. Sridhar, A.H. Mahadi, C. Hatton, T.J. Piggot, P.J. Wotherspoon, A.C. Leney, D.G. Ward, M. Jamshad, V. Spana, I.T. Cadby, C. Harding, G.L. Isom, J.A. Bryant, R.J. Parr, Y. Yakub, M. Jeeves, D. Huber, I.R. Henderson, L.A. Clifton, A.L. Lovering, and T.J. Knowles, *Evidence for phospholipid export from the bacterial inner membrane by the Mla ABC transport system.* Nat Microbiol, 2019. **4**(10): p. 1692-1705.

162. Sievers, F. and D.G. Higgins, *Clustal Omega for making accurate alignments of many protein sequences.* Protein Sci, 2018. **27**(1): p. 135-145.

163. Madeira, F., N. Madhusoodanan, J.H. Lee, A. Eusebi, A. Niewielska, A.R.N. Tivey, R. Lopez, and S. Butcher, *The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024.* Nucleic Acids Research, 2024. **52**(W1): p. W521-W525.

164. Pei, J., M. Tang, and N.V. Grishin, *PROMALS3D web server for accurate multiple protein sequence and structure alignments.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W30-4.

165. Pei, J. and N.V. Grishin, *PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information.* Methods Mol Biol, 2014. **1079**: p. 263-71.

166. Waterhouse, A.M., J.B. Procter, D.M. Martin, M. Clamp, and G.J. Barton, *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-91.

167. Nouri, H.R., A. Karkhah, A. Varasteh, and M. Sankian, *Expression of a Chimeric Allergen with High Rare Codons Content in Codon Bias-Adjusted Escherichia coli: Escherichia coli BL21 (DE3)-Codon Plus RIL as an Efficient Host.* Curr Microbiol, 2016. **73**(1): p. 91-8.

168. Wulfing, C., J. Lombardero, and A. Pluckthun, *An Escherichia coli protein consisting of a domain homologous to FK506-binding proteins (FKBP) and a new metal binding motif.* J Biol Chem, 1994. **269**(4): p. 2895-901.

169. Gatzeva-Topalova, P.Z., A.P. May, and M.C. Sousa, *Structure and mechanism of ArnA: conformational change implies ordered dehydrogenase mechanism in key enzyme for polymyxin resistance.* Structure, 2005. **13**(6): p. 929-42.

170. Weininger, U., C. Haupt, K. Schweimer, W. Graubner, M. Kovermann, T. Bruser, C. Scholz, P. Schaarschmidt, G. Zoldak, F.X. Schmid, and J. Balbach, *NMR solution structure of SlyD from Escherichia coli: spatial separation of prolyl isomerase and chaperone function.* J Mol Biol, 2009. **387**(2): p. 295-305.

171. Kropinski, A.M., A. Mazzocco, T.E. Waddell, E. Lingohr, and R.P. Johnson, *Enumeration of bacteriophages by double agar overlay plaque assay.* Methods Mol Biol, 2009. **501**: p. 69-76.

172. Acs, N., M. Gambino, and L. Brondsted, *Bacteriophage Enumeration and Detection Methods.* Front Microbiol, 2020. **11**: p. 594868.

173. Kost, T.A., J.P. Condreay, and D.L. Jarvis, *Baculovirus as versatile vectors for protein expression in insect and mammalian cells.* Nature Biotechnology, 2005. **23**(5): p. 567-575.

174. Summers, M.D. and G.E. Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell-Culture Procedures.* Texas Agricultural Experiment Station Bulletin, 1987(1555): p. 1-56.

175. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.

176. Routledge, S.J., C.J. Hewitt, N. Bora, and R.M. Bill, *Antifoam addition to shake flask cultures of recombinant Pichia pastoris increases yield.* Microb Cell Fact, 2011. **10**: p. 17.

177. Zumstein, L., *Dialysis and ultrafiltration.* Curr Protoc Mol Biol, 2001. **Appendix 3**: p. Appendix 3C.

178. Fairhead, M., *Parallel rapid expression and purification of proteins for crystallography (PREPX): large scale 1 L cultures V.1* protocols.io, 2023.

179. Pflugrath, J.W., *Practical macromolecular cryocrystallography.* Acta Crystallogr F Struct Biol Commun, 2015. **71**(Pt 6): p. 622-42.

180. Jang, K., H.G. Kim, S.H.S. Hlaing, M. Kang, H.W. Choe, and Y.J. Kim, *A Short Review on Cryoprotectants for 3D Protein Structure Analysis.* Crystals, 2022. **12**(2).

181. Moraes, I., G. Evans, J. Sanchez-Weatherby, S. Newstead, and P.D. Stewart, *Membrane protein structure determination - the next generation.* Biochim Biophys Acta, 2014. **1838**(1 Pt A): p. 78-87.

182. Yamamoto, M., K. Hirata, K. Yamashita, K. Hasegawa, G. Ueno, H. Ago, and T. Kumasaka, *Protein microcrystallography using synchrotron radiation.* IUCrJ, 2017. **4**(Pt 5): p. 529-539.

183. Fromm, S.A., T.A. Bharat, A.J. Jakobi, W.J. Hagen, and C. Sachse, *Seeing tobacco mosaic virus through direct electron detectors.* J Struct Biol, 2015. **189**(2): p. 87-97.

184. Kato, K., N. Miyazaki, T. Hamaguchi, Y. Nakajima, F. Akita, K. Yonekura, and J.R. Shen, *High-resolution cryo-EM structure of photosystem II reveals damage from high-dose electron beams.* Commun Biol, 2021. **4**(1): p. 382.

185. Bromberg, R., Y. Guo, D. Borek, and Z. Otwinowski, *High-resolution cryo-EM reconstructions in the presence of substantial aberrations.* IUCrJ, 2020. **7**(Pt 3): p. 445-452.

186. DeRosier, D.J., *Correction of high-resolution data for curvature of the Ewald sphere.* Ultramicroscopy, 2000. **81**(2): p. 83-98.

187. Henderson, R., *The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic-Resolution Microscopy of Unstained Biological Molecules.* Quarterly Reviews of Biophysics, 1995. **28**(2): p. 171-193.
188. Lee, H.J., H.S. Lee, T. Youn, B. Byrne, and P.S. Chae, *Impact of novel detergents on membrane protein studies.* Chem, 2022. **8**(4): p. 980-1013.
189. Tribet, C., R. Audebert, and J.L. Popot, *Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(26): p. 15047-15050.
190. Jeffrey, G.A., *An Introduction to Hydrogen Bonding*. 1997: Oxford University Press.
191. Teale, F.W. and G. Weber, *Ultraviolet fluorescence of the aromatic amino acids.* Biochem J, 1957. **65**(3): p. 476-82.
192. Yammine, A., J. Gao, and A.H. Kwan, *Tryptophan Fluorescence Quenching Assays for Measuring Protein-ligand Binding Affinities: Principles and a Practical Guide.* Bio Protoc, 2019. **9**(11): p. e3253.
193. Amininan, A.K., D.; Wang, X.; Konijnenberg, A.; Dudkina, N.; Ilitchev, A., *Cryo-EM sample optimization using the VitroEase Buffer Screening Kit [White Paper].* FEI/Thermo Fisher 2021: p. 8.
194. Drulyte, I., R.M. Johnson, E.L. Hesketh, D.L. Hurdiss, C.A. Scarff, S.A. Porav, N.A. Ranson, S.P. Muench, and R.F. Thompson, *Approaches to altering particle distributions in cryo-electron microscopy sample preparation.* Acta Crystallographica Section D-Structural Biology, 2018. **74**: p. 560-571.
195. Vinothkumar, K.R. and R. Henderson, *Single particle electron cryomicroscopy: trends, issues and future perspective.* Quarterly Reviews of Biophysics, 2016. **49**: p. 1-25.
196. Huq, A., P.A. West, E.B. Small, M.I. Huq, and R.R. Colwell, *Influence of water temperature, salinity, and pH on survival and growth of toxigenic Vibrio cholerae serovar 01 associated with live copepods in laboratory microcosms.* Appl Environ Microbiol, 1984. **48**(2): p. 420-4.
197. Hansen, M.L.R.W., A. Clausen, C.S. Ejsing, and J. Risbo, *Modulation of the Lactobacillus acidophilus La-5 lipidome by different growth conditions.* Microbiology-Sgm, 2015. **161**: p. 1990-1998.
198. Shaw, N., K. Heatherington, and J. Baddiley, *The glycolipids of Lactobacillus casei A.T.C.C. 7469.* Biochem J, 1968. **107**(4): p. 491-6.
199. Holzl, G. and P. Dormann, *Structure and function of glycoglycerolipids in plants and bacteria.* Prog Lipid Res, 2007. **46**(5): p. 225-43.
200. Raetz, C.R. and W. Dowhan, *Biosynthesis and function of phospholipids in Escherichia coli.* J Biol Chem, 1990. **265**(3): p. 1235-8.
201. Sohlenkamp, C. and O. Geiger, *Bacterial membrane lipids: diversity in structures and pathways.* Fems Microbiology Reviews, 2016. **40**(1): p. 133-159.
202. Wang, J.L., W.J. Ma, and X.Y. Wang, *Insights into the structure of Escherichia coli outer membrane as the target for engineering microbial cell factories.* Microbial Cell Factories, 2021. **20**(1).
203. Szklarczyk, D., R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A.L. Gable, T. Fang, N.T. Doncheva, S. Pyysalo, P. Bork, L.J. Jensen, and C. von Mering, *The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest.* Nucleic Acids Res, 2023. **51**(D1): p. D638-D646.

204. Tate, S.S. and S.P. Datta, *The Equilibrium of the Reaction Catalysed by Citrate Oxaloacetate-Lyase.* Biochem J, 1965. **94**(2): p. 470-7.

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

206. Meng, E.C., T.D. Goddard, E.F. Pettersen, G.S. Couch, Z.J. Pearson, J.H. Morris, and T.E. Ferrin, *UCSF ChimeraX: Tools for structure building and analysis.* Protein Science, 2023. **32**(11).

207. Pei, J., B.H. Kim, and N.V. Grishin, *PROMALS3D: a tool for multiple protein sequence and structure alignments.* Nucleic Acids Res, 2008. **36**(7): p. 2295-300.

208. Gucwa, M., J. Lenkiewicz, H.P. Zheng, M. Cymborowski, D. Cooper, K. Murzyn, and W. Minor, *Checkmymetal 2.0: A Macromolecular Metal-Binding Validation And Modeling Tool.* Acta Crystallographica a-Foundation and Advances, 2022. **78**: p. A55-A55.

209. Jurrus, E., D. Engel, K. Star, K. Monson, J. Brandi, L.E. Felberg, D.H. Brookes, L. Wilson, J.H. Chen, K. Liles, M.J. Chun, P. Li, D.W. Gohara, T. Dolinsky, R. Konecny, D.R. Koes, J.E. Nielsen, T. Head-Gordon, W.H. Geng, R. Krasny, G.W. Wei, M.J. Holst, J.A. McCammon, and N.A. Baker, *Improvements to the APBS biomolecular solvation software suite.* Protein Science, 2018. **27**(1): p. 112-128.

210. Stansfeld, P.J., J.E. Goose, M. Caffrey, E.P. Carpenter, J.L. Parker, S. Newstead, and M.S. Sansom, *MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes.* Structure, 2015. **23**(7): p. 1350-61.

211. Hall, J.A. and A.M. Pajor, *Functional reconstitution of SdcS, a Na+-coupled dicarboxylate carrier protein from Staphylococcus aureus.* J Bacteriol, 2007. **189**(3): p. 880-5.

212. Ahmad, S., L. Jose da Costa Gonzales, E.H. Bowler-Barnett, D.L. Rice, M. Kim, S. Wijerathne, A. Luciani, S. Kandasaamy, J. Luo, X. Watkins, E. Turner, M.J. Martin, and C. UniProt, *The UniProt website API: facilitating programmatic access to protein knowledge.* Nucleic Acids Res, 2025.

213. Kubo, A. and R.S. Stephens, *Substrate-specific diffusion of select dicarboxylates through Chlamydia trachomatis PorB.* Microbiology (Reading), 2001. **147**(Pt 11): p. 3135-40.

214. Soules, K.R., A. Dmitriev, S.D. LaBrie, Z.E. Dimond, B.H. May, D.K. Johnson, Y. Zhang, K.P. Battaile, S. Lovell, and P.S. Hefty, *Structural and ligand binding analyses of the periplasmic sensor domain of RsbU in Chlamydia trachomatis support a role in TCA cycle regulation.* Mol Microbiol, 2020. **113**(1): p. 68-88.

215. Omsland, A., B.S. Sixt, M. Horn, and T. Hackstadt, *Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities.* FEMS Microbiol Rev, 2014. **38**(4): p. 779-801.