Structural and Functional Characterisation of a Cytosine Transporter

Caitlin Emma Hatton

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Life Sciences

School of Life Sciences
University of Warwick

March 2021

Word Count: 40745

Table of Contents

Acknowledgements *5*

Declaration *6*

Abstract *7*

Abbreviations *8*

Figures, Tables, and Appendices *11*

Figure List **11**

Table List **14**

Appendices **15**

Chapter 1-Introduction *16*

1.1. Biological Membranes **16**

1.2. Membrane Transport **20**

1.3. Alternating Access Model **25**

1.4. LeuT-Fold and the Rocking Bundle Mechanism **28**

1.4.1. LeuT 29
1.4.2. MhsT 33
1.4.3. dDAT 34
1.4.4. SERT 35
1.4.5. vSGLT 36
1.4.6. SiaT 37
1.4.7. BetP 38

1.5. NCS1 family **41**

1.5.1. Mhp1 42
1.5.2. CodB 50

1.6. Conserved Features and Mechanism of Sodium-Dependent Symport **51**

1.7. Studying membrane proteins **52**

Chapter 2-Materials and Methods *55*

2.1. Microbiology and DNA Manipulation **55**

2.1.1. Preparation of Competent Cells 56
2.1.2. Transformation of Competent Cells 56
2.1.3. Cloning of *codB* into modified pWaldo 56
2.1.4. Site-Directed Mutagenesis of pWaldo::*codB* (*S. bovis*) 59
2.1.5. Site Directed Mutagenesis of pWaldo::*codB* (*P. vulgaris*) 60

2.2. Protein Expression **62**

2.2.1. Testing Small-Scale Expression of *codB* 62
2.2.2. Large-Scale Expression of *codB* 63
2.2.3. Expression of 3C protease 64

2.3. Membrane Preparation and Solubilisation **64**

2.3.1. Isolation of *E. coli* Membranes 64
2.3.2. Fluorescence Size Exclusion Chromatography (FSEC) of CodB 64

2.4. Protein Purification **65**

2.4.1. Purification of CodB 65
2.4.2. Expression and Purification of 3C protease 66

2.5. Crystallisation and Structural Determination of CodB **66**

2.5.1. Vapour Diffusion Crystallisation of CodB 66

2.5.2. LCP Crystallisation of CodB 67
2.5.3. Data Collection, Processing and Structural Determination of CodB 67

2.6. Biophysical Methods 68
2.6.1. Circular Dichroism of CodB (CD) 68
2.6.2. Isothermal Calorimetry (ITC) 68
2.6.3. Microscale thermophoresis (MST) 68

2.7. Thermostability Assays 69
2.7.1. GFP-Thermostability Assay (GFP-TS) 69
2.7.2. CPM Assay 70

2.8 Transport Assays 70
2.8.1. In-Cell Transport Assay 70
2.8.2. Proteoliposome Preparation and Transport Assay 71

Chapter 3-Expression, Solubilisation, and Purification of CodB 74

3.1. Introduction 74

3.2. Results 75
3.2.1 Cloning of *CodB* into pWaldo 75
3.2.2. Expression of CodB.GFP fusion 76
3.2.3. Solubility of CodB 79
3.2.4. Purification of CodB 81
3.2.5. Circular Dichroism of CodB 83

3.3. Conclusions and Discussion 84

Chapter 4-Structural Characterisation of CodB 86

4.1 Crystallisation, Data collection, and Structure Solution of CodB 86

4.2. LCP Structure of CodB in an Outward-Open Conformation 91

4.3. CodB compared with Mhp1 100
4.3.1 Substrate binding 103
4.3.2. Sodium binding 110

4.4 CodB compared with the LeuT superfamily 113

4.5 Inward-Open Model of CodB 119

4.6. Conclusions and Discussion 125

Chapter 5-Functional Characterisation of CodB 127

5.1. Introduction 127

5.2. Results 128
5.2.1. Investigating wt CodB Ligand Binding 128
5.2.2. Investigating wt CodB Transport of ^{3}H-5-cytosine 134
5.2.3. Mutagenesis of Trp108 and Phe204 to investigate their role in cytosine binding and transport 139
5.2.4. Mutagenesis of Gln105 and Ser206 to investigate their role in cytosine binding and transport 142

..... 147
5.2.5. Mutagenesis of Asn280 to investigate its role in cytosine binding 148
5.2.6. Mutagenesis of Asn275, Thr278, Thr279 to investigate their role in sodium binding 151

5.3. Conclusions and Discussion 154

Chapter 6-Discussion 159

6.1. CodB and Other Cytosine Permeases 159

6.2. CodB, Mhp1, and the NCS1 Family 162

6.2.1. Solute Coordination 163
6.2.2. Sodium Coordination 166
6.2.3. Interhelix Interactions 168
6.3. CodB and the LeuT-fold 170
6.4. Concluding Remarks 172
Bibliography 173
Appendix 1 181
Appendix 2 186
Appendix 3 193

Acknowledgements

To my work colleagues:

Firstly, I'd like to thank my supervisor Alex for keeping me calm and grounded when I was stressed and panicked, without you I think my life expectancy would be a few years shorter, and I'd certainly have many more grey hairs. My thanks extend to my co-supervisor Tim, for all those times you laughed with/at me.

I'd like to thank the rest of the Cameron Group, Debs thank you for the uncountable number of times you helped me with experiments, especially in the radioactive lab. Patrick and Aury, thank you for all those times when I needed a coffee break because experiments weren't working, or to commiserate the inevitable problems of working with membrane proteins, or just departmental gossip.
Thank you to the Knowles group at the University of Birmingham, I really enjoyed my time with you.

To my family:

I must thank my mam, Kate, for her unconditional and unwavering love and support all these years and my dad, Pete, for always being so passionate and excited for everything I cared about. To my sister, Corinne, for forever reminding me that a world outside academia exists.

To my grandfather, for that year you spent teaching me maths because I'd lost my confidence, you made me understand and really showed me how elegant calculus can be. To my grandmother, for those trips to the local bakery to buy gingerbread men, and for all your endless stories.

This thesis is dedicated to my grandparents Joan and Gordon for placing my education above all else, none of this would have been possible without either of you.

Declaration

This project is my own work except where indicated. All text, figures, tables, data or results which are not my own work are indicated and the sources acknowledged.

Abstract

CodB, the putative cytosine transporter from *Escherichia coli*, is a member of the LeuT-fold and evolutionarily clusters with the hydantoin transporter Mhp1 from *Microbacterium liquefaciens.* Mhp1 was the first active secondary-transporter structurally characterised in all conformations of the alternating-access model. The aim of this thesis was to structurally and functionally characterise CodB to expand understanding of transport mechanism and the role of sodium in the nucleobase cation symporter 1 (NCS1) family. Chapter 3 describes the expression, solubilisation, and purification of CodB from different bacterial species to identify suitable constructs for downstream experiments. CodB from *Proteus vulgaris* was acknowledged as the best candidate for structural and functional studies due its ability to be solubilised and purified in appropriate quantities. Chapter 4 details the structural determination and analysis of CodB to a final resolution of 2.4 Å, this structure is an outward-open conformation bound to both cytosine and sodium. This structure identifies the amino acids required for cytosine and sodium coordination as well as highlighting similarities and differences of CodB with other LeuT-fold proteins. Chapter 5 focuses on ligand binding and transport kinetics of CodB. Cytosine has been confirmed as a ligand for CodB and CodB mediated transport was confirmed to be sodium dependent, whilst other potential ligands and binding partners were screened;. Amino acids in the substrate and cation sites were mutated to understand their role in binding and transport.

Abbreviations

β-OG	Beta-Octyl-glucosidase
ABC	ATP-binding cassette
AdiC	Arginine: agmatine antiporter from *E. coli*
AEBSF	4-benzenesulfonyl fluoride hydrochloride
APC	Amino acid/Polyamine/Organocation
ApcT	Amino acid transporter from *Methanocaldococcus jannaschii* and *Geobacillus kausto-philus*
AU	Asymmetric unit
BCCT	Betaine/choline/carnitine transporter
BetP	Betaine transporter from *Corynebacterium glutamicum*
BH	Benzyl-hydantoin
c-terminus	Carboxyl-terminus
CaiT	γ-butyrobetaine: carnitine exchanger from *Proteus mirabilis* and *E. coli*
CD	Circular dichroism
CHS	Cholesteryl hemisuccinate
CIP	Calf intestinal phosphate
CL	Cardiolipin
CMC	Critical micelle concentration
dDAT	Dopamine Transporter from *Drosophila melanogaster*
DDM	n-Dodecyl β-D- maltopyranoside
DDT	Dichlorodiphenyltrichloroethane
DM	n-Decyl β-maltopyranoside
DNA	Deoxyribonucleic acid
EL	Extracellular loop
Fab	Fragment antigen-binding
FI	Fluorescence Intensity
FRET	Förster resonance energy transfer
FSEC	Fluorescence size exclusion chromatography
GABA	γ-aminobutyric acid
GadC	Glutamate: GABA exchanger from *E. coli*
GFP	Green fluorescent protein

GFP-TS	GFP-thermostability assay
Glt_{Ph}	Na^+- aspartate symporter
GLUT1	Glucose transporter 1
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IDT	Integrated DNA Technologies
IMH	Indolyl methyl-hydantoin
IPTG	Isopropyl β-D-1 thiogalactopyranoside
ITC	Isothermal calorimetry
K_d	Equilibrium dissociation constant
LacY	Lactose permease
LB	Luria-Bertani
LCP	Lipidic cubic phase
LDAO	*N,N* Dimethyldodecylamine *N*- oxide
LeuT	Leucine Transporter from *Aquifex aeolicus*
MD	Molecular dynamics
MES	2-ethanesulfonic acid
MFS	Major facilitator superfamily
Mhp1	Na^+ coupled hydantoin transporter from *Microbacterium liquefaciens*
MhsT	Multi-hydrophobic substrate transporter from *Bacillus halodurans*
MST	Microscale thermophoresis
MW	Molecular weight
n-terminus	Amino-terminus
NCS1	Nucleobase-cation-symporter-1
Neu5Ac	N-acetylneuraminic acid
NhaA	Na^+/ H^+ antiporter A
NM	N-Nonyl-β-D-maltopyranoside
NMH	5-(2-naphthylmethyl)-L-hydantoin
NSS	Neurotransmitter sodium symporter
OD_{600}	Optical density at 600nm
PBS	Phosphate-buffers saline
PCR	Polymerase chain reaction

PDB	Protein data bank
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PutP	Proline transporter from *E. coli*
RNA	Ribonucleic acid
rpm	Rotations per minute
S.E.M.	Standard error of the mean
SDS	Sodium dodecyl sulphate
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SERT	Serotonin transporter from *H. sapiens*
SiaT	Sialic acid transporter from P*roteus mirabilis*
SOC	Super optimal broth with catabolite repression
SSRIs	Selective serotonin reuptake inhibitors
SSS	Solute sodium symporters
TCA	Tricyclic antidepressants
T_m	Melting temperature
TM	Transmembrane helix
TRIS	Trisaminomethane
UM	n-undecyl-β-d-maltopyranoside
vSGLT	Sodium galactose co-transporter from *Vibrio parahaemolyticus*

Figures, Tables, and Appendices

Figure List

Figure 1.1. Schematic representation of the fluid mosaic model as proposed by Singer and Nicolson in 1972.

Figure 1.2. Chemical structure of a general phospholipid.

Figure 1.3. Structure of CL, PE, and PG.

Figure 1.4. Schematic of types of membrane proteins.

Figure 1.5. Schematic illustration of secondary active transporters with example transporters.

Figure 1.6. Membrane topology of common secondary-active transporter folds.

Figure 1.7. Alternating-access mechanisms of substrate transport across the membrane.

Figure 1.8. Schematic of transport in LeuT.

Figure 1.9. Transport cycle of MhsT.

Figure 1.10. Transport model of SERT and ibogaine.

Figure 1.11. Schematic of the eight conformations of BetP structurally characterised.

Figure 1.12. Mhp1 topology.

Figure 1.13. Residues required for coordination of IMH in the Mhp1 binding site.

Figure 1.14. Comparison motions in the substrate binding site of Mhp1.

Figure 1.15. Comparison movements in the sodium binding site of Mhp1.

Figure 1.16. Schematic representation of the alternating access model of Mhp1.

Figure 2.1. Schematic representation of CodB.GFP protein expression.

Figure 3.1. 1% agarose gel demonstrating pWaldo has been cut with restriction enzymes.

Figure 3.2. Quantifying CodB (*E. coli)* expression.

Figure 3.3. Quantifying CodB (*S. enterica)* expression.

Figure 3.4. Quantifying CodB (*P. vulgaris)* expression.

Figure 3.5. Quantifying CodB (*C. lundense)* expression**.**

Figure 3.6. Quantifying CodB (*S. bovis)* expression.

Figure 3.7. Concentration of CodB plotted against elution volume when solubilised in DDM.
Figure 3.8. Concentration of *S. enterica* CodB plotted against elution volume when solubilised in different detergents.
Figure 3.9. Concentration of *P. vulgaris* CodB plotted against elution volume when solubilised in different detergents.
Figure 3.10. Purification of CodB.
Figure 3.11. SEC profile of CodB in various conditions on a S200 column.
Figure 3.12. UV CD of purified wt CodB.

Figure 4.1. CodB crystals and x-ray diffraction.
Figure 4.2. Cartoon representation of CodB AU and individual protomer.
Figure 4.3. Amino acid sequence of CodB and secondary structure topology map.
Figure 4.4. CodB inverted repeats.
Figure 4.5. CodB is found in an outward-open conformation.
Figure 4.6. Cytosine binding site of CodB.
Figure 4.7. Cytosine binding site of CodB in atomic detail.
Figure 4.8. Sodium binding site of CodB.
Figure 4.9. DDM modelled onto CodB.
Figure 4.10. Phospholipid modelled on CodB.
Figure 4.11. Sequence alignment of CodB and Mhp1 using MUSCLE algorithm.
Figure 4.12. Superposition of CodB and 2JLN.
Figure 4.13. CodB and Mhp1 Structural differences in position of TM10.
Figure 4.14. Close up of π-stacking interactions in CodB and Mhp1.
Figure 4.15. Hydrogen bonding network of Q121 and Q42 in Mhp1 with the corresponding residues in CodB.
Figure 4.16. Orientating substrate in the substrate binding pocket is dictated by hydrogen bonds donated by Gln105 in CodB and Asn318 in Mhp1.
Figure 4.17. Water coordination in CodB and the corresponding residues in Mhp1.
Figure 4.18. Hydrogen bonding network between TM4 and TM8 in CodB and Mhp1.
Figure 4.19. Hydrogen bonding of Asn275 and S34 in CodB and the corresponding residues in Mhp1.

Figure 4.20. TM8 and TM4 interaction mediated by Asp281 and Thr142 in CodB and the equivalent residues in Mhp1.
Figure 4.21. Hydrogen bonding network of TM1, 5, and 8 in CodB and Mhp1.
Figure 4.22. Na1 site in CodB and LeuT.
Figure 4.23. Na3 site in CodB and SiaT.
Figure 4.24. Inward-open model of CodB.
Figure 4.25. Putative cytosine binding site of inward-open CodB compared to the outward-open structure.
Figure 4.26. Comparison of the sodium binding site in the outward and inward structures.
Figure 4.27. Comparison of Asp281 interactions in the outward-open structure and the inward-open model.
Figure 4.28. Comparison of the hydrogen bonding network of Asn282, Val26, and Ser154 in the outward-open structure and inward facing model.

Figure 5.1. GFP-TS results of wt CodB.GFP with nucleobases.
Figure 5.2. GFP-TS results of wt CodB.GFP with cytosine-related compounds.
Figure 5.3. Binding affinity of CodB.GFP wt to cytosine.
Figure 5.4. Binding affinity of CodB.GFP wt to 5-fluorocytosine.
Figure 5.5. Isothermal calorimetry of CodB and cytosine.
Figure 5.6. Time course of 3H cytosine uptake by CodB.GFP.
Figure 5.7. Inhibition of ^{3}H-5-cytosine uptake in the presence of 0.1 mM inhibitor.
Figure 5.8. CodB liposomes prepared by the extrusion method.
Figure 5.9. CodB proteoliposomes prepared by rapid dilution.
Figure 5.10. Interactions of Trp108 and Phe204 with cytosine and the impact of mutants Trp108Ala, Phe204Ala on cytosine binding.
Figure 5.11. Thermoshift experiments of detergent solubilised CodB.GFP wt, Trp108Ala, Phe204Ala.
Figure 5.12. Uptake of ^{3}H-5-cytosine by CodB.GFP mutants Trp108Ala and Phe204Ala, normalised to WT.
Figure 5.13. Interactions of Gln105 and Ser206 with cytosine and the impact of mutants Gln105Ala, Ser206Ala on cytosine binding.
Figure 5.14. Thermoshift experiments of detergent solubilised CodB.GFP wt, Gln105Ala, Ser206Ala.

Figure 5.15. Binding affinity of CodB.GFP mutant Ser206Ala.
Figure 5.16. Uptake of ^{3}H-5-cytosine by CodB.GFP mutants Gln105Ala and Ser206Ala, normalised to WT.
Figure 5.17. Interactions of Asn280 with cytosine and the impact of the Asn280Ala mutation on cytosine binding.
Figure 5.18. Thermoshift experiments of detergent solubilised CodB.GFP wt and Asn280Ala.
Figure 5.19. Uptake of ^{3}H-5-cytosine by CodB.GFP mutant Asn280Ala, normalised to WT.
Figure 5.20. Interactions of Asn275, Thr278, Thr279 with Na^{+} and TM1, and the impact of Asn275Ala, Thr278Ala, Thr279Ala mutations on sodium coordination.
Figure 5.21. Thermoshift experiments of detergent solubilised CodB.GFP wt, Asn275, Thr278Ala, and Thr279Ala.
Figure 5.22. Inhibition of ^{3}H-5-cytosine uptake in the presence of 0.1 mM DDM.

Figure 6.1. MUSCLE sequence alignment of CodB, CodB from *E. coli* and VPA1242.
Figure 6.2. Models of cytosine, 5-fluorocytosine, 5-methylcytosine, uracil, and thymine in the CodB solute binding pocket.
Figure 6.3. Chemical structures of substrates for the NCS1 family and related compounds.

Table List

Table 2.1. Bacterial strains used for expression.
Table 2.2. Media components.
Table 2.3. Plasmids used with properties.
Table 2.4. PCR conditions to correct pWaldo::*codB* (*S. bovis*).
Table 2.5. Conditions for optimal protein expression of CodB constructs.
Table 2.6. PCR conditions to generate mutants of pWaldo::*codB P. vulgaris.*
Table 2.7. Primer sequences for site directed mutagenesis of pWaldo::*codB*.

Table 4.1 Data collection statistics of CodB.
Table 4.2. Refinement statistics for CodB:cytosine complex.
Table 4.3. Sequence identity of CodB and other LeuT superfamily members.

Appendices

Appendix 1. IDT Gene optimised DNA sequences of CodB and RONN disorder plots.

Appendix 2. MUSCLE alignment of all characterised NCS1.

Appendix 3. Inhibition of ^{3}H-5-cytosine uptake in the presence of 0.5 mM inhibitor.

Chapter 1-Introduction

1.1. Biological Membranes

All living cells require a biological membrane to separate the cell from the external environment allowing the cell to control and maintain their intracellular environment. These membranes are essential for life. The molecular structure of a biological membrane means it is partially permeable enabling the cell to maintain a stable intracellular environment that is fundamental for life. Biological membranes can be described using the fluid mosaic model in Figure 1.1 (Singer and Nicolson, 1972). This model describes the membrane as proteins that are associated with a lipid bilayer matrix, these proteins are stabilised by hydrophobic interactions with the lipids. This lipid bilayer forms a large planar sheet-like structure that forms a continuous barrier around the bacterium or organelle in eukaryotes. This barrier is essential for life, and due to its partially-permeable nature does not allow the diffusion of many essential nutrients across the membrane. The proteins present in the lipid bilayer carry out the processes that associated with the membrane such as signal transduction, cell communication, toxin secretion, excretion of waste, the uptake of nutrients, and many other fundamental biochemical processes, such as oxidative phosphorylation. Approximately 20-30% of all genes encode for integral membrane proteins (Krogh *et al.*, 2001), highlighting the vital role they play in cellular biology.

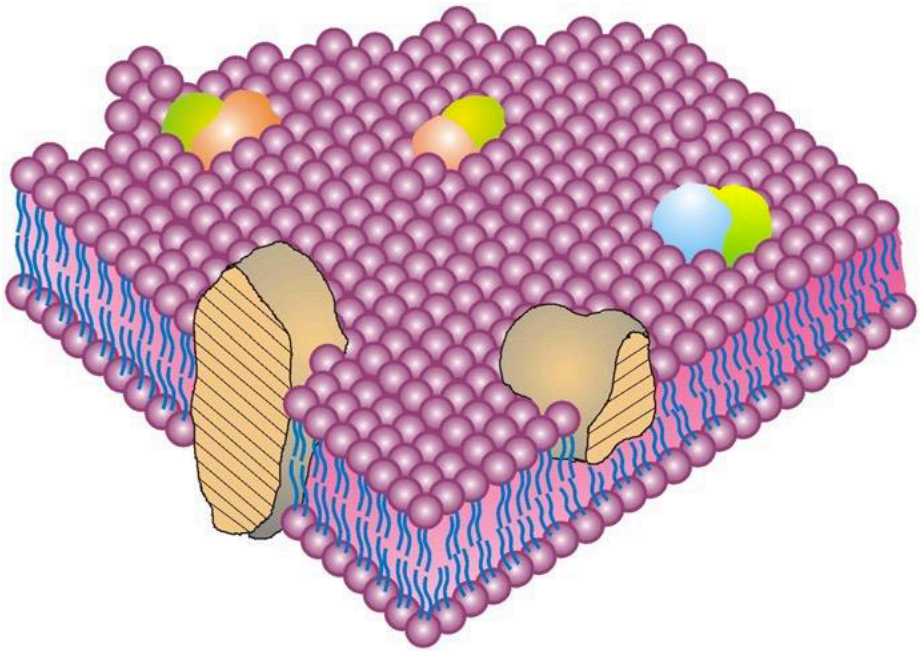


Figure 1.1. Schematic representation of the fluid mosaic model as proposed by Singer and Nicolson in 1972. This cross-sectional model demonstrates that the membrane consists of a sea of lipids providing the environment for membrane proteins to exist, the proteins are then able to laterally move within this bilayer structure. The hydrophobic tails of the lipids are secluded from the aqueous exterior and interior. This figure was taken from (Singer and Nicolson, 1972).

The fluid mosaic model is too simple for accurately describing membrane structure, it was assumed that lipid and protein was randomly distributed and a leaflet was relatively homogenous. In reality, protein is the largest constitute of the membrane not the lipid; areas of the membrane can be specialised by the clustering of specific proteins and lipids as physical properties of the lipids are exploited to determine functionality and influence the fluidity, curvature, and depth of the membrane.

There are three types of membrane lipid: phospholipids, glycolipids and cholesterol. Phospholipids are the most abundant and are composed of a glycerol-3-phosphate attached to two acyl-chains, at positions C1 and C2, that can vary in length (Figure 1.2). The substituent denoted by X determines role the lipid plays in the bilayer. For example, cardiolipin (CL), phosphatidylethanolamine (PE), and phosphatidylglycerol (PG) are the phospholipids found in *Escherichia coli* with their respective general structure shown in Figure 1.3 (Sohlenkamp C, 2016). PE is the main lipid in *E. coli* with its primary role being to spread out negative charge caused by the anionic phospholipids; many membrane transporters do not fold correctly in the absence of PE. As the chemical structure shows the phospholipid is amphipathic with a hydrophilic head group and hydrophobic fatty acid chains. The amphipathic

nature of these molecules allows them to spontaneously form micelles or bilayers by shielding the acyl-chains and allow the polar head groups to be solvated. These bilayers are very impermeable in nature, and it is very energetically unfavourable for ions and hydrophilic molecules to pass through them. However, as described above the membrane needs to be partially permeable and selective so whilst the lipids act as barrier the proteins are responsible for all the other roles of the membrane.

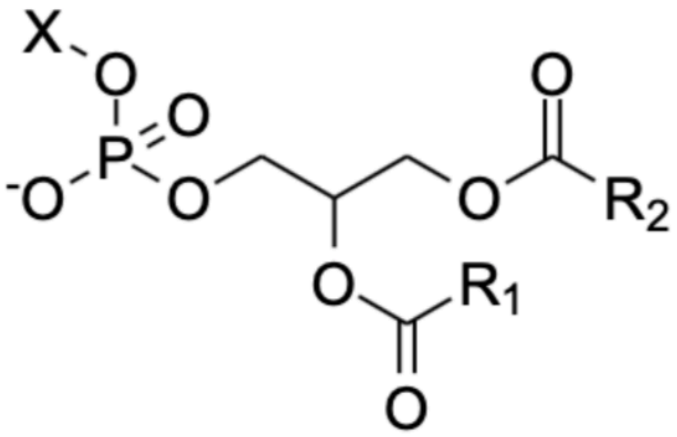


Figure 1.2. Chemical structure of a general phospholipid. X can be a variety of substituents whilst R_1 and R_2 are fatty acids that can vary in length.

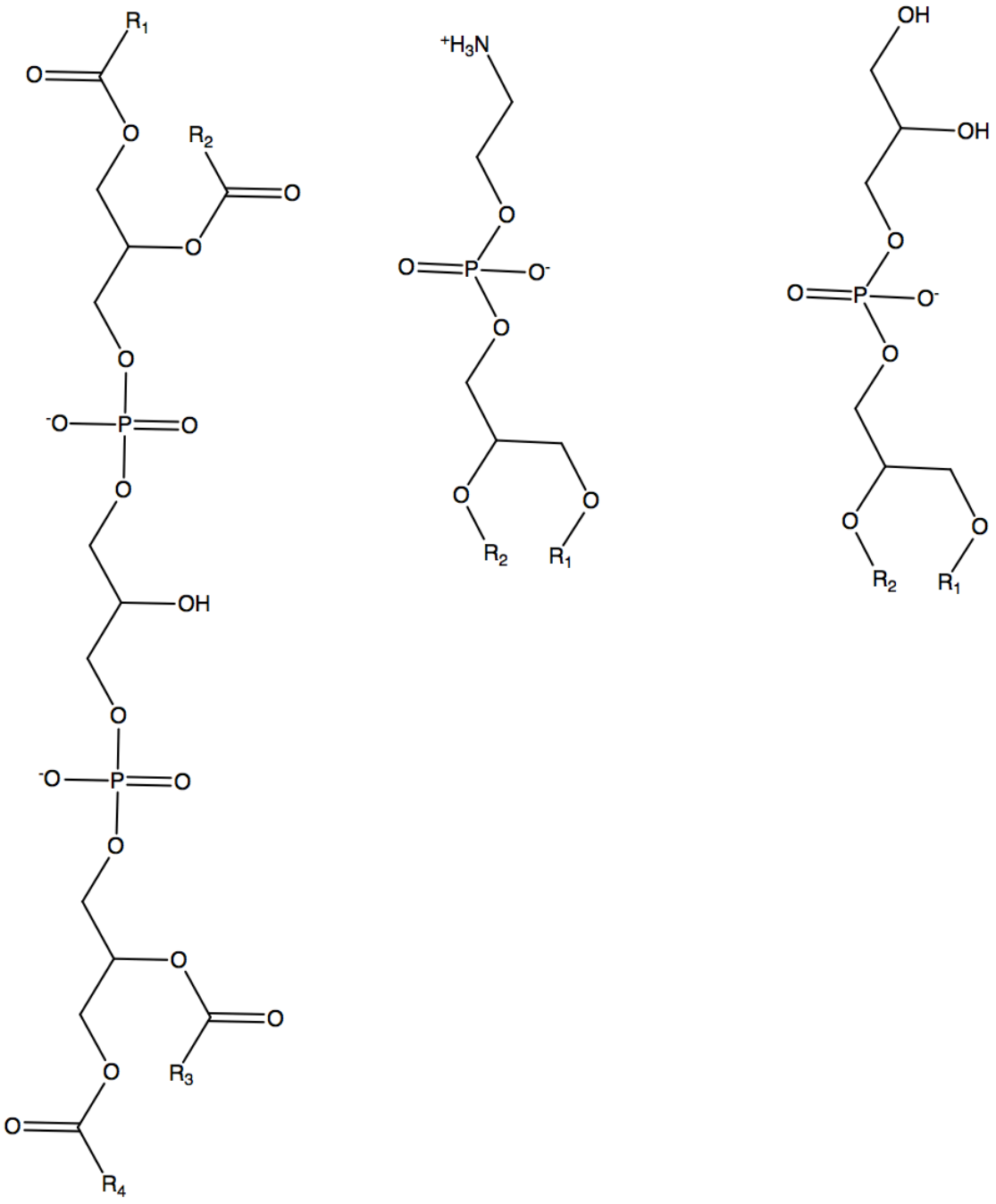


Figure 1.3. Structure of CL, PE, and PG. CL (left) contains four alkyl groups and consequently the structural variety within this class of lipid is enormous. PE (centre) is made up of two alkyl groups that can vary in length and structure and is the main phospholipid in bacteria. PG (right) has a very similar structure to PE, but PE contains an ethanolamine group attached to the head group instead of the glycerol group.

Membrane proteins can be broadly categorised into integral membrane proteins that are permanently attached to the membrane via transmembrane regions, or peripheral membrane proteins that are not permanently attached to the membrane as shown in Figure 1.4.

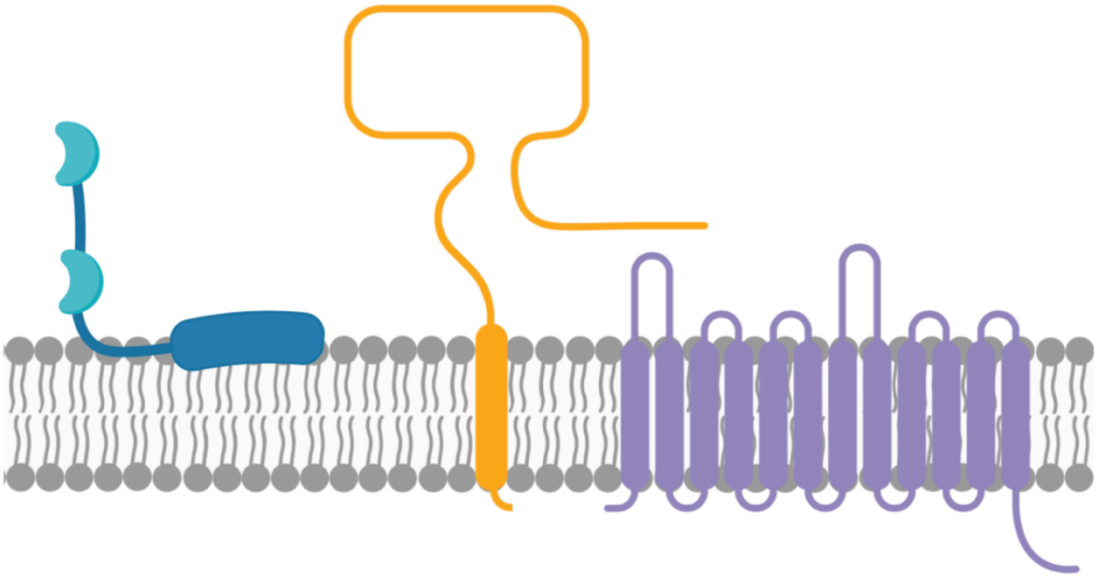


Figure 1.4. Schematic of types of membrane proteins. Membrane associated (blue), single transmembrane (orange) and multi-transmembrane spanning (purple). Membrane is illustrated in grey. This figure was generated in BioRender ™.

Integral membrane proteins have a hydrophobic region(s) that sits within the bilayer and a hydrophilic region(s) that is exposed to the internal or external environment. Integral membrane proteins can be further split into α-helical proteins, or into β-barrel proteins, β-barrel proteins are only found in the outer membrane of gram-negative bacteria and the outer membrane of mitochondria and chloroplasts.

1.2. Membrane Transport

As described in the previous section, the lipid bilayer is incredibly effective as a barrier, however, molecules need to cross the membrane in both directions. Lipophilic molecules can do this easily as they can dissolve in the lipid bilayer, but for polar molecules or ions it is more complex. Crossing the membrane is energetically unfavourable and these molecules or ions require the aid of membrane transport proteins to allow this transport to happen.

Membrane transport can happen via passive transport, where the energy driving transport originates from the gradient of the molecule/ion, or through active transport, active transport requires an energy source. Membrane transport proteins can be split into two major classes: channels or transporters. Channels, such as Na^+ or K^+ channels (Zhang *et al.*, 2012), are open to both the internal and external environment simultaneously whilst transporters are not. Channels allow passive transport via facilitated diffusion allowing molecules to diffuse the membrane uninterrupted. In contrast, transporters have specific binding sites and require multiple conformational changes to facilitate transport and require an energy source.

Transporters can be classified into primary, or secondary active transport. Primary transporters use the energy derived from adenosine 5-triphosphate (ATP) hydrolysis, such as the ATP-binding cassette (ABC) transporter superfamily (Wilkens, 2015), whilst secondary transporters, shown in Figure 1.5, use an ion gradient as the free energy source to drive the transport of a different molecule. Secondary active transporters can transport their solute and co-solute in the same direction, known as symporters, or in opposite directions, known as anti-porters. Uni-porters, such as Glucose transporter 1 (GLUT1) (Deng *et al.*, 2014), share the same kinetics as secondary transporters but instead use facilitated diffusion and do not require a co-ion.

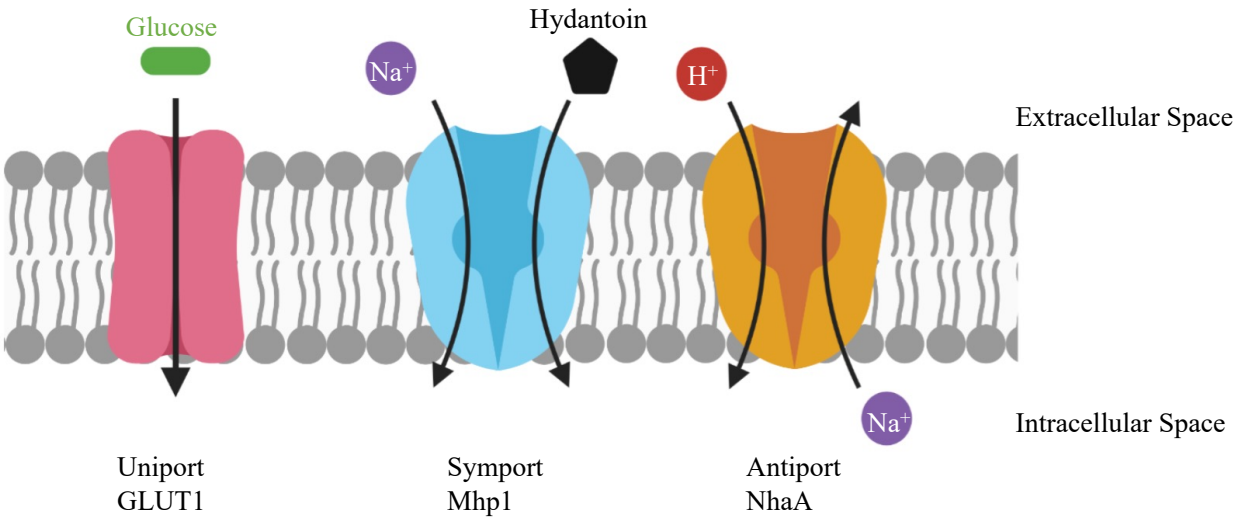


Figure 1.5. Schematic illustration of secondary active transporters with example transporters. Uniporters (left) transport solutes down a concentration gradient. Symporters (centre) transport a solute in the same direction as a co-ion using the co-ion electrochemical gradient as an energy source. Antiporters (right) transport a solute in the opposite direction of the electrochemical gradient of the co-ion. This figure was produced using BioRender ™.

Secondary transporters can be classified into families based on their sequence and functional similarities (Hediger *et al.*, 2004; Saier *et al.*, 2015); as more structures are released it is becoming increasingly clear that despite very low or no sequence/function similarities these proteins are using the same 3D folds (Abramson and Wright, 2009; Theobald and Miller, 2010).

Secondary active transporters are grouped into superfamilies based on their 3D structure:

1. Major facilitator superfamily (MFS)/Lactose permease (LacY)
2. ATP/ Adenosine 5-diphosphate (ADP) Carrier
3. Na^+/ H^+ antiporter A (NhaA)
4. Bacterial Leucine Transporter (LeuT)
5. Na^+- aspartate symporter (Glt_{Ph})
6. Uracil Transporter (UraA)

Despite the variations in the 3D structure all these folds have a common feature of repeated structural motifs (Figure 1.6), these repeated structural folds are thought to have arisen from a gene duplication event followed by fusion and sequence divergence. The symmetry of these inverted repeats is essential for the molecular mechanisms of transport.

The MFS fold was the first fold to be characterised (Abramson *et al.*, 2003), and is thought to make up 25% of all transporters (Pao, Paulsen and Saier, 1998; Quistgaard *et al.*, 2016), this fold consists of 12 helices that split into 2 repeating parallel units of 6 transmembrane helices (TM) i.e. TM1 is equivalent to TM7. The six TM unit can be split into 2 x 3TM that are inverted repeats of each other (Shi, 2013). Solute binds in a cavity at the centre of the transporter between the parallel structural repeats, rocking motions around the solute binding site act as a pivot enable access to either side of the membrane.

The NhaA fold is built from a 2x 5TM inverted repeats with a distinctive discontinuous helix at TM4 and its equivalent partner of TM9, these discontinuous helices cross each other in the centre of the structure to stabilise each other (Hunte *et al.*, 2005; Hu *et al.*, 2011; Lee *et al.*, 2013). The UraA fold consists of 2x 7TM

inverted repeats with the unique beta-strand in the unwound TM3 and its symmetric equivalent of TM10 (Lu *et al.*, 2011; Yu *et al.*, 2017).

The LeuT fold is characterised by a two-fold pseudosymmetric 5 helix invert, i.e TM 1 is structurally equivalent to TM6 despite no sequence identity. A hallmark of the LeuT superfamily is the discontinuous helices of TM1 and TM6, the two shorter helices are connected by an unwound section exposing main-chain hydrogen bonds to provide an environment for solute and co-ion to bind (Yamashita *et al.*, 2005). The LeuT-fold is found across many transporters that show no sequence or functional similarity and it is hypothesised that all members of the LeuT superfamily arise from a common ancestor (Pao, Paulsen and Saier, 1998).

All of these folds, apart from Glt_{Ph}, have family members that are symporters or antiporters, demonstrating that symport and antiport can be achieved with the same protein architecture.

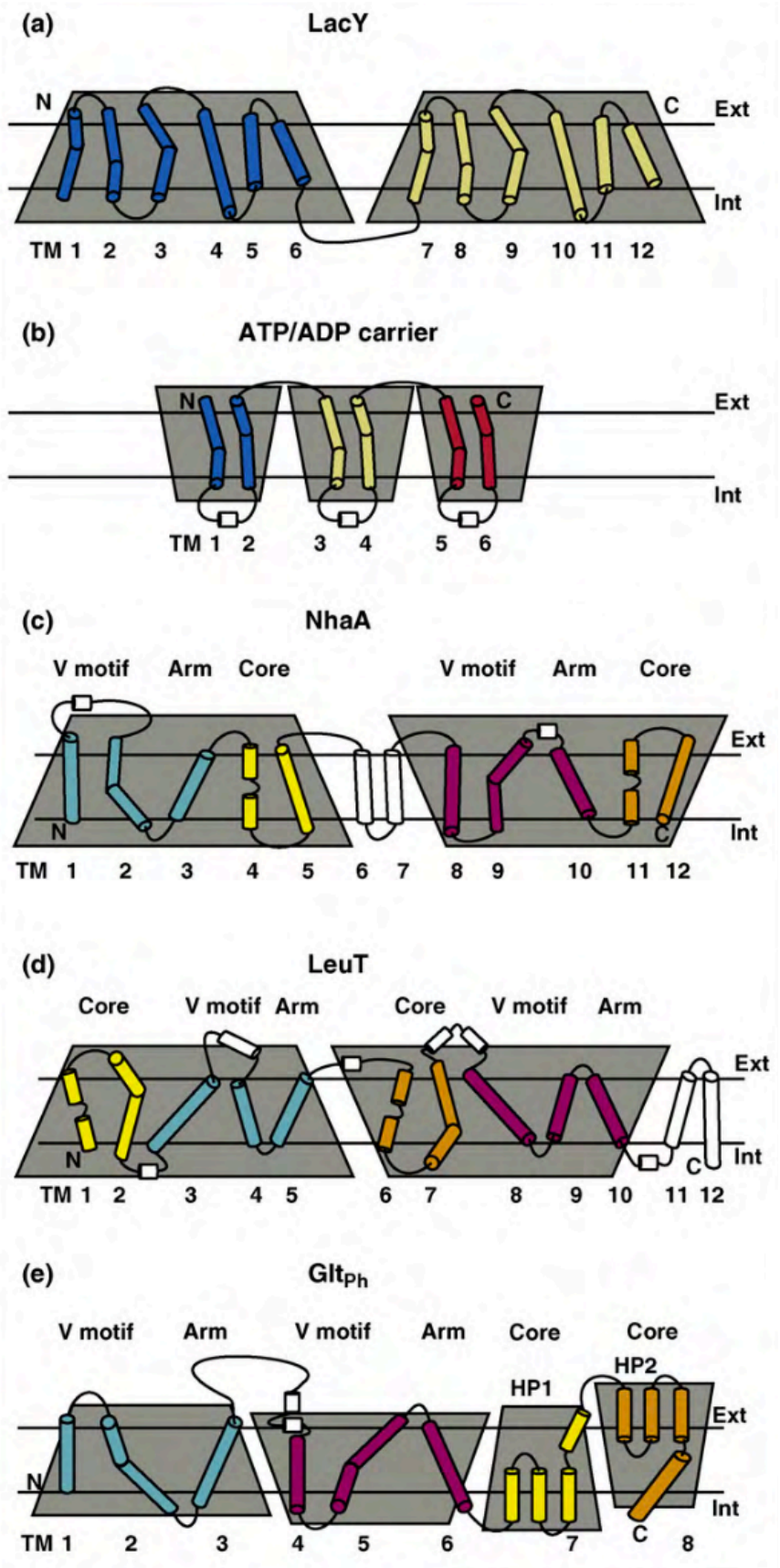


Figure 1.6. Membrane topology of common secondary-active transporter folds. This figure has been taken from (Boudker and Verdon, 2010) Structural repeats are indicated by grey trapezoids. Helices are illustrated as cylinders, with unwound regions depicted as lines.

1.3. Alternating Access Model

In 1966, Jardetzky proposed the alternating access model to describe the movement of small molecules across biological membranes (Jardetzky, 1966). The alternating access model must fulfil 4 conditions (Tanford, 1983):

1. The transporter must exhibit at least 2 distinct conformational states to provide access to each side of the membrane sequentially, the transporter is never accessible to both sides of the membrane simultaneously.
2. The two distinct conformational states have different binding constants for solute, affinity is high on the uptake side but low on the discharge side to enable solute release.
3. During transition, the solute site must move relative to the protein structure or vice-versa.
4. The transporter must recycle back to its original conformation at the end of the transport cycle.

Through the wealth of structures now solved the alternating access can now be split into the rocker-switch mechanism, the rocking-bundle mechanism and the elevator mechanism.

The rocker-switch mechanism describes how solute is able to bind deep within the protein structure between two structurally similar domains i.e. at the interface between TM1-6 and TM7-12 in the MFS-fold. Substrate binding induces the transporter rearrange the domains around the substrate binding site. TM1, 4, 7, 10 are important for substrate biding and are recognised as gating helices. (Drew and Boudker, 2016)

The rocking-bundle mechanism, alternatively the gated pore mechanism, is how the LeuT-fold are able to transport solutes, the precise mechanisms of members will be discussed in detail later in this chapter. In contrast to the rocking-switch, the substrate binds between the interface of structurally different domains; substrate binding induces a closure of the extracellular gate, a rigid-body movement of one domain against the other to shift the transporter into the inward-facing conformation.

Opening of the intracellular gate allows substrate to exit the substrate binding site. The gates in the LeuT fold are more extensive and variable than the MFS.

The elevator model was originally thought just to exist in the glutamate transporter (Boudker *et al.*, 2007) family but with the release of structures from different functional families and folds this mechanism is utilised more than originally thought. The elevator mechanism picks up the solute from one side of the membrane, a gate closes and prevents solute from being released again, the transport domain then undergoes a vertical movement within the membrane and a gate on the other side of the opens and allows substrate to dissociate away. Transporters with the NhaA-fold are thought to use this elevator model to transport solute, as well as Glt_{Ph} (Hu *et al.*, 2011; Lee *et al.*, 2013; Zhou *et al.*, 2014).

The transport mechanism of most transporters are either the rocker-switch or rocking-bundle mechanism, as such these mechanisms are very similar to each other. A schematic of these difference transport mechanisms can be found in Figure 1.7.

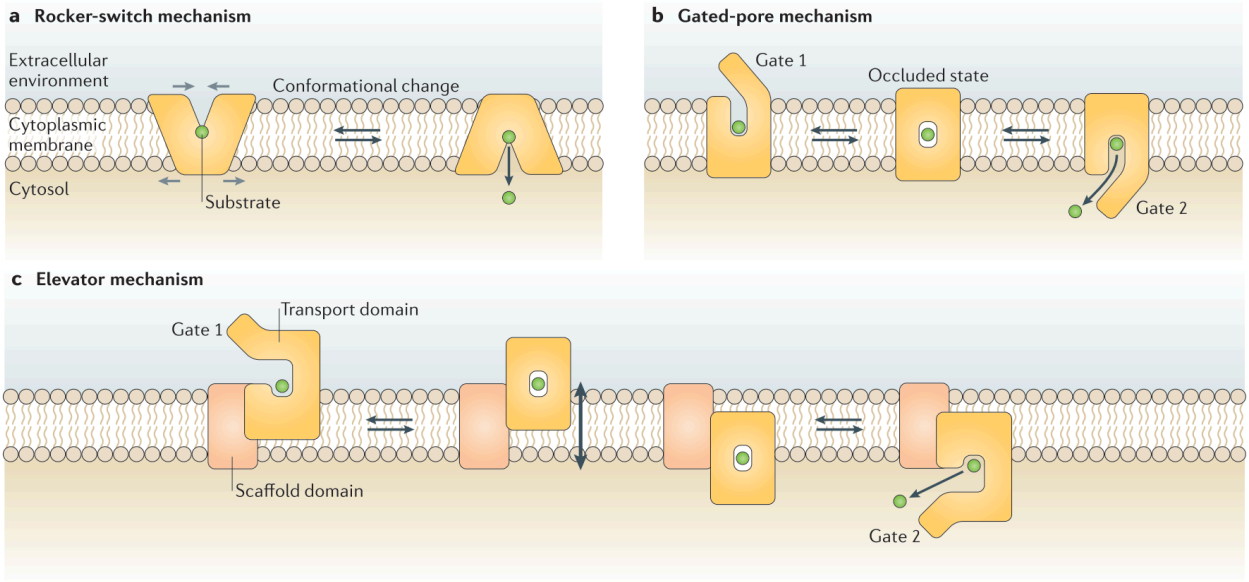


Figure 1.7. Alternating-access mechanisms of substrate transport across the membrane. a) The rocker switch mechanism results in the rearrangement of protein domain around the substrate binding site upon substrate binding to enable access to the alternative side of the membrane. b) The gated pore or rocking bundle mechanism binds substrate in the middle of the membrane and closes the extracellular gate to lock substrate into the binding pocket. A varying degree (protein dependent) of rigid body rocking of the asymmetric domains relative to each other occurs in the transition from outward-facing to inward-facing. The intracellular gate opens to enable substrate access to the interior; the transporter to cycle back to the outward-facing state. c) The elevator mechanism of transport picks up substrate, closes the extracellular gate, allowing vertical translation of the mobile transport domain relative to the membrane and scaffold domain to transport substrate to the other side of the membrane, the intracellular gate opens releasing substrate allowing the transporter to cycle back to the outward-facing state. This figure has been taken from (Slotboom, 2014).

1.4. LeuT-Fold and the Rocking Bundle Mechanism

At the time of writing, structural information exists for these members of the LeuT-fold categorised into the following functional families:

1. Neurotransmitter sodium symporter (NSS) family:
   - LeuT- Leucine Transporter from *Aquifex aeolicus* (Section 1.4.1.) (Yamashita *et al.*, 2005)
   - MhsT - Multi-Hydrophobic Substrate Transporter from *Bacillus halodurans* (Section 1.4.2.) (Malinauskaite *et al.*, 2014)
   - dDAT- Dopamine Transporter from *Drosophila melanogaster* (Section 1.4.3.) (Penmatsa, Wang and Gouaux, 2013)
   - SERT- Human Serotonin Transporter (Section 1.4.4) (Coleman *et al.*, 2019)
2. Solute Sodium Symporters (SSS) family:
   - vSGLT- Sodium Galactose Co-Transporter from *Vibrio parahaemolyticus* (Section 1.4.5.) (Faham *et al.*, 2008)
   - SiaT- Sialic Acid Transporter from *Proteus mirabilis* (Section 1.4.6) (Wahlgren *et al.*, 2018)
3. Betaine/Choline/Carnitine Transporter (BCCT) family:
   - BetP- Betaine Transporter from *Corynebacterium glutamicum* (Section 1.4.7.) (Ressl *et al.*, 2009)
   - CaiT-γ-butyrobetaine: carnitine exchanger from *P. mirabilis* and *E. coli* (Schulze *et al.*, 2010; Tang *et al.*, 2010)
4. Amino acid/Polyamine/Organocation (APC) superfamily
   - AdiC- Arginine: agmatine antiporter from *E. coli* (Fang *et al.*, 2009)
   - ApcT-Amino acid transporter from *Methanocaldococcus jannaschii* and *Geobacillus Kausto-philus* (Shaffer *et al.*, 2009; Jungnickel, Parker and Newstead, 2018)
   - GadC- Glutamate: GABA exchanger from *E. coli* (Ma *et al.*, 2012)
5. Nucleobase-Cation-Symporter-1 (NCS1) family
   - Mhp1- Na+ coupled hydantoin transporter from *Microbacterium liquefaciens* (Section 1.5.1) (Weyand *et al.*, 2008; Shimamura *et al.*, 2010)

The LeuT-fold is large functionally diverse superfamily, the release of structures of different family members with many in multiple conformations has allowed

understanding of conserved and differing features of the mechanism of transport. The original LeuT structure in an outward-occluded conformation revealed unexpected inverted structural repeats (Yamashita *et al.*, 2005), investigation of this pseudo-symmetry resulted in the conception of the rocking bundle mechanism before other conformations were solved (Forrest *et al.*, 2008; Forrest, Krämer and Ziegler, 2011).

Mhp1 was the first member of the LeuT superfamily to solved in multiple conformations and confirmed the rocking bundle mechanism; the only to date where mutations have not been required to force the transporter into a different conformation (Weyand *et al.*, 2008; Shimamura *et al.*, 2010; Weyand *et al.*, 2011; Simmons *et al.*, 2014) (Section 1.5.1). Only LeuT-fold sodium-dependent symporters will be discussed in this thesis.

1.4.1. LeuT

The leucine transporter (LeuT) from *A. aeolicus* is the founding member of the LeuT superfamily and was solved crystallographically in 2005 in a complex with one molecule of leucine and 2 Na^+ ions in an outward-occluded state (Yamashita *et al.*, 2005). LeuT is a member of NSS family, high-profile members of this family are the sertonin, γ-aminobutyric acid (GABA), noradrenaline, and dopamine transporters that have fundamental roles in the function of the nervous system (Masson *et al.*, 1999).

The initial structure of LeuT revealed how substrate and co-ion were specifically bound and the presence of extracellular and intracellular gates.

LeuT can be broken into a “core” or “four-helix bundle” consisting of TM1, 2, 6, 7 and a “scaffold” of TM3, 4, 8, and 9. The structure of LeuT showed discontinuous helices in TM1 and TM6, unwinding of these helices exposed the hydrogen bonds that would normally be involved with mainchain interactions and provides dipoles for the amino and carboxyl group binding of the leucine solute. The amino-group interacts with residues in TM1 and TM6, whilst the carboxyl-group is coordinated with a sodium-ion in the Na1 site, TM1, and the hydroxyl-group from the Tyr108 side chain (TM3). The hydrophobic side chain of leucine is found in a hydrophobic pocket of residues in TM3, TM6, and TM8.

The LeuT structure also yielded the positions of two Na^+ ions bound at the core of LeuT and have a role in stabilising the core, along with the unwound sections of TM1 and TM6 and the leucine substrate. One Na^+ binds in the Na1 site, and the other in the Na2 site. The Na1 site is formed of Na^+ bound in an octahedral arrangement with the leucine solute, and residues in TM1, 6, and 8. The Na2 site is formed of Na^+ coordinating with two carbonyls from TM1, one carbonyl from TM8, and using a Thr354 and Ser355 hydroxyl side chain from TM8.

The intimate interactions of Na^+ and leucine substrate suggest that solute and co-ion are coupled, with the suggestion that Na^+ binding is required to organize the solute binding site as the coordination of the two Na^+ is required to stabilise TM7 and TM8 which in turn stabilise TM1 and TM6.

The occluded nature of this structure means that the extracellular and intracellular gates are simultaneously closed. Closer inspection shows that access to the extracellular from the binding sites is hindered by a Tyr108 and Phe253. Phe253 is part of interactions to hold LeuT in the binding pocket, and through neighbouring Thr254 is also connected to the Na1 site. Phe253 is also forming a cation-π interaction with Arg30 in TM1 (a highly conserved residue), Arg30 is forming a H_2O mediated salt bridge with Asp404 in TM10 (also highly-conserved). This is believed to be part of the extracellular gate, and begins to show how the substrate, Na1 site and extracellular gate are coupled together.
A second highly conserved pair of residues form a salt-bridge at the intracellular face of the protein are predicted to be the intracellular gate.

LeuT has substrate flexibility and through multiple crystal structures and complimentary functional data show that amino acids Gly, Ala, Leu, Ile, Tyr, and Met can all bind and be transported (Singh *et al.*, 2008). A structure of LeuT in complex with Trp, forced LeuT into an outward-facing conformation, and functional studies showed that Trp was a competitive inhibitor of LeuT and prevented the transport cycle to occur. The LeuT and Trp complex shows that Trp108 does not interact with Trp due to steric hinderance as a consequence of the larger volume of Trp; this results in LeuT being solvent accessible and opening up the binding pocket

of about 3 Å. There is a second Trp molecule bound in the "S2 site" that disrupts the salt bridge between Asp404 and Arg30, with the suggestion that this is a second low-affinity binding site for substrate to move into, be dehydrated and then move into the primary binding site. This is a plausible idea, however, when 30 mM leucine was co-crystallised with LeuT there was no density at this S2 site. This S2 site is an area of controversy, Molecular dynamic (MD) simulations, Förster Resonance Energy Transfer (FRET) experiments and mutagenesis suggest that this S2 site is biologically relevant and that substrate allosterically binds in the S2 site to induce release of substrate (Zhao *et al.*, 2010, 2011). Detergent is also capable of binding in this S2 site (Quick *et al.*, 2009). The supposed importance of the S2 site doesn't correlate with crystal structures and most importantly transport data doesn't either.

Before any other conformations were discovered Forrest and colleagues modelled LeuT into an inward-facing conformation by exploiting the internal symmetry of LeuT. Rotation of the four-helix bundle relative to the scaffold domain enables access to the extracellular or intracellular side (Forrest *et al.*, 2008).

In 2012, two structures of LeuT were released where LeuT was mutated and stabilised with antibodies to form an outward-open and inward-open conformation (Krishnamurthy and Gouaux, 2012). LeuT Trp108Phe mutant adopts an outward-open state, that is substrate-free but still bound to two Na^+, TM1b, TM2 and TM6a are able to pivot allowing LeuT to adopt an outward structure. It appears that substrate interactions provide restraints on TM1b and TM6b. In this structure, the "thin" gate formed by the Tyr108 and Phe258 is disrupted due to the mutation forcing an opening to the extracellular solution and consequently, Arg30 and Asp404 are no longer able to form a water-mediated salt bridge. Arg30 and Phe253 have a coordinated movement allow them to preserve their cation-π interaction, interestingly the phenyl ring of Phe253 has now moved into the position of the second Trp molecule in the LeuT Trp complex. Na^+ is still found in both the Na1 and Na2 site, and it appears that the presence of the sodium ions is able to hold the intracellular "thick" gate closed by coordinating the "core" and "scaffold" domains. This structure seems to suggest that sodium binding precedes substrate binding.

LeuT with mutations to disrupt the Na2 site forced LeuT into an inward-facing conformation. The most obvious difference is the placement of TM1a, which is tilted 45 ° relative to the plan of the membrane, TM6b moves independently and rotates away by about 17 °. TM1b and TM6a tilt to block the extracellular and Extracellular Loop (EL) 4 moves into the extracellular cavern to block the extracellular passage.

Leucine binding to an outward-open structure induces the extracellular ends of TM1 and 6 to pivot around their unwound sections like a pivot of about 9 °, allowing the extracellular gate to form to occlude the substrate. EL4 moves into the extracellular cavity. The intracellular section of TM1 is able to move independently and bends at a pivot point above the Na2 site allowing sodium release and determining how the release of sodium and opening of the transporter to the cytoplasm is linked. Disturbance of the Na2 site enables leucine release. (Figure 1.8)

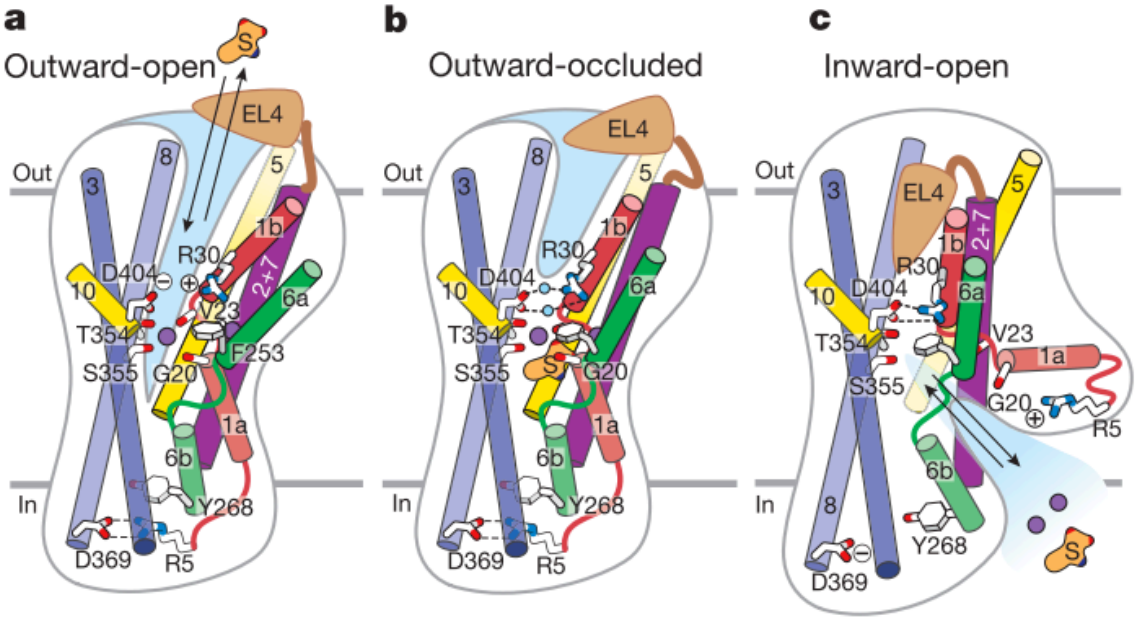


Figure 1.8. Schematic of transport in LeuT. The structural changes of LeuT as it cycles from outward-open (a) to outward-occluded (b) and inward-open (c). This figure has been taken from (Krishnamurthy and Gouaux, 2012).

LeuT is similar to the human neurotransmitter transporters, SERT, and has been shown to bind selective serotonin reuptake inhibitors (SSRI's). Functional studies showed that these SSRI's are non-competitive inhibitors of transport, whilst the SERT was known to be competitively inhibited by SSRI's and therefore Gouaux has

suggested that whilst LeuT is useful for studying how these proteins are able to transport it should not be used as a model for how SSRI's bind. Crystal structures with Tricyclic Antidepressants (TCA) (one category of SSRI's) bound have been released and show that these TCA molecules bind above the substrate binding pocket, at the proposed S2 site, and obstruct the Arg30 and Asp404 interactions and therefore the extracellular gate is unable to form properly and consequently transport cannot happen. However, data published is contradictory on which SSRI's are able to bind and to what extent they inhibit transport (Singh, Yamashita and Gouaux, 2007; Zhou *et al.*, 2007, 2009).

1.4.2. MhsT

MhsT, another member of the NSS family from *B. halodurans*, was crystallised in an inward-occluded conformation with sodium ions in the Na1 and Na2 sites in complex with a tryptophan molecule (Malinauskaite *et al.*, 2014) using vapour diffusion and lipidic cubic phase (LCP). The substrate binding site and Na1 site are similar to the leucine binding site, i.e. coordinated by residues from TM1, 3, 6, and 8, and Na1 site in LeuT. In contrast to other LeuT-fold transporters the Na2 site is coordinated by 6 ligands compared to 5. As before, this site uses 2 carbonyl groups from TM1, a carbonyl from TM8 and 2 hydroxyl side chains from Ser323 and Ser324, with a water molecule as the sixth ligand. Unlike the inward-open conformation of LeuT, the intracellular gate is closed and is more similar to the outward-facing states. TM5 is unwound in this structure allowing water access to the Na2 site form the intracellular space.

From these structures it seems that extracellular closure is not as intimately linked with the opening of the intracellular cavity TM5 is unwound in the MhsT structures forming a solvent pathway allowing water permeation from the intracellular space to the Na2 site. A conserved $GlyX_9Pro$ motif is thought to be responsible for unwinding of TM5. Mutations in the helix-breaking motif in MhsT reduce the ability for MhsT to transport whilst having no effect on binding. Hydration of the Na2 site enables further Na^+ hydration and release into the cytoplasmic space.

Comparing the LCP structure to the vapour diffusion shows that the LCP structure is capable of better crystal packing and that TM5 forms a proper helix with some high

structural flexibility. The authors suggest that TM5 is capable of forming a complete helix upon substrate and sodium release. (Figure 1.9)

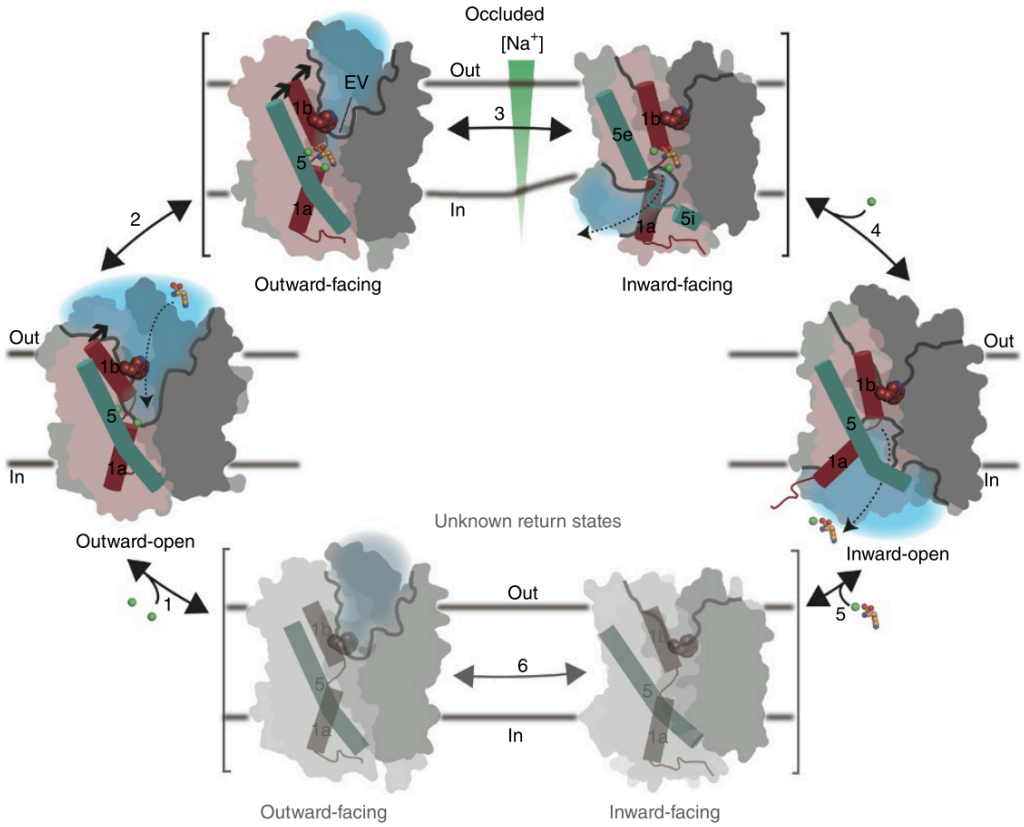


Figure 1.9. Transport cycle of MhsT. Na^+ binding at the Na1 and Na2 sites stabilizes the outward-open state (1) and allows the substrate to bind triggering substrate site occlusion (2). Closure of the hydrophobic extracellular pocket facilitates transition into an inward-facing state in which TM5 unwinding provides a solvation pathway for the Na2 site (3) and an opportunity for Na^+ to escape (4). Na^+ release from Na2 allows TM5 to reform and TM1a to swing out to release the substrate with Na^+ from Na1 in the inward-open state. Finally, the transporter switches back to the outward-open state through yet-unknown occluded return states (5 and 6). Scaffold domain is found in dark grey, with the bundle in pink. Solvent accessible regions are depicted in blue with green spheres representing Na^+ and substrate in orange. This figure has been taken from (Malinauskaite *et al.*, 2014).

Computational docking of this structure showed that a second tryptophan molecule can bind into a site that is equivalent to the S2 site of LeuT, in the crystal structure of MhsT this extracellular cavity is collapsed. Functional binding studies showed that the detergent choice of MhsT affected the stoichiometry of substrate: protein binding, and the concentration of detergent was also critical (Quick *et al.*, 2018).

1.4.3. dDAT

In 2013, the first Eukaryotic LeuT-fold was structurally determined in complex with an antidepressant inhibitor, the dopamine transporter from *D. melanogaster*

(Penmatsa, Wang and Gouaux, 2013). dDAT required 5 single point mutations and fragment antigen-binding (Fab) stabilisation; was found in complex with nortriptyline, 2 sodium ions, a chloride ion, and a cholesterol molecule in an outward-open conformation.

Nortriptyline was found at the interface of TM1, 3, 6, and 8 with residues in TM10 also interacting with the drug, nortriptyline binding prevents the formation of the occluded state by sterically hindering the extracellular gate. Densities for sodium ions aligned with Na1 and Na2 in LeuT, the Na1 site is hydrated by a water molecule acting as a ligand. The Na2 site is found towards to the cytoplasmic face of dDAT and is coordinated with carbonyl groups of residues in TM1 and 8 with the hydroxyl side chains of Asp420 and Ser421 on TM8.

Cholesterol binds in a cavity between TM5, 7, and 1a, if dDAT was to exhibit the same conformation of TM1a as LeuT in the inward-state then this cholesterol site would be disrupted, Gouaux and colleagues suggested that this cholesterol would stabilise dDAT in the outward facing conformation priming dDAT for ion and solute binding.

1.4.4. SERT

SERT, the human serotonin transporter, has been structurally characterised by both x-ray crystallography and single particle electron microscopy. SERT required antibodies fragments to facilitate reconstruction. SERT was found in complex with ibogaine, a non-competitive inhibitor and presumably bound to 2 Na^+ and a Cl^- in outward-open, outward-occluded, and inward-open states (Coleman, Green and Gouaux, 2016; Coleman *et al.*, 2019).

In the outward-open structure cholesteryl hemisuccinate (CHS) can be found bound adjacent to TM1a like dDAT with presumably the same stabilisation effect, this cannot be seen in the inward-open structure.

Due to the low-resolution of these structures it would not be possible to conclusively identify the sodium and chloride sites. However, due to the wealth of knowledge of the sodium sites in structural homologues discussed earlier, the Na1 and Na2 site can be postulated with a reasonable degree of confidence. The highly conserved Na2 site is predicted to be found at the interface of TM1 and TM8 as all other LeuT-fold

sodium symporters, these Na2 site is formed of 2 carbonyl groups from TM1, a carbonyl from TM8 and the side chains of Ser438 and Asp437 in TM8.

Comparing these 3 structures a model of transport can be built. Formation of the occluded state requires the movements in TM1b, 5 ,6, 7, and 10. Moving from the occluded structure to the inward-open structure, TM1b shifts and tilts by 22 ° whilst TM1a moves by 40 °, movement and bending of TM1 is coordinated with the unwinding of TM5 at a conserved GlyX$_9$Pro motif, as seen in MhsT, the shift of TM5 allows sodium in the Na2 site to exit SERT and enter the cytoplasm (Figure 1.10) illustrating how intimately coupled sodium in the Na2 site and conformational changes are.

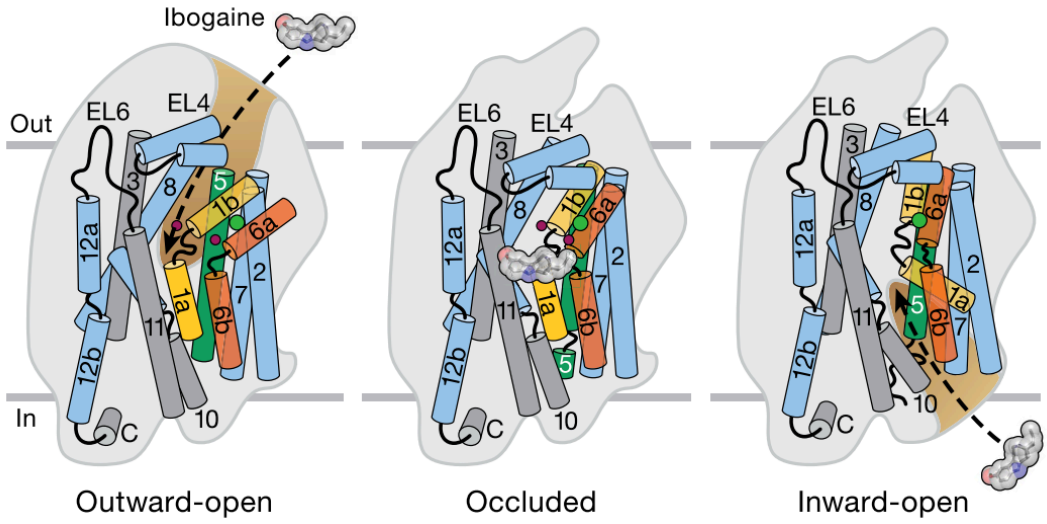


Figure 1.10. Transport model of SERT and ibogaine. Illustration of the conformational changes of SERT to transport ibogaine. Sodium and chloride ions are shown in red and green respectively. This figure has been taken from (Coleman *et al.*, 2019).

1.4.5. vSGLT

vSGLT is a sodium:galactose symporter from *V. parahaemolyticus* in the SSS family. This was the second (unexpected) member of the LeuT superfamily. vSGLT also has the LeuT fold but has 14 helices with an extra helix at the amino-terminus (n-terminus) and carboxyl-terminus (c-terminus), for ease the additional N-terminal helix will be referred to as TM-1. vSGLT was found in an inward-occluded and inward-open conformations bound to galactose and sodium (Faham *et al.*, 2008) and free of ligand and co-ion (Watanabe *et al.*, 2010). Like LeuT, vSGLT has “thin gates” made of amino acids.

Galactose is bound using residues from TM1, 2, 6, 7, and 10 and uses a π-stacking interaction to hold galactose in the binding site, then hydrogen bonds by local residues to orientate substrate. The authors identified a plausible Na^+ site between TM1 and 8 and is equivalent to the Na2 site in LeuT. Once again, this site uses two carbonyl groups from the unwound section of TM1, a carbonyl from TM8, and hydroxyl side chains from Ser364 and Ser365. Conserved Ser365 is proposed to interact with the Na^+ site and a Ser265Ala mutant abolished transport. This Ser365 residue is equivalent to Thr341 in the proline transporter (PutP) and is known to be essential for Na^+ binding (Jung, Hilger and Raba, 2012). The authors identified that both the hash-motif and 4-helix bundle rotated 3 ° away from each other with a 13 ° kink in helix 1 when moving from the inward-occluded and inward-open structure. This opens up the intracellular cavity disrupting sodium and substrate binding site. Based on the observation that the Na2 site is conserved regardless of the stichometry of ions this Na2 site appears to be the site that regulates sodium and substrate release.

MD simulations of vSGLT suggest that Na^+ does not remain in the binding site for long after the conformational change into the inward-facing structure and the Na^+ interacts with Asp189 (TM5) upon exit. As Na^+ leaves vSGLT the intracellular gate Tyr263 is able to exhibit a new rotamer conformation, as a consequence opening the intracellular cavity to solvent and galactose is able to leave vSGLT. If Na^+ was lightly restrained in the binding site then Tyr263 never adopts this alternative conformation. Tyr263 interacts with Asn64 in the unwound section of TM1 near the Na^+ binding site and directly interacts with galactose; demonstrating a clear link of how Na^+, galactose, and the intracellular gate are linked.

1.4.6. SiaT

SiaT is a sodium-dependent sialic-acid transporter from *P. mirabilis* crystallised to 1.95 Å and is a another member of the SSS family (Wahlgren *et al.*, 2018). SiaT is found in an outward-open conformation and is composed of 13 helices, TM0 is found at the N-terminus, with TM1-10 folding in the LeuT fold with TM1 and TM6 unwinding at the substrate binding site. The solute binding site is formed via 8 residues from TM1, 2, 3, and 6 binding sialic acid N-acetylneuraminic acid (Neu5Ac) with 7 water molecules coordinating substrate binding.

2 Na^+ were modelled in SiaT with a sodium ion in the conserved Na2 site formed of 2 carbonyls from the unwound section of TM1, a main-chain carbonyl in TM8, and the hydroxyl side chains of Ser342 and Ser343.The authors observed a second sodium ion in a unique binding site that is different from Na1 in LeuT they termed Na3. This Na3 site is close to Na2 and notably not in contact with the substrate. Mutations in the Na2 site abolished transport, whilst mutations in the Na3 site produced a variety of results, some mutations abolished transport, some increased, and some had little to no effect. Based on sequence alignments it appears that this Na3 site could be found in 19.6% of SSS.
MD simulations demonstrated that Neu5Ac was stably held in the binding site when both sodium ions were bound; removal of Na^+ from the Na2 site destabilises Nau5Ac binding. Simulations in the Na2 site show that the ion is stable in Na2 in the presence of substrate and is more mobile upon removal of substrate. The ion in Na3 appears to be independent of solute binding.

SiaT was modelled onto vSGLT to simulate an inward-facing state. Moving from an outward-facing state to the inward-facing state, TM10 moves over the binding pocket to interact with TM1 and TM2, closure forms an outer gate above the solute binding pocket composed of hydrophobic residues in TM1, TM2 and the TM9-TM10 loop. The inner is formed of a conserved Arg3, Arg44, and Glu176 forming salt bridges, along with Arg40 (corresponding to Arg5 in LeuT), Arg101, and Ile105 interacting. To open the inner gate, all of these interactions are disrupted whilst TM8 and TM9 move away from the intracellular pocket. Arg260 (corresponding to Tyr265 in LeuT) is located below the Na3 site interacting with TM1, TM6, and TM8, and it has been suggested that this residue may prevent premature Na^+ release from the Na3 site.

1.4.7. BetP

BetP from *C. glutamicum* has been crystallised in a variety of conformations (Figure 1.12) and is a member of the betaine/choline/carnitine transporter (BCCT) family (Kappes, Kempf and Bremer, 1996; Ressl *et al.*, 2009; Ziegler, Bremer and Krämer, 2010; Perez *et al.*, 2012). BetP was found in complex with betaine substrate as a functional trimer and known to use $2Na^+$ for transport. BetP exhibits the LeuT-fold

of 10 TM helices arranged in the two-fold pseudosymmetry but has the two additional helices (TM11 and 12 in LeuT) at the N-terminus, for consistency these helices will be donated TM-2, -1 respectively and TM3-TM12 will be acknowledged as TM1-10. At the C-terminus BetP has a helix that extends into the cytoplasm able to sense cytoplasmic K^+, transport of betaine is coupled to osmotic stress that is sensed by this C-terminal domain. Density for betaine was observed in a Trp box consisting of Trp side chains from TM2 and 6; however, a conserved GxGxG motif in the unwound section of TM1 is essential for allowing transport to happen by providing flexibility to TM1. Sodium was not visible in the structure due to the low resolution; based on structural alignment with LeuT, Na1 was predicted to be interacting with the betaine substrate and TM1, with the second Na^+ coordinated with residues from TM1, 6, and 8. However, when MD simulations were performed sodium did not stay in either putative binding site, and mutagenesis of the residues involved did not change sodium dependence/uptake.

A mutation in the flexible hinge of TM1 changed the specificity of BetP and allowed choline to bind instead. and most importantly electron density for a Na^+ could be seen at the conserved Na2 site in a occluded betaine bound state (Perez *et al.*, 2012). This Na2 site was coordinated via two carbonyls from TM1, a carbonyl from TM8, and hydroxyls from Thr467 and Ser468; when subjected to MD simulations this sodium ion stayed in the binding site. This, coupled with mutations of the conserved Thr/Ser reducing the ability of sodium to bind and betaine transport to occur begins to confirm that this is probably the Na2 site. The authors suggested that the second sodium ion binds at a site of the symmetry equivalent of Na2 site and is bound between TM3, and 6. From MD simulations this alternative Na1' site also required a H_2O molecule to act as a ligand to stabilise this site and mutations in this binding site did affected the ability of BetP to bind sodium and is only able to be coordinated in an occluded conformation. The authors also observed sodium binding cooperativity within a single protein chain, and that this cooperativity can be abolished by a mutation of the Ser368 residue in the Na2, one idea behind this cooperativity is that the bundle and hash-motif remove relative to each other when switching conformations, by sodium binding in one of these sites, found at the interface between the bundle and hash-motif, could stabilise the other site.

Like in LeuT and vSGLT the inward-open structure has a disrupted Na2 site and optimal ligand binding could only be achieved when BetP was found in a “closed” transition state between the inward and outward states. Interestingly, residues involved in ligand binding were found in identical conformations for the closed substrate-free and closed substrate-bound. This would mean that the conformational landscape for this is very flat and would allow BetP to cycle back to outward-facing in the absence of substrate (Khafizov *et al.*, 2012).

Binding of sodium in the Na2 site stabilises the transporters to stay in the outward-open state allowing substrate to bind. Substrate binding induces the occluded state, and stops the Na^+ from leaving the Na2 site. Closure of TM5 induces a better fit for the Na1’ in BetP, and molecular dynamics simulations show it is still solvent accessible in the outward-open substrate bound state. BetP uses TM5 and TM10 to act as the intracellular and extracellular “thin” gates. Binding of sodium in the Na2 site is coupled to the opening and closing of TM5, whilst binding in the Na1’is coupled to the opening and closing of TM10. BetP forming a closed structure is evidence that these “thin” gates are able to move independently of each other.

Upon transition from the outward structure to the inward, TM1 bends at the discontinuous section as in LeuT, and TM1a is titled by about 15 °. This movement is not as severe in LeuT, and is more comparable to vSGLT. BetP undergoes a rigid body movement of the hash-motif with respect to the 4-helix bundle of about 13 °, LeuT undergoes a rigid body rotation of about 10 °. The authors suggest that position of sodium sites dictate the extent of helix flexing contributing to mechanism compared to gating hinge motions. All of the conformations of BetP are summarised in Figure 1.11.

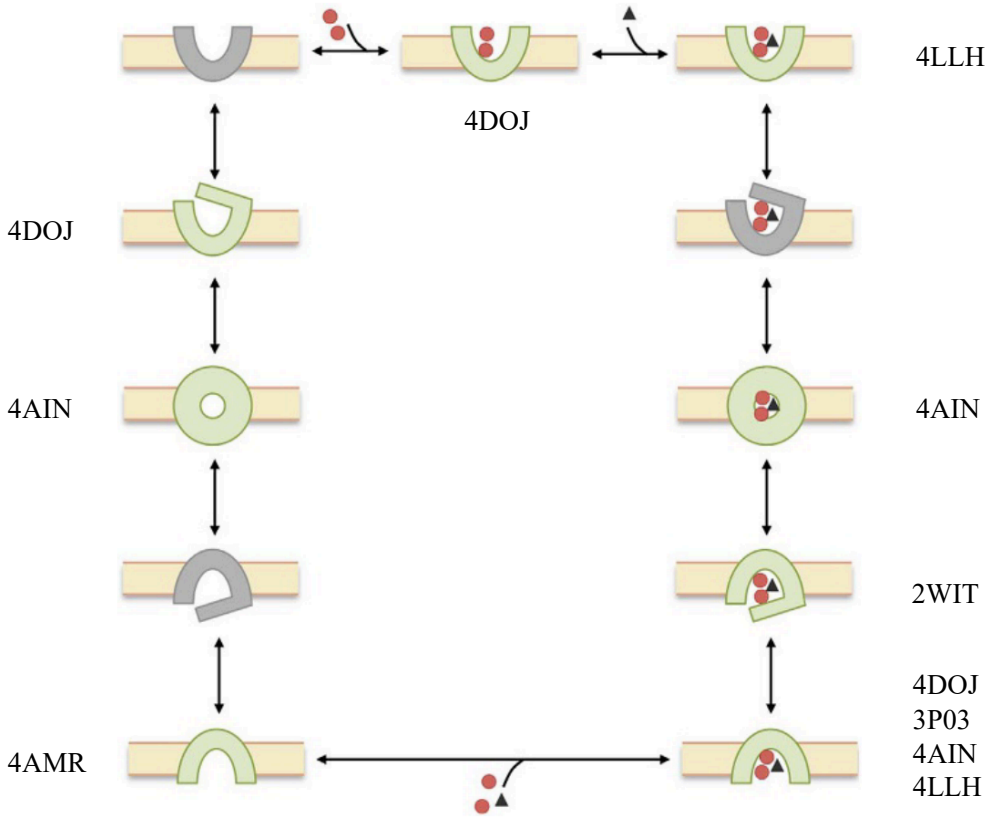


Figure 1.11. Schematic of the eight conformations of BetP structurally characterised. This figure has been taken and adapted from (Perez *et al.*, 2014) with corresponding PDB codes.

1.5. NCS1 family

The NCS1 family of transporters consists of over 2500 sequenced proteins across Gram-positive, and Gram-negative bacteria, along with archaea, yeast, fungi, and plants (Ahmed, 2017; Elbourne *et al.*, 2017). Members of this family are characterised as secondary transporters that use Na^+ or H^+ as the co-ion and nucleobases, nucleosides, hydantoins, and other related compounds as the solute. These proteins are often involved with salvage pathways, and substrates can be used as nitrogen sources or in the case of nucleobases and nucleosides built into Deoxyribonucleic acid (DNA) or Ribonucleic acid (RNA). There are three bacterial members of this family that have been previously characterised. Mhp1 from *M. liquefaciens* (Suzuki and Henderson, 2006; Weyand *et al.*, 2008; Shimamura *et al.*, 2010; Weyand *et al.*, 2011), CodB, from *E. coli* (Danielsen *et al.*, 1992), and PucI from *Bacillus subtilis* (Ma *et al.*, 2016). Mhp1, CodB, and PucI are highly specific for their substrate whilst the plant proteins display more promiscuity. Despite advancement in understanding substrate coordination by Mhp1 structures, it is still

not possible to predict substrate specificity from sequence alone. The sodium binding site in Mhp1 is poorly conserved across the NCS1 family and so structural determination of a different family member would be useful to understand cation coordination and the role this plays in solute transport.

1.5.1. Mhp1

Mhp1 from *M. liquefaciens* allows the uptake of indolyl methyl- and benzyl-hydantoins in a Na^+ dependent manner and was the first secondary transporter to be captured in multiple conformations via high-resolution macromolecular X-ray crystallography (Suzuki and Henderson, 2006; Weyand *et al.*, 2008; Shimamura *et al.*, 2010; Weyand *et al.*, 2011; Simmons *et al.*, 2014) . Indolyl methyl-hydantoin (IMH) and benzyl-hydantoin (BH) are metabolic intermediates for L-tryptophan, phenylalanine, and tyrosine. Mhp1 has been captured in an outward-open, occluded, and inward-open conformation (Weyand *et al.*, 2008; Shimamura *et al.*, 2010; Weyand *et al.*, 2011) comparing these conformations have provided insight into how these transporters work. Identifying the conformational changes of Mhp1 has been instrumental in providing the alternating access model and the mechanism of ion coupling. Mhp1 consists of 12 transmembrane helices.TM 1-10 are arranged in a pseudo two-fold symmetry where TM 1-5 are equivalent to 6-10, TM11 and 12 are separate and are not well conserved across the NCS1 family. Mhp1 can be broken down into structural domains, a “4-helix bundle” of TM1, 2, 6, and 7 and a “hash” motif of TM 3, 4, 8, and 9 (Figure 1.12). The outward-open structure of Mhp1 is substrate-free but is Na^+ bound, indicative that sodium binding precedes substrate binding with an extracellular solvent accessible cavity and TM10 is relatively straight (Weyand *et al.*, 2008).

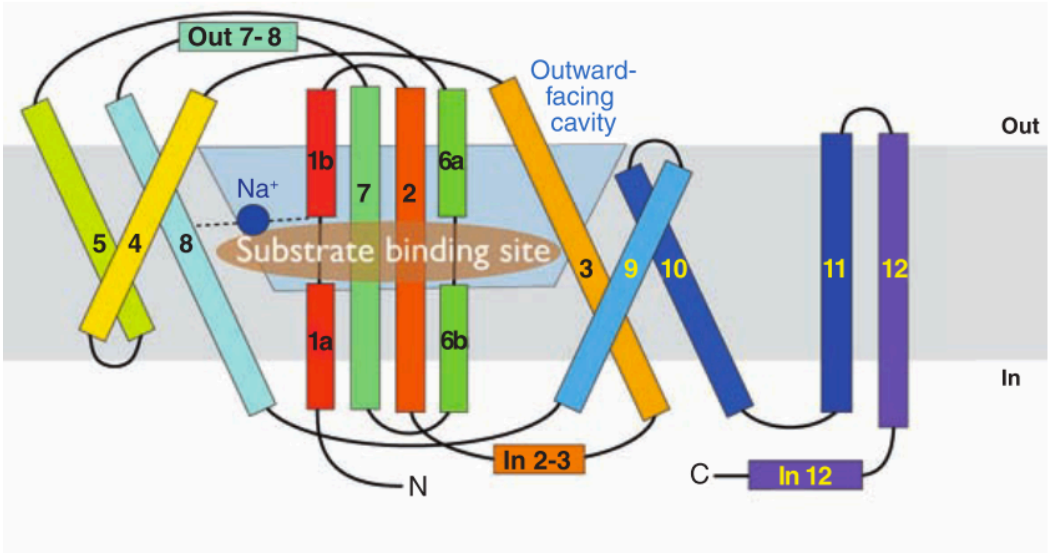


Figure 1.12. Mhp1 topology. Transmembrane helices 1-10 can be broken down into structural components, the 4-helix bundle of TM1,2,6,7, the hash motif of TM3,4,8,9, and with TM 5,10 as flexible helices that are termed "thin gates". TM3 and TM8 pack onto each other in 3D space. This figure was taken from (Weyand *et al.*, 2008).

Functional studies investigating hydantoin binding and/or transport and crystal structures of Mhp1 in complex with IMH and other hydantoin derivatives have elucidated the residues involved (Simmons *et al.*, 2014). IMH is found between the hash and bundle motifs with the hydantoin moiety interacting predominantly with the hash-motif. This hydantoin ring is found parallel to Trp117 (TM3) forming a π-stacking with the indole side-chain, a conservative mutation of Trp177Phe reduced IMH uptake, whereas a Trp117Ala mutation appeared to completely abolish transport showing how important this residue is. Trp220 (TM6) forms a π-stacking interaction by packing perpendicular to the indole moiety, interestingly this residue can be substituted to Ala and ~80% of transport is retained. Residues Gln121 (TM3), Gly219 (TM6), and Asn318 (TM8) provide the hydrogen bonds required to orientate the IMH correctly. Gln121 forms a hydrogen bond network with Gln42 (TM1) and the hydantoin functional group, with the theory that Gln42 is required to keep Gln121 in the correct position to form the hydrogen bonds. Functional knockout of Gln42 significantly reduces Mhp1's ability to transport substrate. Gly219 is found in the discontinuous helix of TM6 and utilises its carbonyl group to hydrogen bond with the indole ring. Mutagenesis of glycine into larger side chains reduces the hydantoin uptake. In crystal structures, Asn318 is unambiguously forming good hydrogen bonds with the hydantoin with complimentary data showing that even a functional conservative mutation of Gln318Asp knocks transport ability to almost

50% of wt. Mhp1 is also able to bind 5-(2-naphthylmethyl)-L-hydantoin (NMH) but is not able to transport it. From the crystal structure it would seem that NMH is an inhibitor as it holds Mhp1 in an outward-open conformation, whilst the mutation of Leu363 (TM10) into Ala allows Mhp1 reinstates transport solidifying this hypothesis. Leu363 and NMH sterically hinder and prevent TM10 from closing. From assessing the functional and structural data it would suggest that TM10 must be closed to allow transport. A close-up of the active site and summation of mutagenesis results is shown in Figure 1.13.

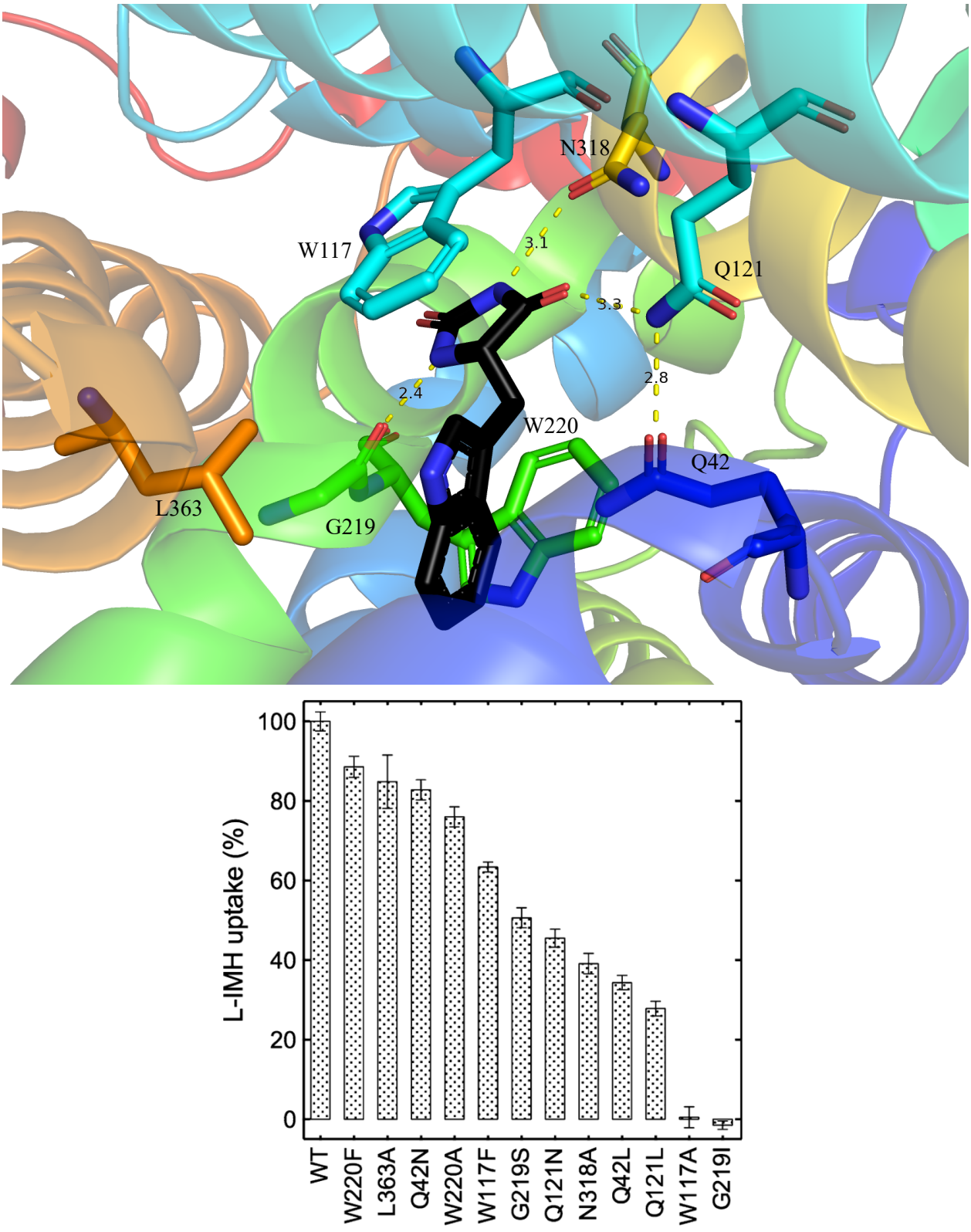


Figure 1.13. Residues required for coordination of IMH in the Mhp1 binding site. Molecular interactions of Mhp1 and IMH (top), IMH is held in the binding site by π-stacking with Trp117 and Trp220 with additional stabilisation and specificity from interactions with Gln121, Gly219, and Asn318. PDB code: 4D1A. Inhibition of IMH uptake for Mhp1 mutants (bottom) was taken from (Simmons *et al.*, 2014). Uptake of 14C-L-IMH was measured for 15 s in *E. coli* expressing Mhp1.

Before the crystal structure of Mhp1 was known it was suspected to be a proton symporter (Suzuki and Henderson, 2006), due to its sequence differences from other members of the functional family. The sodium binding site of Mhp1 is conserved with the Na2 site of LeuT and is found in all sodium-coupled symporters of the LeuT-fold; however, the role of sodium ion binding and how this is coupled to transport is unclear. The Na^+ binding site is thought to reside at the interface of TM1 and TM8, holding the "hash" motif and 4-helix bundle together. Na^+ coordinates in a square pyramidal arrangement with carbonyl groups of Ala38 (TM1), Ile41 (TM1), and Ala309 (TM8), along with Ser312 (TM8) and Thr313 (TM8) side chains. It is postulated that the Gln42 and Asn318 could be the residues that couple solute and co-ion. From functional studies, hydantoin and sodium are intimately coupled as the presence of hydantoin increases the affinity of Mhp1 for sodium (Weyand *et al.*, 2008).

Based on crystal structures, in the inward-facing structure both the hydantoin and sodium binding sites are disturbed and neither substrate or sodium are bound (Shimamura *et al.*, 2010); when compared to the outward-occluded and outward-open structures TM10 is clearly bent whilst TM5 is straight (Weyand *et al.*, 2008, 2011). Work by Calabrese and colleagues conclude that Mhp1 is capable of binding to BH substrate in the absence of sodium whilst Mhp1 is in the inward-facing state (Calabrese *et al.*, 2017). Figure 1.14 and Figure 1.15 shows how residues interacting with IMH and Na^+ are positioned in the outward-open, occluded, inward-open structures.

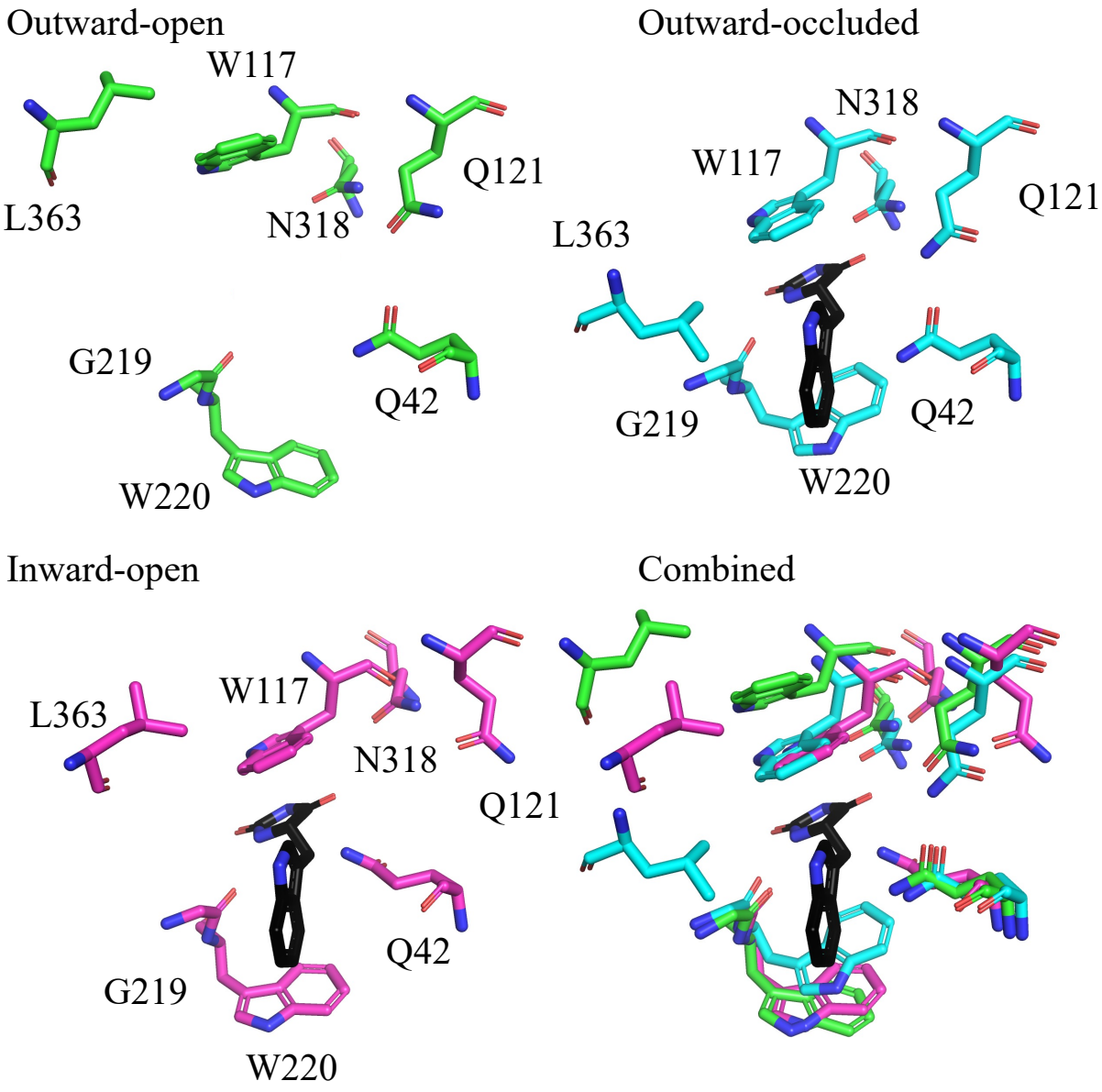


Figure 1.14. Comparison motions in the substrate binding site of Mhp1. In the outward-open (green), occluded (blue), and inward-open (pink) structures of Mhp1 with IMH coloured in black. IMH is the unmoving reference point. IMH is not bound to Mhp1 inward-open structure but pictured for reference. In the outward-open structure Trp117 and Trp220 are not in the prime position to interact with IMH, and move in towards to IMH in the occluded structure, Trp220 moves away from IMH in the inward-open structure. The positions of Asn318 and Gln121 are relatively unchanged in the outward-open structure and occluded structure, but in the inward structure Gln121 pull away from IMH. Leu363 undergoes the largest conformational change to the bending of TM10, In the outward structure TM10 is straight so Leu363 is far away from the binding site, bending of TM10 to produce the occluded structure places Leu363 into the IMH binding site, in the inward-open structure Leu363 moves away from the binding site.

The current idea of the alternating access model is as follows. Mhp1 is able to be conformationally switching between outward-open and inward-open, however purified Mhp1 was shown to be residing in an inward-facing conformation and only moving into an outward-facing state in the presence of both sodium and BH substrate (Calabrese *et al.*, 2017). Binding of sodium between the hash-motif and bundle domain induces stabilisation into the outward-facing conformation and primes the substrate binding pocket. Moving from the outward-open structure to the outward-occluded structure, binding of hydantoin induces the conformational shift with Trp117 and Trp220 rotating to pack better with the substrate. From molecular dynamics it seems that TM10 is always transiently moving between "open" and "closed" states, and upon the correct orientation of substrate is able to be held into the closed state. Binding of hydantoin induces TM10 to bend over the substrate binding pocket and transport is only possible when TM10 is in the closed state. Moving from the occluded to the inward-open structure rotation of the hash-motif relative to the bundle and straightening of TM5 disrupts Na^+ and hydantoin binding by making the cavity for Na^+ too large and moving Trp117 into the hydantoin binding site whilst moving Gln121 and Asn318 away. Molecular dynamics simulations suggest that sodium unbinds then hydantoin.

Rotation of the 4-helix bundle relative to the hash motif impairs the sodium binding site. Sodium is found at the interface of TM1 (4-helix bundle) and TM8 (hash motif), TM8 moves 4.5 Å away from TM1 in the inward structure resulting in Ala309, Ser312, Thr313 too far away from the Na^+ to coordinate.

The major conformational shift of Mhp1 is the rigid body rotation of the "hash" motif relative to the "4-helix bundle" of 30 ° whilst TM5 and TM10 bend and act as thin gates as depicted in Figure 1.16 whilst the hash-motif acts as a "thick gate". In contrast to other members of the LeuT superfamily Mhp1 does not have movement of TM1 and TM6 as part of its transport mechanism.

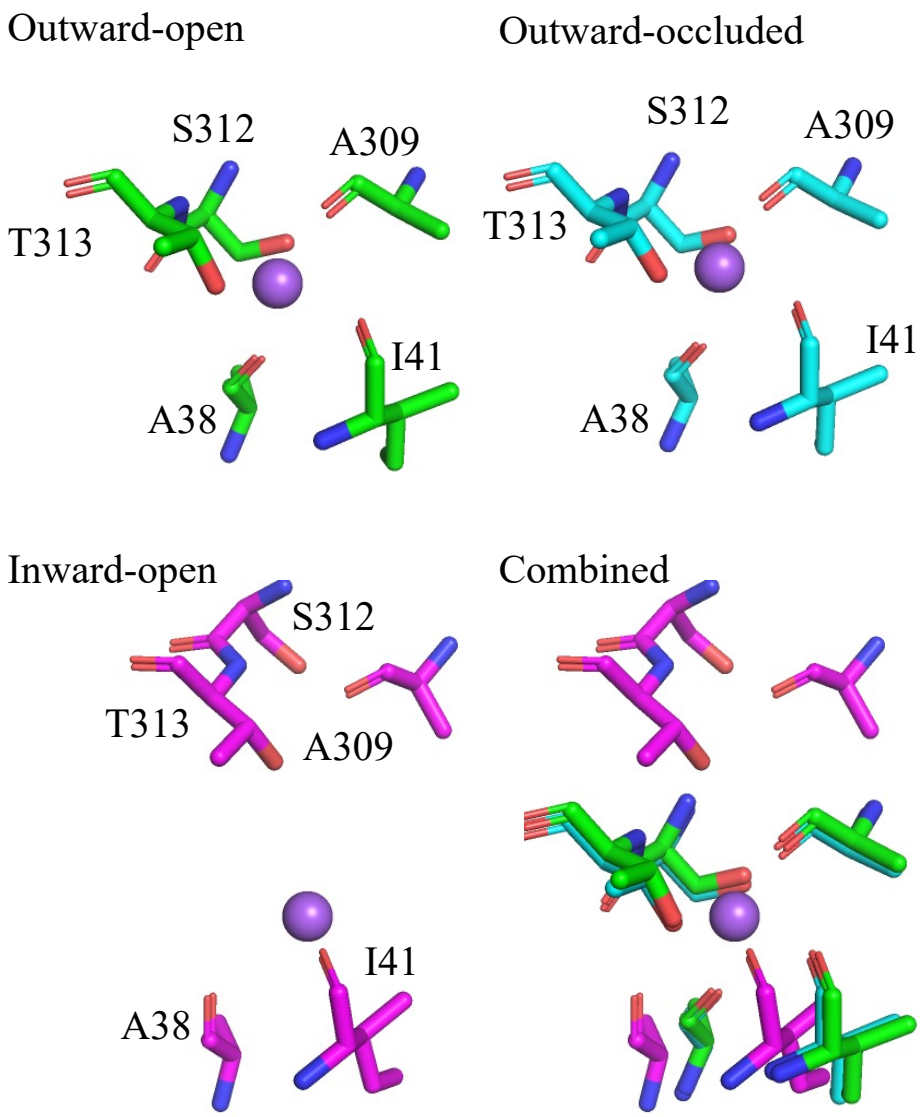


Figure 1.15. Comparison movements in the sodium binding site of Mhp1. In the outward-facing (green), occluded (blue), and inward-facing (pink) structures of Mhp1 with Na^+ coloured in purple. The Na^+ is the unmoving reference point. Na^+ is not bound to Mhp1 inward-open structure but pictured for reference. In the outward-open and occluded structures the sodium site is identical, however, in the inward-structure sodium is unable to bind due to the movement of Ala309, Ser312, and Thr313 means they are not in a position to coordinate with sodium as they have been moved away from the binding site.

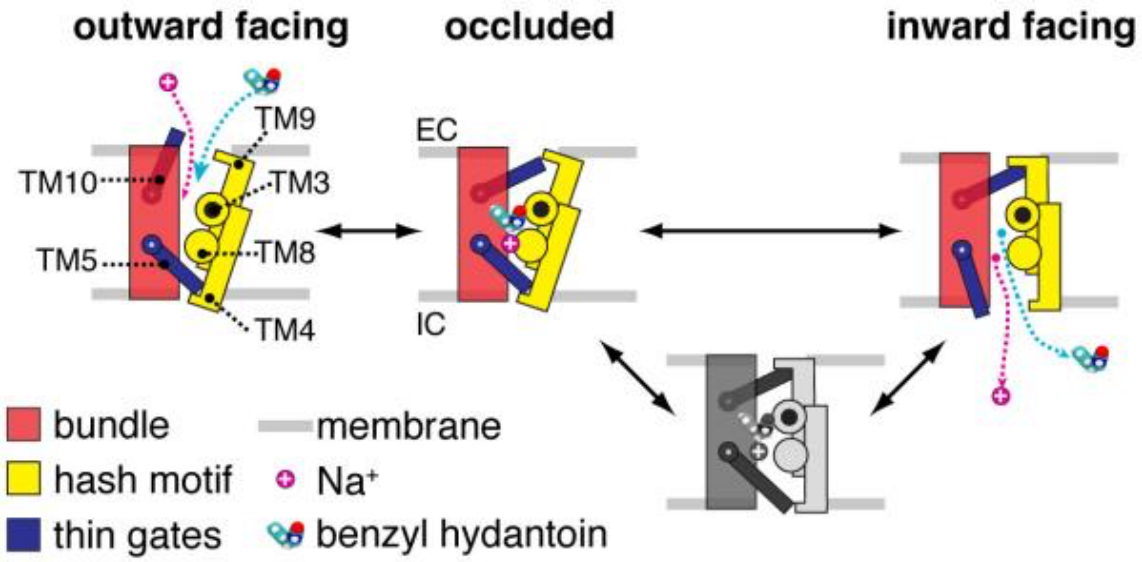


Figure 1.16. Schematic representation of the alternating access model of Mhp1. Coloured diagrams represent the outward-facing, occluded, and inward-facing crystal structures. The black and white diagram shows a theoretical state that be an intermediate between the inward and occluded structures. This figure has been taken from (Shimamura *et al.*, 2010). EC=Extracellular space, IC=Intracellular space.

Whilst Mhp1 has furthered understanding of secondary transport it has been unable to answer all questions, particularly those revolving around the role of Na^+ binding and how this results in the transport of substrate across the membrane.

1.5.2. CodB

CodB is another bacterial member of the NCS1 family and the LeuT superfamily and has been annotated as a cytosine transporter in *E. coli* (Danielsen *et al.*, 1992). CodB is conserved across gram-positive and gram-negative bacteria. CodB is found in an operon with CodA a characterised cytosine deaminase; this operon is regulated by nitrogen and purine with an increase in transcription in the absence of nitrogen. *E. coli* can utilise cytosine as a source of nitrogen by using CodA to deaminate cytosine into uracil and ammonia. CodB exhibits between 15-24% sequence identity to other characterised members of the NCS1 and evolutionarily clusters closest with Mhp1 and the plant transporters (Elbourne *et al.*, 2017). CodB has 22.8% sequence identity and 58.1% sequence similarities with Mhp1 with some conservation of residues at the proposed sodium binding site. Mhp1 unfortunately produces crystals that diffract at a low resolution and it was hoped that CodB may give better diffracting crystals and a higher resolution structure to elucidate the role of sodium binding. Cytosine is

also more soluble than hydantoin and it should be easier to do functional studies. Even if this homologue does not give better resolution, it can still be useful to compare with Mhp1 to determine unique and shared characteristics, and if CodB is able to be crystallised in a different conformation this would provide more insight into the alternating access model. By characterising another member of the NCS1 we can begin to identify shared features. By studying bacterial homologues such as Mhp1 or CodB then we can begin to identify shared features between the LeuT superfamily and increase our knowledge and understanding regarding the LeuT superfamily with potential therapeutic benefits to structural homologs.

1.6. Conserved Features and Mechanism of Sodium-Dependent Symport

Despite sequence variation, all members of the LeuT superfamily exhibit the same 3D fold of 10 TM helices found in 2x 5 TM inverted repeats with TM1 and TM6 being unwound allowing main chain dipoles to coordinate solute and co-ion. Regardless of the chemical variation in the solute and differences in stoichiometry of sodium-ions the overall mechanism of binding and transport is comparable with details varying. All sodium-symporters utilise the exposed dipoles of the unwound section of TM1 and TM6 to coordinate their solute and/or co-ion. All of transporters discussed above, with BetP the exception, bind their solute/inhibitor between the "core/4-helix bundle" and "scaffold/hash motif".

The Na2 site is extremely well conserved, found at the interface between the 4-helix bundle and the hash-domain and has been shown to be the more important site to coupling the sodium gradient to transport. Based on multiple outward-open crystal structures it appears that sodium binding precedes solute binding, and that sodium binding stabilises the outward-facing conformation to allow solute to bind. Mhp1 was shown to exist in an inward-facing conformation, requiring titration of both Na^+ and substrate to move into the outward-facing state (Calabrese *et al.*, 2017). MD simulations in vSGLT showed that Na^+ leaving the Na2 site allows substrate to leave the binding site, but when sodium is restrained in the Na2, substrate cannot diffuse away. Inward-open structures of Mhp1, vSGLT, BetP, and LeuT have shown that substrate and sodium sites are disturbed.

The other sodium sites are less important for transport and their roles are unclear. The Na1 site in LeuT can be found in other members of the NSS family(Yamashita *et al.*, 2005; Penmatsa, Wang and Gouaux, 2013; Malinauskaite *et al.*, 2014; Coleman, Green and Gouaux, 2016), but not in BetP or SiaT which also use 2 Na^+. In BetP the second Na^+ is thought to coordinate at the structural symmetric equivalent of Na2 (the authors termed this Na1'), i.e. between TM3 and TM6, the authors commented that Lys110 in Mhp1 is found at this position which is a positively charged side chain and forms bridges with residues in TM3 and TM6 (Khafizov *et al.*, 2012). SiaT uses a unique Na3 site to bind Na^+; examining sequence alignments it appears that this Na3 site could be found in 19.6% of SSS, the corresponding residues in LeuT, Mhp1, SERT are all hydrophobic (Wahlgren *et al.*, 2018).

Structural repeats in membrane transporters facilitate molecular mechanisms of transport. The structural repeats observed in LeuT enabled the conception of the rocking bundle mechanism. Before other conformations were discovered, a model of LeuT moving from an outward-facing to inward-facing state involved the conformation switching of helices. Inverted repeat 1 had the arrangement as inverted repeat 2 and vice versa whilst maintaining their membrane topology resulting in a rotation of the 4-helix bundle relative to the hash-motif. This conformation switching was confirmed by Mhp1 structures, moving from the outward-facing state to the inward-facing state results in a movement of 30 ° with bending of TM5 and TM10 acting as intracellular and extracellular gates respectively. LeuT and other family members also exhibit this rotation of the 4-helix bundle relative to the hash-motif with each member varying the amount of "rocking". LeuT superfamily members exhibit flexing of their discontinuous helices, to varying amounts with BetP being the most severe and almost no flexing of these helices in Mhp1 with other members somewhere in-between. It has been suggested that the variation of rocking vs flexing varies with the position of sodium ions.

1.7. Studying membrane proteins

Due to the hydrophobic and fragile nature of membranes our knowledge of them is still relatively limited, with many questions still to be answered. It is estimated that 20-30% of all genes in genomes encode membrane proteins with 10% of bacterial genomes

involved with membrane transport systems (Krogh *et al.*, 2001). In human's membrane protein dysfunction has been associated with a variety of health problems, including, but not limited to, depression, anxiety, heart disease, cancer, stroke, and diabetes (Masson *et al.*, 1999) . Because of this, 60% of human drug targets are membrane proteins with particular emphasis on G-protein coupled receptors (Cherezov, Abola and Stevens, 2010). Membrane proteins are found natively in such low concentrations that they must be heterologous expressed, but an overexpression of membrane proteins is often toxic to the cell. Membrane proteins must be extracted from membrane using detergent to stabilise their large hydrophobic regions, due to the intrinsic instability of these proteins they are often difficult to purify, and then because of this instability either struggle to crystallise or cannot form the crystal contacts that would provide high resolution data to give a high-resolution structure. Functionally, membrane proteins are difficult to study because many methods that are suitable for soluble proteins are not for membrane proteins.

Membrane proteins are difficult to study and characterise because of problems with their expression, stability, and purification before structural or functional determination can even be considered (Carpenter *et al.*, 2008; Bill *et al.*, 2011; Kang, Lee and Drew, 2013). As such, the number of membrane protein structures are much lower than their soluble counterparts but there is a currently an upwards trajectory on the number of high-resolution structures being solved. As of writing there are 1232 unique membrane protein structures in the protein data bank (PDB); the number of structures being solved in increasing at an exponential rate. (http://blanco.biomol.uci.edu/mpstruc/)

Despite recent successes there are still several "bottlenecks:

1. Poor protein expression
2. Solubilization
3. Protein purification and stability
4. Crystallization
5. Diffraction data collection
6. Structure solution

The ultimate aim of this project is to understand the physiological role of CodB by generating a 3D atomic model, determining kinetic substrate uptake. All of this will link structure to function and increase our knowledge of the LeuT superfamily and the importance of sodium binding in secondary active transporters.

1. What is the unique 3D atomic structure of CodB and how does this relate to function?
2. Why are residues around the sodium binding site of Mhp1 and putative binding site of CodB conserved?
3. What are the binding and transport kinetics of CodB and what can this protein bind and/or transport?
4. How does sodium drive transport of cytosine?
5. How does the structure of CodB relate to Mhp1 and what does this add to expanding our knowledge of the rocking-bundle model?

Chapter 2-Materials and Methods

2.1. Microbiology and DNA Manipulation

Strain	Genome Type	Antibiotic Resistance	Reference
E. coli NEB5α	fhuA2 a(argF-lacZ)U169 phoA glnV44 a80a(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	N/A	(Anton and Raleigh, 2016)
E.coli Lemo21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo(Cam^R) λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo = pACYC184-PrhaBAD-lysY	Chloramphenicol	(Schlegel *et al.*, 2012)
E. coli BL21 (RIPL)	F-ompT hsdS(rB- mB-) dcm+ Tetr galλ(DE3) endA hte [argU proL Camr] [argU ileY leuW Strep/Specr]	Tetracycline Chloramphenicol Streptomycin	(Kleber-Janke and Becker, 2000)

Table 2.1. Bacterial strains used for expression.

Media	Ingredients
Luria-Bertani (LB)	10 g/L tryptone; 5 g/L NaCl; 10 g/L yeast extract.
Super optimal broth with catabolite repression (SOC)	2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.
PASM-5052	50 mM Na_2HPO_4 , 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 2mM $MgSO_4$, 0.5 % glycerol, 0.05% glucose, 0.2 % lactose, 200 µg/ml of each of 17 amino acids (no C, Y, and M), 135 µg/ml methionine, and 100 nM vitamin B_{12}. 10 µM $FeCl_3$, 4 µM $CaCl_2$, 2 µM $MnCl_2$, 2µM $ZnSO_4$, and 0.4 µM $CoCl_2$, 0.4 µM $CuCl_2$, 0.4 µM $NiCl_2$, 0.4 µM Na_2MoO_4, 0.4 µM Na_2, 0.4 µM SeO_3, 0.4 µM H_3BO_3.

Table 2.2. Media components.

2.1.1. Preparation of Competent Cells

NEB5α and Lemo21(DE3) were streaked on a agar plate or 25 µg/mL chloramphenicol agar plate respectively and incubated overnight at 37 °C, then 1 colony was picked and grown in 100ml of LB in a 500ml conical flask until Optical Density at 600nm (OD_{600}) =0.5 at 37 °C, 180 (rotations per minute) rpm. Lemo21(DE3) cells were grown in the presence of chloramphenicol at a final concentration of 30 µg/mL. Cells were chilled on ice for 10 minutes, then centrifuged at 2600 g at 4 °C for 10 minutes. The cells were resuspended in 2 ml of ice cold 0.1 M $CaCl_2$ and incubated on ice for 10 minutes. The cells were centrifuged for 10 minutes, at 4 °C, 2600 g then resuspended in 2 ml of ice cold 0.1 M $CaCl_2$ and 20% glycerol. Cells were left on ice for a further 10 minutes and then aliquoted into 50 µL fractions and then snap frozen in liquid nitrogen and stored at -80 °C.

2.1.2. Transformation of Competent Cells

NEB5α competent cells were transformed with plasmid to ligate DNA fragments or amplify plasmids. Lemo21(DE3) competent cells were transformed for protein expression (Schlegel *et al.*, 2012). Competent cells were thawed on ice for 10 minutes, then 0.5 µL of plasmid was added and the mixture was incubated on ice for 30 minutes. Cells were heat shocked at 42 °C for 30 seconds for NEB5α and 10 seconds for Lemo21(DE3) and allowed to recover on ice for 5 minutes. 900 µL of SOC media was added to the cells and incubated at 37 °C, 900 rpm for 1 h. 100 µL of incubated cell culture was plated on an agar plate with 50 µg/mL kanamycin or 50 µg/mL kanamycin and 30 µg/mL chloramphenicol respectively and incubated at 37 °C overnight.

2.1.3. Cloning of *codB* into modified pWaldo

codB gene was identified in multiple species of bacteria, basically on suitability for expression, the genes were codon-optimised and synthesised by Integrated DNA Technologies (IDT) (Appendix 1). Species chosen were *Escherichia coli*, *Proteus vulgaris, Salmonella enterica*, *Selonomonas bovis*, and *Clostridium lundense*. The protein expression vector is a modified pWaldod plasmid, with the TEV protease site replaced by 3C protease site (Gifted by Deborah Brotherton, Cameron Group).

Protein Sequence CodB *E. coli*:

MSQDNNFSQGPVPQSARKGVLALTFVMLGLTFFSASMWTGGTLGTGLSYH
DFFLAVLIGNLLLGIYTSFLGYIGAKTGLTTHLLARFSFGVKGSWLPSLLLGG
TQVGWFGVGVAMFAIPVGKATGLDINLLIAVSGLLMTVTVFFGISALTVLSV
IAVPAIACLGGYSVWLAVNGMGGLDALKAVVPAQPLDFNVALALVVGSFIS
AGTLTADFVRFGRNAKLAVLVAMVAFFLGNSLMFIFGAAGAAALGMADIS
DVMIAQGLLLPAIVVLGLNIWTTNDNALYASGLGFANITGMSSKTLSVINGII
GTVCALWLYNNFVGWLTFLSAAIPPVGGVIIADYLMNRRRYEHFATTRMMS
VNWVAILAVALGIAAGHWLPGIVPVNAVLGGALSYLILNPILNRKTTAAMT
HVEANSVE

Protein Sequence CodB *P. vulgaris* (85% Identity to *E. coli*):
MSQDNNYSQGPVPISARKGGLALTFVMLGLTFFSASMWTGGALGTGLSFND
FFLAVLIGNLLLGIYTAFLGFIGSKTGLTTHLLARYSFGIKGSWLPSFLLGGTQ
VGWFGVGVAMFAIPVGKATGIDINLLIAVSGILMTITVFFGISALTVLSIIAVP
AIAILGSYSVYLAIHDMGGLSTLMNVKPTQPLDFNLALAMVVGSFISAGTLT
ADFVRFGRNPKVAVVVAIIAFFLGNTLMFVFGAAGAASLGMADISDVMIAQ
GLLLPAIVVLGLNIWTTNDNALYASGLGFANITGLSSKKLSVINGIVGTVCAL
WLYNNFVGWLTFLSAAIPPVGGVIIADYLMNKARYNTFNIATMQSVNWVA
LLAVAIGIVAGHWLPGIVPVNAVLGGAISYAVLNPILNRRTARQAEISHAG

Protein Sequence CodB *S. enterica* (84% Identity to *E. coli*):
MSQDNNYSQGPVPLAARKGVIPLTFVMLGLTFFSASMWTGGTLGTGLSYND
FFLAVLFGNLLLGIYTAFLGYIGAKTGLSTHLLARYSFGVKGSWLPSLLLGGT
QVGWFGVGVAMFAIPVSKATGIDANILIAVSGLLMTLTIFFGISALTILSIIAVP
AIVILGSYSVWLAVSGVGGLEHLKTIVPQTPLDFSSALALVVGSFVSAGTLTA
DFVRFGRHAKSAVLIAMVAFFLGNSLMFIFGASGAAAVGQADISDVMIAQG
LLLPAIVVLGLNIWTTNDNALYASGLGFANITGLSSRTLSVVNGIIGTVCALW
LYNNFVGWLTFLSSAIPPIGGVIIADYLLNRRRYADFNTVRFIPVNWIAILSVA
LGIAAGHYVPGIVPVNAVLGGVFSYILLNPLFNRSLAKSPEVSHAEQ

Protein Sequence CodB *C. lundense* (58% Identity to *E. coli*):
MSTQNTNYDHDFSLTVVPEGEKKGFLSMLVVMLGFTFFSASMWTGGKLGT
GLDMKTFALAVLSGNLILGAYTGALAYISCETGLSTHLLSRYSFGEKGSYLV
SFLLGGTQIGWFGVGVAMFALPVQKVTGINPYILVLIAGLLMTSSAYFGMKT
LTIISILAVPSIAVLGSYSAINAVNSIGGFSVLMNYQPKETLAFATALTMCVGS
FISGGTLTPDFTRFAKTKKVGVLTTVIAFFLGNSLMFIFGAVGAAATGKSDIS
EVMFLQGLILPAILILGLNIWTTNDNAIYSSGLGFSNITKIPKNKLVIVNGIVGT
LAAMWLYNNFVGWLTFLSSAIPPIGGVILADFFIVNRKMYGKFEETKFKNVN
WNAIVSWTIGTVAAEVIPGITPLYGVLGGAISYIIIGKAMKSKEIKERHEAAA

Protein Sequence CodB *S. bovis* (50% Identity to *E. coli*):
MEKKIDNDFSLTPVSKEGRRGLISMMAIMLGFTFYTGTMLTGGRLGTSLTFG
DLALVLFVGDFILGAYTALLAYMAGKTGLSTHLLARYAFGEKGSYLVSGIL
GLTQVGWFGVSVVMLALPISKVFGLDVTPVILICGALMVTTAYFGVKSLTIL
SVIAVPAIAVLGCYSSSISIAEVGGIGALMNATDVSSMSLTLALSLVVGSFISG
GTLTPDFARFSRTPRIAVVSTVAAFFIGNILMFAFGAIGGLAAGMPDISDVMI
AQGLVISGIVILGLNIWTTNDNTIYAASLAFSNITKMPKKHWVLINGFLSTVF
AMVLYNHFISLLSFLSSIIPPLGAVMIMDYFFLNRKAYAGAFSEAKFAVVNVP
AVLAVVAGGIFGHLPAGIGCLNAVFGAMLTYGIFTEIKVWLVRRREERAAA
GLRKVA

Plasmid	Antibiotic resistance	Properties	Reference
pWaldo::*codB* (modified with a 3C cleavable site instead of TEV)	Kanamycin	-CodB (Various bacterial species) -3C cleavable site -C-terminal GFP -C-terminal His-tag	(Waldo *et al.*, 1999)
pMAL::*3Cprotease*	Ampillicin	-3C protease -C-terminal Maltose binding protein tag -C-terminal His-tag	(Alexandrov, Dutta and Pascal, 2001)

Table 2.3. Plasmids used with properties.

Each IDT gene was supplied as a dry pellet, 100 µL of TE buffer was added and the mixture was briefly vortexed to resuspend the pellet in a final concentration of 10 ng/ µL. This was then incubated at 50 ^{o}C for 20 minutes, vortexed again and finally briefly centrifuged. The IDT genes and pWaldo vector were then cut with NdeI and BamHI at 37 ^{o}C, 2.5 h. 10 µL of the pWaldo cut with BamHI and NdeI reaction was taken away and 1 µL of Calf Intestinal Phosphate (CIP) was added to the remaining mixture and incubated at 37 ^{o}C for 1 h (Waldo *et al.*, 1999; Pédelacq *et al.*, 2006). These reactions were run on a 1% agarose gel for 45 minutes, 110 V, 400 mA. Bands of interest were cut out and purified using the Invitrogen GeneJET Gel Extraction Kit™. IDT genes and pWaldo vector were ligated overnight at 16 ^{o}C and then heat inactivated at 65 ^{o}C for 10 minutes. The reactions were then transformed into NEB5α competent cells. Colonies were grown overnight in 5 mL of LB

supplemented with 50 µg/mL kanamycin and their DNA was miniprepped by the Invitrogen GeneJET Plasmid Miniprep Kit™. Plasmids were sequenced using the T7 forward primer and GFP reverse primer.

2.1.4. Site-Directed Mutagenesis of pWaldo::*codB* (*S. bovis*)

To correct an single base in the pWaldo::*codB* (*S. bovis*) the following was set up:

5 µL of Pfu reaction buffer

2 µL of pWaldo::*codB* (*S. bovis*) plasmid

1.25 µL of forward primer

1.25 µL of reverse primer

1 µL of dNTPs at 10mM

15.25 µL of H_2O

0.25 µL of Pfu enzyme

This plasmid was then amplified using the polymerase chain reaction (PCR) with the following conditions.

Cycles	Temperature (°C)	Time
1	95	2 minutes
20	95	30 seconds
20	72	20 seconds
20	72	9 minutes
1	72	10 minutes

Table 2.4. PCR conditions to correct pWaldo::*codB* (*S. bovis*).

1 µL of Dpn1 was added to the mixture post-PCR and briefly spun in a microfuge then incubated at 37 °C for 5 minutes. 2 µL of this PCR mixture was transformed into NEB5α cells as section 2.2.2. Colonies were grown overnight in 5 mL of LB supplemented with 50 µg/mL kanamycin and their DNA was miniprepped by the Invitrogen GeneJET Plasmid Miniprep Kit™. Plasmids were sequenced using the T7 forward primer and Green Fluorescent Protein (GFP) reverse primer.

2.1.5. Site Directed Mutagenesis of pWaldo::*codB* (*P. vulgaris*)

Site directed mutagenesis was done with Quikchange ™ (Aligent) kit

To generate mutants of pWaldo::*codB* the following was set up:

5µL of 10x reaction buffer

1 µL of pWaldo::*codB*

1.25 µL of forward primer

1.25 µL of reverse primer

1 µL of dNTPs at 10mM

1.5ul of Quik Solution

39 µL of H_2O

1 µL of enzyme

Cycles	Temperature (°C)	Time
1	95	2 minutes
18	95	20 seconds
18	60	10 seconds
18	68	3 minutes
1	68	5 minutes

Table 2.5. PCR conditions to generate mutants of pWaldo::*codB (P. vulgaris).*

Mutants Trp108Ala, Phe204Ala, Thr278Ala, Thr279Ala were made by an undergraduate student Mehalah Spencer.

CodB mutant	Primer sequence 5' to 3' Forward primer (Top) Reverse primer (Bottom)
F33A	CATGTTAGGTCTGACATTCGCTTCCGCATCTATGTGGACC GGTCCACATAGATGCGGAAGCGAATGTCAGACCTAACATG
G112A	CGCAAACATAGCGACGGCTACACCAAACCATCC GGATGGTTTGGTGTAGCCGTCGCTATGTTTGCG
Q105A	CTGGGGGGGACTGCGGTTGGATGGTTTGGTGTAG CTACACCAAACCATCCAACCGCAGTCCCCCCCAG
W108A	GGGACTCAGGTTGGAGCGTTTGGTGTAGGCGTCGCTATG CATAGCGACGCCTACACCAAACGCTCCAACCTGAGTCCC
F204A	CGATGGTCGTGGGATCAGCCATTAGCGCTGGTACACTTAC GTAAGTGTACCAGCGCTAATGGCTGATCCCACGACCATCG
S206A	CGATGGTCGTGGGATCATTCATTGCCGCTGGTACACT AGTGTACCAGCGGCAATGAATGATCCCACGACCATCG
N275A	CTATCGTGGTCCTGGGTTTGGCTATTTGGACCACAAATGACA TGTCATTTGTGGTCCAAATAGCCAAACCCAGGACCACGATAG
T278A	CTGGGGTTTGAATATTTGGGCCACAAATGACAATGCC GGCATTGTCATTTGTGGCCCAAATATTCAAACCCCAG
T279A	CTGGGGTTTGAATATTTGGACCGCAAATGACAATGCCCTTTACG CGTAAAGGGCATTGTCATTTGCGGTCCAAATATTCAAACCCCAG
T279S	TGGGTTTGAATATTTGGACCTCAAATGACAATGCCCTTTAC GTAAAGGGCATTGTCATTTGAGGTCCAAATATTCAAACCCA
T278AT279A	CTGGGTTTGAATATTTGGGCCGCAAATGACAATGCC GGCATTGTCATTTGCGGCCCAAATATTCAAACCCAG
N280A	GTCCTGGGTTTGAATATTTGGACCACAGCTGACAATGCCCTTTAC GTAAAGGGCATTGTCAGCTGTGGTCCAAATATTCAAACCCAGGAC
N282A	ATATTTGGACCACAAATGACGCTGCCCTTTACGCCTCAGGA TCCTGAGGCGTAAAGGGCAGCGTCATTTGTGGTCCAAATAT

Table 2.6. Primer sequences for site directed mutagenesis of pWaldo::*codB*. Mutated codons are shown in blue.

1 µL of Dpn1 was added to the mixture post-PCR and briefly spun in a microfuge then incubated at 37 °C for 5 minutes. 2 µL of this PCR mixture was transformed into

NEB5α cells as section 2.1.2. 1 colony was grown in 5 mL of LB media supplemented with 50 µg/mL kanamycin at 37 °C, 200 rpm, overnight. DNA was miniprepped by the Invitrogen GeneJET Plasmid Miniprep Kit ™. Plasmids were sequenced using the T7 forward primer and GFP reverse primer.

2.2. Protein Expression

2.2.1. Testing Small-Scale Expression of *codB*

A single colony of *E. coli* Lemo21(DE3) transformed with pWaldo::*codB* was used to inoculate 5 mL LB, supplemented with 50 µg/mL kanamycin and 30 µg/mL chloramphenicol, was grown overnight at 37 °C, 180 rpm. 2.5 mL of this overnight culture was then used to inoculate 125 mL of PASM-5052 media. (Studier, 2005) This culture was split into 12x 5 mL cultures with [rhamnose] of 0 mM, 0.1 mM, 0.25 mM, 0.5 mM, 0.75 mM, 1 mM. All cultures were grown at 37 °C, 180 rpm and grown until OD_{600} reached 0.5. The remaining 65 mL was split into 12x5 mL cultures and rhamnose at 0 mM, 0.1 mM, 0.25 mM, 0.5 mM, 0.75 mM, 1 mM. 50% of the small cultures were induced with 0.4 mM Isopropyl β-D-1 thiogalactopyranoside (IPTG) and then grown overnight at 25 °C, 180 rpm (Lee *et al.*, 2014). Figure 2.1 shows the Lemo21(DE3) expression system. Cultures were diluted 1:20 in Phosphate-Buffers Saline (PBS) and OD_{600} was measured. For each culture, 2x1 mL was taken, and spun at 13,000 rpm, 4 °C, 10 minutes. For each culture, 1 of these pellets was resuspended in 100 µL of PBS and a GFP reading taken, enabling quantification of protein concentration. The other pellet was then resuspended in 100 µL of 1x SDS (Sodium Dodecyl Sulphate) loading dye and sonicated for 10 seconds at 70% and diluted 1:10 in 1x SDS loading dye, then run on a 12% Sodium Dodecyl Sulphate Polyacrylamide gel electrophoresis (SDS PAGE) gel for 45 minutes, 400 mA, 200 V.

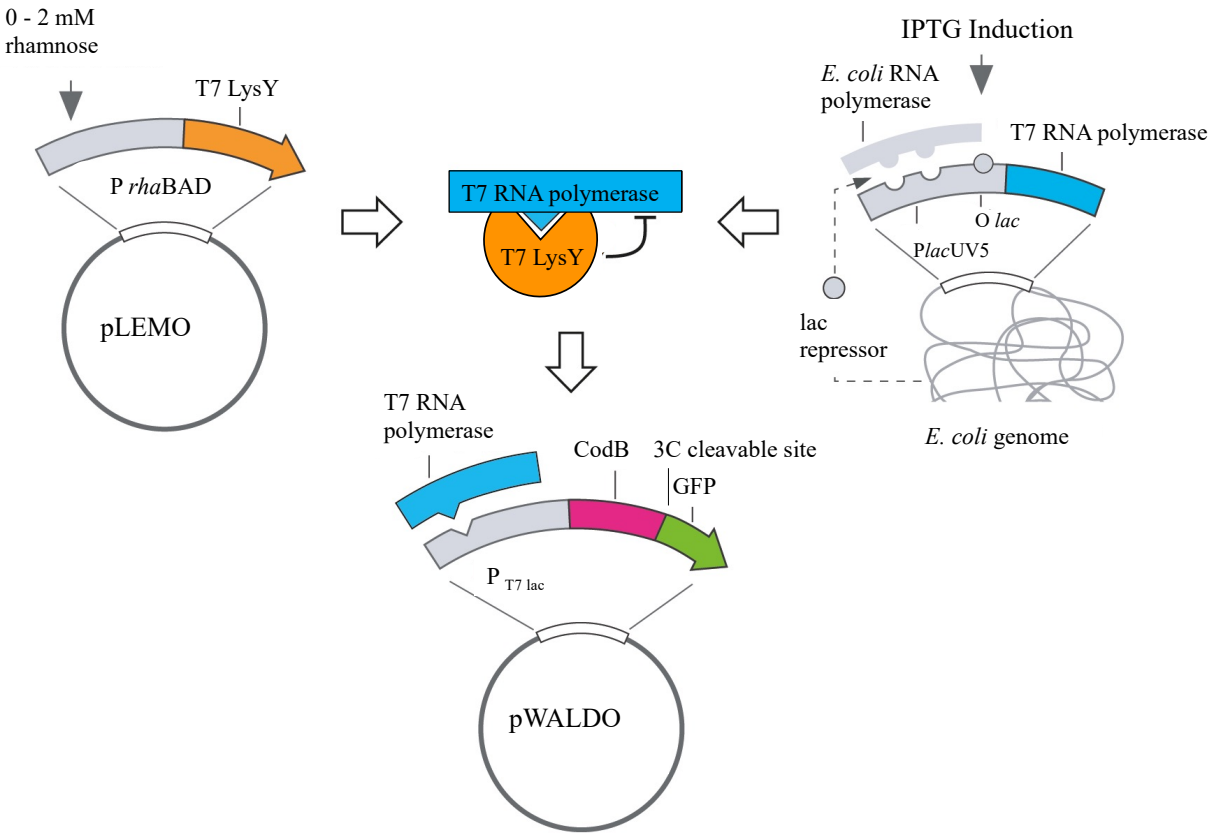


Figure 2.1. Schematic representation of CodB.GFP protein expression. The expression of T7 lysozyme (T7 LysY) is under the control of the tuneable rhamnose promoter (P *rha*BAD). T7 lysozyme inhibits T7 RNA polymerase and by altering the rhamnose concentration gene expression can be controlled. This figure has been adapted from (Lee *et al.*, 2014).

2.2.2. Large-Scale Expression of *codB*

A single colony of *E. coli* Lemo21(DE3) was incubated in 20 mL of LB (50 µg/mL kanamycin and 30 µg/mL chloramphenicol) and grown overnight at 37 °C, 180 rpm. 20 mL of this overnight culture was then used to inoculate 1 L of PASM-5052 media. For each construct the following condition was used:

Construct	Condition
pWaldo::*codB E. coli*	0.25 mM rhamnose
pWaldo::*codB S. enterica*	0.5 mM rhamnose
pWaldo::*codB P. vulgaris*	0.1 mM rhamnose and 0.4 mM IPTG
pWaldo::*codB C. lundense*	0.5 mM rhamnose
pWaldo::*codB S.bovis*	0.25 mM rhamnose

Table 2.7. Conditions for optimal protein expression of CodB constructs

Culture was grown at 37 °C, 180 rpm until OD_{600} reached 0.5. All cultures were moved to 25 °C, 180 rpm and pWaldo::*codB P. vulgaris* was induced with 0.4 mM IPTG. All cultures were grown overnight, after growth the cells were harvested by centrifugation at 5000 g for 15 minutes and snap frozen in liquid nitrogen, then stored -80 °C. Gels and GFP fluorescence were taken as section 2.2.1.

2.2.3. Expression of 3C protease

0.5 µL of 3C protease. pMAL vector provided by the Cameron group was transformed into BL21 (RIPL) as section 2.2.2. 1 colony was used to inoculate 200 mL of LB and was grown at 37 °C, 180 rpm overnight. 10 mL of overnight culture was added to 1 L of LB and grown at 37 °C, 180 rpm until OD_{600} = 0.5, 1 mM IPTG was added to induce protein expression, then cultures were grown at 18 °C, 180 rpm overnight. Cells were harvested by centrifugation at 5000 g for 15 minutes and pellets snap frozen in liquid nitrogen, then stored -80 °C.

2.3. Membrane Preparation and Solubilisation

2.3.1. Isolation of *E. coli* Membranes

Cell pellets were resuspended in PBS with 1 mM $MgCl_2$, DNAase, and 0.5 M 4-benzenesulfonyl fluoride hydrochloride (AEBSF) and disrupted at 25 kPsi x3. Cell lysate was centrifuged at 24,000 g, 4 °C for 12 minutes to remove insoluble cell debris, and the supernatant was subjected to ultracentrifugation at 200,000 g, 4 °C for 45 minutes. Membrane pellets were resuspended in PBS, 15 mL per 1 l of culture, snap frozen in liquid nitrogen, and then stored -80 °C. Each fraction was diluted 1:10 and run on a 12% SDS PAGE gel and GFP fluorescence measurements taken as section 2.2.1.

2.3.2. Fluorescence Size Exclusion Chromatography (FSEC) of CodB

100 µL of membrane suspension was solubilised with 100 µL of 10 % detergent and 800 µL of 20 mM trisaminomethane (TRIS) pH 7.5, 150 mM NaCl for 1 h 30 minutes at 4 °C with mild agitation. The solubilised membrane suspension was ultracentrifuged at 100,000g for 45 minutes and loaded onto a Superose 6, 30 100 Size Exclusion Chromatography (SEC) column preloaded with 20 mM TRIS pH

7.5, 150 mM NaCl, 0.03% n-Dodecyl β-D-maltoside (DDM) and fractioned into 180 µL aliquots, their GFP fluorescence was taken. (Riehm *et al.*, 2005; Kawate and Gouaux, 2006; Backmark *et al.*, 2013)

2.4. Protein Purification

2.4.1. Purification of CodB

45 mL of membrane suspension, produced in section 2.3.1 was solubilised in 155 mL Solubilisation buffer (1x PBS, 150 mM NaCl, 1% DDM) for 2 h at 4 °C with mild stirring, then ultracentrifuged at 200,00g, 45 mins, 4 °C to remove any insoluble material. Imidazole was added to the supernatant to give a final concentration of 20 mM and mixed with 15 mL of loose Ni resin (pre-equilibrated with wash buffer 1) for 3 h, 4 °C with mixing. Supernatant and resin slurry was transferred to a gravity flow column, flowthrough was collected and resin was washed with 100 mL of wash buffer 1 (1x PBS, 150 mM NaCl, 20 mM imidazole 0.1% DDM) then 100 mL of wash buffer 2 (1x PBS, 150 mM NaCl, 30 mM imidazole 0.03% DDM). Compacted resin was made into a slurry with wash buffer 2 and 15 mg of 3C protease was added; the column was then left rolling overnight at 4 °C.

The column was left static for 1 h, then washed with 50ml of wash buffer 2 and fractionated into 5 x 10 mL fractions (fractions 1 → 5) with CodB in the flowthrough. The column was then eluted with 50 mL of elution buffer (1x PBS, 150 mM NaCl, 250 mM imidazole 0.03% DDM) to remove any bound 3C protease and GFP into 5 x 10 mL fractions (fractions 6 → 10). Fractions 1→ 3 were then pooled together and filtered through a 0.45 µm filter (Sartorius ™) and loaded onto a 5 mL HisTrap column (GE healthcare), pre-equilibrated with wash buffer 2 to remove any residual 3C protease and GFP, then chased with 10 mL of wash buffer 2 and fractionated into 9 x 5ml fractions (RI 1 →RI 9). Fractions RI 2 → RI 7 were then pooled and concentrated with a 100 kDa molecular weight concentrator at 4 °C until a final protein concentration was between 8-10 mg/mL, and stored at 4 °C or flash frozen in liquid nitrogen and stored at -80 °C. If CodB needed to be purified in the presence of cytosine then 1 mM cytosine was added to all buffers.

2.4.2. Expression and Purification of 3C protease

Cell pellets were resuspended to 150 mL in 40 mM sodium phosphate pH 7.8, 200 mM NaCl, 1 mM Dichlorodiphenyltrichloroethane (DDT), 20 mM Imidazole, 5% glycerol and disrupted at 25 kPsi twice. Cell lysate was centrifuged at 17,000 g, 4 ^{o}C for 30 minutes to remove insoluble cell debris, lysate was passed through a 0.4 µM filter and left to bind to 2x 5 mL HisTrap ™ columns for 2h at 4 ^{o}C. Columns were washed with 50 mL of 40 mM sodium phosphate pH 7.8, 200 mM NaCl, 1 mM DDT, 20 mM Imidazole, 5% glycerol, then eluted into 5x10 mL fractions in 40 mM Sodium phosphate pH 7.8, 200 mM NaCl, 1 mM DTT, 250 mM Imidazole, 5% glycerol. Fractions containing 3C protease were pooled and dialysed overnight at 4 ^{o}C into 40 mM sodium phosphate pH 7.8, 200 mM NaCl, 1 mM DTT, 5% glycerol, repeated into 40 mM sodium phosphate pH 7.8, 200 mM NaCl, 1 mM DTT, 10% glycerol. Diluted 1:1 in 40 mM sodium phosphate pH 7.8, 200 mM NaCl, 1 mM DDT, 30% glycerol. 1 mL Aliquots were snap frozen in liquid nitrogen and stored at -80 ^{o}C

2.5. Crystallisation and Structural Determination of CodB

2.5.1. Vapour Diffusion Crystallisation of CodB

500 µL of purified CodB was loaded onto a Superdex 200 column (GE healthcare), pre-equilibrated with SEC buffer (20 mM TRIS pH 7.5, 150 mM NaCl, 0.7% NM). CodB was eluted from the column over a 24 mL volume. Purified CodB in the presence of 1 mM cytosine and NM was concentrated to 10 mg/mL was crystallised in a variety of conditions found in commercially available screens, MemGold ^{TM}, MemGold 2 ^{TM}, MemTrans ^{TM}, MemChannel ^{TM} (Molecular Dimensions) using sitting drop vapour diffusion method. Screens were set up using a mosquito nanolitre pipetting robot (TTP labtech). using 96 well MRC plates with 0.1 µL of protein (10 mg/mL) solution mixed with 0.1 µL of reservoir solution. Screens were sealed with ClearVue sealing sheets (Molecular Dimensions) and allowed to equilibrate at 20 °C or 4 °C. Initial ‘hits’ were optimised using hanging drop vapour diffusion method, in 24 well plates (Molecular Dimensions) with 1 µL protein solution mixed with 1 µL reservoir solution. Crystallisation optimisation was carried using an additive screen of MemAdvantage™ (Molecular Dimensions) and detergent screen (Hampton

Research) manually in a 24-well handing drop method. Conditions were added in a 1:10 dilution.

2.5.2. LCP Crystallisation of CodB

Purified CodB in the presence of 1 mM cytosine and DDM was concentrated up to 32 mg/mL, then extruded into monoolein until mixture is clear.

Volume of protein (µL) / (0.94/0.66) = Volume of Monoolein (µL).

LCP mixture was subjected to crystallisation screening in MemMeso™ and MemGold Meso ™ (Molecular Dimensions). Screens were set up using a mosquito nanolitre pipetting robot (mosquito) using 96 well glass plates with 50 nL of protein (10 mg/mL) solution mixed with 800 nL of reservoir solution and stored crystallisation plates at 20 °C or 4 °C. (Caffrey, 2009)

2.5.3. Data Collection, Processing and Structural Determination of CodB

To prepare crystals for data collection they were snap-frozen in liquid nitrogen. Crystals were taken to Diamond Light Source and shot on beamline i24 whilst crystals were maintained at cryogenic temperatures.

Wavelength=0.9688

Oscillation=10 °

Transmission=10.05 %

Exposure= 0.010 s

All datasets were processed in Dials Xia2 pipeline (Winter, 2010; Waterman *et al.*, 2016). Processed data was then scaled and merged in Aimless in the CCP4 suite before being transferred to Phenix (Adams *et al.*, 2010; Afonine *et al.*, 2012; Evans and Murshudov, 2013). Mhp1 was used as the model for Rosetta to build a CodB model that was used as the molecular replacement model to generate an electron density map (Dimaio *et al.*, 2013). Mhp1 was manually rebuilt into CodB, multiple rounds of refinement and model building was done in Phenix and Coot respectively (Emsley *et al.*, 2010; Afonine *et al.*, 2012). Protein structure figures have been produced by PyMol (http://www.pymol.org) apart from electron density figures using CCP4 mg (McNicholas *et al.*, 2011). Structural alignments were done in PyMol using cealign algorithm unless specified otherwise.

2.6. Biophysical Methods

2.6.1. Circular Dichroism of CodB (CD)

Purified CodB was concentrated to 2 mg/mL put into dialysis overnight at 4 °C either 20 mM TRIS pH 7.5, 150 mM Na_2SO_4, 0.03% DDM or 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM. A spectrum was recorded between 190 nm to 330 nm with 100 µL of sample at 20 °C in a 96-well format with a pathlength of 0.28 cm. (Greenfield, 2007)

2.6.2. Isothermal Calorimetry (ITC)

CodB was concentrated to 8 mg/ml (184 µM) and put into dialysis overnight in 20 mM TRIS pH 7.5, 150 mM NaCl, 10 % glycerol, 0.03% DDM overnight at 4 °C, cytosine was diluted into dialysis buffer to 5 mM. 300 µL of CodB was placed into the cell with 5 mM in the syringe, measurements were performed at 10 °C. (Rajarathnam and Rösgen, 2014)

2.6.3. Microscale thermophoresis (MST)

GFP labelling strategy:

CodB was purified as in section 2.4.1, but instead of cleaving off the initial His resin using 3C protease, CodB was instead eluted with 1x PBS, 150 mM NaCl, 250 mM imidazole 0.03% DDM. Fractions containing CodB.GFP fusion protein were pooled and put into dialysis overnight at 4 °C in 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM. CodB.GFP was aliquoted into 1 mL fractions, snap frozen and stored at -80 °C until use. CodB.GFP was diluted to 100 nM in 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM, 0.05% Tween20. 10 µL of CodB.GFP was added to either 2 mM cytosine in 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM, 0.05% Tween20 or 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM, 0.05% Tween20. Samples were spun briefly and loaded into Monolith™ NT.115 MST Premium Coated Capillaries. Capillaries were loaded into Monolith NT.115. (Jerabek-Willemsen *et al.*, 2011)

Thiol labelling strategy:

0.4 nmoles CodB and 4 nmoles of AlexaFlour488 were mixed together and left at 4 °C overnight in the dark. Mixture was applied to a G25 column pre-equilibrated with

20 mM TRIS pH 7.5, 150 mM Nacl, 0.03% DDM. CodB was diluted to 100 nM in 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM, 0.05% Tween20. 10 µL of CodB was added to either 2 mM cytosine in 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM, 0.05% Tween20 or 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM, 0.05% Tween20. Samples were spun briefly and loaded into Monolith™ NT.115 MST Premium Coated Capillaries. Capillaries were loaded into Monolith NT.115.

2.7. Thermostability Assays

2.7.1. GFP-Thermostability Assay (GFP-TS)

To generate a melting curve:

150 µL of *E. coli* membrane with overexpressed CodB was diluted 1:10 in 20 mM TRIS pH 7.5, 150 mM NaCl, 1% DDM, 1% Beta-Octyl-glucosidase (β-OG), 1 mM substrate and left mixing at 4 ^{o}C for 1 h then aliquoted into 150 µL fractions. Aliquots were subjected to various temperatures, 4 ^{o}C, 20 ^{o}C, 25 ^{o}C, 30 ^{o}C, 35 ^{o}C, 40 ^{o}C, 45 ^{o}C, 50 ^{o}C, 60 ^{o}C for 10 minutes then spun at 16,000g for 30 minutes. 100 µL of supernatant was transferred to a 96 well black plate and GFP measurements were taken. The apparent melting temperature (T_m) for each titration was calculated by plotting the normalised average GFP fluorescence intensity from two technical repeats at each temperature and fitting the curves to a sigmoidal dose–response equation (variable slope) by GraphPad Prism software. Values reported are the averaged mean ± Standard error of the mean (S.E.M.) of the fit from n = 2 independent titrations. (Nji *et al.*, 2018)

To generate a K_d:

150 µL of *E. coli* membrane was solubilised as before but substrate was added at a final concentration between 0-1000 µM. Aliquots were put at 35 ^{o}C for 10 minutes and spun at 16,000g for 30 minutes. 100 µL of supernatant was transferred to a 96 well black plate and GFP measurements were taken. Binding curve was fitted by nonlinear regression (one site, total binding) by GraphPad Prism software, and the values reported are the averaged mean ± S.E.M. of the fit from n = 3 independent titrations. (Nji *et al.*, 2018)

2.7.2. CPM Assay

50 µL of 20 mM TRIS pH 7.5. 150 mM NaCl with 3x critical micelle concentration (CMC) of detergent (0.03% DDM, 0.09% n-Undecyl-β-D-Maltopyranoside (UM), 0.3% n-Decyl-β-D-Maltopyranoside (DM), 0.7% NM, 0.59% β-OG, or 0.66% *N,N* Dimethyldodecylamine *N*- oxide (LDAO)) and CodB at a final amount of 2.5 µg and 0.05 mg/mL CPM dye was set up. Samples were placed in an Agilent Technologies Stratagene Mx3005P ™ with a starting temperature of 20 °C increasing to 95 °C at 1 °C per minute. Fluorescence was measured at 470 nm. (Alexandrov *et al.*, 2008)

2.8 Transport Assays

2.8.1. In-Cell Transport Assay

For Time Course Assay:

CodB was expressed in 10 mL along with Lemo21(DE3) cells. Cultures were then spun at 2600 g, 10 minutes, 20 °C, supernatant was removed and the cell pellet was resuspended in 5 mL 5 mM 2-ethanesulfonic acid (MES) pH 6.6, 150 KCl. This was repeated three times. Cells were resuspended to OD_{600} = 2, to 200 µL with 5 mM MES pH 6.6, 150 mM NaCl or 5 mM MES pH 6.6, 150 mM choline chloride. 1 µL of 6.25 µM ^{3}H-5-cytosine (American Radiolabelled Chemicals) was added to cells and incubated at 37 °C, 900rpm and at times points 30seconds, 1 minute, 2 minutes, 5, minutes, 10 minutes, or 20 minutes. Cells were centrifuged at 13000rpm, 1 minute, 20 °C. Supernatant was removed and the pellet was resuspended in 200 µL stop buffer (5 mM MES pH 6.6, 150 mM KCl, 1 mM cytosine) and added to a 1 mL reservoir of stop buffer, 0.2 µm Whatman cellulose nitrate membrane filter under vacuum followed by immediate washing with 4 x 2 mL 0.1 M LiCl. All filters were dissolved in 10 ml Emulsifier Safe scintillation fluid and counted using a Tri-Carb A4810TR Liquid Scintillation Analyzer (Perkin Elmer). (Ma, 2010; Hu *et al.*, 2011; Jackson, 2012)

For Inhibition Assay:

CodB was expressed in 10 mL along with Lemo21(DE3) cells. Cultures were then spun at 2600 g, 10 minutes, 20 °C, supernatant was removed and the cell pellet was resuspended in 5 mL 5 mM MES pH 6.6, 150 KCl. This was repeated three times.

Cells were resuspended to OD_{600} = 2, to 200 µL with 5 mM MES pH 6.6, 150 mM NaCl, 0.1 mM compound. 1 µL of 6.25 µM ^{3}H-5-cytosine was added to cells and incubated at 37 °C, 900rpm for 1 minute. Cells were centrifuged at 13000rpm, 1 minute, 20 °C. Supernatant was removed and pellet was resuspended in 200 µL stop buffer (5 mM MES pH 6.6, 150 mM KCl, 1 mM cytosine) and added to a 1 mL reservoir of stop buffer, 0.2 µm Whatman cellulose nitrate membrane filter under vacuum followed by immediate washing with 4 x 2 mL 0.1 M LiCl. All filters were dissolved in 10 ml Emulsifier Safe scintillation fluid and counted using a Tri-Carb A4810TR Liquid Scintillation Analyzer (Perkin Elmer).

2.8.2. Proteoliposome Preparation and Transport Assay

Extrusion method:

0.8 g of Soybean Lipids (phosphatidylcholine) was added to 40 mL of MMK buffer (10 mM 3-N-morpholinopropanesulfonic acid (MOPS) pH7, 5 mM $MgCl_2$, 100 KCl) then vortexed for 1h until dissolved. Lipids were sonicated for 1 second, rested for 1 second x 15 then rested for 1 minute. This was repeated 6x. Lipids were centrifuged for 3 minutes at 15000g, and freeze thawed in liquid nitrogen x 8. Empty liposomes were stored at -80 °C until use.

Thawed liposomes were passed through an extruder 20x with a 0.4 µM filter and then a 0.2 µM. 1 mL of liposomes, 90 µL 20% sodium cholate and 40 µL CodB at 2 mg/mL were mixed together and left at room temperature for 30 minutes. Proteoliposome were added to a PD10 column pre-equilibrated with MMK buffer, proteoliposomes were eluted into a 2.5 mL fraction and dialysed in 1 L of MMK buffer overnight at 4°C. Empty liposomes were prepared in an identical manner, but CodB was not added. NaCl added to 150 µL of proteoliposomes at a final concentration of 150 mM, and proteoliposomes were activated by shaking at 37 °C, 5 minutes, 900 rpm. 1 µL of 12.5 µM ^{3}H-5-cytosine was added to proteoliposomes and proteoliposomes were incubated at 37 °C, 5 minutes, 900 rpm, 150 µL of ice cold MMK buffer was added and proteoliposomes were applied to a G25 column preequilibrated with MMK buffer. 700 µL of MMK was applied and liposomes were collected into 2.5 mL of Emulsifier Safe scintillation fluid and count using a Tri-Carb A4810TR Liquid Scintillation Analyzer (Perkin Elmer). (Sekiguchi, 2014)

Destabilisation method:

500 µL of CodB at 2 mg/mL was loaded onto a Superdex12 column pre-equilibrated with 20 mM Tris pH 7.5, 150 mM NaCl, 0.28% DM. CodB was eluted over column volume. Meanwhile, liposomes made with *E. coli* polar lipids provided by Patrick Becker were extruded through a 0.4 µM filter x13. Liposomes were diluted to 5mg/mL in 100 potassium phosphate pH 7.5, 2 mM β-mercaptoethanol and destabilised with 0.12 % triton. 140 µL of CodB in the most concentrated fraction and 140 µL of 100 potassium phosphate pH 7.5, 2 mM β-mercaptoethanol were added to liposomes to produce proteoliposomes and empty liposomes respectively. proteoliposomes and liposomes were incubated at room temperature for 10 minutes, 7.2 mg biobeads were added and suspension was incubated with mixing for 1 h at RT. Biobeads were changed and incubated for 1 h x2. A final change of biobeads was preformed and liposomes were left at 4 ^{o}C overnight. Proteoliposome suspensions were ultracentrifuged at 220,000g, 4 ^{o}C, 1 h, then resuspended in 40 µL to 100 mg/mL. These suspensions were then snap frozen and stored at -80 ^{o}C until use. Liposomes were diluted to 1 mL in 100 potassium phosphate pH 7.5, 2 mM β-mercaptoethanol, then extruded through a 0.4 µM filter x13, then ultracentrifuged at 18 ^{o}C, 1 h, 220,000g. Liposome pellet was resuspended in 10 µL of 100 potassium phosphate pH 7.5, 2 mM β-mercaptoethanol, then diluted 1:200 in 10 mM HEPES pH 7.4, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 100 mM NaCl. To activate liposomes, they were incubated at 37 ^{o}C, 900 rpm for 5 minutes, then 1 µL of 12.5 µM ^{3}H-5-cytosine was added and proteoliposomes were incubated at 37 ^{o}C, 900 rpm for 5 minutes. 150 µL of cold 10 mM HEPES pH 7.4, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 100 mM NaCl was added and the suspension was applied to a G25 column and chased with 150 µL of buffer, the colum was washed with a further 700 µL into 2.5 mL of Emulsifier Safe scintillation fluid and count using a Tri-Carb A4810TR Liquid Scintillation Analyzer (Perkin Elmer).(Jung *et al.*, 1998)

Rapid Dilution method:

E. coli polar lipids (Avanti)were resuspended to 40 mg/mL in 20 mM TRIS pH 7.5, 150 mM NaCl, 0.03% DDM to 250 µL, 2 mL of CodB at 0.5 mg/mL was added and the suspensions were left on ice for 1 h. Proteoliposomes were spun at 6000g for 10 minutes at room temperature, 2 mL of these proteoliposomes were loaded onto an

S200 column pre-equilibrated with 20 mM TRIS pH 7.5, 150 KCl. (Hughes *et al.*, 2019)

Chapter 3-Expression, Solubilisation, and Purification of CodB

3.1. Introduction

For any functional, structural, or biochemical study the protein of interest must be expressed, purified, soluble, and stable in suitable quantities to allow said experiment. As stated in chapter 1, the aim of this project was to use macromolecular x-ray crystallography to determine the 3D structure of CodB. Good-quality diffraction is dependent on well-ordered crystals, which is in turn dependent on a well-behaved stable protein. As described in Chapter 1, there is an upwards trajectory of solving membrane structures and currently this field is still growing exponentially. This has been possible in part because of recent developments the ability to screen at various bottlenecks.

Producing protein crystals requires lots of pure protein and it well established that membrane proteins are natively expressed in very low quantities; purification from a native source in sufficient quantities is not possible apart from specific examples, such as bacteriorhopsin. Especially in the context of CodB as its expression is induced in the presence of cytosine and therefore native protein expression would not produce enough protein for downstream experiments. Most crystallography now requires the use of heterologous protein expression to produce enough protein. After expressing protein, solubility testing is required to identify which detergents are appropriate. Good detergent flexibility allows a larger screen for crystal conditions and a higher chance of good diffraction. Detergents that have large micelles often produce poor-diffraction and smaller detergent micelles generally produce high-resolution diffraction as the protein can form better crystal contacts (Sonoda *et al.*, 2010). After this screening it can be decided which constructs are worth pursuing and purifying for future crystal screens. Ultimately, there is not enough time or space or resources to screen a protein in the infinite different conditions to grow protein crystals, and consequently some proteins will just not crystallise; therefore the best approach is to try and screen a few different constructs at the start of a project and use constructs that perform well at each of these bottlenecks (Sonoda *et al.*, 2010). This chapter focuses on the heterologous expression, solubility testing and purification of CodB from different bacterial species to identify constructs suitable for future experiments.

3.2. Results

3.2.1 Cloning of *CodB* into pWaldo

CodB in *E. coli* has previously been characterised as a cytosine transporter; other genes believed to be CodB in different bacterial species were screened using the following criteria:

1. Must be in an operon with CodA, a cytosine deaminase also characterised initially in *E. coli*.
2. Must have a Thr-Thr motif that aligns with Ser312 and Thr313 in Mhp1.
3. RONN (Yang *et al.*, 2005) was used to predict protein disorder, if a protein sequence was predicted to be disordered the gene was discounted (Appendix 1).

Finally, 5 genes were selected for expression testing with a range of sequence identity (85% to 50%) to CodB found in *E. coli.*

To identify if IDT gene inserts and pWaldo were cut with restriction enzymes BamHI and NdeI the agarose gel in Figure 3.1 was run to confirm that pWaldo vector were cut, thus allowing inserts to be ligated into the plasmid.

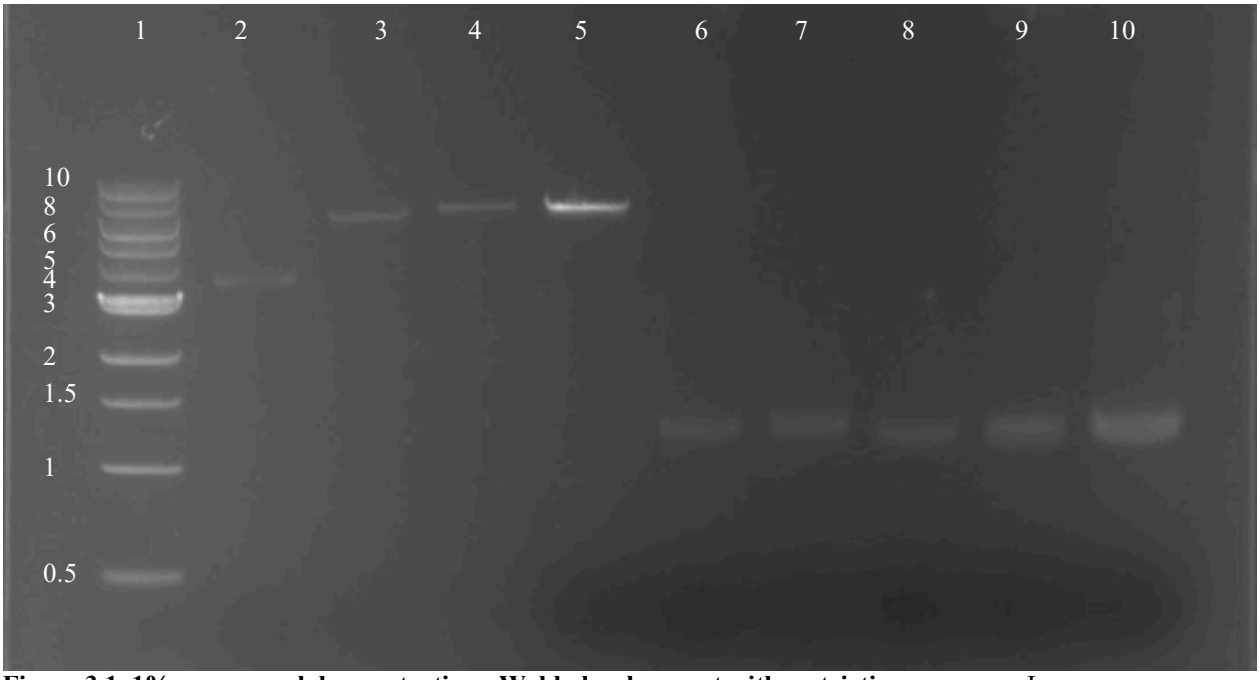


Figure 3.1. 1% agarose gel demonstrating pWaldo has been cut with restriction enzymes. Lanes were loaded as: 1) 1kBP ladder; 2) 474 ng of uncut pWaldo; 3) 474 ng of pWaldo cut with NdeI; 4) 474 ng of pWaldo cut with BamHI; 5) 3 µg of pWaldo cut with NdeI and BamHI; 6) CodB IDT gene (*E. coli*); 7) CodB IDT gene (*S.bovis*); 8) CodB IDT gene *(P. vulgaris*); 9) CodB IDT gene (*C. lundense*); 10) CodB IDT gene (*S. enterica*).

After the ligation reaction, reaction mixture was transformed into NEB5α *E. coli*, colonies were grown, plasmid was harvested, and the sequence of the construct was confirmed by Sanger sequencing. All constructs were correct, apart from *S. bovis* which had a single base pair mutation that resulted in an Ile278Phe mutation. This was corrected by site-directed mutagenesis and confirmed by Sanger sequencing.

3.2.2. Expression of CodB.GFP fusion

All constructs were transformed into LEMO21 (DE3) *E. coli* cells and 5 mL cultures were grown as described in Chapter 2 (Lee *et al.*, 2014). Expression was investigated and quantified in triplicate with Figure 3.2-3.6 showing amounts of expression and a 12% SDS gel for each construct to show that CodB.GFP fusion was expressed as opposed to GFP alone. All constructs showed expression. CodB.GFP is found between 40 kDa and 63kDa markers and runs faster than the combined 70 kDa that would be expected. However, due to membrane proteins running faster than their soluble counterparts this was expected.

For the *E. coli* and *S. bovis* construct it appeared that on average the best condition for expression was 0.25 mM rhamnose added when the culture was set up with no IPTG induction. For the *C. lundense* and *S. enterica* constructs the condition that on average gave the most protein expression per OD unit was 0.5 mM rhamnose added when the culture was started with no IPTG induction. Finally, for the *P. vulgaris* construct the best expression condition on average was determined to be when 0.1 mM rhamnose was added at the beginning of growth with 0.4 mM IPTG induction when the culture reached $OD_{600} = 0.5$.

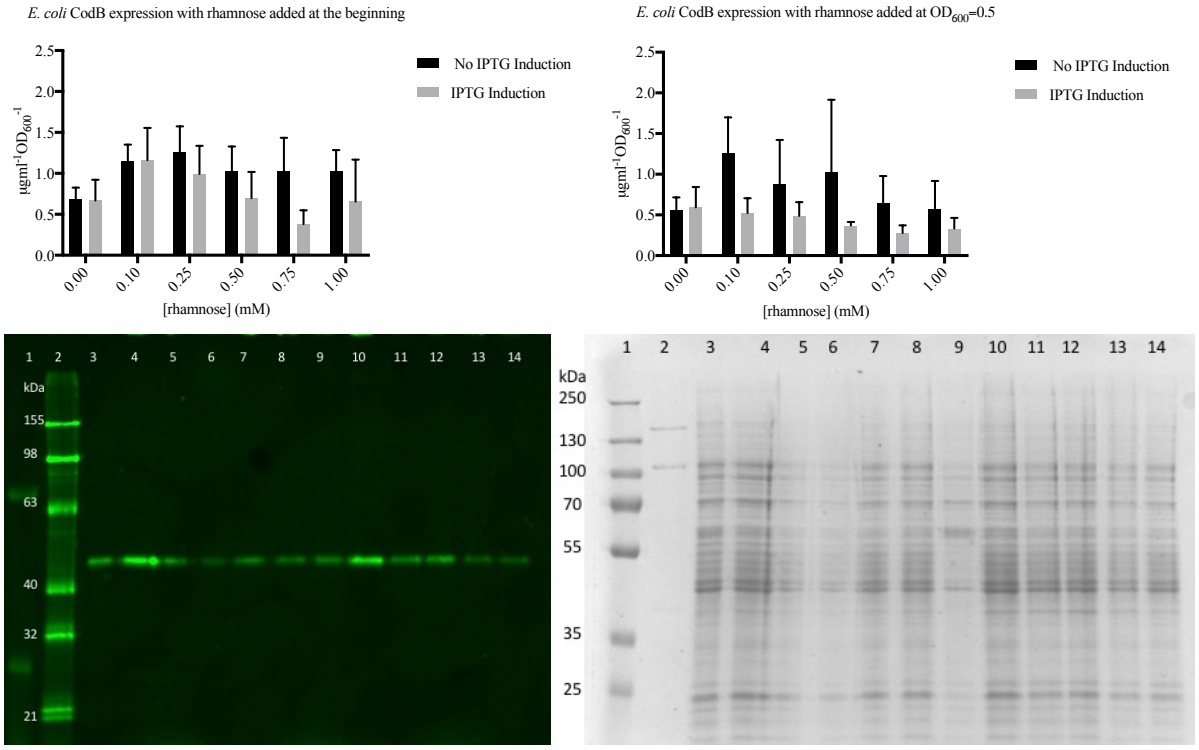


Figure 3.2. Quantifying CodB (*E. coli)* expression. Top left is quantifying GFP expression of CodB when rhamnose is added at the beginning, whilst top right is when rhamnose is added when OD_{600} is 0.5. All conditions were run on a 12% SDS PAGE gel. 1) Protein ladder; 2) Fluorescent protein ladder; 3) 0 mM rhamnose added; 4) 0.1 mM rhamnose added; 5) 0.25 mM rhamnose added; 6) 0.5 mM rhamnose added; 7) 0.75 mM rhamnose added; 8) 1 mM rhamnose added; 9) 0 mM rhamnose and 0.4 mM IPTG added; 10) 0.1 mM rhamnose and 0.4 mM IPTG added; 11) 0.25 mM rhamnose and 0.4 mM IPTG added; 12) 0.5 mM rhamnose and 0.4 mM IPTG added; 13) 0.75 mM rhamnose and 0.4 mM IPTG added; 14) 1 mM rhamnose and 0.4 mM IPTG added. Bottom left panel shows the GFP fluorescence whilst bottom left shows the Coomassie stained SDS PAGE gel.

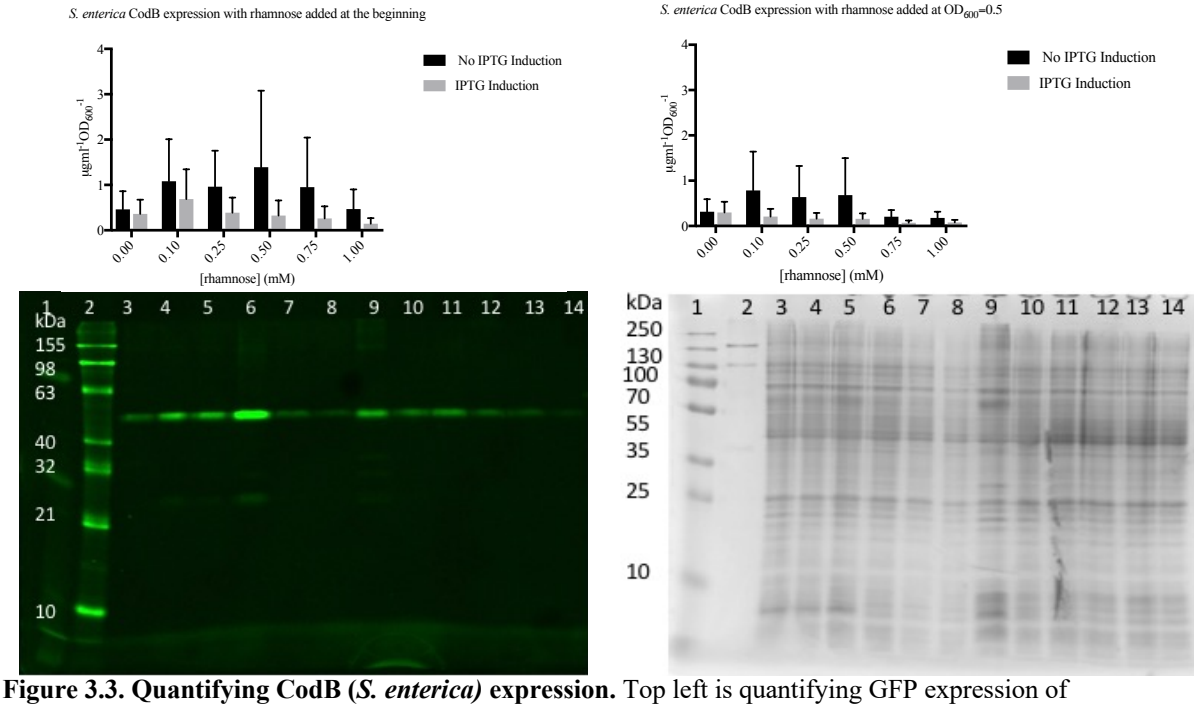


Figure 3.3. Quantifying CodB (*S. enterica)* expression. Top left is quantifying GFP expression of CodB when rhamnose is added at the beginning, whilst top right is when rhamnose is added when OD_{600} is 0.5. All conditions were run on a 12% SDS PAGE gel. Samples were loaded as Figure 3.2.

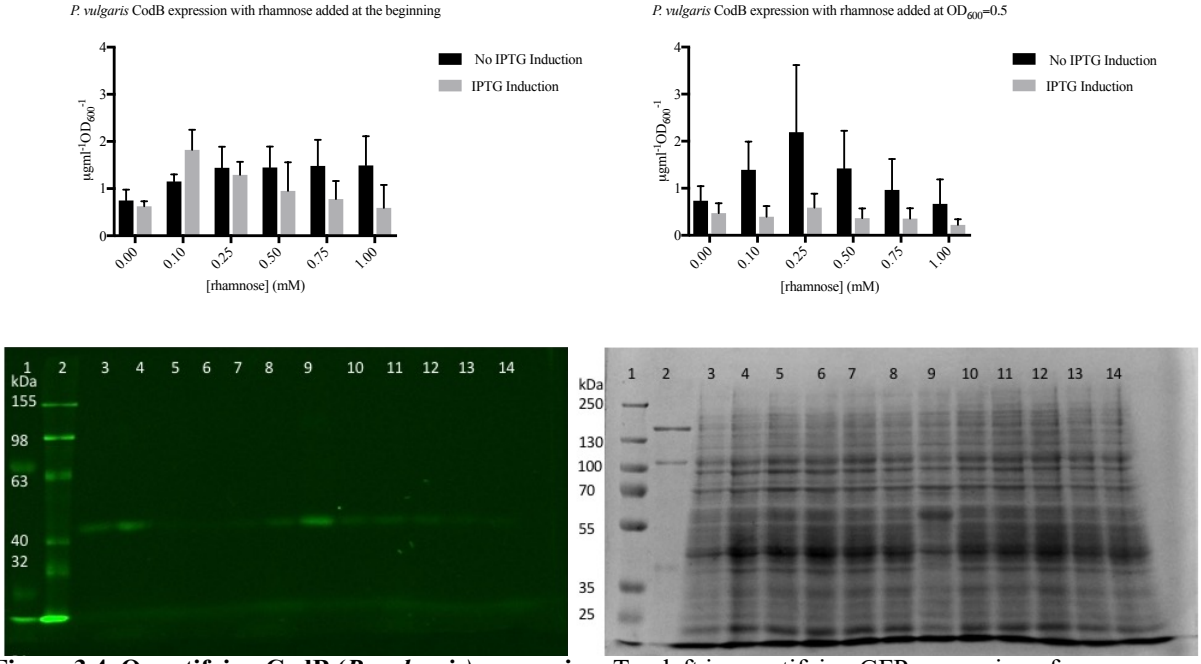


Figure 3.4. Quantifying CodB (*P. vulgaris)* expression. Top left is quantifying GFP expression of CodB when rhamnose is added at the beginning, whilst top right is when rhamnose is added when OD_{600} is 0.5. All conditions were run on a 12% SDS PAGE gel. Samples were loaded as Figure 3.2.

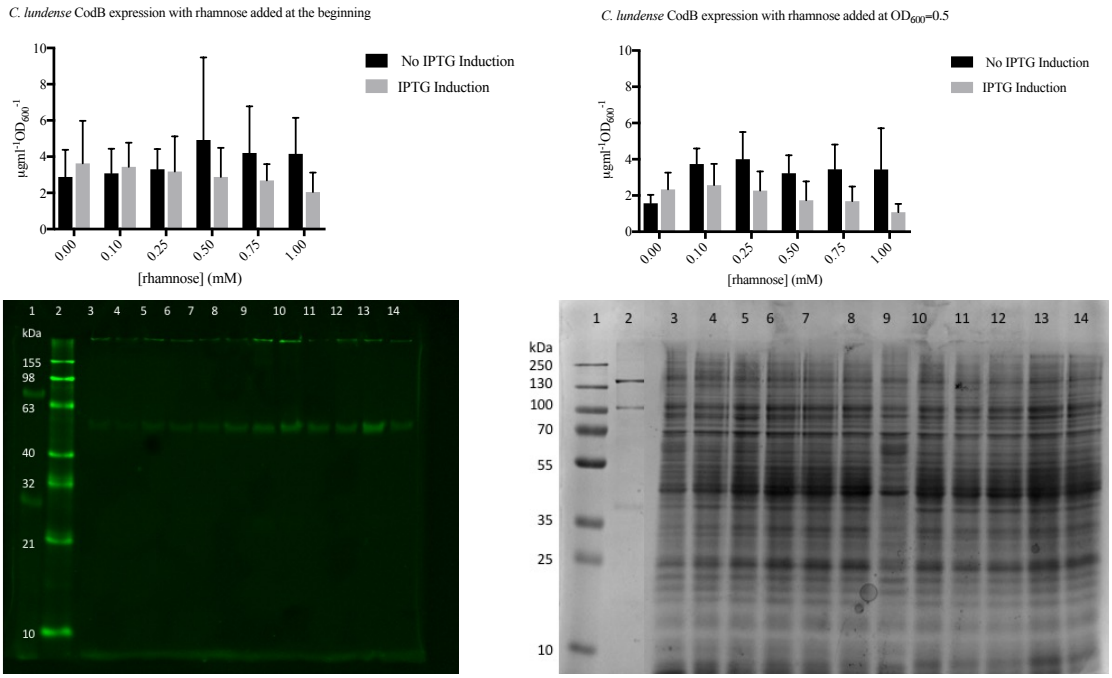


Figure 3.5. Quantifying CodB (*C. lundense)* expression. Top left is quantifying GFP expression of CodB when rhamnose is added at the beginning, whilst top right is when rhamnose is added when OD_{600} is 0.5. All conditions were run on a 12% SDS PAGE gel. Samples were loaded as Figure 3.2.

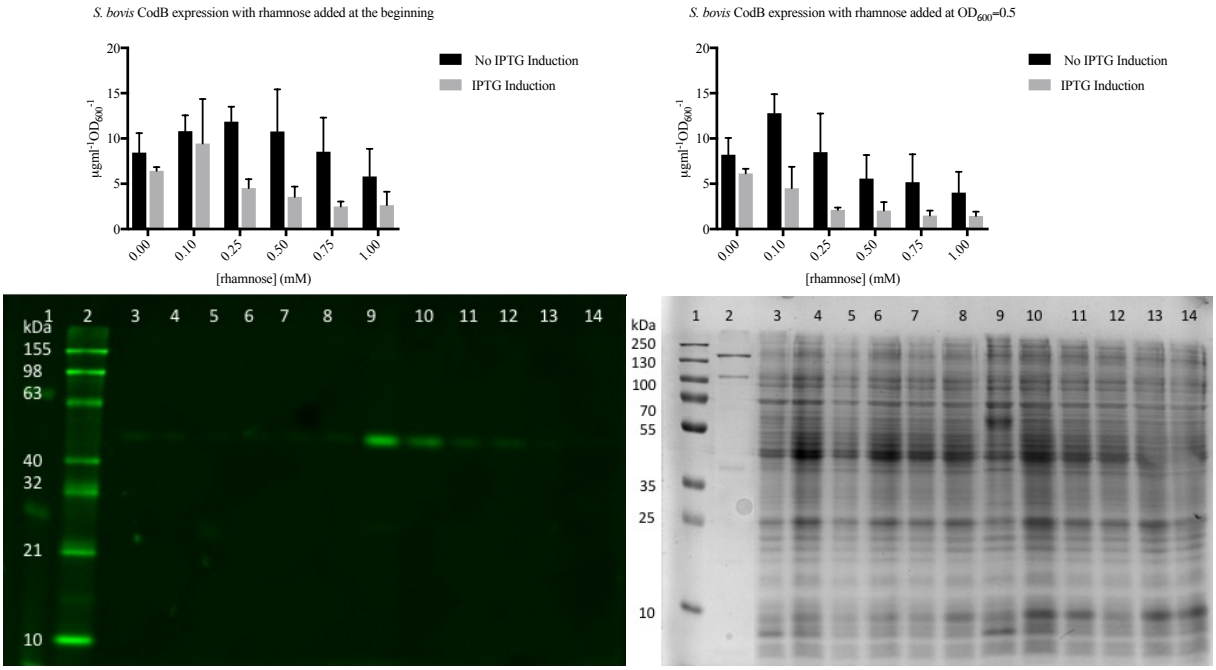


Figure 3.6. Quantifying CodB (*S. bovis)* expression. Top left is quantifying GFP expression of CodB when rhamnose is added at the beginning, whilst top right is when rhamnose is added when OD_{600} is 0.5. All conditions were run on a 12% SDS PAGE gel. Samples were loaded as Figure 3.2.

3.2.3. Solubility of CodB

After all constructs were shown to have some expression, the next step was test solubility to identify if constructs are viable for purification and down-stream investigation. Each construct was grown in 1 L PASM-5052 with their respective optimised expression condition (Section 3.2.2), then cells were harvested, and membranes isolated as described in Chapter 2. Solubilised membrane was subjected to FSEC to investigate the solubility of CodB (Kawate and Gouaux, 2006; Drew *et al.*, 2008). Initially CodB was solubilised in DDM before testing other detergents.

DDM is a non-ionic detergent that is often used as the first detergent for solubilising membrane to allow purification and is considered to be a mild detergent. Logically this was the first detergent to test as if CodB is not soluble in DDM micelle then it is not worth carrying this construct forward for purification. Figure 3.7 shows that the *P. vulgaris* and *S. enterica* constructs solubilised in DDM by their ability to form defined mono-dispersed peaks with no suggestion of aggregation. However, the remaining constructs did not. Therefore, *P. vulgaris* (Figure 3.8) and *S. enterica* (Figure 3.9) constructs were then tested in different detergents, DM, NM, and LDAO.

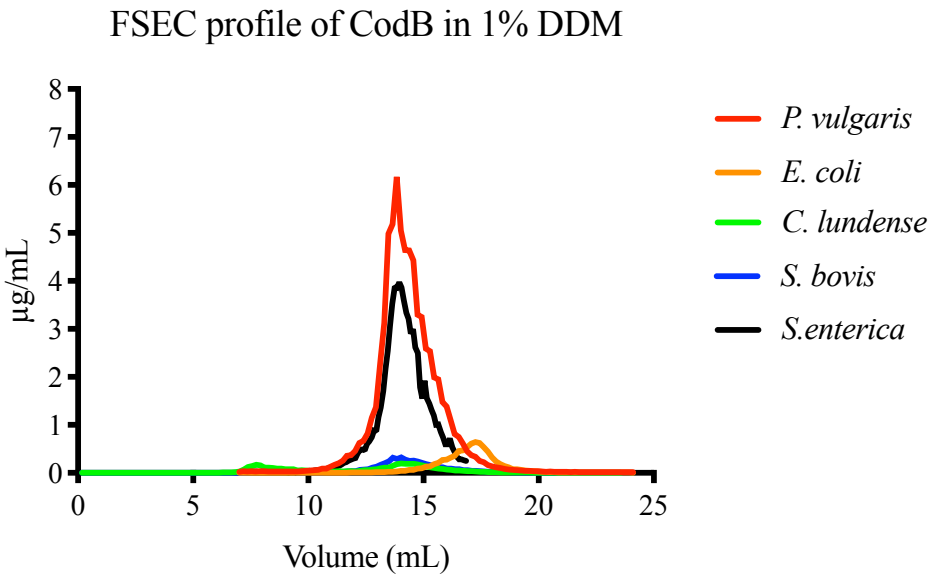


Figure 3.7. Concentration of CodB plotted against elution volume when solubilised in DDM. *P. vulgaris* and *S. enterica* CodB looked the best in terms of solubilisation and showed an obvious peak at approximately 14 mL showing that they both are soluble in DDM. Other constructs did not show this.

The *S. enterica* CodB construct as shown in Figure 3.8 is soluble in DM, and NM producing a single mono-dispersed peak with little aggregation but is not soluble in LDAO, no peak is visible for LDAO. The *P. vulgaris* construct in Figure 3.9 is the same and is soluble in DM and NM but not in LDAO. DM and NM have the same overall structure as DDM but varying in the length of the acyl chain. DDM has the longest chain, and the largest detergent micelle whilst NM has the shortest chain and the smallest detergent micelle.

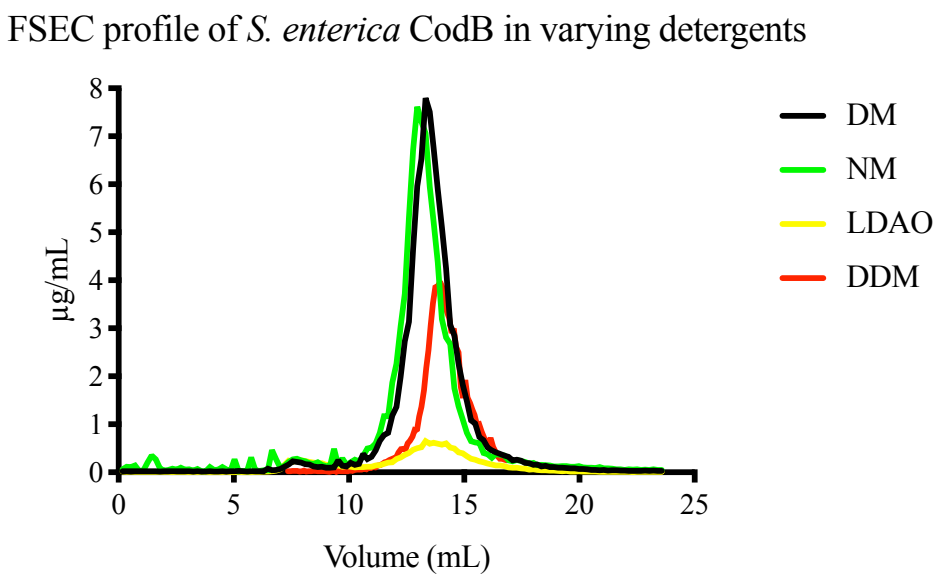


Figure 3.8. Concentration of *S. enterica* CodB plotted against elution volume when solubilised in different detergents. CodB solubilises in DM, NM, and DDM to produce a single mono-dispersed peak but does not solubilise well in LDAO.

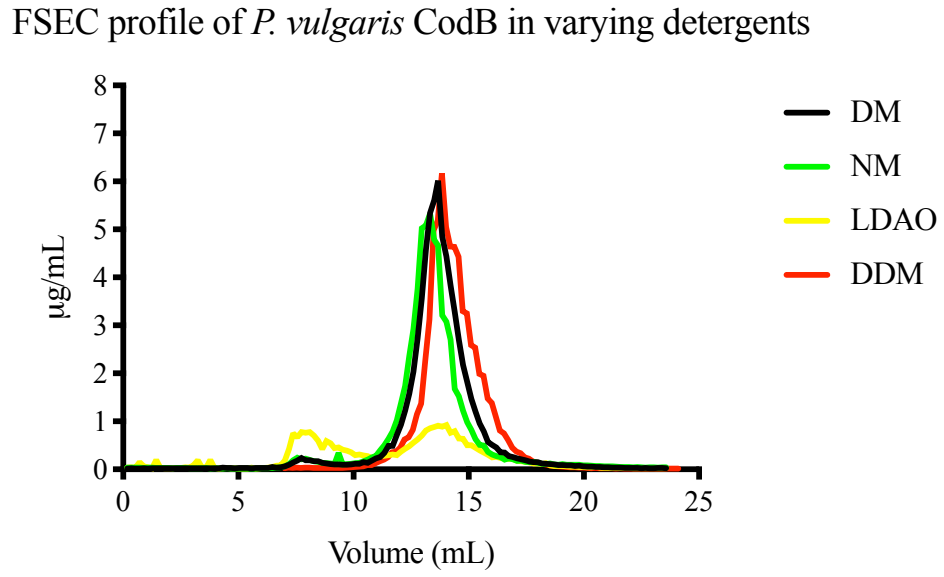


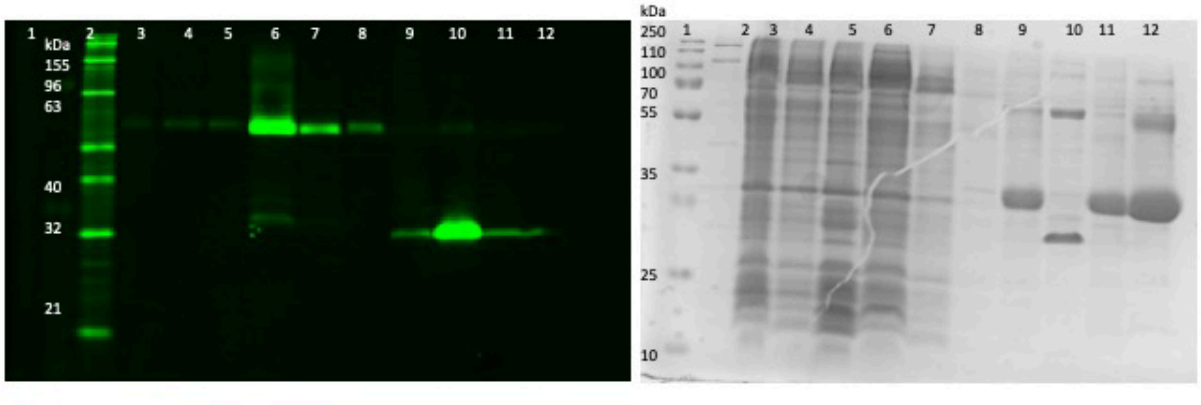
Figure 3.9. Concentration of *P. vulgaris* CodB plotted against elution volume when solubilised in different detergents. CodB solubilises in DM, NM, and DDM to produce a single mon-dispersed peak but does not solubilise well in LDAO.

3.2.4. Purification of CodB

Based on expression trials and solubility testing, CodB from *P. vulgaris* and *S. enterica* were suitable candidates for testing purification. CodB.GFP from *P. vulgaris* fusion was expressed in LEMO21 (DE3) *E. coli* then membranes isolated and solubilised; allowing CodB to be purified using His-affinity chromatography as described in the Chapter 2. Post-3C protease on column cleavage, CodB was applied

to a new His-affinity column to remove any residual 3C or GFP. Fractions were analysed using SDS-PAGE gels (Figure 3.10) showing that CodB running with a molecular weight (MW) of approximately 32 kDa compared to a predicted MW of 43 kDa. If CodB needed to be analysed by SEC or exchanged into a different buffer as described as above then protein was concentrated and subjected to SEC, CodB in a DDM micelle eluted from the S200 column at ~10 mL, compared to ~13 mL when in the smaller NM micelle (Figure 3.11). Addition of 1 mM cytosine during purification allowed a greater yield of CodB to be purified.

Figure 3.10. Purification of CodB. GFP fluorescent SDS PAGE gel (left) and Coomassie (right) Lane 1- Protein Ladder, Lane 2- Fluorescent Ladder, Lane 3- Membrane, Lane 4- Solubilized membrane, Lane 5- Post ultracentrifugation, Lane 6- Ni NTA flowthrough, Lane 7- 20 mM imidazole



wash, Lane 8- 30 mM imidazole wash Lane 9- Post-3C cleavage wash, Lane 10- Post-3C cleavage elution, Lane 11- Post-Reverse IMAC, Pre-SEC, Lane 12- Post SEC.

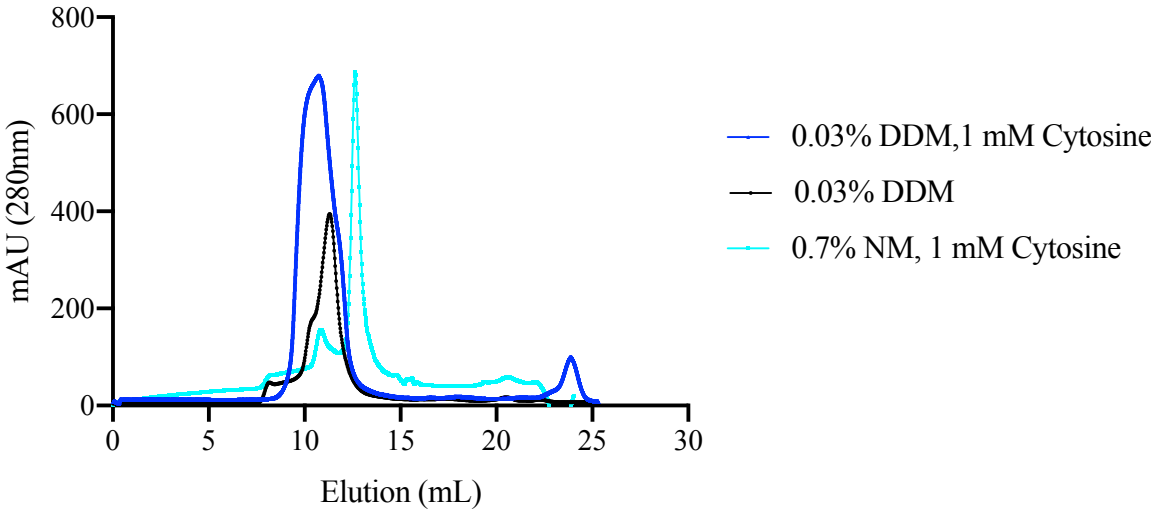


Figure 3.11. SEC profile of CodB in various conditions on a S200 column. All runs were done in 20 mM TRIS pH 7.5, 150 mM NaCl. Addition of 1 mM cytosine provided a stabilising effect allowing a greater yield of purification.

3.2.5. Circular Dichroism of CodB

Circular dichroism (CD) spectra can be a quick and easy method of evaluating the secondary structure of pure protein and check that protein is correctly folded before further characterisation. Purified CodB was investigated using circular dichroism (Figure 3.12) in the presence and absence of cytosine, as it was a concern that the absorbance profile of cytosine could impact the CD spectra. From this, CodB demonstrates α-helical structure by the presence of a positive band at 193 nm and negative bands at 208 nm and 222 nm; the addition of cytosine does not change the secondary structure (Figure 3.12).

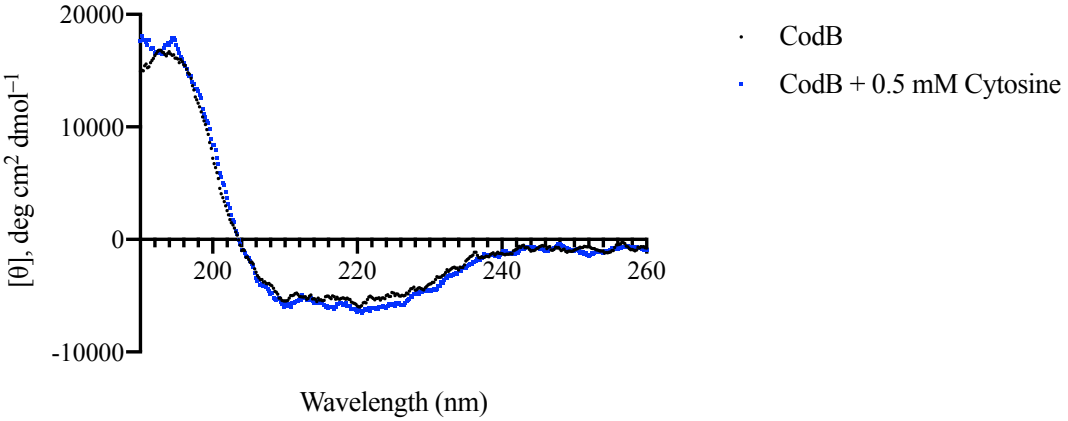


Figure 3.12. UV CD of purified wt CodB. Purified CodB (2 mg/ml) was buffered-exchanged into 20 mM TRIS pH 7.5, 150 mM $NaSO_4$, 0.03% DDM with and without 0.5 mM cytosine.

3.3. Conclusions and Discussion

As described above, two constructs (CodB from *P. vulgaris* and *S. enterica*) of CodB have been successfully expressed in reasonable quantities, are soluble in a variety of detergents, and are able to be successfully purified to move forward for structural and functional characterisation.

Interestingly, the best expression conditions varied across the constructs. The Memstar ™ approach used for protein expression has been optimised for membrane protein production (Lee *et al.*, 2014). Rhamnose is used to dampen overexpression in Lemo21 (DE3) (Schlegel *et al.*, 2012), the PASM-5052 (Studier, 2005) media is auto-inducing and then IPTG is used to induce further overexpression. Membrane protein over-expression is often toxic due to a saturation of the membrane protein biogenesis pathways. Rhamnose allows titratable expression and can be used to set the expression intensity and allow optimum overexpression. This is often used in combination with an auto-inducing media and IPTG, but as shown in section 3.2.2 all constructs except the *P. vulgaris* construct preferred no IPTG induction, all constructs required rhamnose for optimal expression. PASM-5052 media only begins to induce protein expression when the OD_{600} ~1. From this it can be postulated that CodB overexpression early at OD_{600}~ 0.5 is undesirable, but once the cells begin to reach a higher OD they can tolerate more expression. *P. vulgaris* construct did require IPTG induction for optimal protein production but interestingly did not express much better than any other construct highlighting the importance of rhamnose addition.

After expression trials all constructs were grown in a larger culture, membranes isolated and solubilised in DDM. As described in Section 3.2.4 two constructs (*P. vulgaris* and *S. enterica*) were shown to be solubilised in DDM, but the others were not. These two constructs were then shown to be soluble in DM, and NM, but not LDAO, making suitable them for purification, crystallisation, and functional characterisation. DDM is a very "soft" detergent and if CodB could not be solubilised in DDM then it would not be soluble in any other detergent, therefore this was a good starting detergent to test which constructs were worth testing further. However, crystals grown in DDM are unlikely to give high-resolution diffraction and therefore other detergents would be better for crystallisation. DM and NM are very

similar to DDM, they only vary in the length of their tail. The length of the chain can make a huge difference as crystals grown in NM generally give higher resolution diffraction when compared to DDM and DM due to the smaller detergent micelle. It is useful that CodB can be solubilised in multiple detergents. In contrast, LDAO is a harsher detergent and CodB was not soluble in LDAO. n-Octyl-β-D-glucoside (β - OG) is another detergent that is considered to be "harsher" than DDM, DM, and NM, whilst it is possible that CodB could be soluble and stable in OG when it is not in LDAO it is unlikely and was not worth testing with FSEC.

The expression trials and FSEC in combination have shown that *S. enterica* and *P. vulgaris* constructs are worth purification. They were both purified by Immobilized metal affinity chromatography (IMAC), then the GFP and His-tag cleaved by 3C protease, and further purified by SEC as described above. Initial purifications worked reasonably well with CD spectra confirming that CodB is folded correctly with α-helical structure, but there were issues with protein stability, at higher concentrations (>10 mg/mL) CodB would precipitate overnight at 4 ^{o}C. Precipitation could be reduced if SEC was performed in the presence of 1 mM cytosine but did not solve the issue completely. Addition of 1 mM cytosine at all stages of the purification did solve this issue. Problematically, for future functional work CodB needs to be purified without substrate and rapid purification without the SEC, then using protein fresh or storing at -80 ^{o}C was the only viable option.

The flexibility in detergent solubilisation is useful, as it provides more starting points within the crystallisation space and allows more coverage of conditions as less as hopefully allowing higher resolution diffraction if crystals grow. This detergent flexibility is also useful for the functional work as it allows more coverage of conditions. DDM is often used as the initial solubilisation but is too large a micelle to work well for crystallisation in hanging drop or sitting drop and is often better for lipidic cubic phase (LCP) as described in later chapters. However, due to the ability of CodB to be NM or DM micelles this also means that sitting drop and handing drop crystallisation is a viable attempt as well as LCP increasing the chance of CodB crystallisation. In parallel, functional assays to determine kinetic properties of CodB will be done as described in later chapters.

Chapter 4-Structural Characterisation of CodB

4.1 Crystallisation, Data collection, and Structure Solution of CodB

For clarity, unless stated otherwise CodB will refer to CodB from *P. vulgaris*. Following purification in the presence of 1 mM cytosine, CodB was buffer exchanged into NM then subjected to vapour diffusion crystal screening, crystals (Figure 4.1) grew to maximum dimension in 2 weeks, were harvested and flash-frozen. Vapour diffusion crystals diffracted to 3.4 Å and processed to 6 Å with severe anisotropy. Following this, crystals were subjected to dehydration before x-ray analysis, and CodB was also subjected to detergent and additive screening during crystallisation trials; neither of these approaches improved X-ray diffraction. Mhp1 in all conformations was used as the molecular replacement model, the 4-helix bundle and hash-motif of Mhp1 were separated out and used as separate models along with Acrimbaldo and Morda (Rodríguez *et al.*, 2009; Simpkin *et al.*, 2018). This vapour diffusion crystal dataset was not able to be solved, due to issues with poorly diffracting crystals and dataset not being able to solved possibly due to poor-quality data, CodB was subjected to LCP crystal trials. LCP crystallisation has been extremely successful, especially for GPCR's and allows the protein to form better crystal contacts due to the absence of detergent (Pebay-Peyroula *et al.*, 1997; Cherezov, Abola and Stevens, 2010).

CodB was purified in the presence of 1 mM cytosine, concentrated to 32 mg/mL in DDM and extruded into LCP with monoolein and subjected to crystallisation screening. CodB crystals appeared in a variety of conditions and took between 5 days and 3 weeks to grow to maximum dimension (Figure 4.1). Crystals were harvested and flash-frozen in liquid nitrogen. Multiple datasets at various resolutions were collected at beamline i24 at Diamond Light Source.

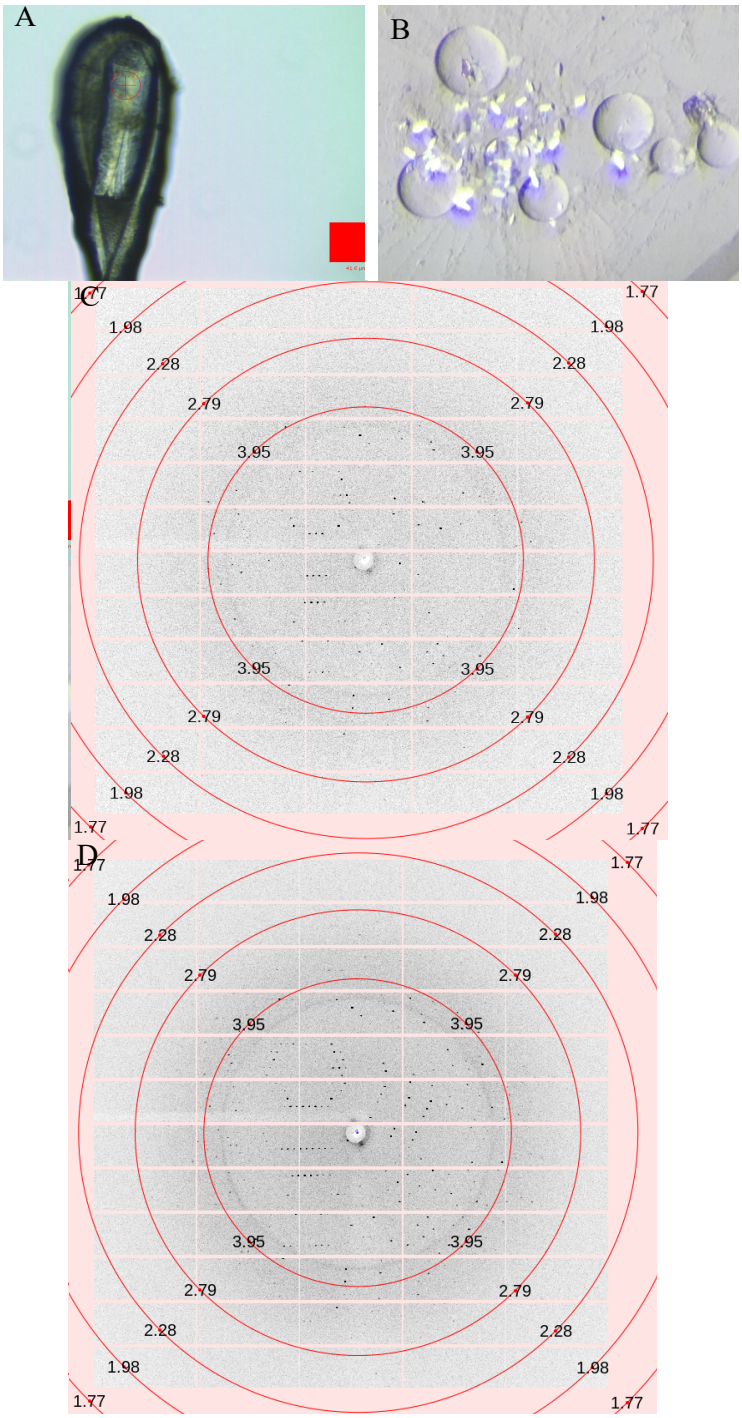


Figure 4.1. CodB crystals and x-ray diffraction. Panel A: Vapour diffusion crystal grown in 0.2 M MOPS 6.5, 0.025 M $MgCl_2$, 32% Peg400 mounted on beamline i24 at DLS. Panel B: LCP crystals of CodB in complex with cytosine before flash freezing. Panel C: Single diffraction image of CodB complexed with cytosine in LCP with maximum diffraction detected at 1.7 Å. Panel D: Diffraction of CodB complexed with cytosine in LCP with maximum diffraction detected at 1.7 Å, images stacked to 10.

Initially, a structure of CodB was solved with data that was processed to 3.6 Å using an outward-facing model of CodB built by Rosetta was used as a model for molecular replacement (Dimaio *et al.*, 2013). The outward-open structure of Mhp1 (PDB code: 2JLN) was manually rebuilt into CodB, as described in Chapter 2. When better diffraction data was collected, with spots detected at a maximum resolution of 1.7 Å, data was processed and scaled with a cut-off of 2.4 Å (Figure 4.1) (Table 4.1), and the structure of CodB at 3.6 Å was refined using the 2.4 Å dataset. Data was processed in space group C2 2 21 to a final R_{free} = 0.246 (3.s.f) with a structure of CodB bound to cytosine to a resolution at 2.4 Å. A summary of datasets collected is in Table 4.1, with model building and refinement statistics in Table 4.2. The structure discussed in detail is from crystals that grew in 0.1 M sodium cacodylate pH 6.5, 0.45 M NaCl, 39% Peg400.

This 2.4 Å structure of CodB was used a model for molecular replacement for the vapour diffusion dataset but did not solve the data. An inward-open model of CodB was built based on the inward-open structure of Mhp1 (PDB code: 2X79) and was also used as search model for molecular replacement, but this approach didn't work either. This inward model is discussed into more detail in section 4.5.

Crystallisation method	Vapour Diffusion	LCP		
Crystallisation condition	0.2 M MOPS pH 6.5 0.025 M $MgCl_2$ 32% Peg400	0.1 M HEPES pH 7 0.4 M NaCl 20% Peg400	0.1 M HEPES pH 7 0.15 M NaCl 30% Peg400	0.1M Sodium Cacodylate pH 6.5 0.45 M NaCl 39% Peg400
Space Group	P622	C222	C2221	C2221
Beamline	I24	I24	I24	I24
Detector	Pilatus3 6M	Pilatus3 6M	Pilatus3 6M	Pilatus3 6M
Wavelength (Å)	0.9686	0.9688	0.9686	0.9686
Resolution (Å)	123.0-4.00 (4.32-4.00)	52.3-3.60 (3.66-3.60)	70.1-3.2 (3.28-3.23)	73.1-2.39 (2.48-2.39)
Cell Dimensions a, b, c (Å) α, β, γ (°)	142.0, 142.0, 411.8 90.0, 90.0, 120.0	108.6, 209.2, 102.5 90.0, 90.0, 90.0	107.6, 207.9, 103.2 90.0, 90.0, 90.0	108.2, 209.0, 102.5 90.0, 90.0, 90.0
Number of Unique reflections	21566 (4348)	13843 (636)	14731 (630)	33490 (1557)
Completeness (%)	99.7 (99.9)	99.4 (91.9)	99.7 (99.0)	99.1 (92.8)
R_{merge}	1.0 (-250)	0.4 (2.0)	0.2 (13.6)	0.3 (3.9)
R_{pim}	0.1 (0.3)	0.3 (0.9)	0.4 (4.2)	0.1 (0.1)
CC1/2	1.0 (0.7)	0.8 (0.2)	0.8 (0.3)	1 (0.5)
I/σ (I)	1.0 (0.0)	2.7 (1.1)	4 (0.9)	14.1 (1.7)

Table 4.1 Data collection statistics of CodB. Values in parentheses refer to data in the highest resolution shell.

	CodB:cytosine 3.6 Å	CodB:cytosine 2.4 Å
Wavelength (Å)	0.9686	0.9686
Resolution range	52.4 - 3.6 (3.7 - 3.6)	58.6 -2.4 (2.5-2.4)
Space group	C 2 2 21	C 2 2 21
Unit cell a, b, c (Å) α, β, γ (°)	 108.5, 209.5, 102.8 90.0, 90.0, 90.0	 108.2, 209.0, 102.5 90.0, 90.0, 90.0
Total reflections	72321 (7329)	580190 (49529)
Unique reflections	13943 (1337)	45754 (4436)
Multiplicity	5.2 (5.3)	12.7 (11.0)
Completeness (%)	95.6 (97.0)	99.4 (98.6)
Mean I/σ (I)	3.5 (0.9)	11.6 (1)
Wilson B-factor	34.5	42.0
R_{merge}	0.5 (1.2)	0.3 (2.8)
R_{meas}	0.6 (1.4)	0.4 (2.9)
R_{pim}	0.3 (0.6)	0.1 (0.9)
CC1/2	0.7 (0.5)	1.0 (0.5)
CC*	0.9 (0.8)	1.0 (0.8)
Reflections used in refinement	13353 (1337)	45487 (4428)
Reflections used for R-free	1335 (133)	2208 (215)
R-work (%)	0.3 (0.4)	20.1 (26.4)
R-free (%)	0.3 (0.4)	24.6 (33.4)
CC(work)	0.6 (0.4)	0.9 (0.8)
CC(free)	0.7 (0.3)	0.9 (0.7)
Number of non-hydrogen atoms Macromolecules Ligands Solvent Protein residues	5711 5695 16 0 793	6139 5899 175 65 803
RMS(bonds)	0.01	0.008
RMS(angles)	1.4	0.9
Ramachandran favoured (%)	95.7	98.8
Ramachandran allowed (%)	3.7	1.1
Ramachandran outliers (%)	0.6	0.1
Rotamer outliers (%)	7.7	1.8
Clashscore	23.2	8.74
Average B-factor Macromolecules Ligands Solvent	19.98 20.04 0.50	52.52 52.00 71.15 49.49

Table 4.2. Refinement Statistics for CodB:cytosine complex. Values in parentheses refer to data in the highest resolution shell.

4.2. LCP Structure of CodB in an Outward-Open Conformation

The final structure of CodB:cytosine complex is an asymmetric unit (AU) composed of two protein chains arranged inverted relative to each other (Figure 4.2). Each protomer is composed of 12 TM helices, with TM1-10 exhibiting the distinctive LeuT-fold with a single molecule of cytosine and a sodium ion bound per protein chain. Analysis using PISA, calculated that 24 residues (6%) were found at the interface between chain A and B, with no hydrogen bonds, salt or di-sulphide bridges present between chains. The combination of this analysis and in the inverted orientation relative to each other means this interface is most likely a crystallographic packing artefact and not biologically relevant.

Each monomer, chain A pictured in Figure 4.2 with sequence and topology in Figure 4.3, has 12 transmembrane helices. TM1-5 (residues 21-176) fold into the first inverted repeat connected by a 17 amino acid extracellular loop to the second repeat of TM6-10 (residues 194-354). This second inverted repeat is connected to TM11 and TM12 via a 19 residue intracellular loop. These inverted repeats can be aligned to each other with an RMSD of 4.33 (3 s.f.) for 144 residues in each inverted repeat (Figure 4.4).

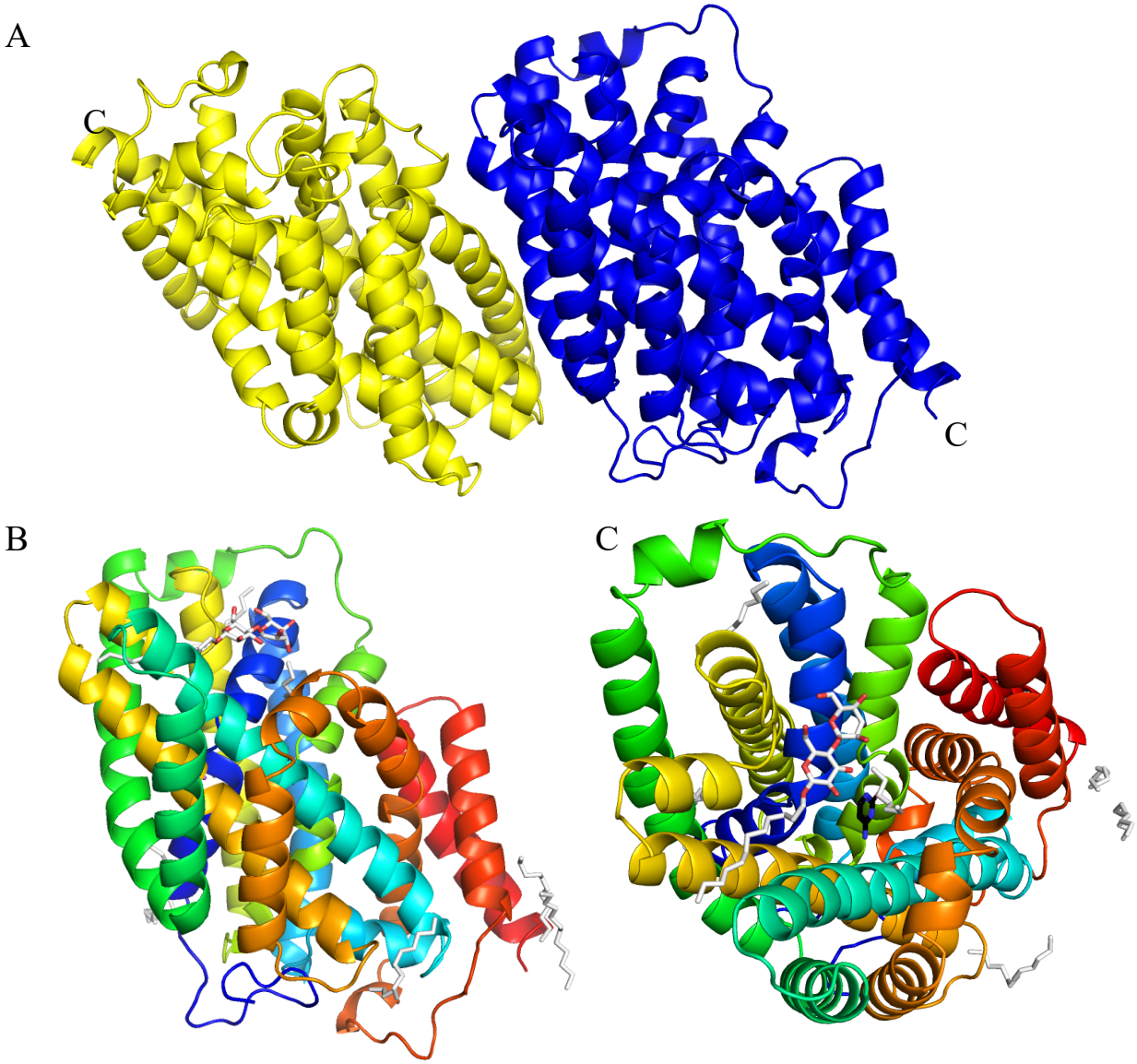


Figure 4.2. Cartoon representation of CodB AU and individual protomer. Panel A. AU of CodB: cytosine complex, individual protein chains are found inverted relative to each other with crystal contacts of 7 residues in TM5 forming the interface, c-terminus is labelled for orientation. Panel B. Chain A of CodB bound to cytosine viewed from the plane of the membrane, cytosine represented in stick form in black, and a sodium ion as a magenta sphere. Pictured in white are detergent and lipid molecules. N-terminus is blue whilst the c-terminus is in red. Panel C. Birds eye view of CodB from the extracellular.

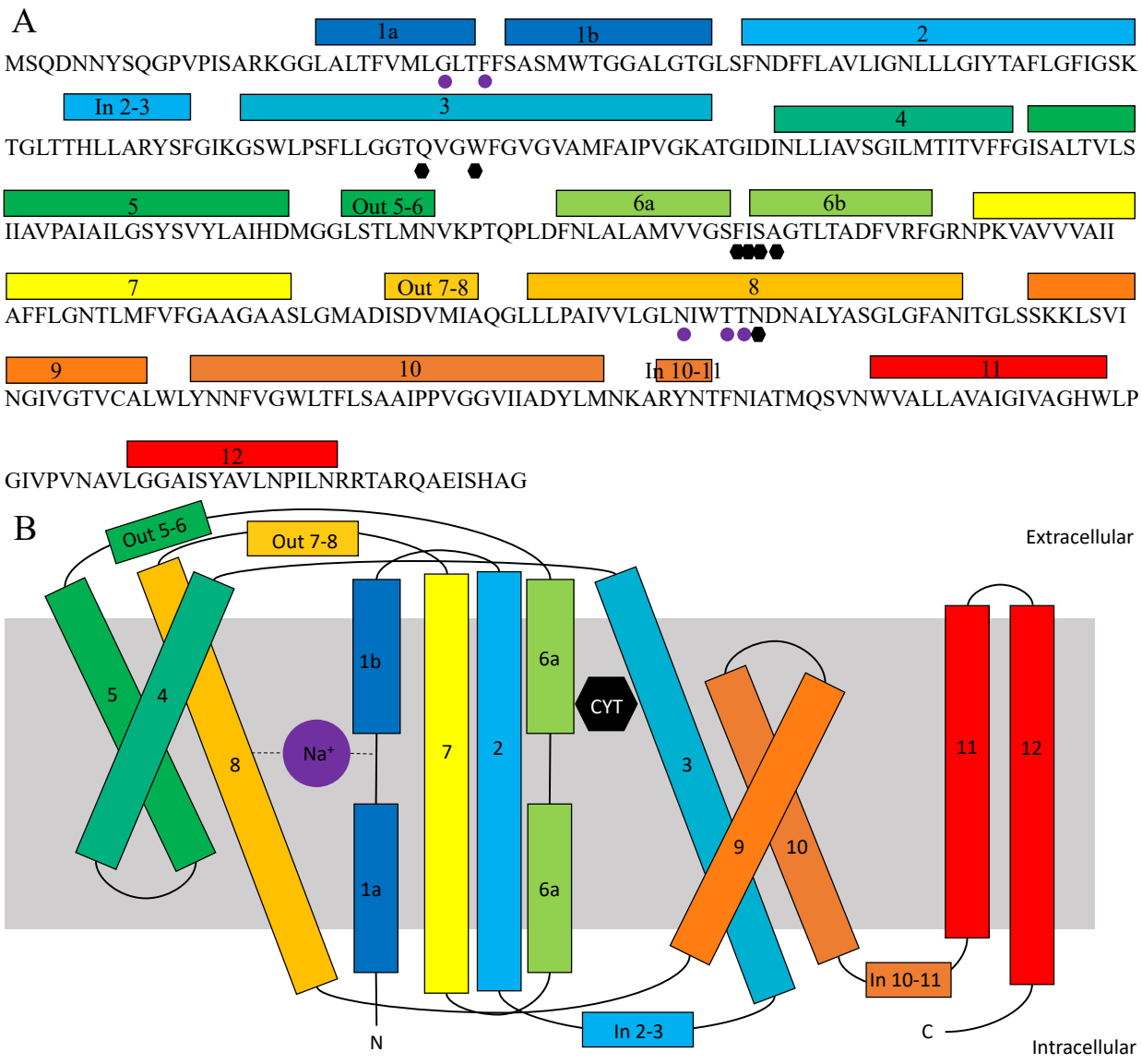


Figure 4.3. Amino acid sequence of CodB and secondary structure topology map. Panel A. Sequence of CodB, TM helices are coloured with the corresponding colouring as cartoon representation in Figure 4.2. Residues implicated in solute binding are denoted with black hexagons with sodium binding with a magenta circle. Panel B. Topology map of CodB, helices are depicted as cylinders with corresponding colouring as cartoon representation in Figure 4.2. Unwound sections are illustrated as lines. Membrane is depicted in grey.

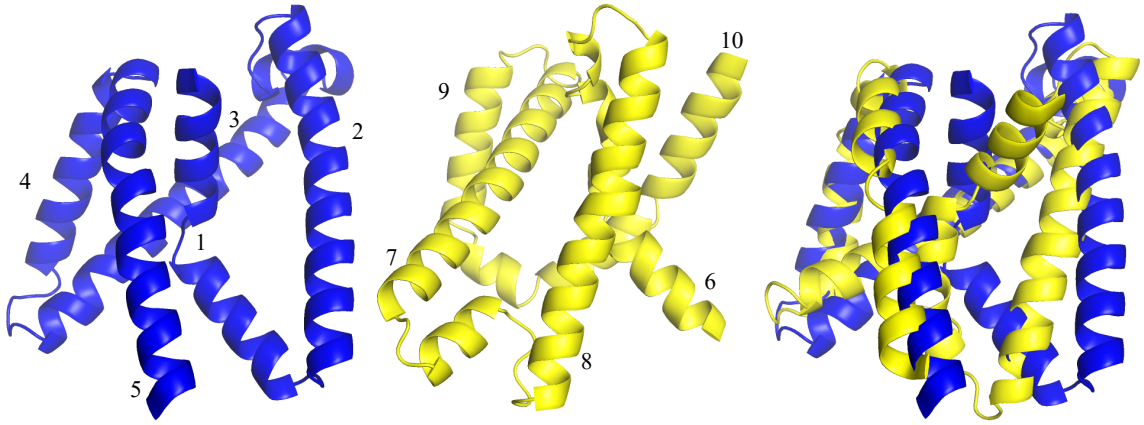


Figure 4.4. CodB inverted repeats. From left to right, in blue, TM1 to TM5 with TM6-TM10 in yellow. Inverted repeats are overlaid onto each other with an RMSD of 4.33.

This CodB:cytosine complex is found in an outward-facing conformation with the solute binding pocket solvent accessible (Figure 4.5). CodB:cytosine complex has average B-factors of 53 with a maximum of 124, a minimum of 25. Broadly, B-factors are lower in the 4-helix bundle and around the solute and cation binding sites with loops having highest B-factors (Figure 4.5). TM10, 11, and 12 have higher B-factors than the rest of the structure, TM10 has less defined electron density and was harder to build than the other helices. The higher B-factors and less-defined electron density correlates with flexibility of TM10, the putative extracellular gate.

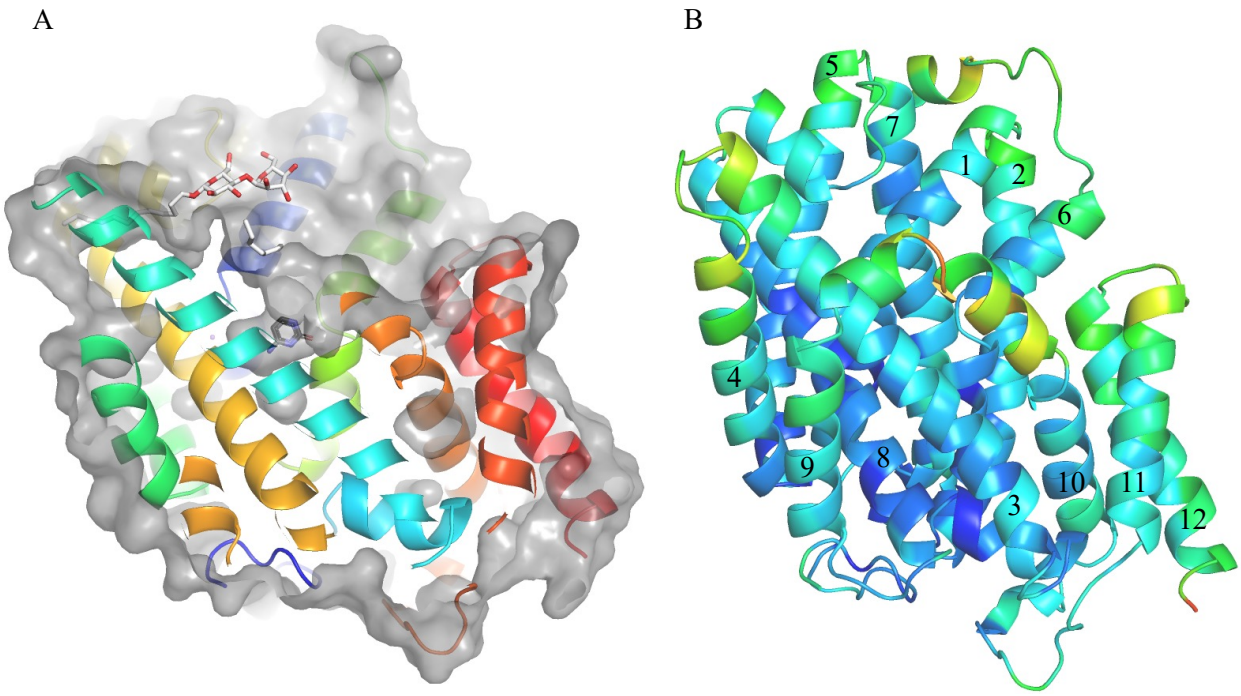


Figure 4.5. CodB is found in an outward-open conformation. Panel A. CodB (coloured as Figure 4.2) with surface (grey) cut away to show that the cytosine binding site is solvent accessible to the extracellular space. Panel B. CodB coloured by B-factors, blue denotes low B-factors, with red having high B-factors.

In both chains, the cytosine binding site (Figure 4.6 and Figure 4.7) shows cytosine is bound at the interface between the hash-motif and 4 helix bundle forming interactions with TM3,6, and 8 by forming a Pi-stacking interaction with residues Trp108 (TM3) and Phe204 (TM6). An Fo-Fc map in panel C Figure 4.6 shows that cytosine fits well into the additional density in the putative solute binding site. From the electron density it is not possible to precisely orientate the cytosine within the binding pocket, the orientation depicted in Figure 4.6 and Figure 4.7 was modelled to optimise hydrogen bonds. Gln105 (TM3) provides specific hydrogen bonding patterns that orientates cytosine within the binding pocket. The carbonyl (position 2) of cytosine is in proximity to be hydrogen bonding to Ala207. Two water molecules can be seen in the site, the first coordinating with the carbonyl of cytosine and the amides of Phe204 (TM6), Ile205 (TM6) and Ser206 (TM6), the second water molecule is hydrogen bonding with primary amine (position 4) of cytosine and the main chain carbonyl of the Trp108 (TM3), and carbonyl of the carboxamide side chain of Asn280 (TM8).

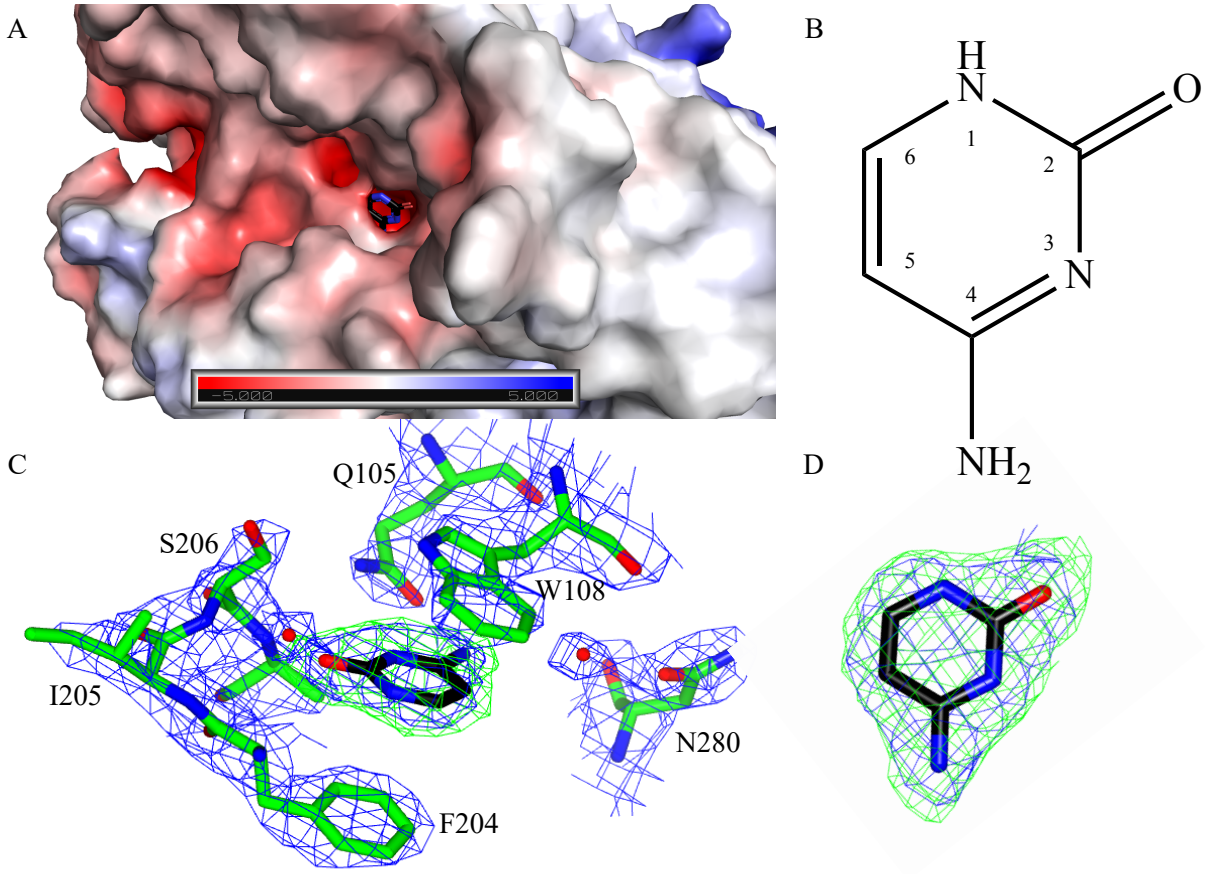


Figure 4.6. Cytosine binding site of CodB. Panel A: Electrostatic surface of CodB with cytosine bound in a solvent accessible polar pocket. Panel B: Cytosine molecular structure with numbering. Panel C: Cytosine and binding site modelled in Fo-Fc map, contoured at 3 σ in green with 2Fo-Fc map in blue contoured at 1 σ. Panel D: Cytosine modelled into electron density without surrounding amino acids.

In chain A, shown in Figure 4.7 panel A, the residue Ser206 is facing away from the cytosine, but is in a position capable of coordinating cytosine via an intermediatory water molecule. Whilst it is in a different rotamer in chain B, shown in panel B, the serine residue could be in position to coordinate with the carbonyl moiety of cytosine although looking at the distances between the cytosine and the hydroxyl of Ser206 it looks unlikely. This Ser206 residue is fairly conserved across the characterised NCS1 family, alternatively a threonine (Appendix 2).

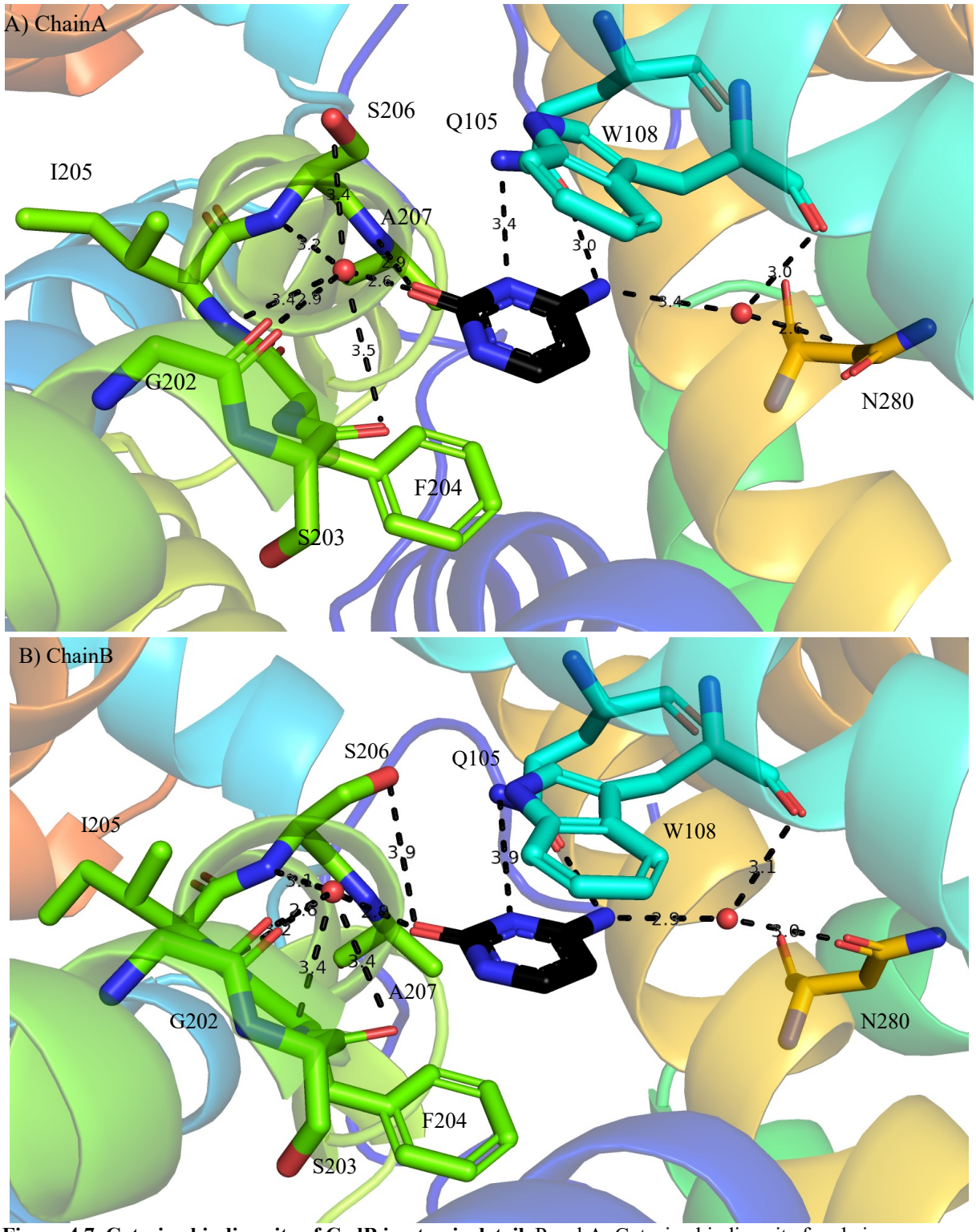


Figure 4.7. Cytosine binding site of CodB in atomic detail. Panel A: Cytosine binding site for chain A, Panel B: Cytosine binding site for chain B. Ser206 is in different rotamer conformations in each protomer whilst the other surrounding side chains are in the same position in both protein chains. Hydrogen bonds are depicted as black dashed lines.

The sodium binding site is found coordinating between the hash-motif and four-helix bundle interacting with the carbonyls of residues Gly29, Phe32 in TM1, Asn275 in TM8 and the hydroxyl side chains of Thr278 and Thr279 in TM8. This aligns with the sodium site in Mhp1, vSGLT, and the Na2 sites in LeuT, BetP, MhsT, and SiaT, dDAT, and SERT. The bond lengths are between 2.2-2.7 Å which are too short for a water molecule or potassium ion (Figure 4.8).

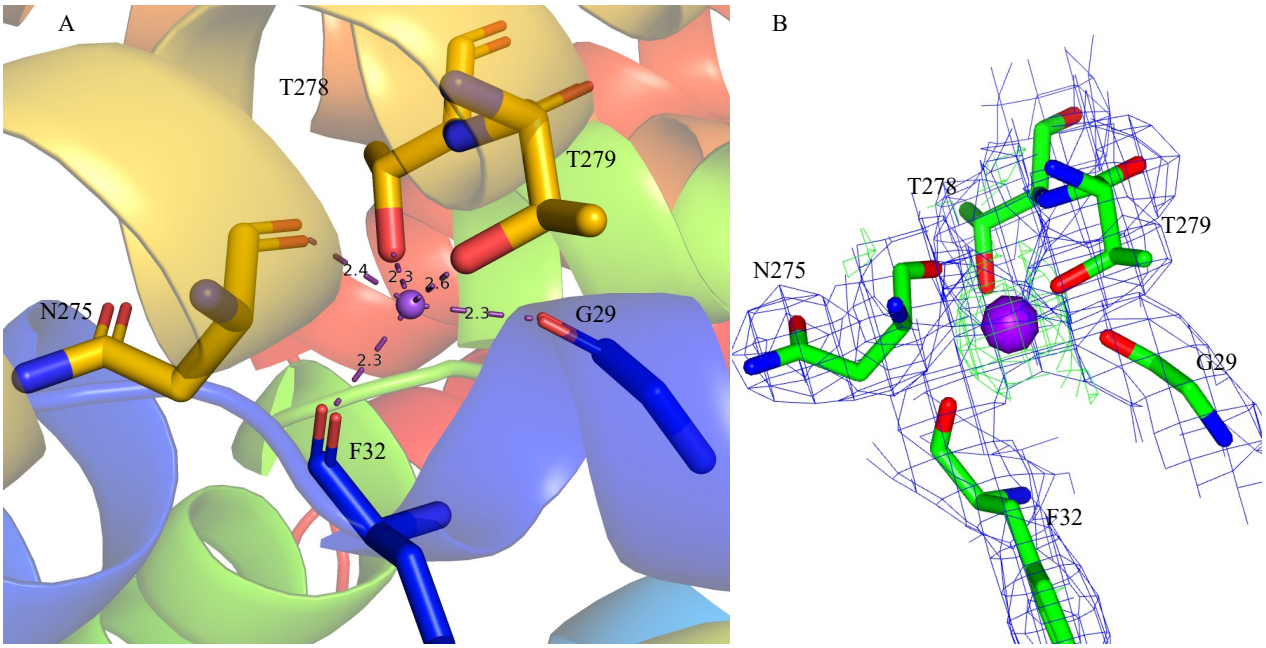


Figure 4.8. Sodium binding site of CodB. Panel A: Sodium in the binding site, with distances between coordinating residues, sodium is depicted as a purple sphere. Panel B: Sodium modelled in Fo-Fc map, contoured at 3 σ in green with coordinating residues in an 2Fo-Fc map in blue contoured at 1 σ. Sodium-oxygen interaction is depicted as purple dashed lines.

CodB has discontinuous non-proteinaceous density on the surface, it is not possible to define this density exactly, and instead monoolein has been modelled into the density. Theoretically, monoolein should have replaced detergent, and potentially any native lipids. Monoolein may not be the culprit of these sections of density but seems the most likely candidate unless more defined density suggests a different molecule is appropriate. In chain A, there are 4 separate sections of electron density that could be consistent with acyl chains and have been modelled as the acyl chains of monoolein. A DDM molecule has been modelled into a cavity formed of TM1b, TM3, and TM8 (Figure 4.9) This DDM molecule, with corresponding density shown in Figure 4.9, is found close to an acyl chain modelled as monoolein, this acyl chain is running parallel in a cavity to part of TM10. A phospholipid has been modelled in

a hydrophobic region of TM11 and TM12 almost parallel to the helices (Figure 4.10). In chain B, the density that corresponds to the DDM and phospholipid not as well defined and instead have been modelled as monoolein. There are 6 molecules of monoolein in chain B.

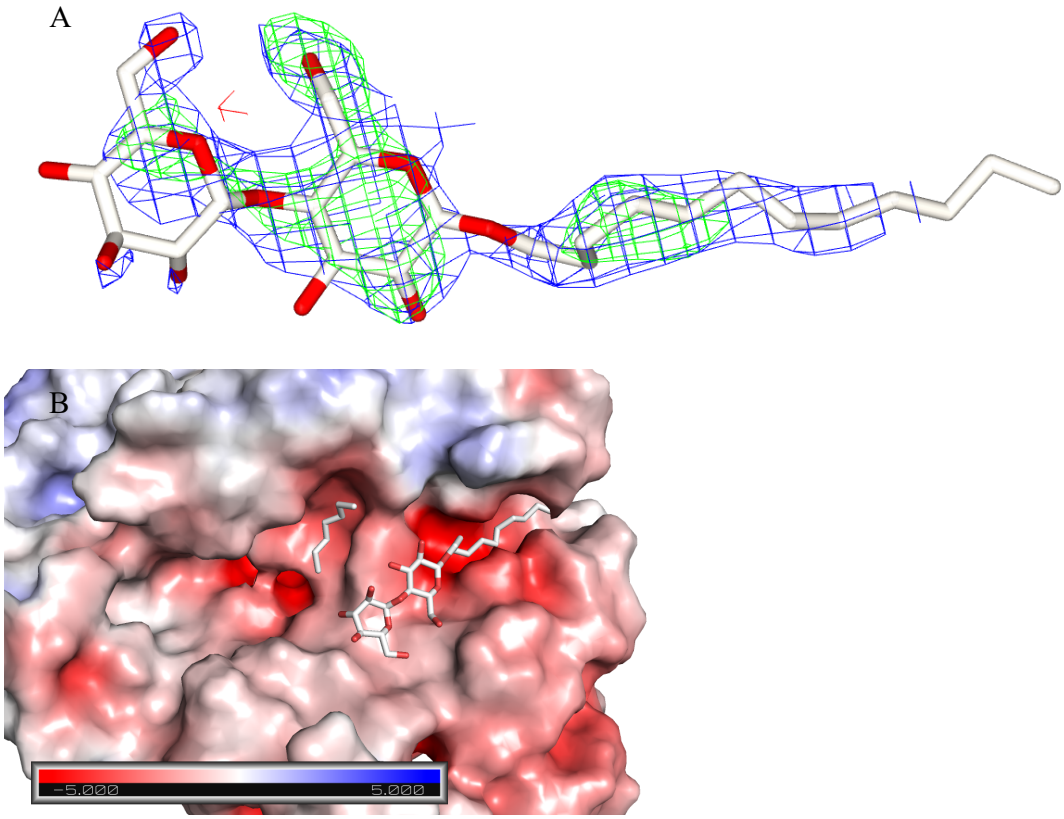


Figure 4.9. DDM modelled onto CodB. Panel A: DDM modelled into a Fo-Fc map coloured in green contoured at 3 σ and an 2Fo-Fc map (post-refinement) in blue contoured at 1 σ. Panel B: Electrostatic surface of CodB bound to DDM in the cavity between TM1b, TM3 and TM8.

It was originally difficult to assign the molecules in these sections of density due to the non-specific shape of them. The majority of the monoolein head groups did not have corresponding electron density and so only their acyl chains were modelled. For the case of the phospholipid, two elongated sections of electron density were close in proximity and building acyl chains worked very easily. When modelling the two phospholipid acyl chains these chains fitted into the density perfectly with no change in torsion angles, it could be a coincidence that two acyl chains from monoolein were this close together but based on the ease that the acyl-chains from a phospholipid fitted it seems sensible to assign this section of electron density as a phospholipid. No density for the head group of a phospholipid was observed (Figure

4.10). Regarding the DDM, this density was difficult to determine, at first a complete monoolein molecule was built but needed three water molecules to hydrate the head group to fit the density and form hydrogen bonding. However, when a DDM molecule was instead built into this density, it fitted the density (Figure 4.9) and formed the hydrogen bonding networks rather well.

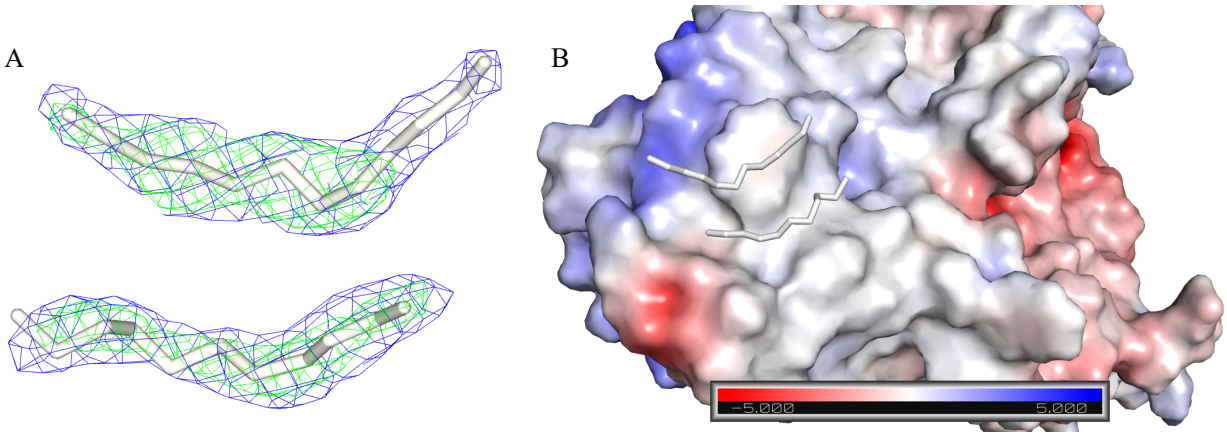


Figure 4.10. Phospholipid modelled on CodB. Panel A: phospholipid acyl chains modelled into Fo-Fc map coloured in green contoured at 3 σ and an 2Fo-Fc map (post-refinement) in blue contoured at 1 σ. Panel B: Electrostatic surface of CodB bound to phospholipid, this phospholipid is found running along TM11 and TM12 at a completely hydrophobic surface.

4.3. CodB compared with Mhp1

CodB and Mhp1 exhibit 20.6% sequence identity and with 52.7 % sequence similarity (Figure 4.11). This structure of CodB in complex with cytosine and sodium is most similar to the outward-open conformation of Mhp1 (PDB code: 2JLN) (Weyand *et al.*, 2008). CodB aligned (sequence independent) with Mhp1 with a RMSD of 3.34 (3 s.f) of 384 residues out of a possible 416. The 4-helix bundle and hash domain were extracted and aligned with a RSMD of 3.04 (3 s.f.) of 104 residues and a RSMD of 2.58 (3 s.f.) of 104 residues respectively (Figure 4.12).

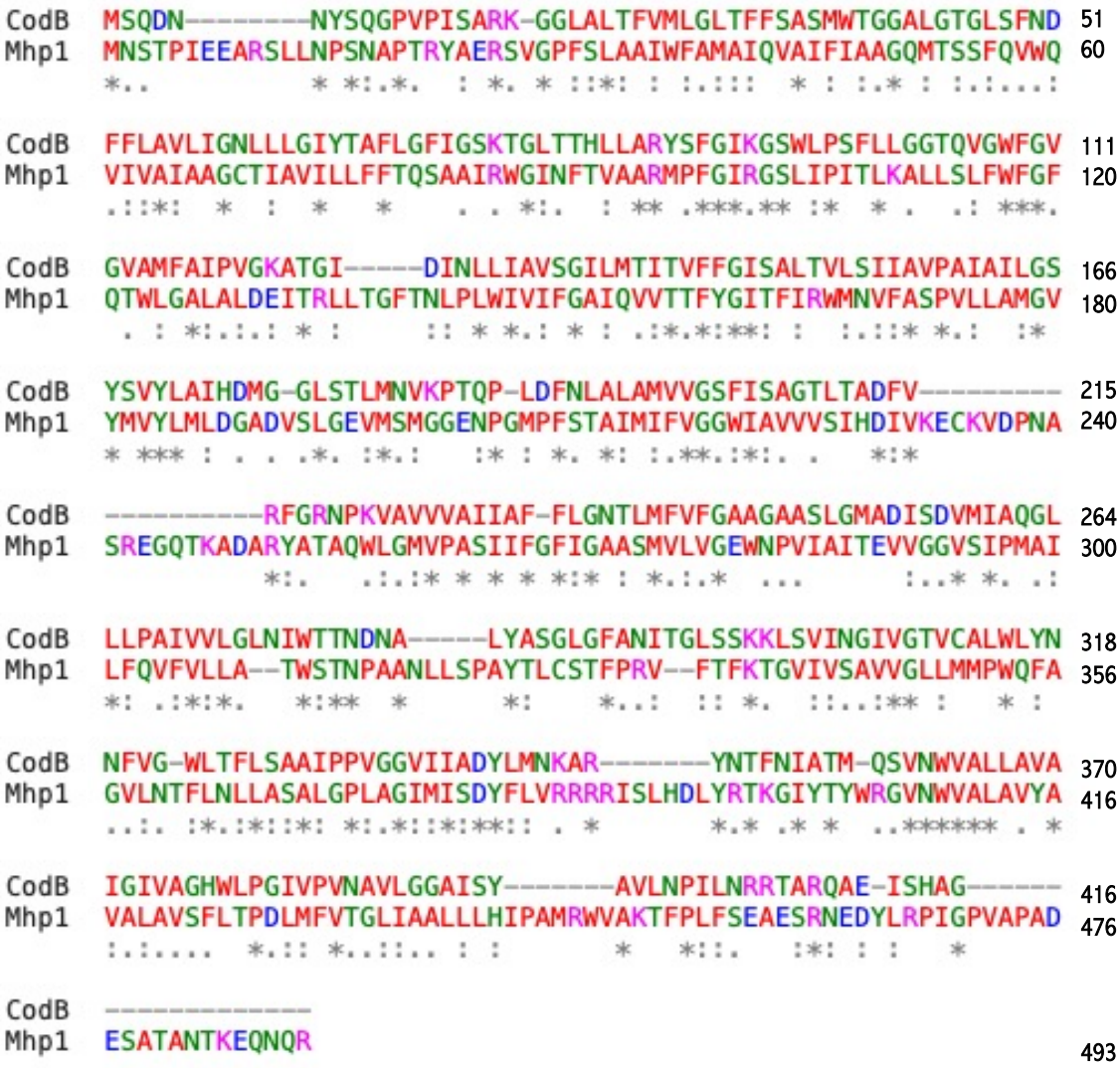


Figure 4.11. Sequence alignment of CodB and Mhp1 using MUSCLE algorithm (Madeira *et al.*, 2019)**.**

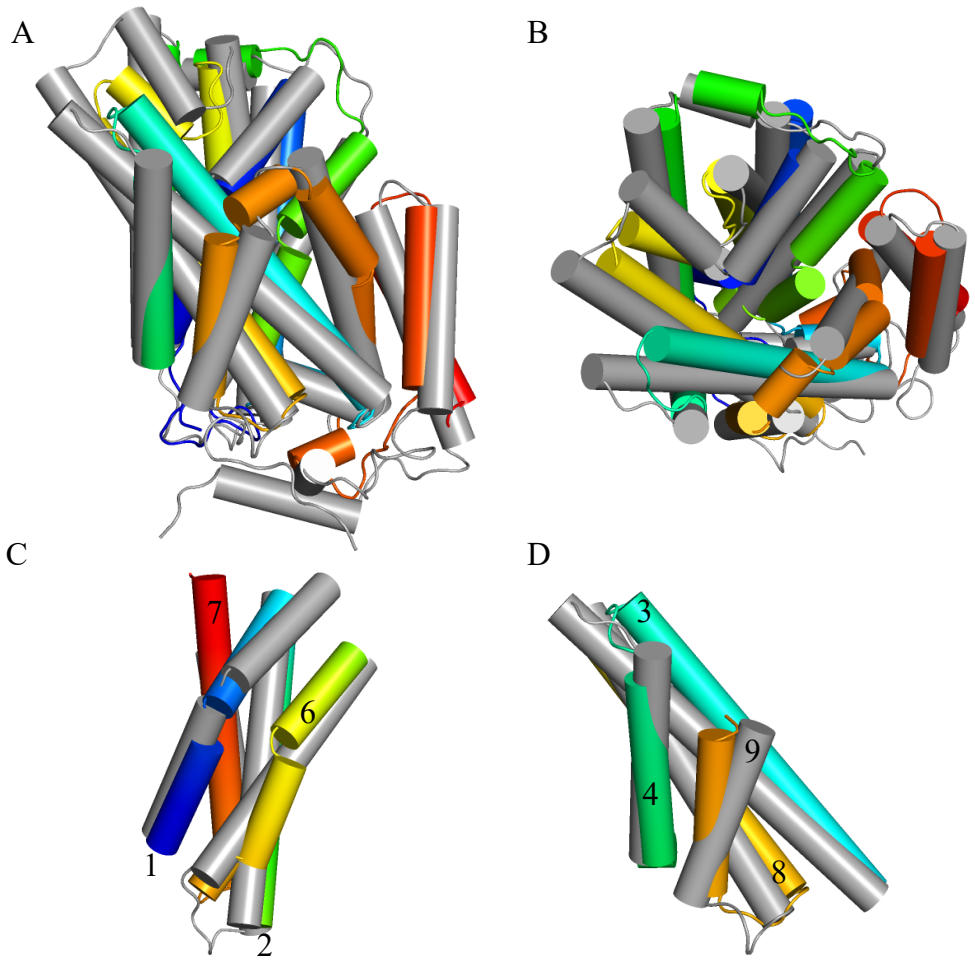


Figure 4.12. Superposition of CodB and 2JLN. Panel A, CodB with helices depicted as cylinders in coloured as Figure 4.3 and Mhp1 in grey. Panel B, CodB visualised in bird-eyes view from the extracellular space. Panel C, alignment of 4-helix bundle of CodB in colour and Mhp1 in grey with helix number denoted. Panel D, Superposition of the hash domain in CodB (colour) and Mhp1 (grey) with helix number denoted.

Overall, the structure of CodB and outward-open Mhp1 (PDB code:2JLN) are very similar, the N-terminal (extracellular side) of TM1 are held in different positions whilst the C-terminal end of TM3 has deviated by a helix width (5.4 Å) away from TM3 of Mhp1 and is straighter in CodB than Mhp1. TM8 has an additional residue in Mhp1 and is slightly kinked compared to CodB. However, the major difference between Mhp1 and CodB is TM9 and TM10. In CodB TM10 is found to be unwound, 3 residues away from a Pro-Pro-Val-Gly-Gly sequence, at the equivalent point this is Gly-Pro sequence in Mhp1. Comparing the position of TM10 in CodB, 2JLN, and outward-occluded IMH bound Mhp1 (PDB code: 4D1A) (Simmons *et al.*, 2014), in CodB TM10 is the held away from the 4-helix bundle whilst in 4D1A TM10 positioned folding over TM1 and 6 closing the binding pocket to the

extracellular space (Figure 4.13). The outward-open conformation of 2JLN has TM10 in a slightly more similar position to CodB then 4D1A but is found between CodB and 4D1A. As mentioned in section 3.3.2 a molecule of DDM and monoolein have been modelled and would sterically clash if TM10 of CodB was found in the position of TM10 in 4D1A (Figure 4.13).

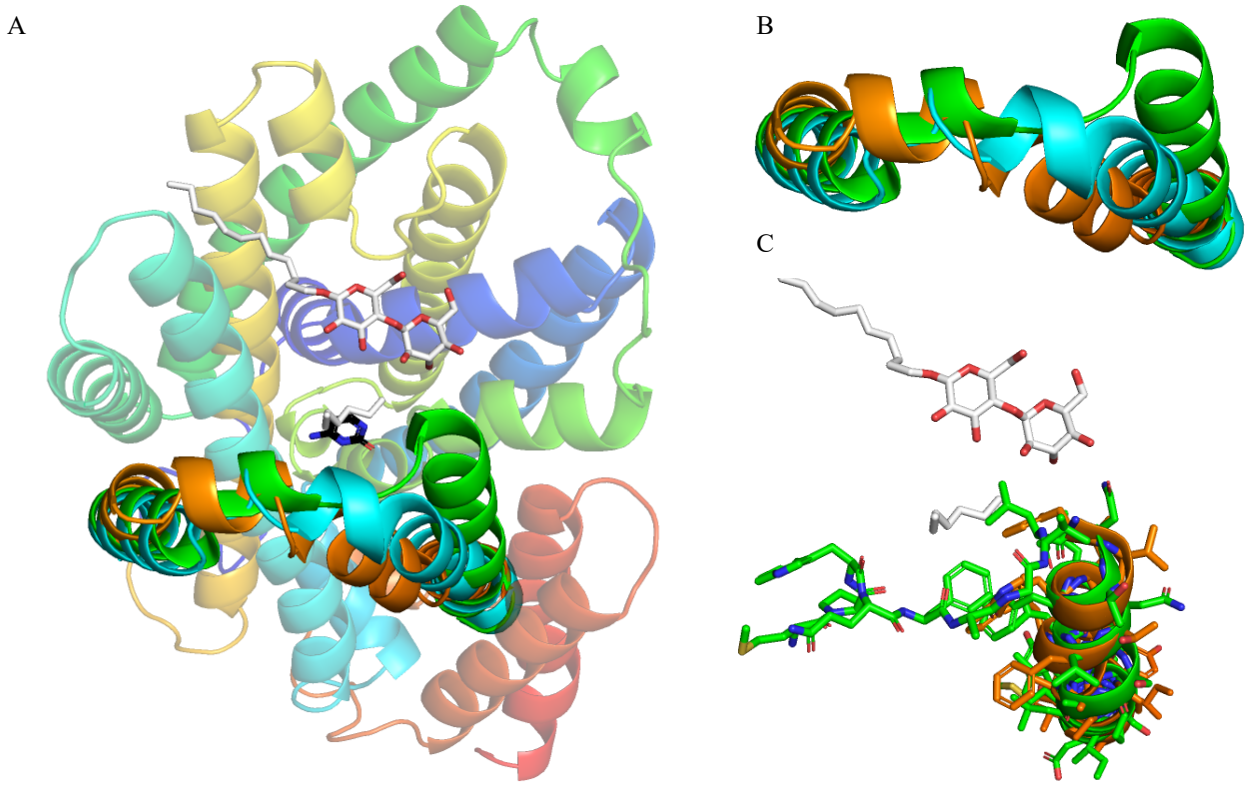


Figure 4.13. CodB and Mhp1 Structural differences in position of TM10. Panel A, CodB, with helices 9 and 10 of Mhp1 aligned 2JLN (cyan), and 4D1A (green) with a clear deviation of the position of TM10, showing that CodB is found in an outward-open structure as TM10 is closer to 2JLN than 4D1A. Panel B, Close of TM9 and TM10 only of CodB (burnt orange), 2JLN, and 4D1A. Panel C, alignment TM10 of CodB onto 4D1A, DDM and monoolein sterically clash with TM10 of 2JLN, and when CodB TM10 is modelled onto 4D1A this also sterically clashes with the monoolein and DDM.

4.3.1 Substrate binding

To enable comparison of the solute binding pocket CodB was aligned with Mhp1 in the occluded state bound to IMH (PDB code: 4D1A) (Simmons *et al.*, 2014). The cytosine and hydantoin binding pocket overlap with each other, both proteins substrate recognition is based on a π-stacking interaction of substrate sandwiched between aromatic residues with amino acids in the binding pocket providing the hydrogen bonds to select for solute, however, the substrate selectivity in between CodB and Mhp1 is quite different. IMH interacts with Mhp1 by directly interacting

with residues in TM3, 6, and 8 whilst CodB only uses residues from TM3 and TM6 and uses a coordinating water to interact with TM8. The hydantoin binding pocket of Mhp1 is dehydrated whilst in CodB cytosine binding is partially coordinated by water molecules, however, this is probably due to the higher resolution of CodB.

IMH is found π-stacking with residues Trp117 and Trp220 in Mhp1, which correspond to Trp108 and Phe204 in CodB (Figure 4.14), cytosine is found π-stacking with Trp108 and Phe204. Due to the smaller volume of cytosine compared to IMH Phe204 is about 2 Å closer to Trp108 than Trp117 and Trp220 in Mhp1 (Figure 4.14). In CodB, Phe204 is using a face-to-face π-stacking interaction to hold substrate whilst Trp220 in Mhp1 is using a face-to-edge π-stacking interaction. Trp108 in CodB is highly conserved across the characterised members of the NCS1 and is found using a face-to-face π-stacking in both Mhp1 and CodB; Phe204 is functionally conserved as an aromatic residue across most of the NCS1 members (Appendix 2).

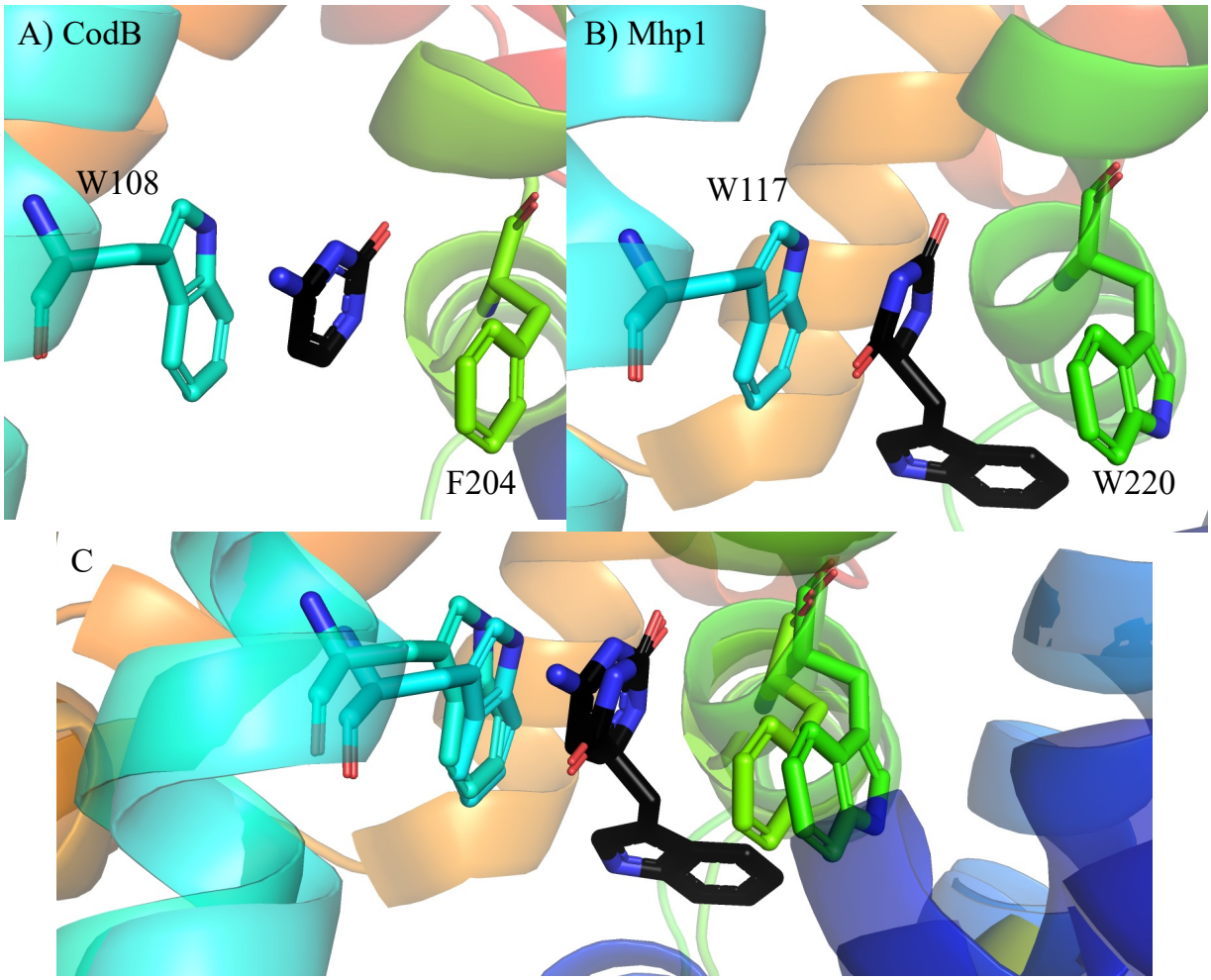


Figure 4.14. Close up of π-stacking interactions in CodB and Mhp1. CodB (Panel A), Mhp1 (panel B) and superposition of CodB and 4D1A (Panel C). TM3 is found in cyan with TM6 in green. Cytosine and IMH are pictured in black. Trp108 in CodB is conserved with Trp117 in Mhp1, whilst Phe204 in CodB is functionally equivalent to Trp220 in Mhp1.

In Mhp1, Gln42 and Gln121 form a hydrogen bonding network to hold Gln121 in position to interact with the carbonyl at position 4 of the hydantoin moiety, in CodB these residues are equivalent to Phe33 and Gly112, Phe33 is found in the position of Gln42 and Gln121 to create a hydrophobic region for the position 5 of cytosine to sit in (Figure 4.15). There is no conservation of any amino acids at Gln42/Phe33 position in contrast to Gln121 being found in a few NCS1 members (Appendix 2).

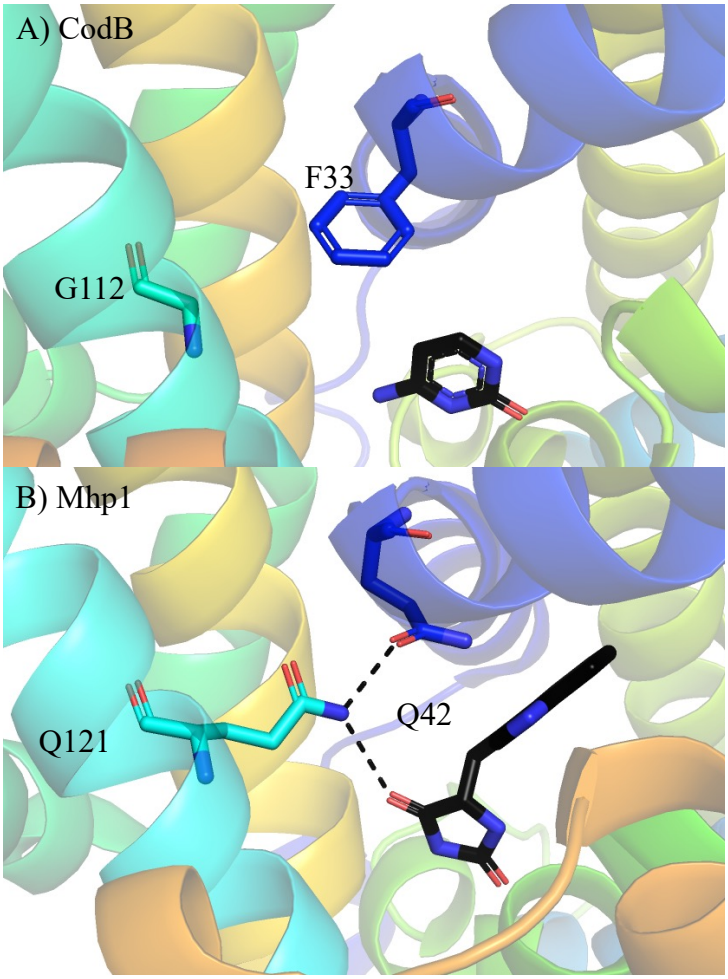


Figure 4.15. Hydrogen bonding network of Q121 and Q42 in Mhp1 with the corresponding residues in CodB. CodB (Panel A) and Mhp1 (Panel B) TM1 is visualised in dark blue with TM3 in cyan. Cytosine and IMH are pictured in black. Hydrogen bonds are depicted as black dashed lines.

Mhp1 uses Asn318 (TM8) to correctly orientate IMH in the binding pocket, in CodB the residue found at this position is Ala283, this side chain does not interact with the cytosine substrate. The corresponding functional amino acid in CodB is Gln105, found on TM3 which is Ser114 in Mhp1, Ser114 does not seem to have role in IMH binding (Figure 4.16). There is no conservation of residues for Ser114/Gln105 in the NCS1 (Appendix 2). Asn318 is well conserved in the NCS1, apart from CodB; sequence alignments did not align any residues in CodB to this position, only structural determination could align Asn318 (Mhp1) to Ala283 (CodB) (Figure 4.16).

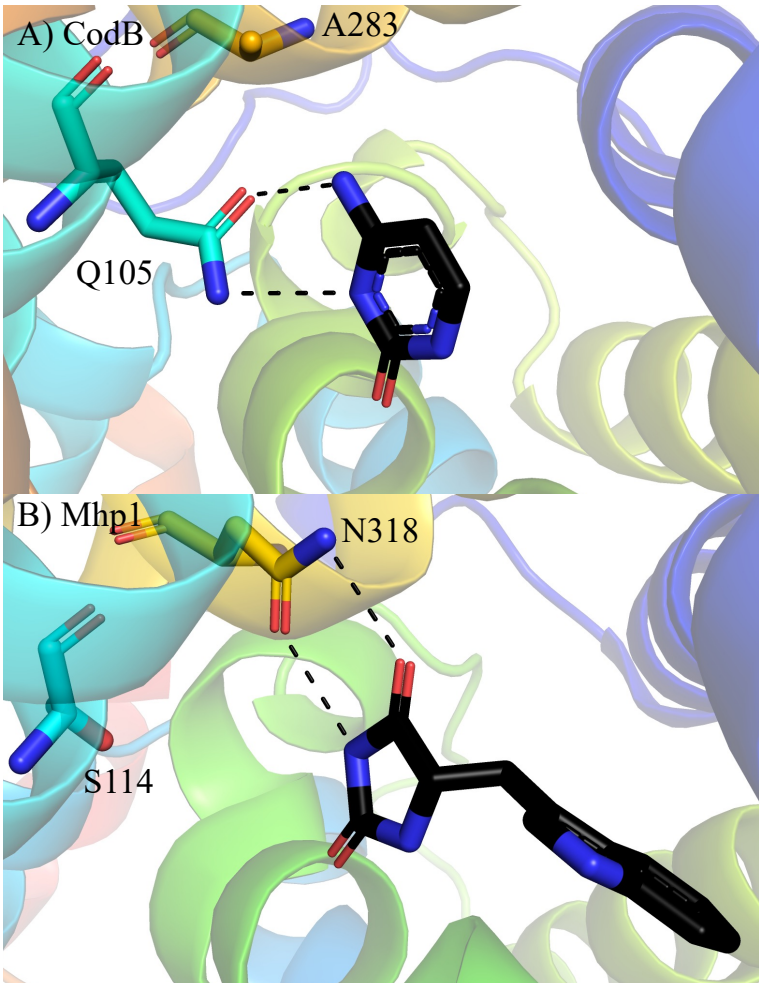


Figure 4.16. Orientating substrate in the substrate binding pocket is dictated by hydrogen bonds donated by Gln105 in CodB and Asn318 in Mhp1. CodB (Panel A) and Mhp1 (Panel B). TM3 is coloured in cyan whilst TM8 is orange. Cytosine and IMH are pictured in black, with hydrogen bonds being depicted as black dashed lines.

Leu363 in Mhp1 is responsible for preventing larger solutes from being transported by sterically hindering with larger substrates, this corresponds to Leu325 in CodB. The conformation of TM10 is very different in CodB and 4D1A and therefore nothing conclusive can be said about the significance of Leu325.
In Mhp1, Gly219 is found in the hydantoin binding pocket with its main carbonyl interacting with the secondary amine at position 1, in CodB this is Ser203 and the main chain carbonyl is hydrogen bonding with secondary amine at position 1 of cytosine. The carbonyl of cytosine is hydrogen bonding with amide bond of Ala207, whereas in Mhp1 no residues are in proximity to interact with hydantoin.

In CodB, water can be visualised in the binding site coordinating with cytosine and surrounding amino acids. Ser206 (TM6) and Asn280 (TM8) in CodB appear to the

coordinating with cytosine through water molecules (Figure 4.17). Mhp1 does not have water molecules modelled in its binding site; this is most likely due to the lower resolution of Mhp1. Ser206 does not have conserved functional residue at the same position in Mhp1 and has Ala222. Whereas asparagine is conserved at the 280/314 position for CodB/Mhp1. The position of this water molecule is very similar to the position of the side chain of Asn318 in Mhp1, from this Asn280 and Asn314 are performing the same role of coordinating a water/side chain enabling the moiety to coordinate with the respective substrate (Figure 4.17).

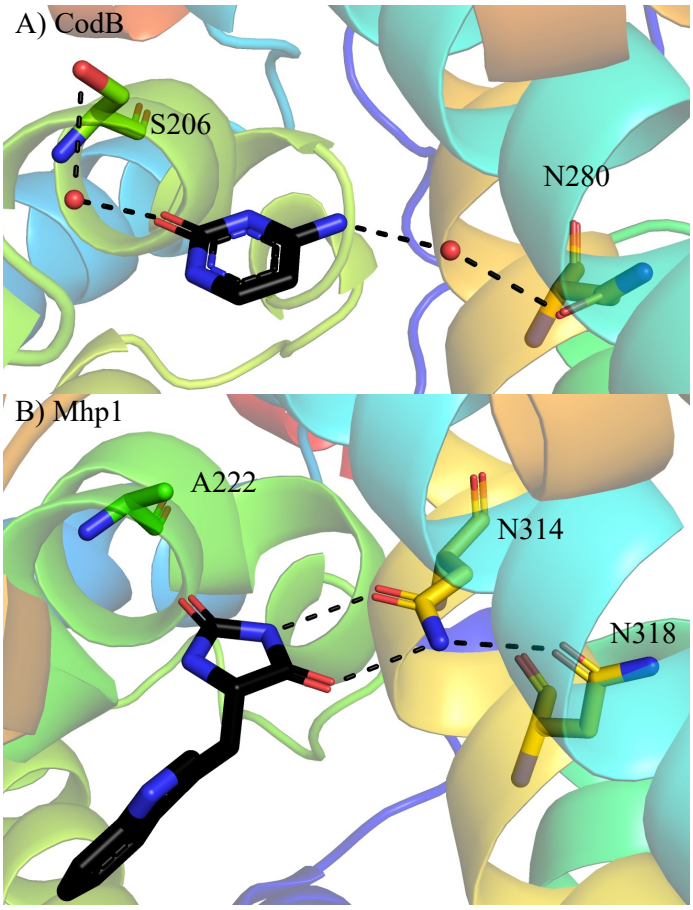


Figure 4.17. Water coordination in CodB and the corresponding residues in Mhp1. CodB (Panel A) and Mhp1 (Panel B). TM6 is coloured in green with TM8 in burnt orange, cytosine and IMH are pictured in black. Hydrogen bonds are depicted as black dashed lines.

In Mhp1, a hydrogen-bonding network of Gln153, Asn314, and Asn318 is shown to be important for IMH binding and transport, with the suggestion that this hydrogen bonding network is able to position Asn318 into the correct position to coordinate with IMH in Mhp1. Whilst there isn't a functionally equivalent residue in CodB for Asn318, Asn280 (CodB) is conserved with Asn314 (Mhp1), with Met139 in CodB corresponding to Gln153. As shown in Figure 4.18, there is an interaction between

the highly conserved Asn280 and less conserved Met139. The Gln153/Met139 position is variable across the NCS1, although this position does conserve residues capable of forming hydrogen bonds as opposed to hydrophobic side chains. This hydrogen bonding network is disrupted in the inward-open structure of Mhp1 (PDB code: 2X79).

Asn280/Asn314 corresponds to Asn374 in Fcy2, an NCS1 purine-cytosine permease in *Saccharomyces cerevisae*. Mutagenesis of Asn374 was shown to reduce uptake of substrate in Fcy2 (Ferreira *et al.*, 1997), along with Asn377 that aligns with Ala283 and Asn318 in CodB and Mhp1 respectively (Appendix 2).

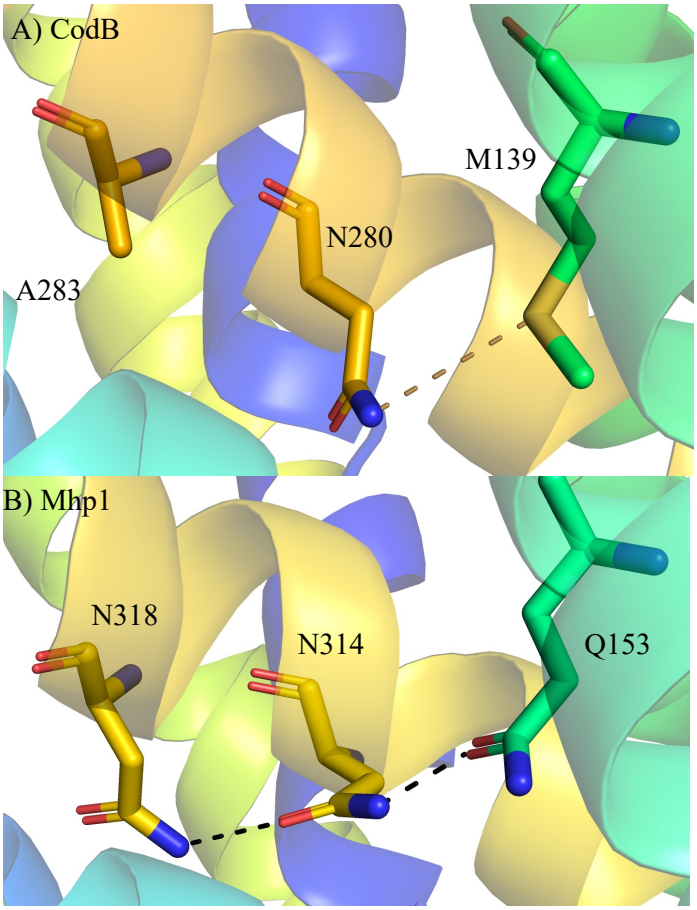


Figure 4.18. Hydrogen bonding network between TM4 and TM8 in CodB and Mhp1. CodB (Panel A) and Mhp1 (Panel B). TM4 is coloured in turquoise and TM8 in orange. In Mhp1, there is a hydrogen bonding network between TM8 and TM4 that enables the Asn318 side to orientate to coordinate IMH in the binding site. In CodB, there is no interactions with the cytosine substrate at this position but interaction between TM8 and TM4 remain intact, admittedly weaker due to Met139. Hydrogen bonds are depicted as black dashed lines, whilst the Asn280 Met139 interaction is visualised by a yellow dashed line.

4.3.2. Sodium binding

The amino acids contributing to sodium binding site of CodB are functionally conserved with Mhp1. CodB and Mhp1 both coordinate their sodium with a square pyramidal arrangement and use 2 main chain carbonyls from the unwound section of TM1, and a main chain carbonyl from TM8 along with hydroxyl side chains from TM8.

Mhp1 uses Ala38, Ile41, Ala309, Ser312, Thr313 to form the sodium binding site whilst CodB coordinates sodium using Gly29, Phe32, Asn275, Thr278, Thr279. Ala38/Gly29 has a variety of smaller side chains at this position across the NCS1 family. The Ile41/Phe32 position has conserved hydrophobic residues across the NCS1, with PucI the exception with a histidine residue, and phenylalanine being conserved in a few of the yeast transporters. A residue of interest between CodB and Mhp1 is Asn275 and Ala309 respectively (Figure 4.19). Alanine has a reasonable amount of conservation at this position with CodB being the only member of the NCS1 family to have an asparagine, a few of the yeast and plant transporters have serine at this position instead. In CodB, the carbonyl of Asn275 is interacting with the Na^+ whilst the side chain is hydrogen bonding with Ser34, found on the unwound section of TM1. Ser34 corresponds to Val43 in Mhp1. The side chain of Ala309 in Mhp1 is found in a hydrophobic pocket of Ile41 and Val43.

Ser312/Thr278 and Thr313/Thr279 use their hydroxyl groups to coordinate sodium in Mhp1/CodB respectively. An overall alignment of CodB and Mhp1 does not highlight the structural differences of TM8, therefore TM8 from CodB and Mhp1 were extracted and aligned (sequence dependent, using the align function in PyMOL) to each other to really investigate TM8 in close detail. The N-terminal ends of TM8 align very well but begin to deviate as Mhp1 has Pro315 that induces kink in TM8 whilst the helix remains straight in CodB.

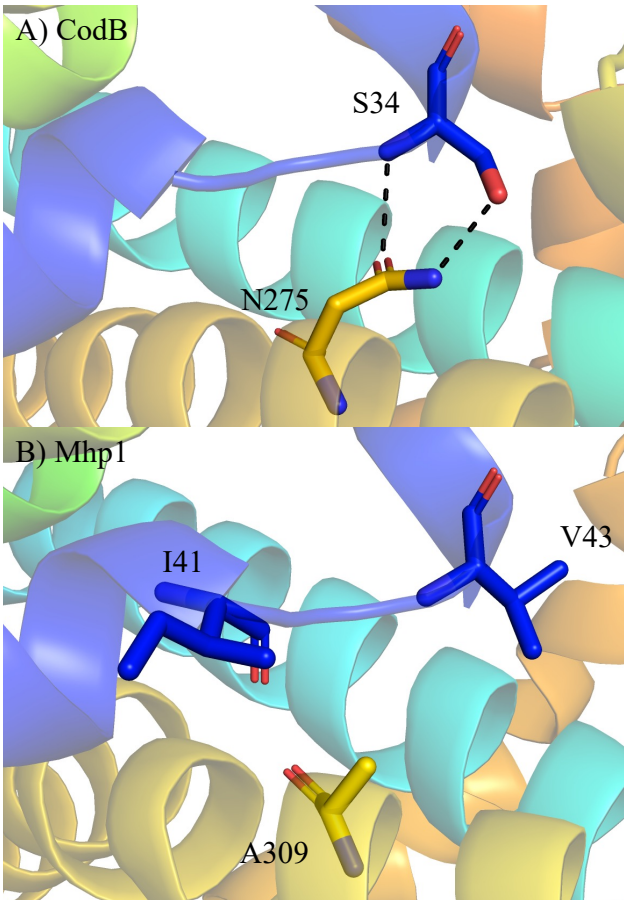


Figure 4.19. Hydrogen bonding of Asn275 and Ser34 in CodB and the corresponding residues in Mhp1. CodB (Panel A) and Mhp1 (Panel B). The side chain of Asn275 is interacting with Ser34 in TM1 whilst there is no interaction in Mhp1, and the side chain of Ala309 is sat in a hydrophobic pocket. TM1 is coloured in blue with TM8 in orange. Hydrogen bonds are depicted as black dashed lines.

Asp281 is found in CodB at the same position as Pro315 in Mhp1, this residue is broadly found as a small hydrophobic residue in the NCS1 family. However, in CodB this is a polar residue and is found hydrogen bonding to Thr142 (TM4), corresponding to Thr157 in Mhp1 (Figure 4.20).

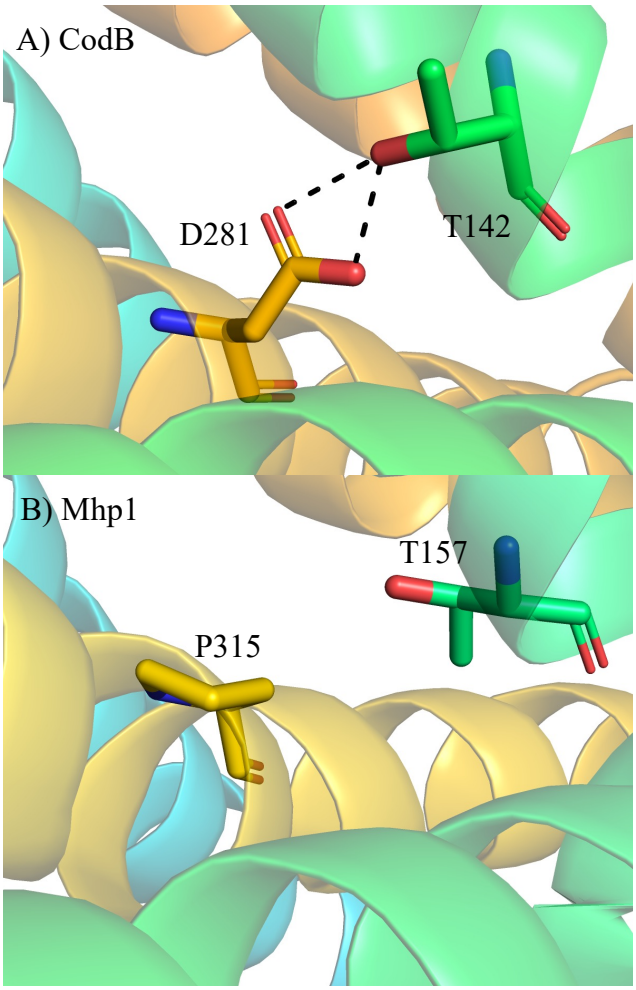


Figure 4.20. TM8 and TM4 interaction mediated by Asp281 and Thr142 in CodB and the equivalent residues in Mhp1. CodB (Panel A) and Mhp1 (Panel B). CodB has Asp281 above the sodium site that is interacting with the part of TM4 that moves. In Mhp1, this hydrogen bond doesn't exist, despite the conservation of Thr at this position. Hydrogen bonds are depicted as black dashed lines.

Ala317 in Mhp1 is well conserved in the NCS1 family, apart from in CodB where from sequence alignments it was unclear which residue would be found at this position, due to the additional residue on TM8 of Mhp1. Ala317 packs onto TM1 with the methyl side chain exposed to the substrate pocket. At the equivalent position in CodB has Asn282 with the side chains interacting with the carbonyl of Val26 on TM1 (Figure 4.21). The side chain hydroxyl of Ser154 on TM5 is also interacting with the amide of Val26 creating a hydrogen bonding network between TM1, 5, and 8. This hydrogen bonding network does not exist in Mhp1 (Figure 4.21).

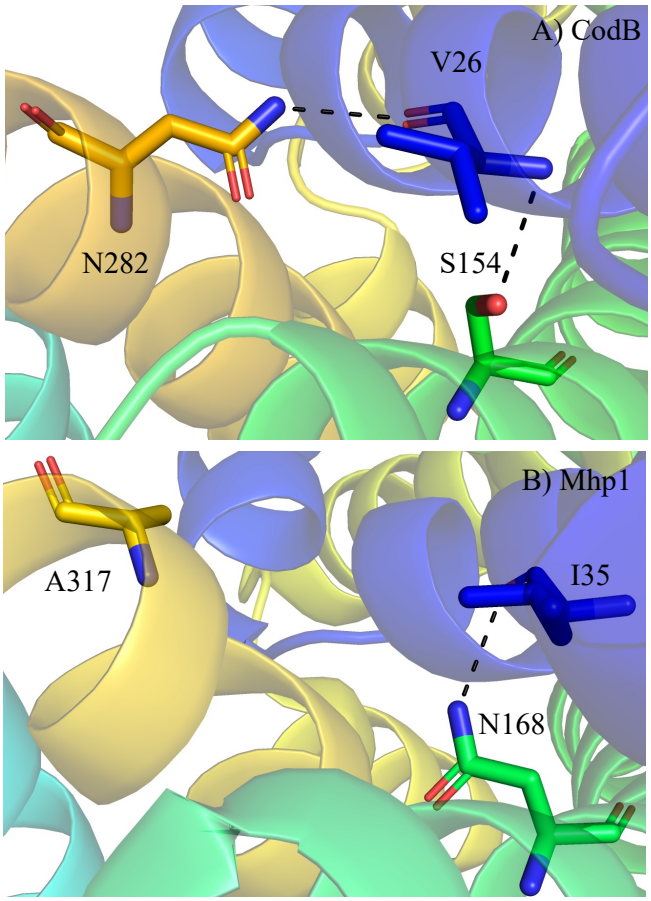


Figure 4.21. Hydrogen bonding network of TM1, 5, and 8 in CodB and Mhp1. CodB (Panel A) and Mhp1 (Panel B). CodB uses a hydrogen bonding network to directly link TM1, 5, and 8 whilst Mhp1 is unable to link TM1 and TM8 together. TM1 is found in blue, TM5 in green, and TM8 in orange. Hydrogen bonds are depicted as black dashed lines.

Asp213 is the most conserved residue in the NCS1 family and is found towards the c-terminal (intracellular) end of TM6. In CodB Asp213 does not appear to be interacting with any nearby residues. In Mhp1, Asp226 is interacting with Lys110 (TM3) in an outward-facing conformation, that is disrupted in the inward-facing state, Lys110 was highlighted by Ziegler and colleagues as significant as the position of this positively charged residues is found at the same position as the predicted Na1' of BetP (Khafizov *et al.*, 2012). In CodB there are only hydrophobic residues around this site.

4.4 CodB compared with the LeuT superfamily

CodB exhibits low sequence identity to other LeuT transporters despite structural similarity (Table 4.3), as calculated by the LALIGN algorithm (https://embnet.vital-it.ch/software/LALIGN_form.html) (Huang and Miller, 1991). Members of the LeuT superfamily transport a wide variety of solutes and utilise different sodium: solute

stoichiometries. All of these homologs bind their substrate at a similar position in the centre of the membrane around the discontinuous helices of TM1 and TM6, different chemical structures of the solute mean that solute binding isn't conserved around the LeuT superfamily. CodB, Mhp1, and vSGLT (Faham *et al.*, 2008) exhibit a π-stacking interaction to bind their substrate and then use surrounding amino acids to orientate substrate within the binding pocket.

Protein	Sequence Identity (%)	Sequence Similarity (%)	Conserved Sodium site
LeuT	16.0	42.9	Gly20 (TM1) MC Val43 (TM1) MC Ala351 (TM8) MC Thr354 (TM8) SC Ser355 (TM8) SC
MhsT	17.7	45.8	Gly24 (TM1) MC Val27 (TM1) MC Ala320 (TM8) MC Ser323 (TM8) SC Ser324 (TM8) SC
dDAT	17.2	44.1	Gly42 (TM1) MC Val45 (TM1) MC Leu417 (TM8) MC Asp420 (TM8) SC Ser421 (TM8) SC
SERT	16.5	47.4	Gly94 (TM1) MC Val97 (TM1) MC Ala434 (TM8) MC Asp437 (TM8) SC Ser358 (TM8) SC
vSGLT	15.4	51.1	Ala62 (TM1) MC Ile65 (TM1) MC Ala361 (TM8) MC Ser365 (TM8) SC Ser356 (TM8) SC
SiaT	18.5	50.1	Ala56 (TM1) MC Leu59 (TM1) MC Ala339 (TM8) MC Ser342 (TM8) SC Ser343 (TM8) SC
BetP	20.1	49.5	Ala147 (TM1) MC Met150 (TM1) MC Phe464 (TM8) MC Thr467 (TM8) SC Ser468 (TM8) SC

Table 4.3. Sequence identity of CodB and other LeuT superfamily members. Amino acids involved in coordinating sodium in the conserved Na2 site are listed with their interaction denoted with MC (main chain interaction) or SC (side chain interaction).

The conserved feature between all of the LeuT superfamily sodium symporters is the Na2 site found at the interface of TM1 and TM8 bridging the bundle and hash domains (Table 4.3). The individual amino acids contributing to sodium binding do vary, however, the interactions of 2 main-chain carbonyls from TM1, a carbonyl on TM8 and 2 side chains from TM8 with sodium in a square pyramidal arrangement are well conserved. All characterised members of the LeuT superfamily have a hydrophobic side chain for the amino acid in TM8 using its carbonyl to coordinate, CodB also uses this carbonyl but is an outlier as it uses the polar side chain of Asn275 to interact with TM1. CodB is the only member of the superfamily that uses a side chain at this position to interact with TM1. CodB is the only member characterised that uses Thr-Thr side chains to bind sodium.

Only one plausible sodium site can be visualised in CodB. When comparing CodB to LeuT (PDB code:2A65) (Yamashita *et al.*, 2005) at the Na1 site, LeuT coordinates Na^+ in this site using an octahedral arrangement with carbonyls of Ala22 (TM1) and Thr254 (TM6) and the side chains of Asn27 (TM1), Thr254 (TM6), and Asn286 (TM7) and the leucine solute. CodB has Thr31, Ser36, Ser203, and Asn236; the residues on TM1 and TM7 could coordinate sodium binding in this site, but the position of TM6 is too far away to be interacting and the cytosine substrate is also too far away to be a ligand for Na^+ in the Na1 (Figure 4.22). CodB does not show any density for a sodium ion or water to be bound at this site; CodB has been purified with sodium throughout purification and the crystallisation condition uses a high sodium concentration and that should saturate all sodium sites. The combination of no observable density at this position and an absence of a suitable coordination are in agreement; there is no evidence to suggest that CodB is capable of coordinating sodium at this position. In LeuT, sodium is necessary to coordinate the negatively carboxy group of the leucine solute, whilst cytosine does not have such a negatively charged group.

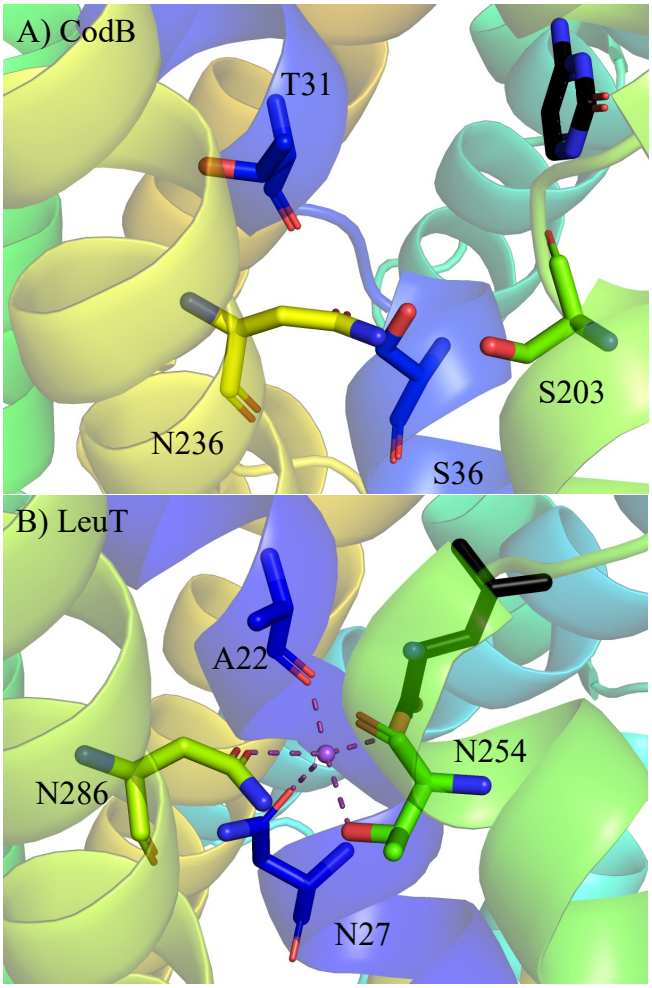


Figure 4.22. Na1 site in CodB and LeuT. CodB (Panel A), LeuT (Panel B). LeuT coordinates sodium in octahedral complex using substrate, residues from TM1, 6, and 7 as ligands. Thr31, Ser36, and Asn236 are in position to coordinate sodium but Ser203 and cytosine are too far away from the site so it seems unlikely that sodium could be coordinated in this position. TM1 is blue, TM6 is green, TM7 in greenish/yellow. Cytosine and leucine substrate are depicted in black with sodium in purple. Sodium-oxygen interactions are pictured using a purple dashed line.

SiaT (PDB code: 5NV9) (Wahlgren *et al.*, 2018) binds a second Na^+ at the Na3 site found between TM1 and TM5. This Na3 site is coupled to the Na2 site via Ser342, Ser342 side chain hydroxyl is a ligand with Na^+ at the Na2 site with its main chain carbonyl coordinating with the Na3 site. There is no observable density at this position in CodB and the structure around this site would not coordinate a sodium ion as TM5 is held too far away from the site and side chains of Asp281 and Asn828 on TM8 are pointing away from the site (Figure 4.23). There is no evidence that CodB has a Na3 site capable of sodium binding for the same reasons as the Na1 site.

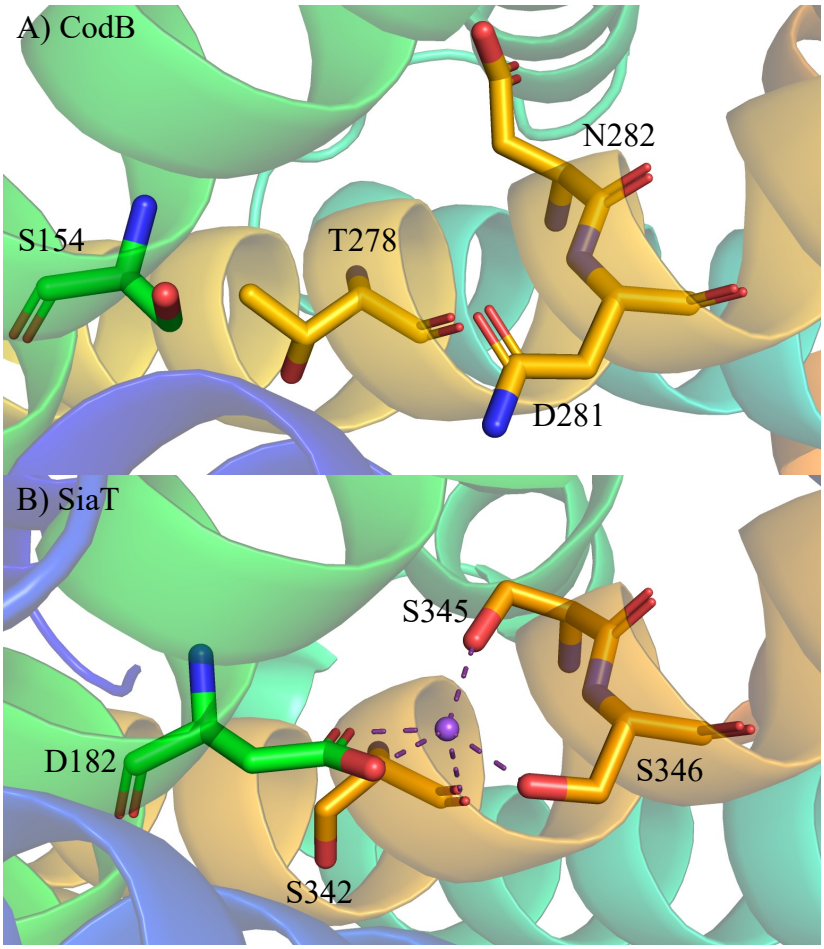


Figure 4.23. Na3 site in CodB and SiaT. CodB (Panel A) and SiaT (Panel B). SiaT coordinates sodium in a Na3 site thought to be playing a regulatory role in transport and is found between TM5 and TM8. TM5 is coloured in green, with TM8 in orange and Na^+ in purple. Sodium-oxygen interactions are pictured using a purple dashed line.

BetP (Khafizov *et al.*, 2012) has a unique Na1' site that is the structural symmetric of the Na2 site found between TM3 and TM6, although no density for sodium could be seen in crystal structures MD simulations and mutagenesis of amino acids in this site do impact sodium binding. BetP uses its betaine substrate, side chain and main chain carbonyl of Thr246 (TM3), the side chain of Thr250 (TM3), and the side chain of Phe380 (TM6) to hold sodium in this position. In CodB there is no density at this position and the residues in this position form a hydrophobic pocket. There is no evidence of Na^+ binding at this site.

Asn280 and Asn282 in CodB are found to be forming hydrogen bonding networks and linking TM8 directly with TM5 or to TM5 through TM1 respectively. LeuT has a comparable interaction by using Ser356 found at the equivalent position on TM8

that interacts with Asn179 on TM4, this interaction is preserved in the inward-facing LeuT structures. No other family members use residues on TM8 to form a hydrogen bonding network between TM1, 5, and 8.

MhsT has a $GlyX_9Pro$ motif that is conserved in LeuT, dDAT, and SERT that is thought to enable the unwinding of the intracellular side of TM5 and allow solvation of the Na2 site and enable Na^+ from this site (Malinauskaite *et al.*, 2014). The authors noted that this motif can be found in Mhp1, and the structure of CodB shows that this $GlyX_9Pro$ motif can also be found in the same position. CodB also has this motif at the structurally symmetric position of TM10 which is unwound. Mhp1 has a $GlyX_{14}Pro$ in TM10 and MhsT has $GlyX_{16}Pro$

4.5 Inward-Open Model of CodB

An inward-open model of CodB was built as described in the materials and methods, based on the inward-open model of Mhp1 (PDB code: 2X79). Aligning both structures by the 4-helix bundle highlights the movement of the hash-motif with TM3 and TM8 moving an entire helix-width away from their original position (~5.4Å) (Figure 4.24). To move CodB into an inward-facing state from an outward-facing state, the hash-motif undergoes a rigid body movement whilst the 4-helix bundle remains static, TM10 and TM5 are modelled as the extracellular and intracellular gates respectively and undergo bending to close and open.

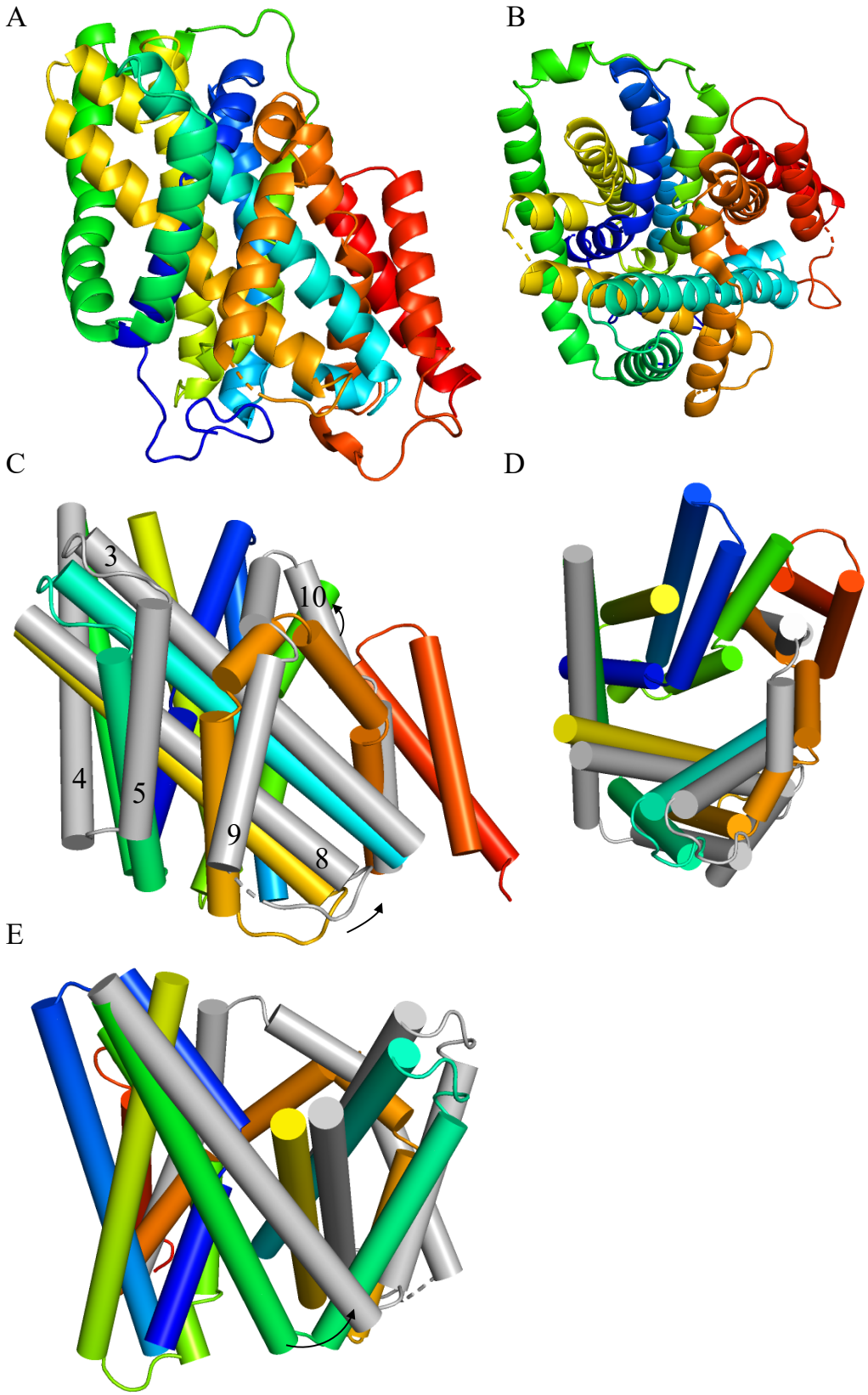


Figure 4.24. Inward-open model of CodB. Panel A, CodB in the plane of the membrane depicted in cartoon representation with helices coloured as Figure 4.3 and from the extracellular space in panel B. Panel's C, D, and E have the crystallographically determined outward-open structure of CodB in colour as Figure 4.3, with the inward-model coloured in grey with helix number denoted for ease. Arrows represent movement from the outward to inward state.

Analysis of the cytosine binding pocket shows that Phe204 and Ser206 are found in the same position, a consequence of TM6 remaining stationary, whilst Trp108 moves out of the binding pocket due to the movement of TM3, however, the movement Gln105 and Asn280 is much more pronounced. Gln105 and Asn280 move out the binding pocket making a space far too large for cytosine to form favourable interactions with CodB (Figure 4.25). Asn280 is found interacting with Met139 and the cytosine substrate in the outward-open structure discussed above, the movement of TM8 results in movement of Asn280 out of the binding pocket whilst still being able to coordinate water and interact with Met139.

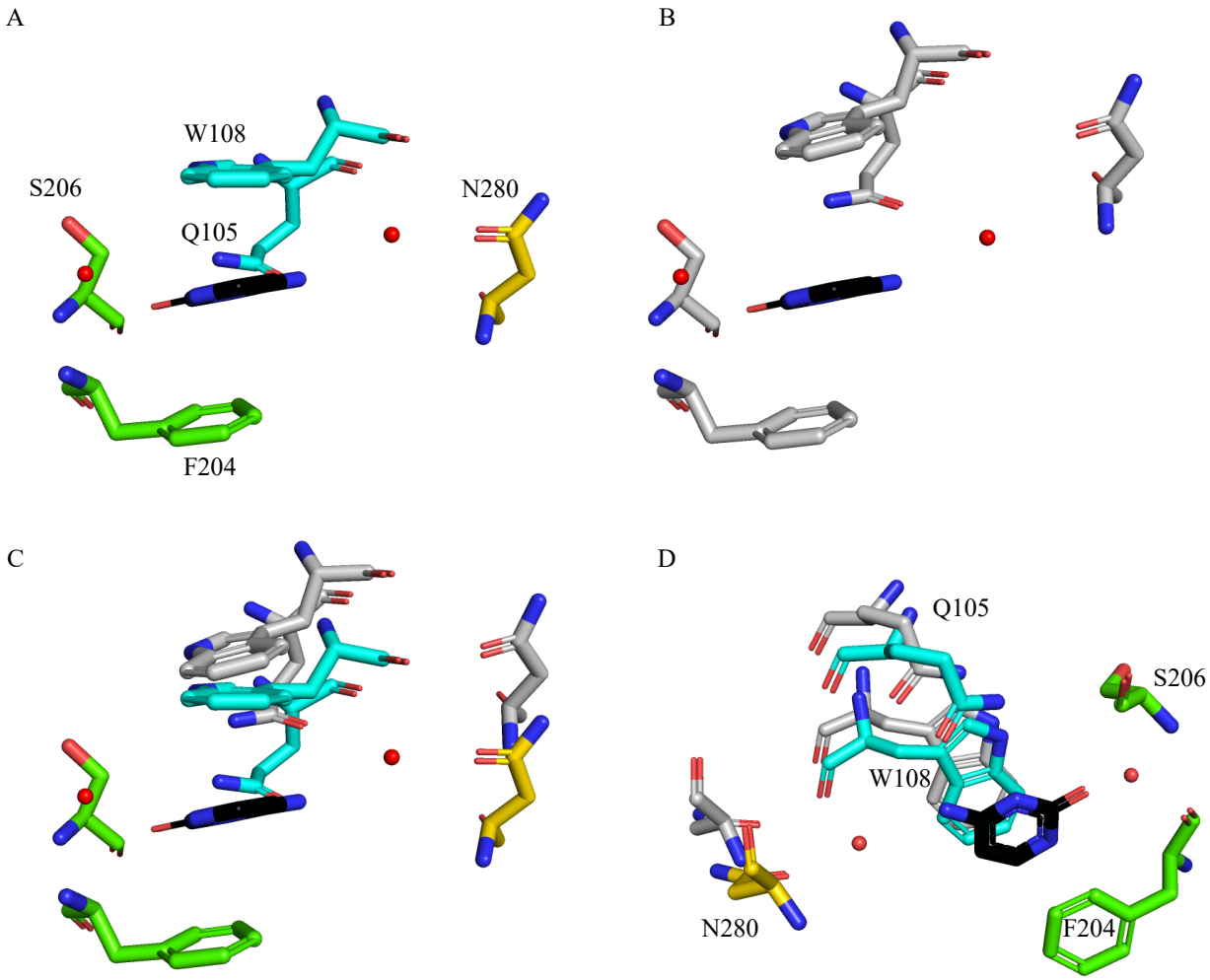


Figure 4.25. Putative cytosine binding site of inward-open CodB compared to the outward-open structure. Panel A, residues involved in cytosine coordination in the outward-open structure. Panel B, inward-open model with cytosine and water pictured for reference. Panel C and D, overlapping of outward-open structure and inward-open model. Outward structure is in colour with inward-model in grey.

Movement of TM8 results in a disruption of the sodium binding site, by pulling Asn275, Thr278, and Thr279 out of position to coordinate sodium (Figure 4.26). Despite the movement of Asn275, the amide of the asparagine side chain is still in position to interact with hydroxyl side chain of Ser34 (TM1), whilst the carbonyl of the side chain has moved too far away to be interacting with the main chain amide of TM1 (Figure 4.26)

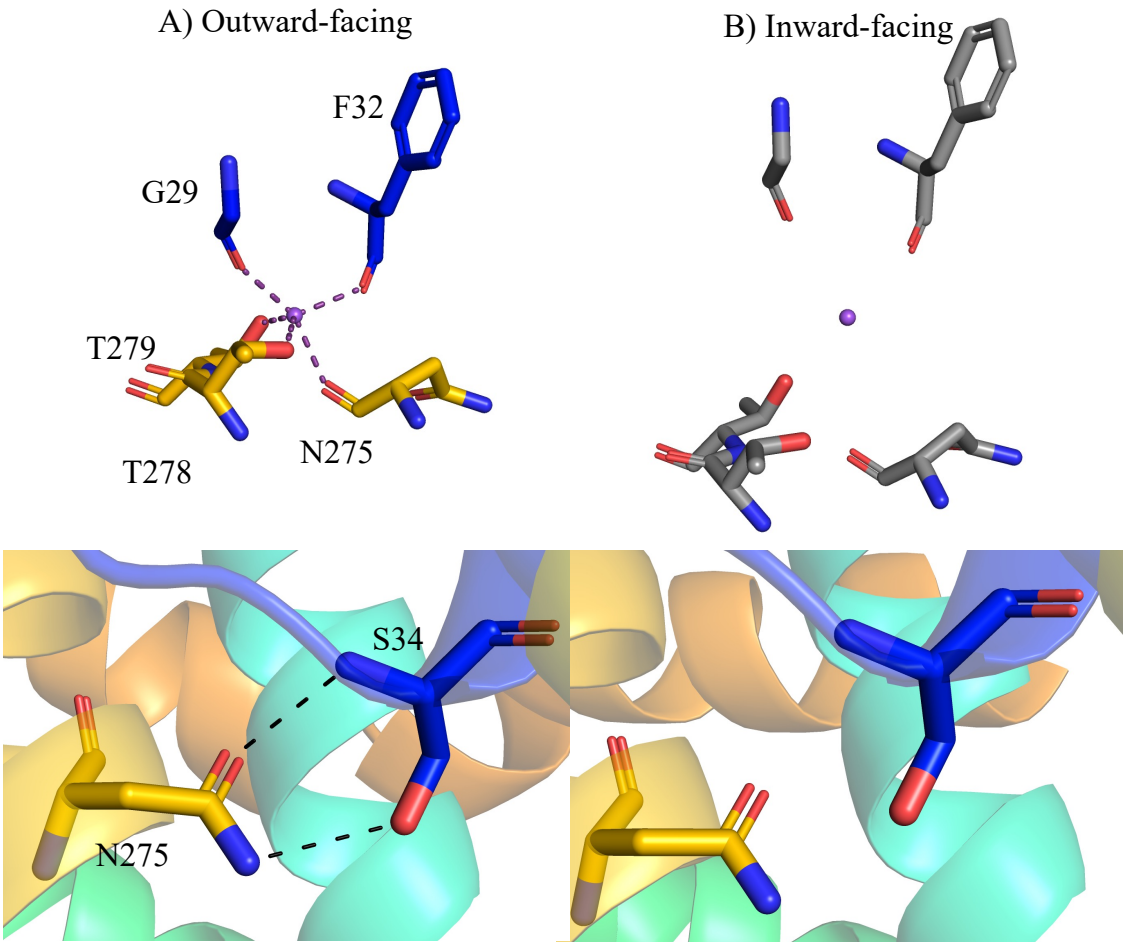


Figure 4.26. Comparison of the sodium binding site in the outward and inward structures. Panel A, Sodium coordination in the outward-open structure. Panel B, The inward-open model, residues involved in sodium coordination have moved away from their original position are not longer able to coordinate sodium correctly, sodium is depicted for reference. Panel C and D hydrogen bonds allowing Asn275 and Ser34 to interact in the outward-open structure and inward-model respectively. Sodium-oxygen interactions are pictured using a purple dashed line, whilst Hydrogen bonds are depicted as black dashed lines.

Asp281 has also moved away from its original position in the outward-facing structure and is no longer in an optimal position to hydrogen bond with the side chain of Thr142 due to the bending of TM5 (Figure 4.27).

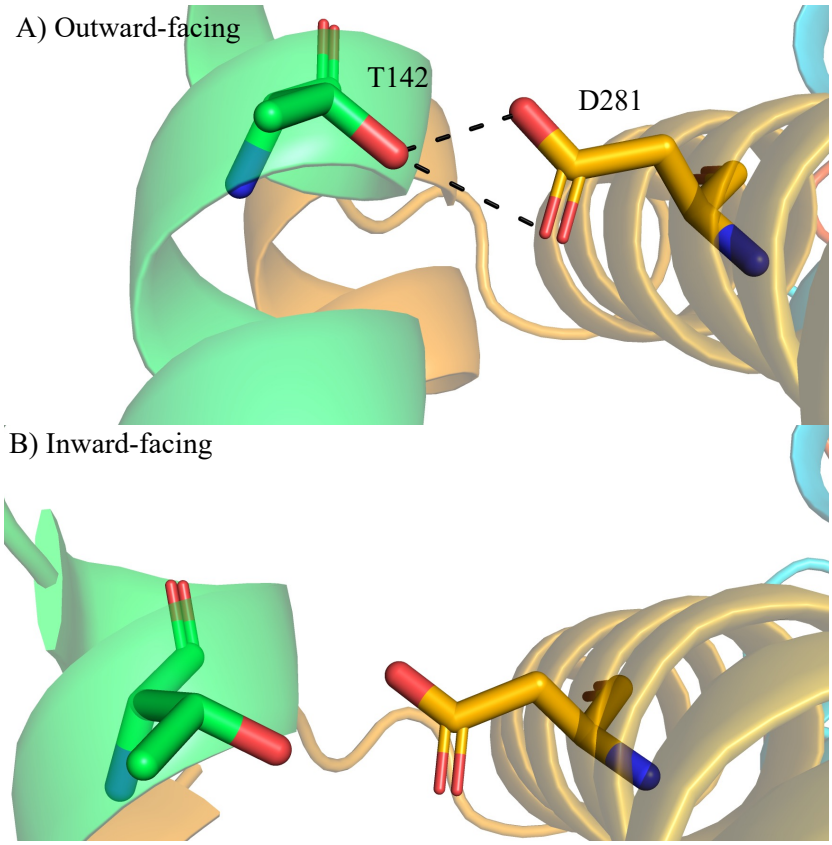


Figure 4.27. Comparison of Asp281 interactions in the outward-open structure and the inward-open model. Outward-open structure (Panel A) and inward-open model (Panel B), Asp281 no longer in an optimal position to interact with Thr142 due to the bending of TM5. Hydrogen bonds are depicted as black dashed lines.

The hydrogen bonding network of Asn282, Val26, Ser154 is completely disrupted in the inward-facing model due to the rotation of TM8 away from TM1 and the bending of TM5 away from TM1 (Figure 4.28), this is the most extreme disturbance of the hydrogen bonding networks.

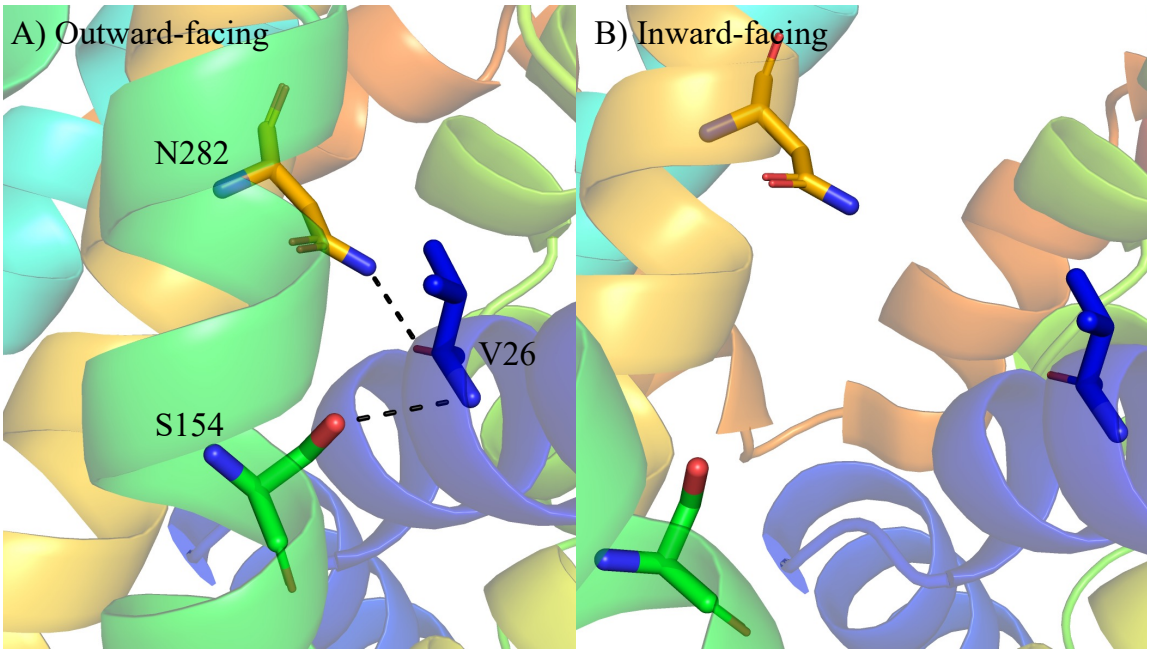


Figure 4.28. Comparison of the hydrogen bonding network of Asn282, Val26, and Ser154 in the outward-open structure and inward facing model. Outward-open structure (Panel A) and inward-open model (Panel B). Rotation of TM 8 away from TM1, and the bending of TM5 results in the movement of Asn282 and Ser154 away from Val26 to disrupt the hydrogen bonding network of TM1,5, and 8. Hydrogen bonds are depicted as black dashed lines.

4.6. Conclusions and Discussion

CodB from *P. vulgaris* has been successfully crystallised using vapour diffusion and LCP crystallisation methods, with both types of crystals diffracting and shown to be protein. The crystals from vapour diffusion were unfortunately low-resolution, however, despite this a dataset was collected, but was severely anisotropic and could not be solved. Crystallisation of CodB in LCP generated several datasets of various resolutions and allowed a structure of CodB to be determined to 2.4 Å.

Solving the phase problem was not as easy as initially anticipated. Mhp1 in an outward-open conformation (2JLN) was used a model for molecular replacement, but did not work, neither did outward-occluded (4D1A) and inward-open (2X79) Mhp1. Mhp1 might not have worked as molecular replacement model due to the low-sequence identity of CodB and Mhp1 (22%). Instead, Rosetta was used to build a model of CodB based on 2JLN, and this model did allow the dataset to be solved, however, this model did not fit the electron density well, and instead 2JLN was rebuilt into CodB due to its good secondary structure; producing a structure of CodB bound to cytosine and sodium in an outward-open conformation.

This structure is found also bound to DDM, phospholipid, and monoolein. The molecule of DDM is particularly interesting as it is found in a cavity between TM10 (the putative extracellular gate) and the 4-helix bundle and could be holding CodB in an artificial crystallisation state by preventing closure of the extracellular gate.

Analysis of the cytosine binding site in CodB and the hydantoin binding pocket of Mhp1 highlights some shared characteristics. Both CodB and Mhp1 utilise π-stacking of aromatics residues to hold their solutes in the binding pocket, however, CodB uses face-to-face interactions and forms a π-stacking "sandwich". In contrast, Mhp1 has a face-to-face interaction for the aromatic on TM3 (Trp117) and an edge-to-face interaction for Trp220 on TM6. CodB and Mhp1 both use surrounding residues to orientate their respective substrates in the solute binding site; CodB also has 2 water molecules coordinated in the site. However, CodB is at a higher-resolution and has a soluble substrate, whilst Mhp1 has an insoluble substrate it is

also at a lower-resolution, meaning water may not be visible in the binding site even if it is present.

This structure of CodB also showed electron density in a site consistent with the Na2 site across the LeuT superfamily. H^+ would not be visible with x-ray diffraction, and K^+ would not be accommodated in the site due the largest diameter of K^+, and K^+-coupled transporters are exporters so K^+ would not be expected as the co-ion. From this Na^+ is the most likely co-ion driving transport and building Na^+ into this site has B-factors consistent with the surrounding residues. Thr278 and Thr279 appear to be coordinating sodium in the Na2 site through their side chains whilst Asn275 is using its main chain carbonyl to hold sodium in place. Asn275 is a residue of interest, looking at the LeuT superfamily, all members apart from CodB have a hydrophobic side chain at this position whilst CodB uses the polar interactions of the Asn275 side chain to interact with TM1. Mhp1 has been shown to require sodium to be capable of binding its BH and IMH substrates. Ala309 is the corresponding residue in Mhp1 and a Ala309Asn mutation has removed the sodium-dependency of Mhp1, meaning that Ala309Asn can bind BH and transport IMH with high affinity in the presence and absence of sodium (Jackson, 2012).

An inward-open model of CodB was built by manually rotating the hash-motif relative to the 4-helix bundle, TM5 was then straightened and TM10 bent over the cytosine binding pocket. This inward-open model was used a model for molecular replacement of the vapour diffusion dataset, but this did not solve the dataset. It is interesting to note that the vapour diffusion crystals were grown in the absence of sodium in the crystallisation condition; addition of sodium did not improve diffraction. This coupled with the observation of the vapour diffusion crystals in a different space group, meaning different crystal packing, it is possible the crystals grown from vapour diffusion have CodB in a different conformation.

From the structure of CodB bound to cytosine the following residues are implicated in cytosine binding, Gln105, Trp108, Phe204, Ser206, and Asn280. These residues will be investigated functionally in Chapter 5. The sodium-binding residues Asn275, Thr278, and Thr279 will be investigated in Chapter 5.

Chapter 5-Functional Characterisation of CodB

5.1. Introduction

As described in Chapter 1, secondary-active transporters are integral membrane proteins responsible for the transport of molecules across biological membranes. Functional characterisation of these transporters is difficult due to the hydrophobic and fragile nature often associated with integral membrane proteins. Enzymatic assays have the benefit of monitoring the production or breakdown of their products/reactants whilst transporters do not chemically alter their substrate. Separating binding events and transport kinetics is required for complete annotation and understanding. Investigating binding is relatively flexible with multiple methods to try. Monitoring a change in tryptophan fluorescence was used successfully with Mhp1 to measure hydantoin binding, however, due to cytosine absorbance overlapping with tryptophan absorbance spectra this could not be used for CodB. Thermoshift experiments, monitoring the stabilisation of the transporter in the presence of putative solute has been utilised successfully (CPM Assay and GFP-TS) and is useful as a preliminary screening tool to identify ligands. Surface plasma resonce (SPR), MST, and Quartz Crystal Microbalance with Dissipation (QCM-D) have risen in popularity as methods for determining binding in soluble proteins and have been used somewhat successfully for membranes proteins, MST and SPR have the benefit of using small amount of pure protein. ITC is still the gold-standard approach for measuring protein binding and has been used successfully for membrane proteins but can require large amounts of purified protein and therefore other methods may be preferred.

Binding assays are an excellent tool for identifying molecules capable of binding but do not give information about whether they are an inhibitor or a solute capable of being transported, therefore transport assays must be done in combination to accurately annotate a transporter. Transporters that use protons as a co-ion can be investigated by pH indicators, but sodium ions are more difficult and generally require radioactivity as a label.

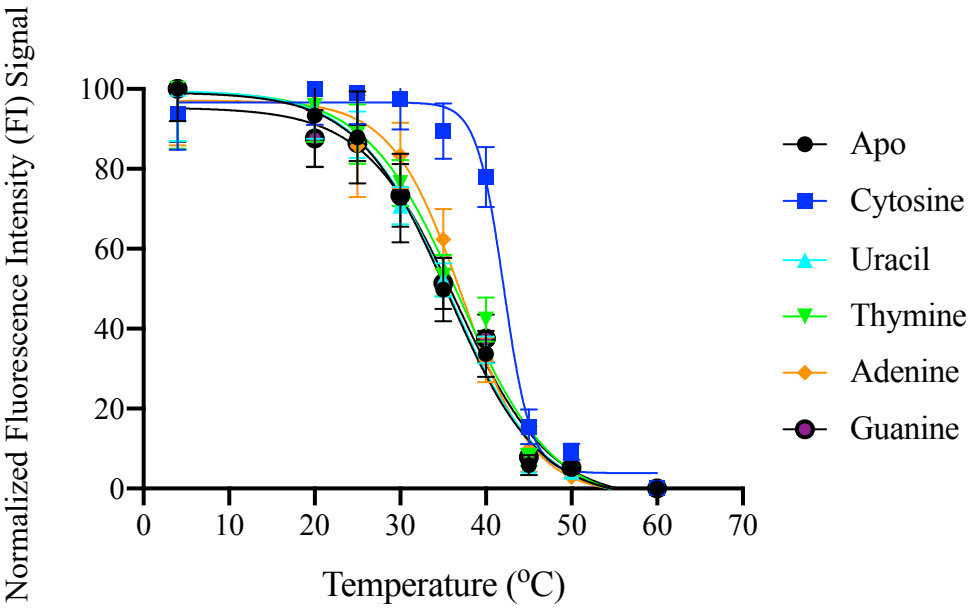
5.2. Results

5.2.1. Investigating wt CodB Ligand Binding

5.2.1.1 GFP-TS to generate melting temperatures and screen for ligands

GFP-TS has been previously utilised as a screening tool for protein-ligand interactions (Nji et al., 2018). Membranes containing the protein of interest overexpressed with a c-terminal GFP tag are solubilised and subjected to thermal denaturation, whilst monitoring the fluorescence of GFP. The melting temperature (T_m) of detergent solubilised CodB.GFP was investigated, in the presence of naturally occurring nucleobases, an increase in the T_m is indicative of protein stabilisation due to the ligand.

Addition of 1 mM cytosine to the membrane solubilisation of CodB resulted in a stabilisation of 6.4 ^oC whilst none of the other nucleobases tested showed any stabilisation, suggesting that they do not bind CodB (Figure 5.1). Next, GFP-TS was used to screen for small molecules that are similar to cytosine (Figure 5.2). Isocytosine, an isomer of cytosine, and 5-methylcytosine did not stabilise CodB, but 5-fluorocytosine did show an increase in the T_m of 6.7 ^oC suggesting that 5-fluorocytosine could be a ligand.



Condition	Melting temperature, T_m (°C)	T_m difference (°C)
Apo	35.7 ± 2.8	n/a
1 mM cytosine	42.1 ± 0.6	+6.4
1 mM uracil	35.8 ± 0.9	+0.1
1 mM thymine	36.9 ± 1.7	+1.2
1 mM adenine	37.3 ± 2.3	+1.6
1 mM guanine	36.6 ± 2.9	+0.9

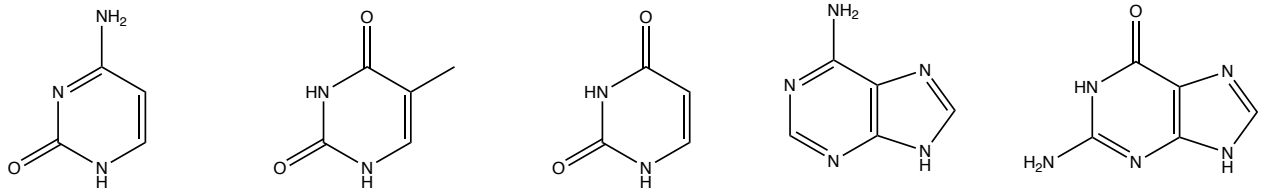
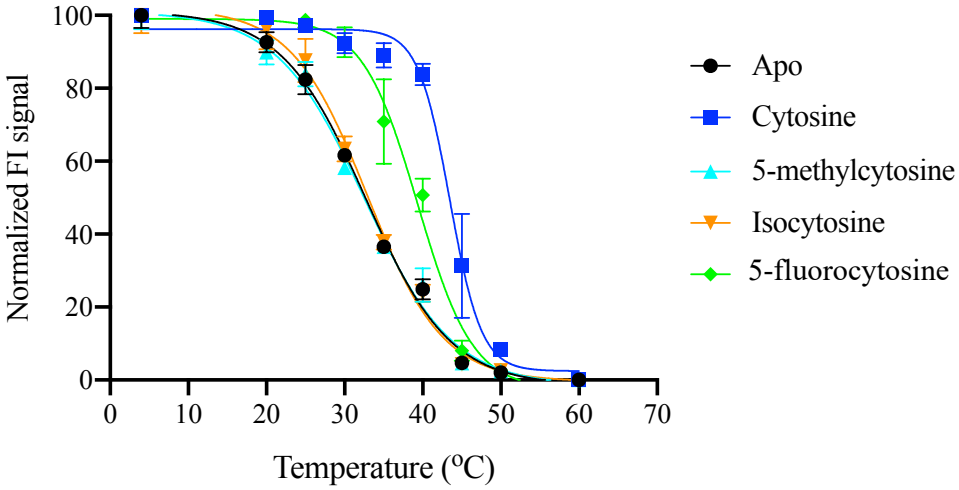


Figure 5.1. GFP-TS results of wt CodB.GFP with nucleobases. The melting temperature of wt CodB.GFP increases in the presence of 1mM cytosine indicative of substrate binding. Addition of other nucleobases did not result in a shift of the melting temperature. Points plotted are the average mean of 2 technical repeats with error bars of the standard error of the mean (S.E.M.)



Condition	Melting temperature, T_m (°C)	T_m difference (°C)
Apo	32.7 ± 0.1	n/a
1 mM cytosine	43.5 ± 2.1	+10.8
1 mM 5-methylcytosine	32.6 ± 0.8	-0.1
1 mM isocytosine	32.8 ± 0.2	+0.1
1 mM 5-fluorocytosine	39.4 ± 1.2	6.7

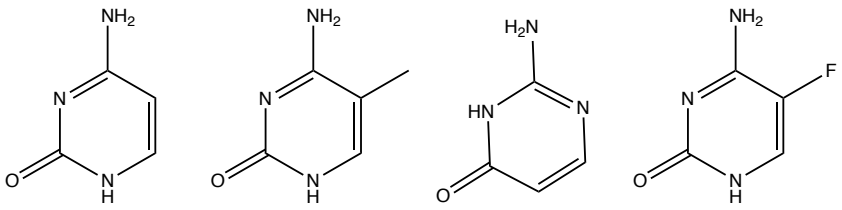


Figure 5.2. GFP-TS results of wt CodB.GFP with cytosine-related compounds. The melting temperature of wt CodB.GFP increases in the presence of 1mM cytosine and 5-fluorocytosine indicative of ligand binding. Points plotted are the average mean of 2 technical repeats with error bars of the S.E.M..

5.2.1.2. GFP-TS to investigate ligand binding

To quantify the binding affinity of CodB wt to bind cytosine and other ligands, GFP-TS was used to generate an equilibrium dissociation constant (K_d). Figure 5.3 shows that CodB is able to bind cytosine with K_d of 51.3 µM ± 9.1, 3.s.f.

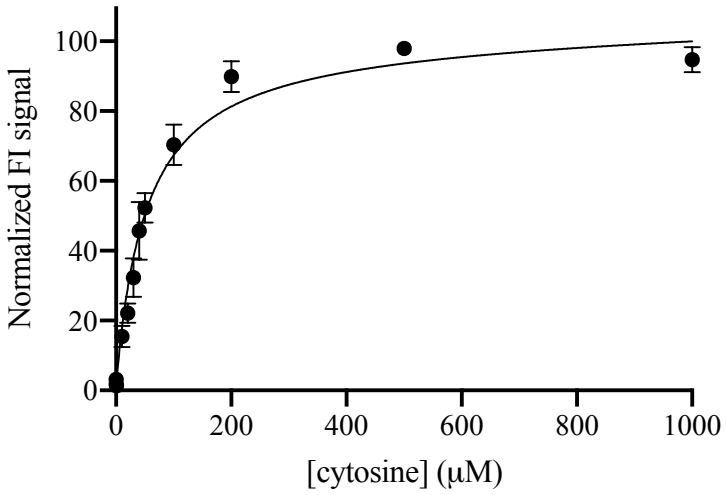


Figure 5.3. Binding affinity of CodB.GFP wt to cytosine. Cytosine was titrated into detergent solubilised membranes with CodB.GFP overexpressed. CodB.GFP can bind cytosine with a K_d ± S.E.M. of 51.3 µM ± 9.1, 3.s.f. Points plotted are the average of 4 independent titrations with error bars of the S.E.M.

5-fluorocytosine, a cytosine analogue and antifungal medication, was used by Danielsen and colleagues to monitor CodA activity in *E. coli*, cells with cytosine deaminase activity are susceptible to cell death by CodA deaminating 5-fluorocytosine into the toxic 5-flourouracil (Danielsen *et al.*, 1992). CodA is found in the same operon as CodB and therefore 5-fluorocytosine is also plausible as a ligand for CodB if CodA is capable of turning over 5-fluorocytosine. The addition of 1 mM 5-fluorocytosine resulted in a stabilisation compared to ligand-free CodB. Therefore 5-fluorocytosine was also investigated by changing the concentration of 5-fluorocytosine to determine a K_d of 285 µM ± 39, 3.s.f. (Figure 5.4).

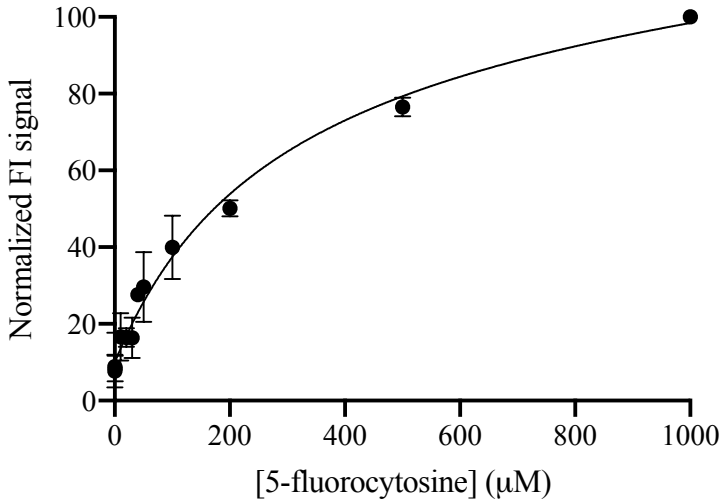


Figure 5.4. Binding affinity of CodB.GFP wt to 5-fluorocytosine. Cytosine was titrated into detergent solubilised membranes with CodB.GFP overexpressed. CodB.GFP can bind cytosine with a K_d ± S.E.M. of 285 μM ± 39, 3.s.f., n=3. Points plotted are the average of 3 independent titrations with error bars of the S.E.M.

5.2.1.3. Isothermal Calorimetry (ITC) to investigate cytosine binding

The ITC machine was provided by Malvern as a demonstration of its capabilities. CodB was prepared as described in materials and methods with extensive dialysis. Binding of cytosine to CodB did induce a response, but binding saturated quickly. Software calculated a K_d of 2.16 μM. Due to the limited amount of time on the machine a K_d of CodB and cytosine binding could not be accurately determined but shows that ITC would be suitable for measuring cytosine and CodB binding.

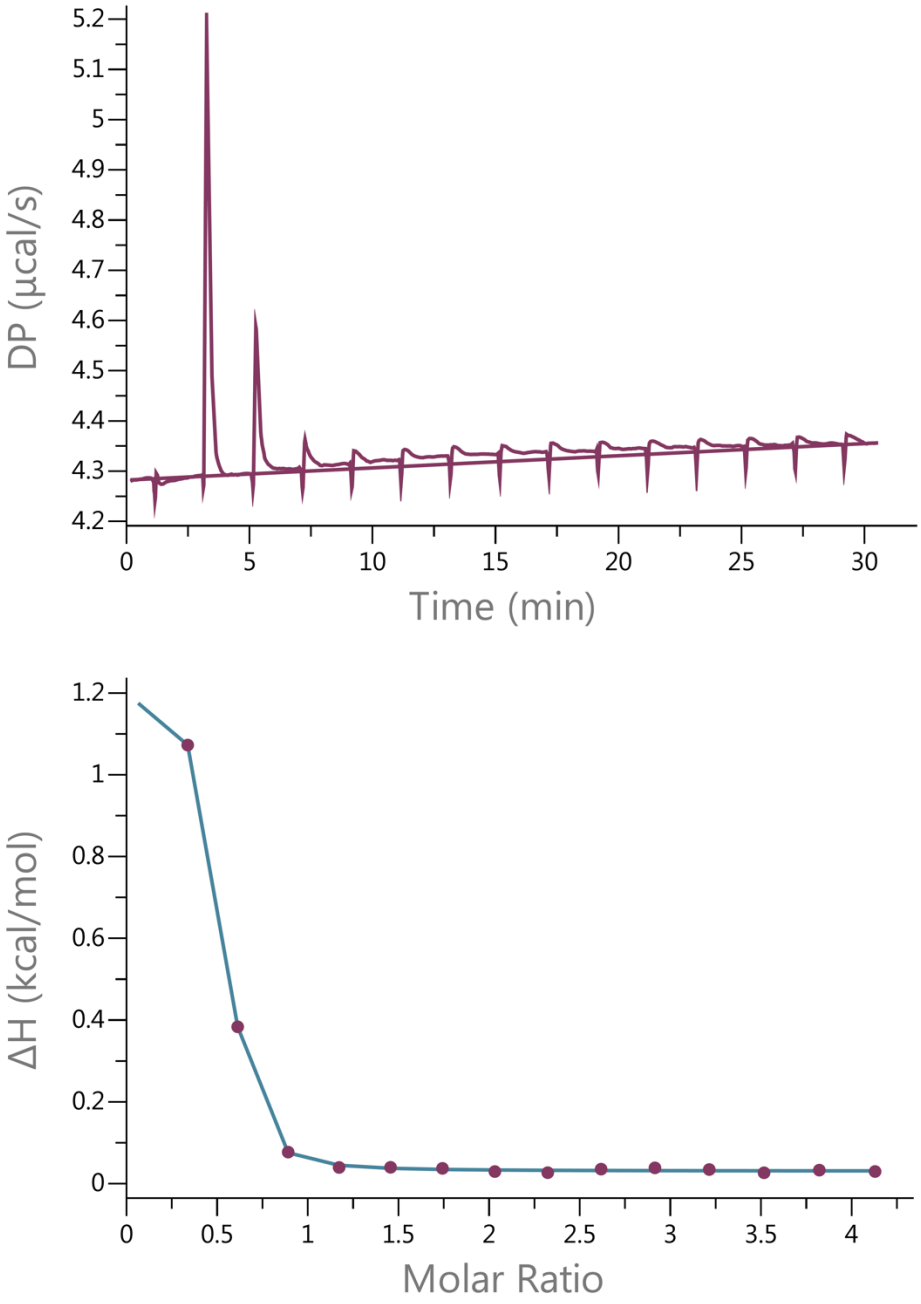


Figure 5.5. Isothermal calorimetry of CodB and cytosine. Raw temperature change is plotted against time in the top graph, with the processed data in the bottom graph.

5.2.1.4. CPM Assay

Purified CodB was prepared into a CPM assay as described in Chapter 2 to investigate detergent stability. CPM binds to exposed cysteine residues, this assay works best if all cysteine residues are found in the core of protein and become exposed to the surrounding solvent when protein unfolds.

CodB did not produce sigmoidal curves as expected (an indicator of protein unfolding), and once the structure of CodB was known, the results were explainable. CodB has one cysteine residue for the maleimide dye to bind to, on the outside of the protein. This assay is dependent on Cys residues being inaccessible to the CPM dye until some unfolding has occurred, whereas in the case of CodB, the single Cys residue is always accessible to the CPM dye, and therefore wt CodB is not suitable for this experiment.

5.2.1.5. Microscale Thermophoresis (MST)

CodB.GFP fusion protein was produced as described in materials and methods, then subjected to MST to investigate if cytosine binding could be quantified. Despite varying standard parameters such as buffer conditions, detergents, CodB.GFP in the presence of 1 mM cytosine did not show a difference compared to ligand-free CodB.GFP. An alternative labelling strategy of a maleimide dye did not resolve the issue.

5.2.2. Investigating wt CodB Transport of ^{3}H-5-cytosine

5.2.2.1. In-cell Radioactive Assay

A time course of ^{3}H-5-cytosine transport by CodB shows ^{3}H-5-cytosine is internalised by *E. coli* overexpressing CodB to a maximum of 400 pmol mg^{-1} of CodB in the presence of 150 mM sodium; appeared to reach a plateau after about 2 minutes (Figure 5.6). Lemo cells not transformed with CodB.pWaldo plasmid were used as background. The same time course was repeated in parallel, but instead 150 mM choline chloride was used instead of 150 mM NaCl. In the absence of sodium, cytosine transport was inhibited to 20 pmol mg^{-1} of CodB. From this, sodium is likely to be the co-ion driving transport.

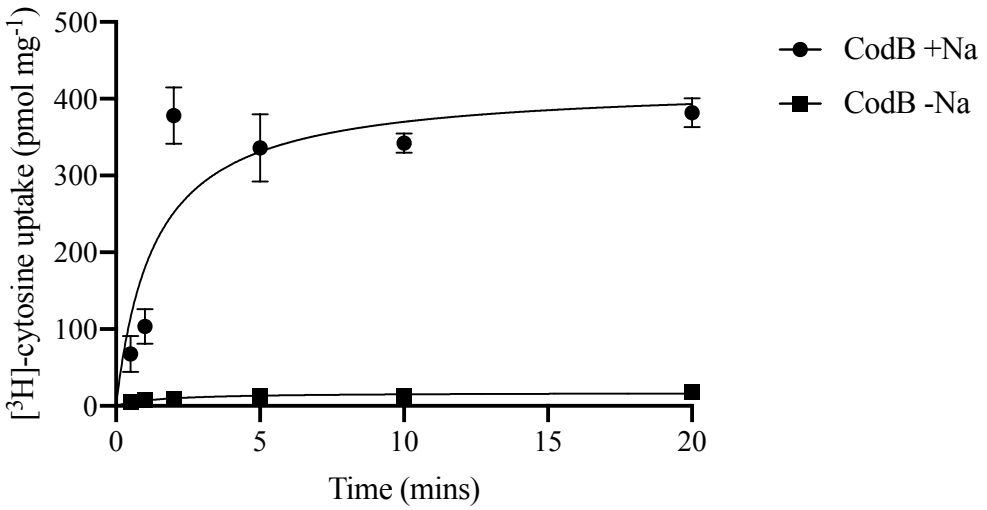
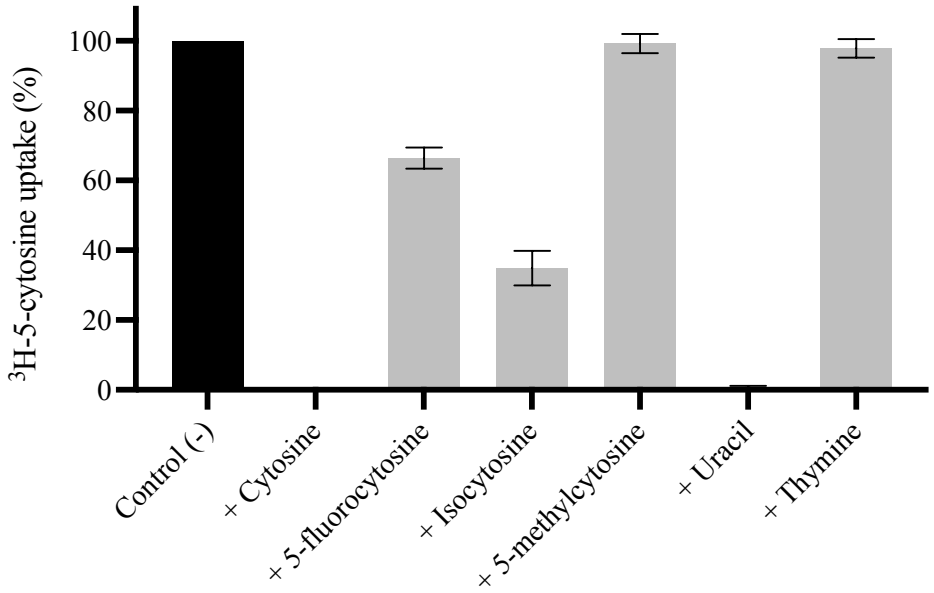


Figure 5.6. Time course of 3H cytosine uptake by CodB.GFP. Lemo21 (DE3) cells were used as a background measurement, with Lemo21(DE3) expressing CodB.

To investigate the substrate specificity of CodB, ^{3}H-5-cytosine uptake in *E. coli* overexpressing CodB was determined in the presence of 0.1 mM potential substrates (Figure 5.7). CodB exhibits a broader range of specificity than expected, especially in the context of the GFP-TS assays (Section 5.2.1.1). As shown in Figure 5.7, addition of 0.1 mM cytosine completely inhibited transport, solidifying the role of CodB in cytosine uptake. Addition of 0.1 mM 5-fluorocytosine reduced transport to 66%, this ability to inhibit cytosine transport suggests that 5-fluorocytosine could be a ligand for CodB, albeit with a lower binding affinity. In the case of isocytosine and uracil, they both reduced cytosine transport to 35% and 1% respectively, indicating they could be ligands for CodB, despite not exhibiting any stabilisation in GFP-TS. 5-methylcytosine and thymine did not show any inhibition and is there is no evidence these compounds would be potential binding partners for CodB.



Compound	^{3}H-5-cytosine uptake (%)
Cytosine	-4.0 ± 0.2
5-fluorocytosine	66.4 ± 1.8
Isocytosine	34.9 ± 2.8
5-methylcytosine	99.2 ± 1.6
Uracil	0.7 ± 0.2
Thymine	97.9 ± 1.5

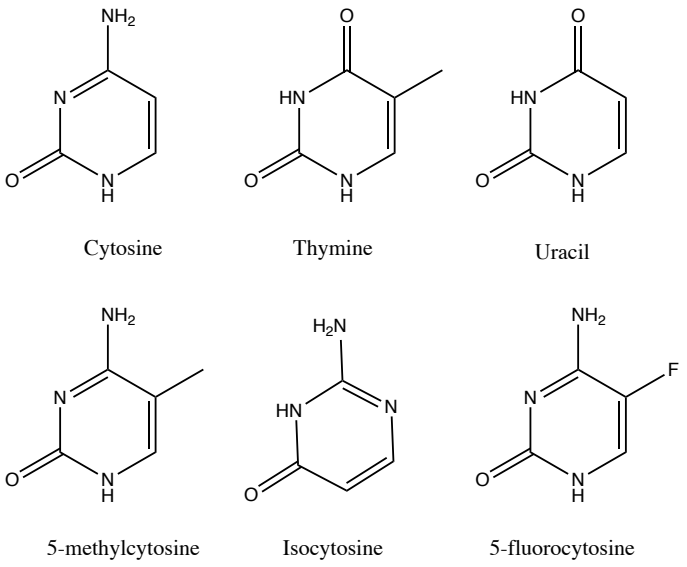


Figure 5.7. Inhibition of ^{3}H-5-cytosine uptake in the presence of 0.1 mM inhibitor. Uptake of ^{3}H-5-cytosine was measured after 1 minute with 0.1 mM inhibitor. Control (-) is uptake of ^{3}H-5-cytosine with no inhibitor, normalised to 100%, results are visualised as % of control (-) with error bars as S.E.M. of triplicate experiments, each from a different culture.

5.2.2.2. Proteoliposome Radioactive Transport Assay

CodB was reconstituted into liposomes as described (Section 2.8.2.). The first method (extrusion) resulted in the successful reconstitution of CodB into proteoliposomes (Figure 5.8), with subsequent transport assays showing very little [^{3}H]-cytosine uptake for empty liposome and proteoliposomes. Conditions for the transport assay were investigated by changing incubation time with the radiolabel, incubating with co-ions Mg^{2+}, Li^+, and K^+ and finally addition of gramicidin. None of these yielded any better results, and there was no noticeable difference between the empty liposomes and proteoliposomes (Figure 5.8).

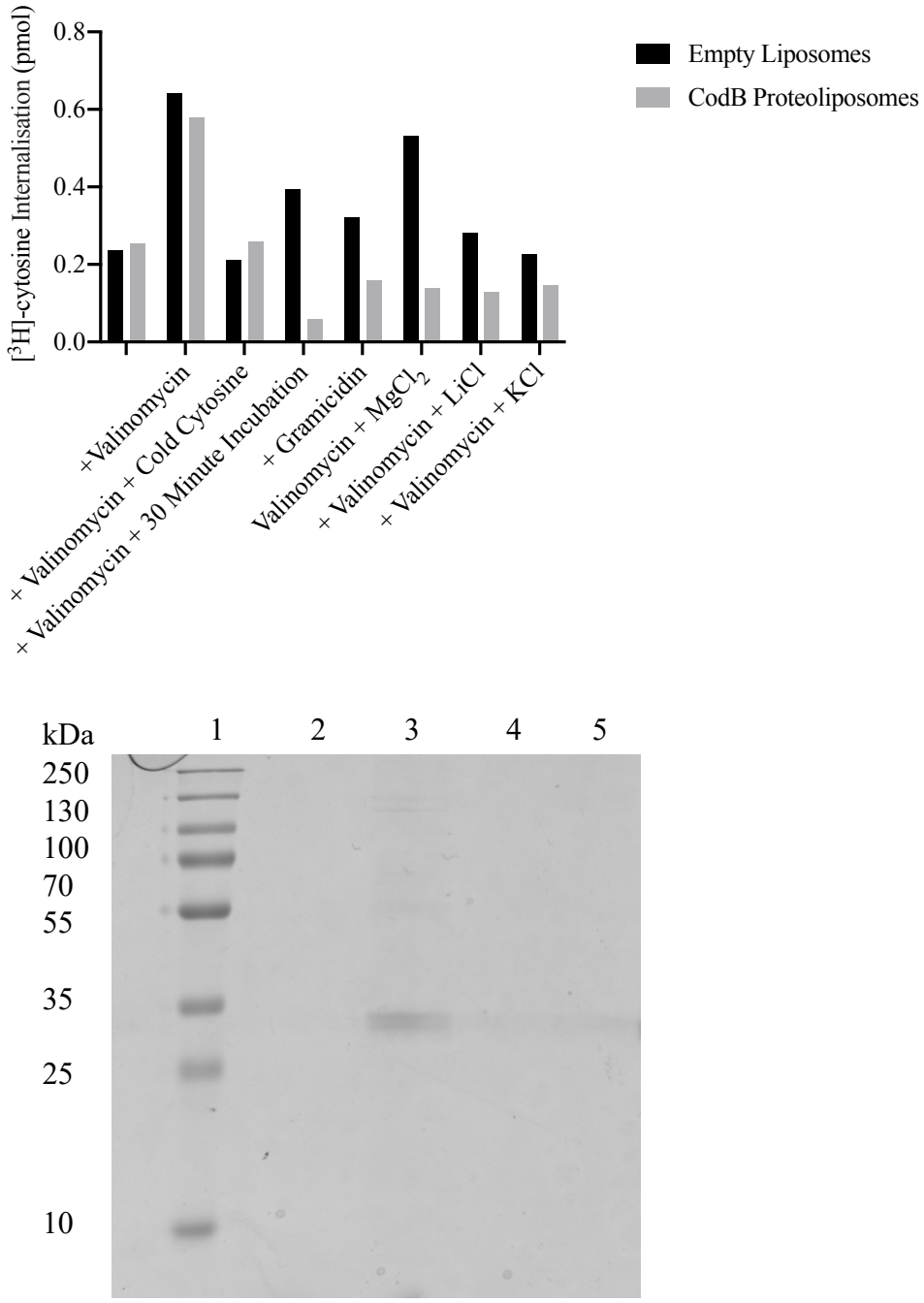


Figure 5.8. CodB liposomes prepared by the extrusion method. Graph showing the uptake of [^{3}H]-cytosine in empty liposomes and CodB proteoliposomes. SDS-PAGE gel of liposomes. Lane 1= protein ladder, lane 2= empty liposomes diluted 1:2, lane 3= CodB proteoliposomes diluted 1:2, lane 4= empty liposomes diluted 1:10, lane 5= CodB proteoliposomes diluted 1:10.

After this, a destabilisation method was attempted, with empty liposomes provided by Patrick Becker. These transport assays did not fare any better than the extrusion method previously tested, and analysis of SDS-PAGE gels do not show that CodB was successfully reconstituted into the liposomes.

Finally, a rapid dilution method was attempted with the help of Dr Gareth Hughes (University of Birmingham). As shown in Figure 5.9, CodB was successfully reconstituted into proteoliposomes by the presence of a single monodispersed peak with SEC. These liposomes were never tested with the transport assay due to time constraints.

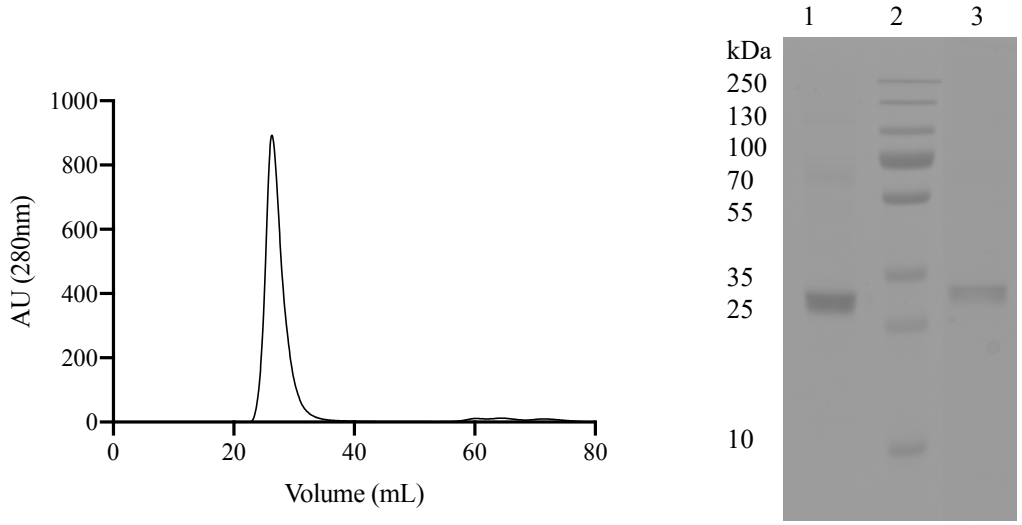


Figure 5.9. CodB proteoliposomes prepared by rapid dilution. CodB proteoliposomes eluted in a monodispersed peak at about ~26 mL with the corresponding SDS PAGE gel. Lane 1= CodB at 2 mg/ml before reconstitution, lane 2=protein ladder, lane 3= CodB proteoliposomes from the most concentrated fraction.

5.2.3. Mutagenesis of Trp108 and Phe204 to investigate their role in cytosine binding and transport

Trp108 and Phe204 are implicated in substrate binding based on sequence alignments with Mhp1; corroborated by the structure of CodB discussed in detail in Chapter 4. Trp108 and Phe204 appear to be interacting by π-stacking interactions with the cytosine substrate (Figure 5.10). Mutagenesis of Trp108 and Phe204 into alanine would remove the π-stacking interaction and severely impair ligand binding.

Mutagenesis of these residues were performed by an undergraduate project student, Mehalah Spencer, that I supervised. Mutants were confirmed by Sanger sequencing, these were expressed and membranes harvested as stated in materials and methods.

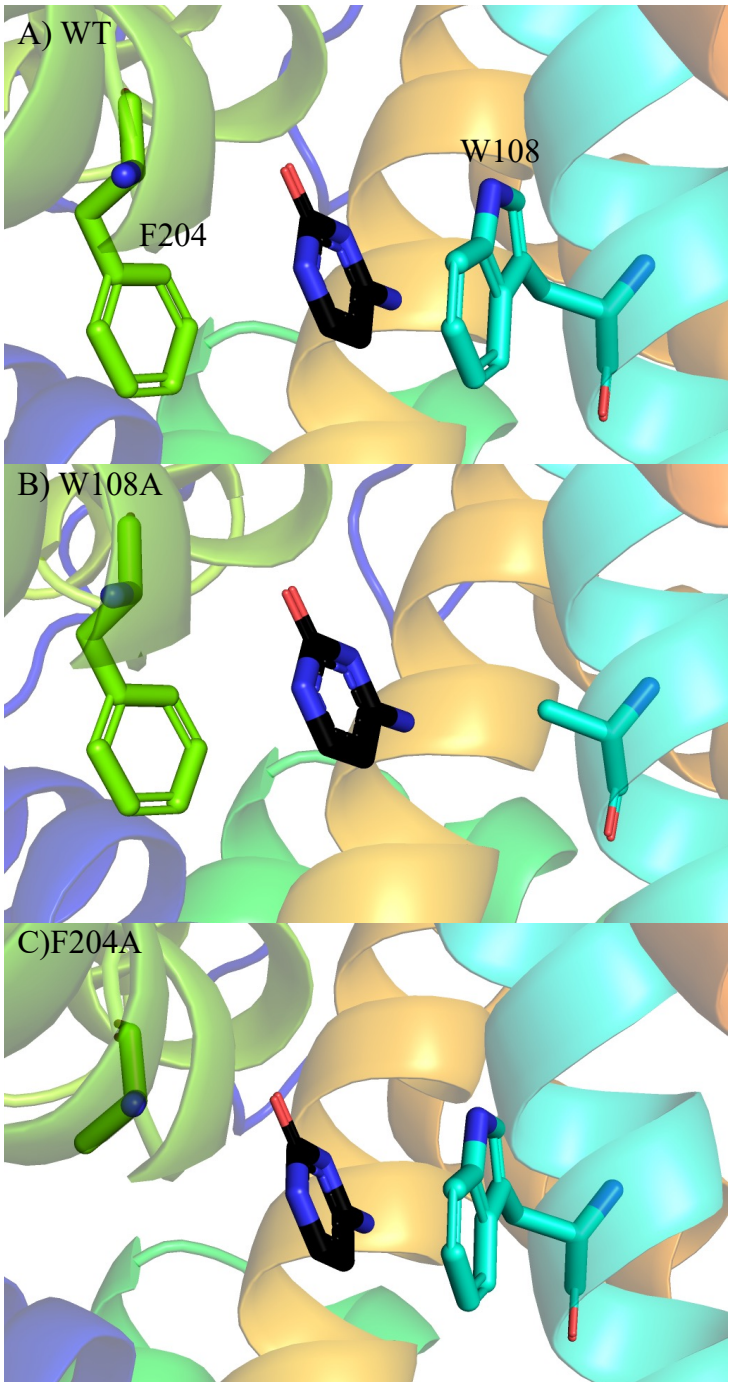
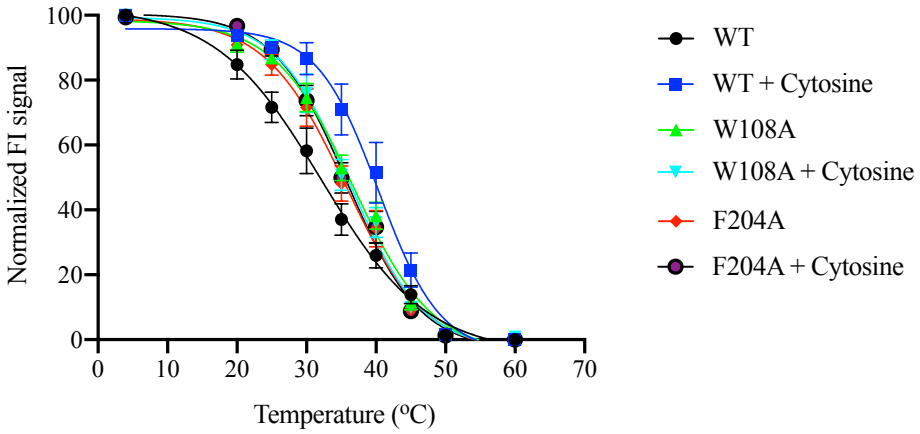


Figure 5.10. Interactions of Trp108 and Phe204 with cytosine and the impact of mutants Trp108Ala, Phe204Ala on cytosine binding. Panel A, wt CodB. Panel B, Trp108Ala. Panel C, Phe204Ala.

5.2.3.1. GFP-TS to investigate substrate binding

These thermoshift experiments were performed by Mehalah Spencer (Figure 5.11), all further work is my own.

Thermoshift assays were used to see if CodB mutants were still stabilised in the presence of 1 mM cytosine. Mutants Trp108Ala and Phe204Ala both exhibit a higher T_m than wt CodB. Wt CodB has a stabilisation of 8.2 °C upon the addition of 1 mM cytosine, whereas Trp108Ala and Phe204Ala do not demonstrate an increase in T_m when exposed to 1 mM cytosine suggesting that these mutants do not bind cytosine or that substrate binding is impaired (Figure 5.11).



Condition	Melting Temperature, T_m (°C)	T_m Difference (°C)
WT	31.9 ± 3.1	n/a
WT + 1 mM Cytosine	40.0 ± 2.8	8.2
W108A	36.6 ± 1.7	n/a
W108A + 1 mM Cytosine	36.0 ± 2	-0.6
F204A	35.5 ± 2.5	n/a
F204A + 1 mM Cytosine	35.7 ± 1.9	0.2

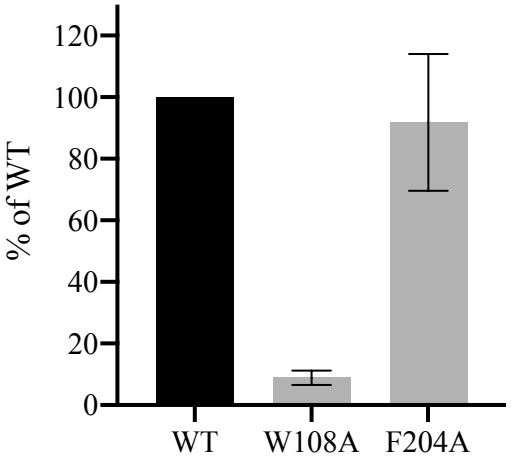
Figure 5.11. Thermoshift experiments of detergent solubilised CodB.GFP wt, Trp108Ala, Phe204Ala. CodB.GFP wt was used as a positive control (in the presence of 1 mM cytosine), and as a negative control (absence of cytosine). T_m of Trp108Ala and Phe204Ala is higher than the T_m of wt, but Trp108Ala and Phe204Ala did not exhibit the same stabilisation compared to wt CodB.GFP. The lack of thermoshift is an indicator that residues Trp108 and Phe204 are important for cytosine binding. Points plotted are the average mean of 2 technical repeats with error bars of the S.E.M..

To quantify the impact of these mutations, GFP-TS was used to generate a K_d for both of these mutants. Binding affinity experiments for wt binding to ligands had much easier interpretation of the data, whereas mutants Trp108Ala and Phe204Ala were more difficult to interpret. Individual titrations all showed different trends, and when combined together the error bars are clearly more variable than the wt. GraphPad prism was able to calculate K_d values, and the curves generated could be

convincing for substrate binding, however, the R^2 values are 0.312 for Trp108Ala and 0.209 for Phe204Ala, suggesting that whilst GraphPad Prism is able to plot a trend, the data do not fit the trend well. Overall, this data is beginning to suggest that these mutants probably abolish substrate binding.

5.2.3.2. Radioactive uptake assay to measure transport of mutants Trp108Ala and Phe204Ala

^{3}H-5-cytosine uptake was almost completely abolished in cells over-expressing Trp108Ala mutant protein as transport was reduced to 8.8 % ± 2.3 compared to WT CodB. However, Phe204Ala retained 91.7 % ± 22.1 of its activity compared to WT (Figure 5.12). From this, Trp108Ala is clearly essential for CodB to bind and transport cytosine. The significance of Phe204Ala is more unclear, in assays the transport activity that Phe204Ala retained was highly variable across experiments, with activity reduced to ~50 % in some replicates, whilst other samples maintained more activity.



Mutant	% of WT Uptake
Trp108Ala	8.8 ± 2.3
Phe204Ala	91.7 ± 22.1

Figure 5.12. Uptake of ^{3}H-5-cytosine by CodB.GFP mutants Trp108Ala and Phe204Ala, normalised to WT. Uptake of ^{3}H-5-cytosine was measured after 1 minute. WT is normalised to 100%, results are visualised as % of WT with error bars as S.E.M. of at least 4 experiments, each from a different culture.

5.2.4. Mutagenesis of Gln105 and Ser206 to investigate their role in cytosine binding and transport

Based on the crystal structure of CodB Gln105 is implicated in cytosine binding as it's side chain appears to be hydrogen bonding directly with the cytosine substrate

(Figure 5.13). Ser206 is near the solute binding site in CodB but does not appear to be interacting with the cytosine substrate and is instead coordinating water, which in turn is interacting with the cytosine. This water molecule is coordinated by several interactions with the backbone of TM6 (Figure 5.13). Mutagenesis of Gln105Ala should result in a loss of the hydrogen bonds of the Gln105 side chain with cytosine, whilst Ser206Ala would remove an interaction to hold the water molecule in the solute site, however, this water is mainly held in position by interacting with exposed dipoles of the main chain of TM6

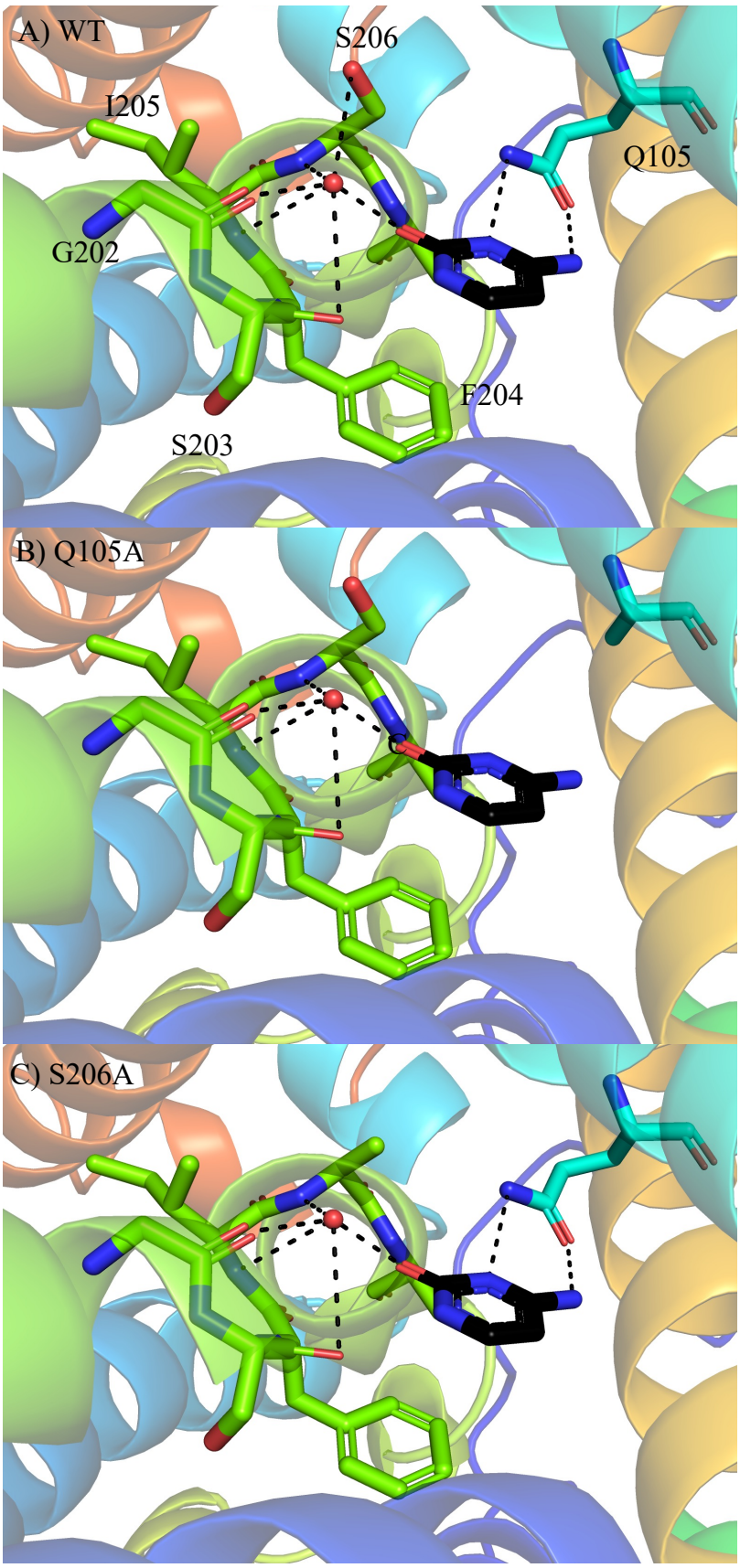
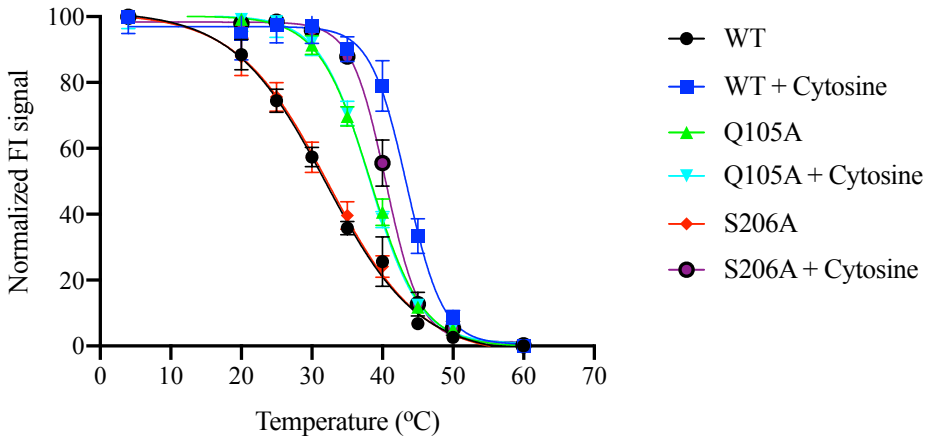


Figure 5.13. Interactions of Gln105 and Ser206 with cytosine and the impact of mutants Gln105Ala, Ser206Ala on cytosine binding. Panel A, wt CodB. Panel B, Gln105Ala. Panel C, Ser206Ala. Hydrogen bonds are depicted as black dashed lines.

5.2.4.1. GFP-TS to investigate substrate binding of Gln105Ala and Ser206Ala

Initial GTP-TS experiments were done as before to determine if these mutants were stabilised in the presence of cytosine (Figure 5.14). Wt CodB exhibits a shift in T_m by 11.5 ^oC whereas Gln105Ala does not exhibit stabilisation in the presence of cytosine but does see a shift in the apo state like Trp108Ala and Phe204Ala. From this it seems that Gln105 is involved in cytosine binding. The apo state of Ser206Ala had a T_m that was comparable to wt CodB and did not exhibit the baseline stabilisation that Gln105Ala, Trp108Ala, and Phe204Ala did. However, Ser206Ala did exhibit a stabilisation upon the addition of 1 mM cytosine of about 8.1 ^oC in independent experiments, the stabilisation of Ser206Ala was always a lower value than the thermoshift seen for wt CodB. From this, Ser206 could be involved with cytosine binding, as mutation of this residue did reduce the stabilisation, however, this could also be argued as simply mutating Ser to Ala could reduce the T_m independently of cytosine binding ability. Therefore, these residues were investigated for their binding affinity.



Condition	Melting Temperature, T_m (ºC)	T_m Difference (ºC)
WT	32.0 ± 1.5	n/a
WT + 1 mM Cytosine	43.4 ± 0.8	11.5
Q105A	38.3 ± 0.7	n/a
Q105A + 1 mM Cytosine	38.2 ± 0.2	-0.1
S206A Apo	32.3 ± 1.9	n/a
S206A + 1 mM Cytosine	40.5 ± 0.6	8.2

Figure 5.14. Thermoshift experiments of detergent solubilised CodB.GFP wt, Gln105Ala, Ser206Ala. CodB.GFP wt was used as a positive control (in the presence of 1 mM cytosine), and as a negative control (absence of cytosine). T_m of Q105A is higher than the T_m of wt, however, Q105A does not demonstrate the same stabilisation compared to wt CodB. The lack of thermoshift implicates Gln105 in a role responsible for cytosine binding. S206A is more ambiguous. Points plotted are the average mean of 2 technical repeats with error bars of the S.E.M..

When attempting to determine a K_d Gln105Ala behaved the same as Trp108Ala and Phe204Ala and a reliable K_d could not be determined, therefore Gln105 is most likely to be involved with cytosine binding. Ser206Ala was easier to interpret, as Ser206Ala clearly demonstrated stabilisation upon the addition of cytosine. Ser206Ala exhibited a K_d of 297 µM ±109 3.s.f. (Figure 5.15). Reduction of the binding affinity of Ser206Ala suggests that the hydroxyl-side chain of Ser206Ala does assist in the coordination of cytosine in the binding site but is not essential to enable substrate binding, which is consistent with the water being held in position mainly by the main chain interactions from TM6 which will have remained.

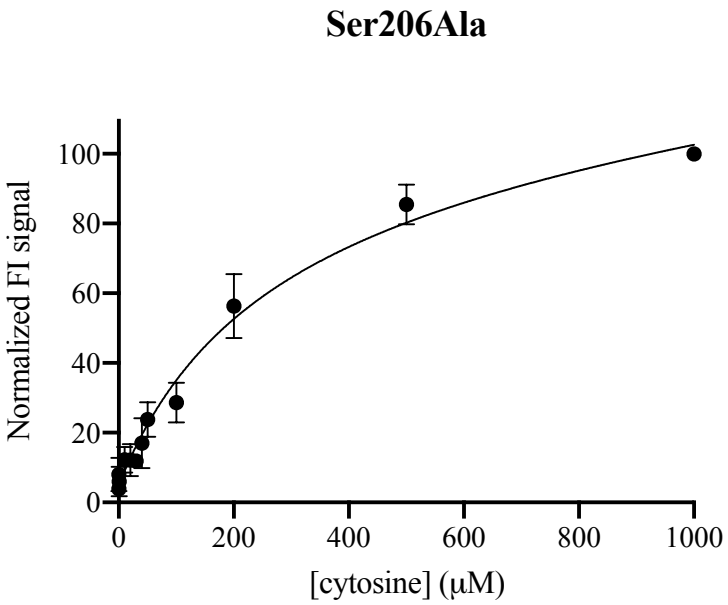
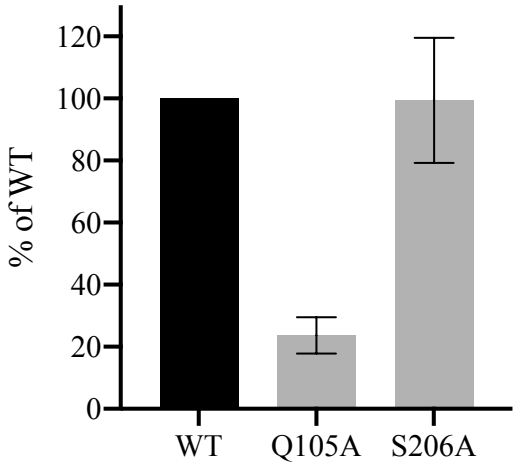


Figure 5.15. Binding affinity of CodB.GFP mutants Ser206Ala. Cytosine was titrated into detergent solubilised membranes with CodB.GFP Ser206Ala overexpressed. Points plotted are the average of 3 independent titrations with error bars of the S.E.M..

5.2.4.2. Radioactive uptake assay to measure transport of mutants Gln105Ala and Ser206Ala

Transport data for the Gln105Ala mutant shows that Gln105Ala reduces transport to 24.3 % ± 7.2 compared to WT CodB (Figure 5.16), suggesting that Gln105 is important for cytosine coordination and transport. The Ser206Ala mutation shows transport activity of 99.4 % ± 20.1 compared to WT CodB (Figure 5.16), suggesting that Ser206 is not especially important for cytosine transport.



Mutant	% of WT Uptake
Gln105Ala	24.3 ± 7.2
Ser206Ala	99.4 ± 20.1

Figure 5.16. Uptake of ^{3}H-5-cytosine by CodB.GFP mutants Gln105Ala and Ser206Ala, normalised to WT. Uptake of ^{3}H-5-cytosine was measured after 1 minute. WT is normalised to 100%, results are visualised as % of WT with error bars as S.E.M. of at least 4 experiments, each from a different culture.

5.2.5. Mutagenesis of Asn280 to investigate its role in cytosine binding

From the structure of CodB bound to cytosine, it seemed that Asn280 is found in the binding pocket of CodB interacting with cytosine via an intermediatory water molecule acting as a bridge, whilst also interacting with a methionine side chain on TM4, from this it seems plausible that Asn280 has a role in cytosine coordination and could be a coupling residue of cytosine binding and the opening of the intracellular gate (Figure 5.17). To investigate the role that Asn280 plays in cytosine coordination, the mutant Asn280Ala was made, this should abolish the hydrogen bonding interaction of the Asn280 side chain with the water molecule coordinated in the solute site; this water is also interacting with the carbonyl of Trp108.

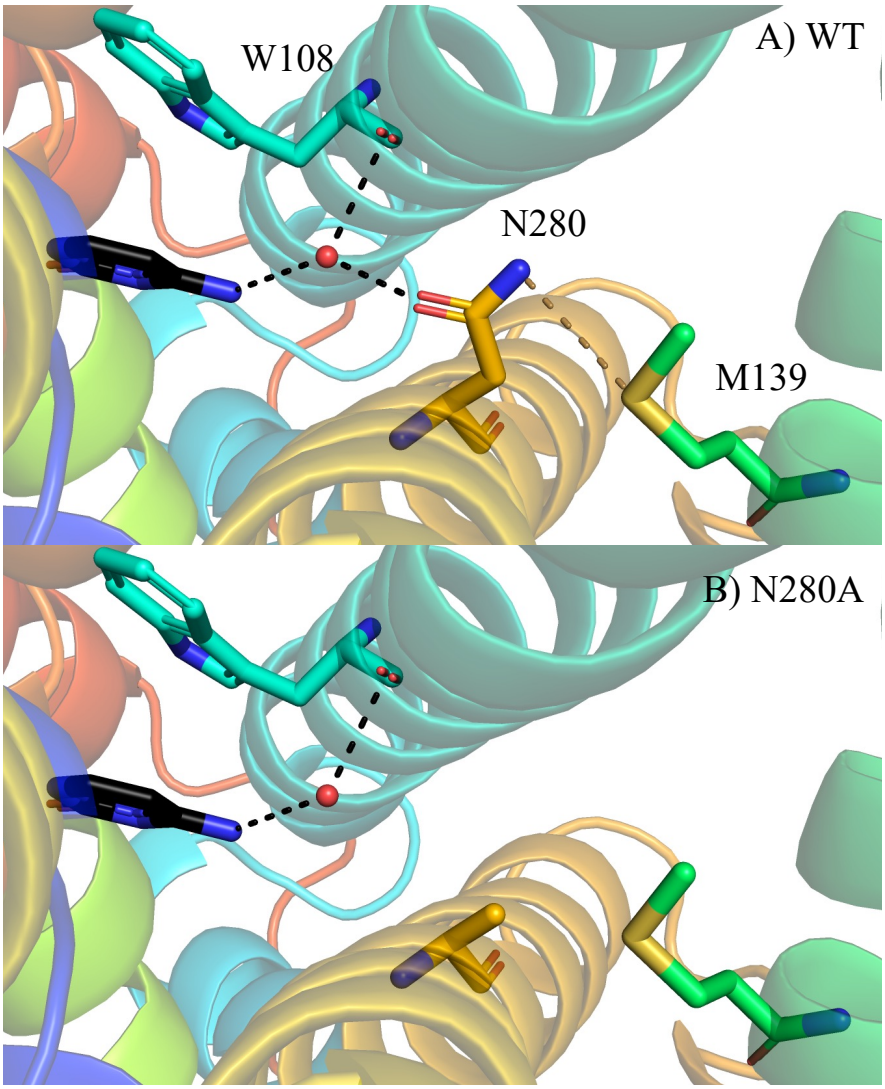
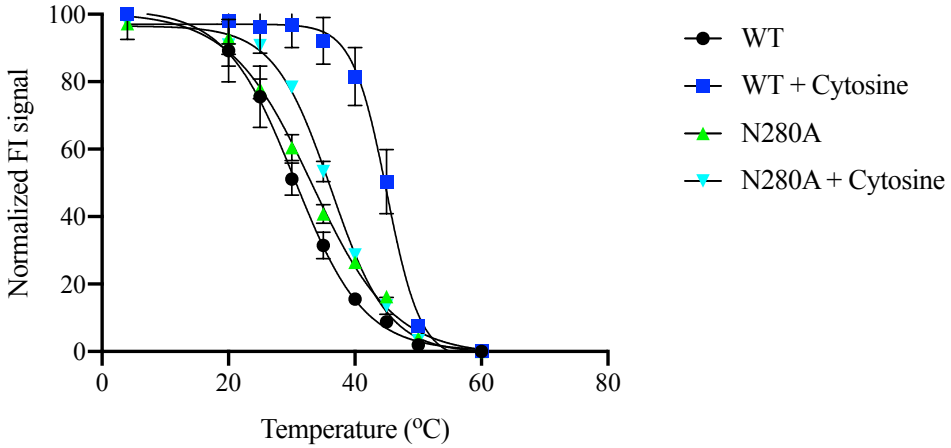


Figure 5.17. Interactions of Asn280 with cytosine and the impact of the Asn280Ala mutation on cytosine binding. Panel A, wt CodB. Panel B, Asn280Ala. Hydrogen bonds are depicted as black dashed lines, with the Asn280-Met139 interaction in yellow dashed lines.

5.2.5.1. GFP-TS to investigate substrate binding of Asn280Ala

GFP-TS was used to investigate if Asn280Ala is stabilised in the presence of 1 mM cytosine and shows that Asn280Ala displays a stabilisation of 3.1 °C compared to 14.4 °C of wt CodB (Figure 5.18). Asn280Ala does not exhibit the same stabilisation as wt CodB suggesting that Asn280 is involved in solute coordination.

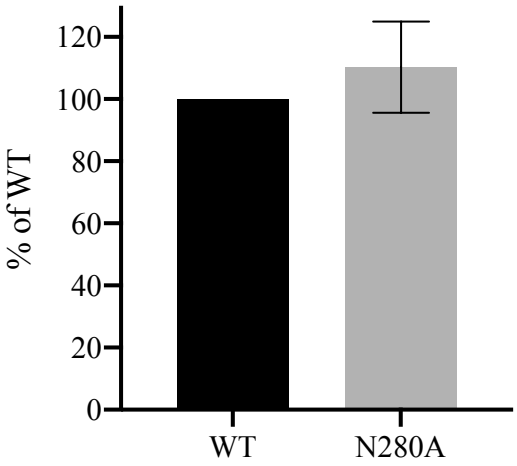


Condition	Melting Temperature, T_m (°C)	T_m Difference (°C)
WT	30.5 ± 1.1	n/a
WT + 1 mM Cytosine	44.9 ± 0.8	14.4
N280A	33.1 ± 1.4	n/a
N280A + 1 mM Cytosine	36.2 ± 0.2	3.1

Figure 5.18. Thermoshift experiments of detergent solubilised CodB.GFP wt and Asn280Ala. CodB.GFP wt was used as a positive control (in the presence of 1 mM cytosine), and as a negative control (absence of cytosine). T_m of Asn280Ala is slightly higher than wt CodB.GFP, and does not exhibit a stabilisation upon the addition of 1 mM cytosine. Points plotted are the average mean of 2 technical repeats with error bars of the S.E.M..

5.2.5.2. Radioactive uptake assay to measure transport of mutant Asn280Ala

The ability of Asn280Ala to transport ^{3}H-5-cytosine was investigated (Figure 5.19). Asn280Ala can uptake ^{3}H-5-cytosine at 110% ± 14.7 compared to wt. From transport data it seems that Asn280 is not required for cytosine transport.



Mutant	% of WT Uptake
Asn280Ala	110.3 ± 14.7

Figure 5.19. Uptake of ^{3}H-5-cytosine by CodB.GFP mutant Asn280Ala, normalised to WT. Uptake of ^{3}H-5-cytosine was measured after 1 minute. WT is normalised to 100%, results are visualised as % of WT with error bars as S.E.M. of at least 4 experiments, each from a different culture.

5.2.6. Mutagenesis of Asn275, Thr278, Thr279 to investigate their role in sodium binding

Previous studies on Mhp1 showed that solute binding is sodium-dependent and that the binding affinity of Mhp1 was severely reduced in the absence of sodium (S. Weyand *et al.*, 2008).

From sequence alignments, Asn275, Thr278, and Thr279 were predicted to be the residues on TM8 involved with sodium coordination (Figure 5.20). Thr278 and Thr279 are interacting via their hydroxyl side chains whilst Asn275 uses main-chain carbonyl to hold sodium in its binding site. Mutation of Thr278 and Thr279 into alanine should reduce the ability of sodium to bind in the Na^{+} site, and probably reduce the ability of cytosine to bind. Asn275 uses its main-chain carbonyl to coordinate sodium and is using its side chain to interact with Ser34 on TM1 (Figure 5.20). CodB is unique in the NCS1 and LeuT family by having a polar residue at this position of TM8, the corresponding residue in Mhp1 is Ala309. An Ala309Asn mutation removed the sodium-dependence of Mhp1; therefore, Mhp1 was able to both bind and transport IMH and BH in the absence of sodium, whilst wt Mhp1 required sodium to bind and transport BH and IMH (Jackson, 2012). In theory, the Asn275Ala mutation in CodB is unlikely to influence sodium coordination, as

Asn275 uses its main-chain carbonyl to interact with sodium, therefore the sodium coordination should be retained. Removing the polar interactions linking TM1 and TM8 could alter CodB's transport activity.

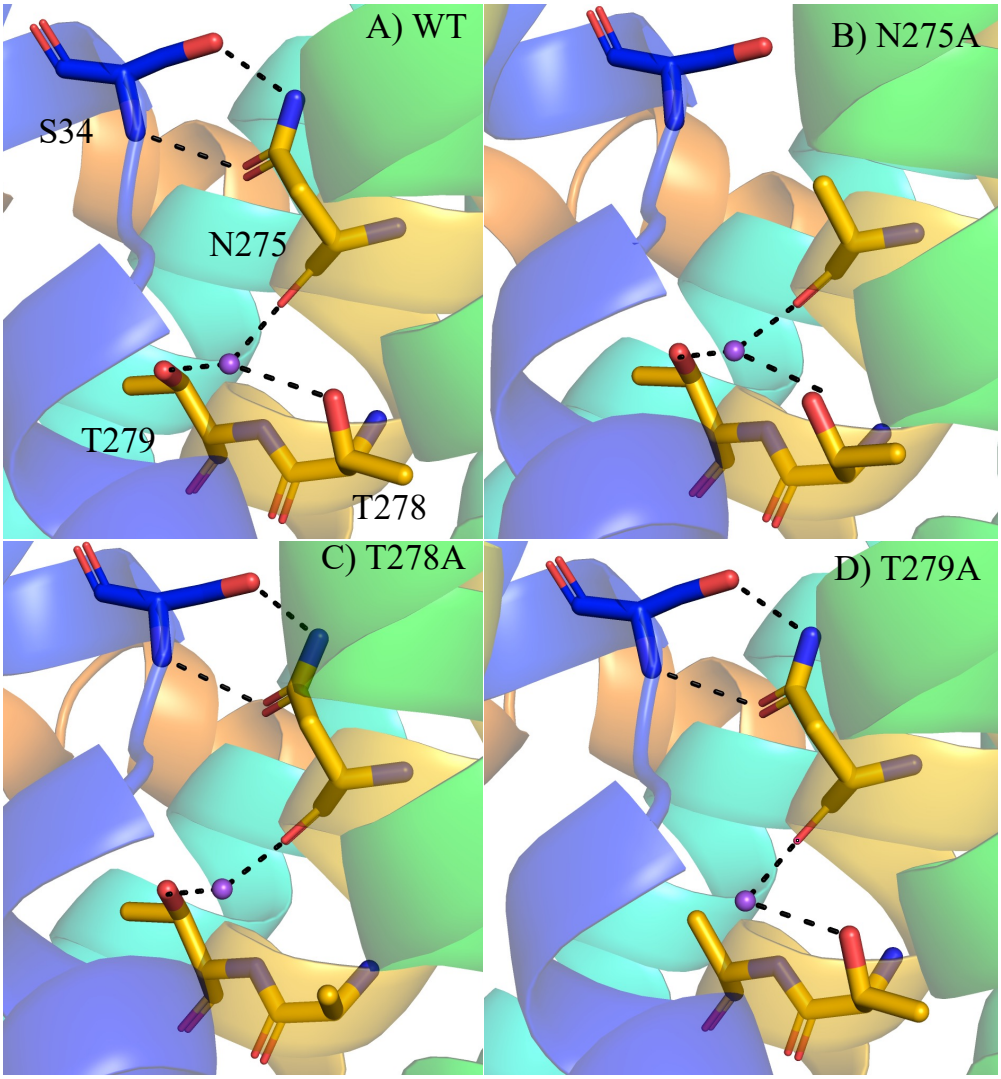
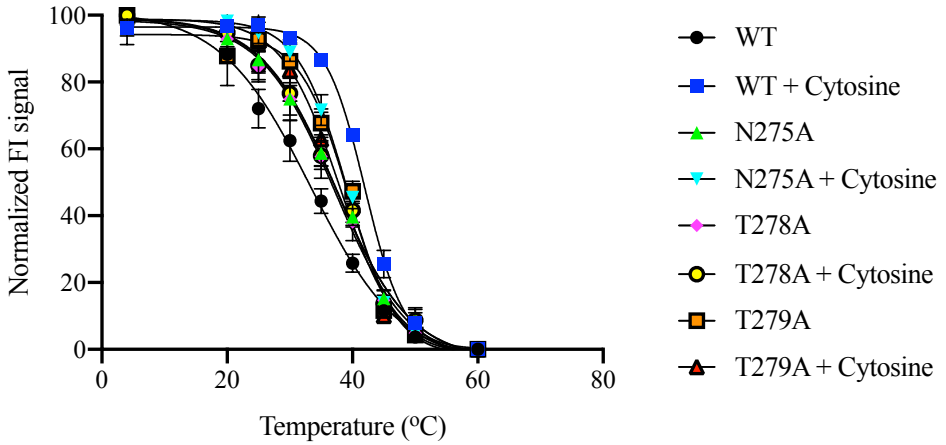


Figure 5.20. Interactions of Asn275, Thr278, Thr279 with Na^+ and TM1, and the impact of Asn275Ala, Thr278Ala, Thr279Ala mutations on sodium coordination. Panel A, wt CodB. Panel B, Asn275Ala. Panel C, Thr278Ala. Panel D, Thr279Ala.

5.2.5.1. GFP-TS to investigate substrate binding

GFP-TS used to investigate if Asn275Ala, Thr278Ala, Thr279Ala mutants were capable of being stabilised upon the addition of cytosine. As in previous experiments, CodB wt in the presence and absence of 1 mM cytosine was used as a positive and negative control. As shown in Figure 5.21 all of these mutants did not exhibit the same stabilisation when compared to wt, suggesting that these mutants

exhibit some loss of ability to binding cytosine, most likely due to a loss of sodium affinity.



Condition	Melting Temperature, T_m (ºC)	T_m Difference (ºC)
WT	33.1 ± 2.4	n/a
WT + 1 mM Cytosine	42.1 ± 0.7	9
N275A	37.2 ± 1.0	n/a
N275A + 1 mM Cytosine	39.1 ± 0.1	1.9
T278A	36.8 ± 2.1	n/a
T278A + 1 mM Cytosine	37.5 ± 1.2	0.7
T279A	39.2 ± 0.9	n/a
T279A + 1 mM Cytosine	38.0 ± 0.3	-1.2

Figure 5.21. Thermoshift experiments of detergent solubilised CodB.GFP wt, Asn275, Thr278Ala, and Thr279Ala. CodB.GFP wt was used as a positive control (in the presence of 1 mM cytosine), and as a negative control (absence of cytosine). T_m of all mutants is slightly higher than wt CodB.GFP, and they do not exhibit a stabilisation upon the addition of 1 mM cytosine. Points plotted are the average mean of 2 technical repeats with error bars of the S.E.M..

Determining a reliable binding affinity for these mutants was difficult; with the same problems as mutants Gln105Ala, Trp108Ala, and Phe204Ala. Consistent with the idea that the cytosine binding ability of these mutants is altered. Therefore, transport assays with these mutants are required to understand the role these residues play in transport and are particularly interesting for Asn275.

5.3. Conclusions and Discussion

The discussion in this chapter will discuss the problems and limitations of the experiments talked about in this chapter. The biological implications of the results here will be discussed in Chapter 6, and how they relate to the structure of CodB and further understanding of transport mechanism. As mentioned before, separating binding events and transport is required to fully understand the alternating access mechanism.

There is an ever-expanding repertoire of techniques to measure protein and ligand interactions. ITC is currently the gold-standard method for monitoring protein and ligand binding events, however, this does require a large amount of purified protein. CodB and cytosine binding was investigated using ITC with CodB demonstrating a response when cytosine was titrated in. However, CodB saturated too quickly and an accurate binding affinity could not be determined in the time that the machine was available, but this could be an avenue for determining binding kinetics for CodB in the future.

MST is an alternative biophysical technique that requires either fluorescence labelled ligand or protein with the benefit of using very small amounts of pure protein in solution (Jerabek-Willemsen *et al.*, 2011). CodB was labelled using GFP or a maleimide dye, neither of these labelling strategies could successfully determine CodB and cytosine interactions. CodB.GFP fusion protein did not produce a good signal-to-noise window using MST, there are plausible reasons for this.

1. There is flexible linker region between CodB and GFP, if CodB does exhibit different thermophoresis due to cytosine binding it may not be transferred to the GFP due to the flexible nature of the linker.
2. The combined MW of CodB.GFP and the DDM micelle is ~140 kDa, the binding of cytosine at 110 Da might not have enough of an impact to alter the thermophoresis of CodB.GFP.

Thermal shift assays have been used monitor protein and ligand binding and can be used for pure protein samples or complex mixtures. CPM is a maleimide dye that increases in fluorescence when bound to thiol groups and can be used with a small

amount of pure protein in a 96-well format. However, due to the only native cysteine of CodB being solvent accessible a CPM assay would never work.

Measuring binding of purified wt CodB has not been achieved in this thesis but using a thermostability assay of detergent solubilised membranes (GFP-TS) to infer binding has been successful as a screening tool of potential ligands and has determined a K_d for wt CodB binding to cytosine (predicted natural ligand) and 5-fluorocytosine providing evidence these are both binding partners.

Monitoring transport is more difficult. LeuT, vSGLT, BetP, and SiaT have been characterised using proteoliposome transport assay's whilst MhsT and Mhp1 have required in-cell transport assays (Yamashita *et al.*, 2005; Faham *et al.*, 2008; Malinauskaite *et al.*, 2014; Perez *et al.*, 2014; Simmons *et al.*, 2014; Wahlgren *et al.*, 2018). Proteoliposome assays and in-cell assays both have their own advantages and limitations. Proteoliposome based assays are ultimately an *in vitro* model and have the benefit that their conditions can be tightly controlled and can be characterised biochemically. However, proteoliposome preparation is non-trivial and there is not a standard method that can be applied to all proteins. Other issues include reconstitution might not mean that the protein is active in the liposome and protein orientation in the proteoliposomes might not be topologically beneficial for the experiment. In-cell based assays have the benefit that protein is orientated correctly in the membrane and its native environment, however, completely controlling the experiment conditions is not possible.

In this thesis, CodB was successfully reconstituted into proteoliposomes, however, these proteoliposomes did not demonstrate an ability up uptake ^{3}H-5-cytosine more than empty liposomes. There are a few explanations for this;

1. CodB might have reconstituted with the majority of protein in the opposite orientation, i.e. TM10 facing into the lumen of the proteoliposome instead of solvent surrounding the proteoliposome.
2. Lipids/ buffer system/ protocol not compatible with CodB activity.
3. As discussed in Chapter 4, CodB has a molecule of DDM bound between TM10 and the 4-helix bundle, the position of this DDM molecule could

inhibit the closure of the extracellular gate and prevent transport from occurring. In theory, reconstitution of CodB into proteoliposomes should remove all detergent but DDM was still bound to CodB in LCP when monoolein should have replaced all detergent, so it's conceivable that DDM was still bound in this site in the proteoliposomes and could be the reason they CodB proteoliposomes did not transport ^{3}H-5-cytosine.

Luckily, in-cell transport assays of wt CodB monitoring ^{3}H-5-cytosine uptake have been successful, proving that cytosine is a solute of CodB. Transport assays were performed with a pH gradient due to the protocol used, however, time course kinetics in the absence of sodium clearly demonstrate that transport is significantly reduced in the absence of sodium and therefore sodium is the most likely co-ion to drive transport. Transport assays at 1 minute with 0.1m mM inhibitors gave unexpected results and contrasted with the ligand screening using GFP-TS. Inhibition of transport by cytosine and 5-fluorocytosine was to be expected as these had been identified as potential ligands using GFP-TS, however, isocytosine and uracil were unexpected, especially as uracil had been investigated previously for CodB from *E. coli*. In previous inhibition assays uracil inhibited CodB from *E. coli* at 15 seconds but did not inhibit uptake of ^{3}H-5-cytosine at 2 minutes (Ma, 2010).

Both the in-cell transport assay and GFP-TS are indirect assays allowing them to be influenced by other variables and have a plethora of other proteins in the samples. In the context of GFP-TS, the potential ligands that were tested could be binding to other proteins in the membrane and therefore sequestered away and not interacting and therefore stabilising CodB. However, at 1 mM the ligands should be at saturating amounts, the native protein expression of the individual proteins should be low enough that this isn't a problem, and finally the ligands investigated are relatively unusual; for example, glucose will be a ligand to many proteins but cytosine and the other-nucleobases investigated probably wouldn't be.

The GFP-TS experiments are indirect assays in nature, as the precipitation of GFP fusion protein is the actual parameter being monitored; therefore, results from this assay should be taken as a recommendation.

The in-cell transport assay has the potential to suffer from the same issues as GFP-TS in terms of ligand sequestering, especially as the concentration of ligands is at 0.1 mM, lower than the GFP-TS experiments. However, the results from the inhibition studies do not suggest that this is a problem as cytosine and uracil both inhibit transport completely. During experiment optimisation 0.5 mM ligand was investigated (Appendix 2) and 5-methylcytosine and thymine were shown to be inhibiting transport to 60-70% of normal, it's more likely that 5-methylcytosine and thymine are not ligands and instead the high concentration of the ligands was slowing transport down due to their concentration as opposed to actually interacting with CodB.

Another problem with the in-cell transport assay is the potential for the membrane of the *E. coli* to be leaky, the addition of lipophilic molecules could intercalate into the membrane and increase the ability of other molecules to cross the partially permeable barrier. As discussed previously, the position of DDM in the CodB structure could be preventing transport, and therefore an assay in the presence of 0.1 mM DDM was run (Figure 5.22), just below the CMC of DDM (the CMC of DDM in 0.2 M NaCl is 0.12 mM) to prevent solubilisation of the membrane but be a high enough concentration to inhibit CodB. The addition of DDM reduced transport to about ~60%, if transport has been reduced completely then it would be likely that DDM was inhibiting transport, but at higher levels of transport it's not clear what is happening, it could be that DDM is not inhibiting transport, or not at a high enough concentration to bind, or inserting into the membrane and allowing the *E. coli* to be more permeable, although background Lemo21(DE3) would also demonstrate an increase in uptake in case, which they didn't.

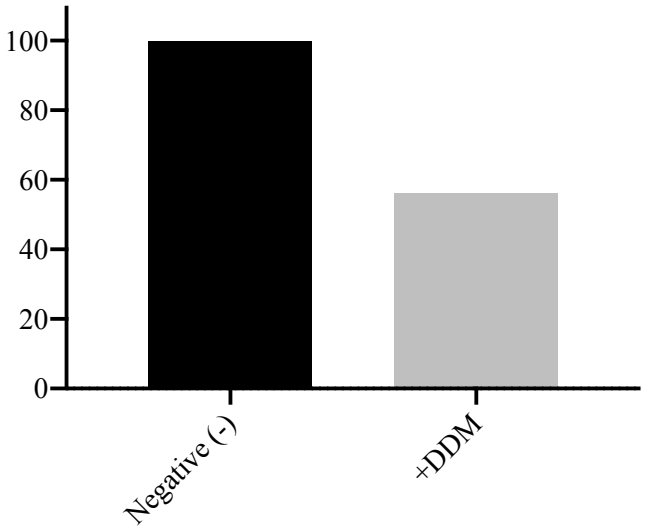


Figure 5.22. Inhibition of ^{3}H-5-cytosine uptake in the presence of 0.1 mM DDM.

Chapter 6-Discussion

6.1. CodB and Other Cytosine Permeases

CodB from *E. coli* was assigned as a putative cytosine permease due to its existence in the same operon as CodA, annotated as a cytosine deaminase (Danielsen *et al.*, 1992), and VPA1242 from *Vibrio parahaemolyticus* has also been characterised as a cytosine permease (Ahmed, 2017). This thesis has structurally and functionally characterised CodB as a sodium-dependent cytosine permease from *P. vulgaris*, a gram-negative opportunistic pathogen. For clarity, unless stated otherwise CodB will refer to CodB from *P. vulgaris*.

The profile of time course experiments of CodB are different to the profiles determined previously for CodB from *E. coli* and VPA1242, for the data presented in this thesis CodB-mediated ^{3}H-5-cytosine uptake saturated at 2 minutes and plateaued, whilst for the *E. coli* and *V. parahaemolyticus* protein uptake peaked at 1 minute and then had fallen by 3 minutes and kept decreasing across the rest of the time-course (Ma, 2010; Ahmed, 2017). However, they are different proteins (78% sequence identity and 73% respectively) with different expression systems; these differences are enough to explain the difference in time-course kinetics, and all experiments have shown that CodB is able to uptake cytosine. The ability of CodB to uptake ^{3}H-5-cytosine reduced to background levels when transport assays were performed in the absence of sodium, this and the structure of CodB suggest that sodium is the most likely co-ion to be driving transport. The ability of BL21(DE3) cells expressing VPA1242 to uptake ^{3}H-5-cytosine did not vary when sodium concentration was reduced suggesting VPA1242 is sodium independent (Ahmed, 2017). The amino acids (Gly29, Phe32, Asn275, Thr278, and Thr279) coordinating sodium in CodB are present in both *E. coli* CodB and VPA1242 (Figure 6.1).

Figure 6.1. MUSCLE sequence alignment of CodB, CodB from *E. coli* and VPA1242.

When investigating the ligand preferences of CodB using GFP-TS, cytosine and 5-fluorocytosine were identified as possible ligands; other nucleobases and cytosine analogues did not demonstrate stabilisation suggesting they were not binding partners. However, during transport inhibition assays, cytosine, 5-fluorocytosine, isocytosine, and uracil did reduce ^{3}H-5-cytosine uptake suggesting they are ligands for CodB, whether these compounds are solutes or inhibitors for CodB will need to be investigated further. 5-fluorocytosine can be modelled into the cytosine binding pocket easily (Figure 6.2), whilst 5-methylcytosine would be sterically hindered by the side chain of Phe33. Uracil can also be modelled into the solute site if the side chain of Gln105 is flipped by 180 ^{o} to optimise hydrogen bonds, thymine can be modelled in the binding pocket in 2 conformations but would be sterically hindering with either side chain of Phe33 or the carbonyl of Ser203 (Figure 6.2). 5-methylcytosine and thymine did not inhibit uptake of ^{3}H-5-cytosine, suggesting they are not ligands capable of binding to CodB; this is consistent with the modelling of the binding pocket and GFP-TS experiments. Previous work on *E. coli* CodB has

shown that the addition of 0.5 mM unlabelled-cytosine inhibition transport, BH also inhibited transport (Ma, 2010), whilst uracil at the 15 second time-point inhibited transport but had recovered to normal levels at the 2 minute time-point. VPA1242 was shown to be inhibited by cytosine, hydantoin, BH, and uracil (Ahmed, 2017). The amino acids thought to be interacting with cytosine (Gln105, Trp108, Phe204, Ser206, Asn280) are conserved in both *E. coli* CodB and VPA1242 (Figure 6.1).

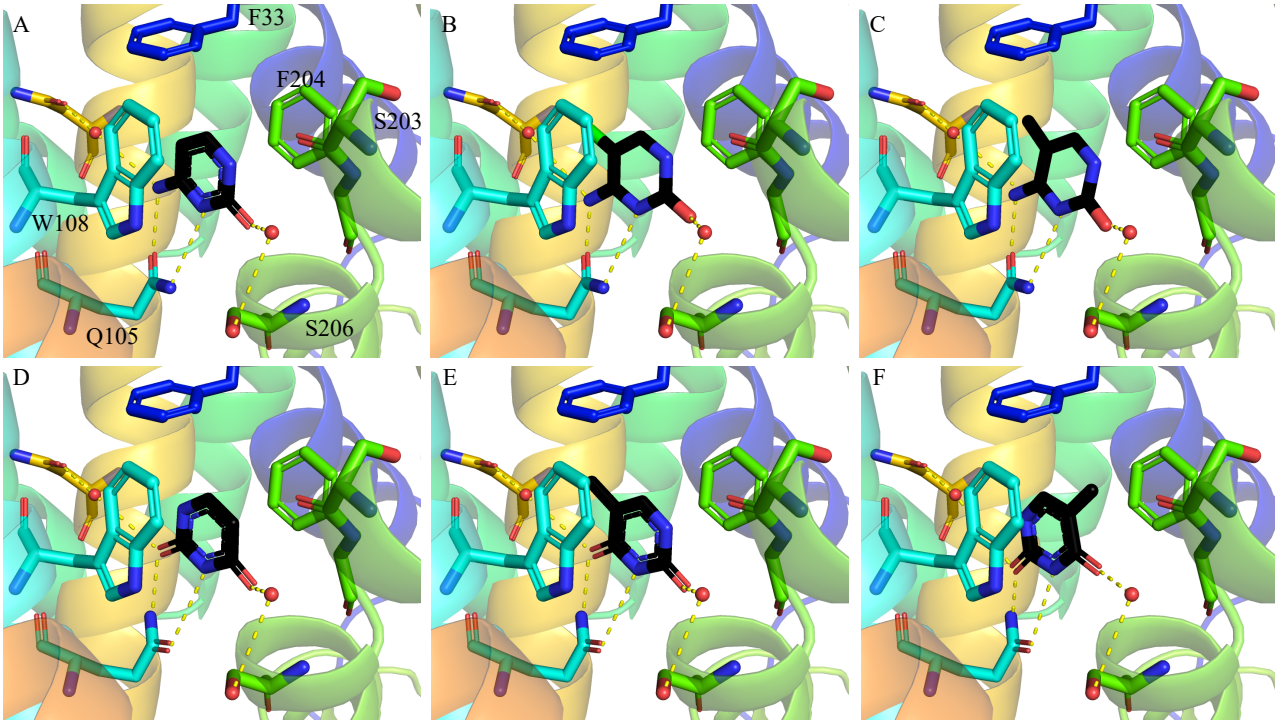


Figure 6.2. Models of cytosine, 5-fluorocytosine, 5-methylcytosine, uracil, and thymine in the CodB solute binding pocket. Panel A, cytosine. Panel B, 5-fluorocytosine can be accommodated in the binding site but due to the hydrophobic pocket formed of Phe33 and the methyl moiety of Thr279 fluorine is not optimal at this position. Panel C, 5-methyl is excluded from binding in CodB as the methyl group would sterically clash with Phe33. Panel D, uracil can also be modelled into the binding site by a side chain flip of Gln105. Panel E and F, thymine can be modelled as uracil was but would be excluded due to steric clashes with Phe33 or Ser203.

For CodB, the substrates that inhibit ^{3}H-5-cytosine uptake all have a 6-membered ring and all have a carbonyl group at the 3 position, with secondary amines at 2 and 4 position, cytosine differs chemically due to the primary amine at the 1 position (Figure 6.3). Uracil was shown to inhibit ^{3}H-5-cytosine uptake, suggesting that uracil binds CodB, the only different between uracil and hydantoin is that hydantoin is a 5-membered ring as opposed to uracil having a 6 membered ring. It's possible that hydantoin would be a ligand for CodB but hydantoins such as allantoin, BH, and IMH will not bind due to their larger volume.

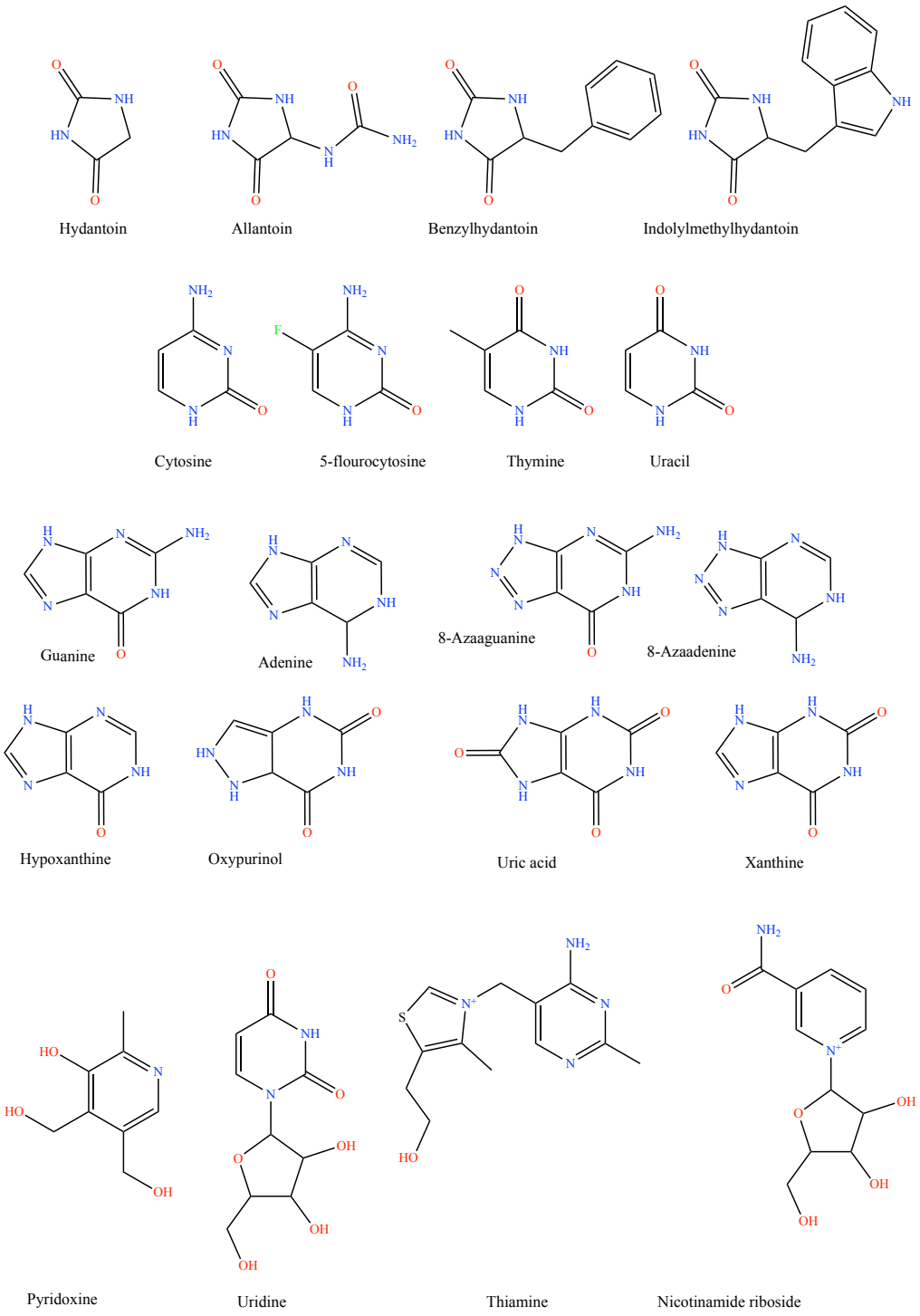


Figure 6.3. Chemical structures of substrates for the NCS1 family and related compounds. These structures were drawn by ChemDraw.

6.2. CodB, Mhp1, and the NCS1 Family

Overall, CodB is very structurally similar to Mhp1 in many aspects as they both exhibit the classic LeuT-fold of 12 TM helices arranged in 2x5TM helices arranged in a pseudo-symmetric repeat with TM1 and the equivalent TM6 being

discontinuous. Both proteins align to each other for their outward-open conformations and have a functionally conserved sodium binding site between TM1 and 8.

CodB and Mhp1 differ the most regarding the position of TM10, in the outward-open structure of Mhp1 (PDB code: 2JLN) this TM is straight whilst in the outward-occluded structure it is bent over the binding pocket. TM10 in CodB is more similar to the outward-open structure, but this TM10 is held even further away from the binding pocket than in the outward-open structure of Mhp1. Closer inspection of the structure suggests that in CodB this TM10 is being held in place by the presence of a molecule of DDM and a monoolein acyl chain. When the TM10 of CodB was modelled onto Mhp1 in both conformations, the side chains would sterically clash with the DDM and monoolein. DDM and monoolein will not be natively interacting with CodB and they are probably crystallisation/purification artefacts; in theory when CodB was placed into the LCP all detergent should have been replaced with monoolein. Based on the observation that DDM could be modelled with relative ease it seems that the DDM was not completely removed and could bind to CodB with a reasonable affinity. The position of DDM between TM10 and the 4-helix bundle could be conformation-locking CodB in an outward-open state, to test this theory the in-cell assay for ^{3}H-5-cytosine uptake was performed in the presence of 0.1 mM DDM, this inhibited transport to 56%, which as discussed in detail in Section 5.3. does not confirm if DDM is conformation locking CodB and preventing transport. For future work it would be interesting to try and purify CodB in a different detergent, ideally with a different chemical structure that would not bind in a similar manner to see if the position of TM10 is a crystallisation artefact.

6.2.1. Solute Coordination

The substrate binding site does differ, both proteins require a π-stacking interaction to hold their solute in place whilst using surrounding residues to orientate the substrate in the pocket. In Mhp1, Trp117 was shown to be important for IMH and BH binding, if replaced by phenylalanine Mhp1 was able to retain most of its transport ability, whilst a Trp117Ala mutation abolished transport. Mutating Trp108 in CodB reduced transport to ~8% compared to wt; Trp108Ala did not show an

increase in T_m in the presence of saturating amounts of cytosine, nor did it produce a reliable K_d. Based on the high conservation of Trp117/Trp108 across the NCS1 family and the reduction in protein activity in both Mhp1 and CodB this residue seems to be essential for activity. In contrast, Trp220 in Mhp1 was not essential for transport despite good conservation at this position across the NCS1 family. Phe204 in CodB uses a π-stacking face-to-face interaction to coordinate substrate, whilst Trp220 uses a face-to-edge π-stacking interaction for substrate binding. Phe204Ala showed the same binding impairment as Trp108Ala using GFP-TS, whilst Phe204Ala retained about 91% of its transport activity. Transport assays of Trp108Ala were far more reproduceable than Phe204Ala. Mhp1 uses more residues in the binding site to coordinate BH and IMH so this might reduce the significance of Trp220, whereas CodB appears to have less interactions with cytosine and might need Phe204 to bind cytosine effectively. Further work to confirm the importance of Phe204 is required to fully understand the role of this residue and why it is conserved across the NCS1 family.

Mhp1 transport is hindered by mutations of Gln42 and Gln121 which form a hydrogen bonding network to coordinate IMH effectively, Gln121 requires Gln42 to orientate the side chain of Gln121 to optimise interactions with solute. CodB does not have comparable interactions at the same structural position and Gln42 and Gln121 are not conserved across the NCS1, instead CodB uses Gln105 which is closer towards to the intracellular end of the helix. The Gln105Ala mutation shows the same results as Trp108Ala and Phe204Ala for GFP-TS assays suggesting that it is an important residue for cytosine binding and coordination, and this was confirmed by transport assays as Gln105Ala mutation reduced ^{3}H-5-cytosine uptake to 25% of wt activity. Gln105 is the only residue directly hydrogen bonding with cytosine and is there is no conservation of any residue across the NCS1 at this position. Gln42 in Mhp1 is Phe33 in CodB; as discussed earlier the large bulky nature of the phenylalanine side chain could be preventing larger substitutions at the 5 position of cytosine.

Ser206 is found close to the cytosine binding site btu does not appear to be interacting directly with the cytosine substrate and instead interacting with a water that is coordinated between cytosine and TM6. The residue Ala222 in Mhp1 has

been previously investigated due its proximity to the solute binding site, which corresponds to Ser206 in CodB, and serine/threonine is highly conserved at this position in the NCS1 family. The Ala222Ser resulted in reduction of BH binding in the presence and absence of sodium, which cannot be explained with wildtype Mhp1 crystal structures (Jackson, 2012). From the crystal structure in this thesis Ser206Ala appears to be interacting with a water molecule that is coordinating cytosine, this water molecule is held in position by interactions with the exposed dipoles of TM6. Mutant Ser206Ala reduced the ability of CodB to coordinate cytosine in GFP-TS assays and reduced the binding affinity of CodB to ~300 μM whilst transport activity of Ser206Ala was retained. From this, it seems that Ser206 increases the binding affinity of CodB but is not essential for transport, which is consistent with the idea that Ser206 is coordinating water in the solute pocket, but the water is highly coordinated by other interactions so the Ser206 interaction is not essential to hold water in place.

Asn280 is also not directly interacting with the cytosine and is using a water molecule as a bridge, this water is also interacting with the carbonyl of Ala107 on TM3. This water molecule in CodB appears to be fulfilling the role of Asn318 in Mhp1. From GFP-TS experiments it seems that Asn280 is involved with cytosine binding as Asn280Ala did not exhibit the same stabilisation when 1 mM cytosine was added; however, transport activity was not impacted by this mutation. From the substrate bound structures of Mhp1 it was clear that Asn318 was important for substrate coordination and knockout mutagenesis confirmed this. Asparagine is the most common residue at this position in the NCS1 but interestingly is not found in CodB and is instead an alanine. In the outward-occluded structure of Mhp1, Asn318 utilises a hydrogen bonding network of Asn314 and Gln153 to orientate the side chain of Asn318 correctly. In CodB, there is a hydrogen bond between Asn280 (corresponding to Asn314), highly conserved in the NCS1, and Met139 (corresponding to Gln153) and a water molecule that is interacting with the cytosine substrate.

The purine-cytosine permease in *S. cerevisae* has Asn374 and Asn377 that have been shown to important for transport as mutagenesis of these residues reduces transport ability of Fcy2 (Ferreira *et al.*, 1997). From sequence alignments Fcy2 has Trp177

and Trp281 that will probably exhibit π-stacking with the cytosine substrate, but there is no residue that would act functionality as Gln105. Asn374 corresponds to Asn280 in CodB, highlighting the importance of this residue in the NCS1 family. In contrast, Asn377 is found in Fcy2 but this is Ala283 in CodB, with no residues in this region interacting with the cytosine substrate. From this, it seems that whilst both CodB and Fcy2 are able to transport cytosine they utilise different amino acids to coordinate substrate binding. Fcy2 exhibits more substrate promiscuity as it can bind and transport adenine, guanine, and hypoxanthine as well cytosine whilst CodB is appears to be more selective.

Differences in the coordination of solute are probably less conserved features of the NCS1 due to the variety of substrates (Ahmed, 2017) (Figure 6.3), whilst the π-stacking sandwich between TM3 and TM6 appears to be more conserved and could be a common feature for the NCS1 family.

6.2.2. Sodium Coordination

All of the LeuT superfamily sodium-linked symporters have the conserved Na2 site that consists of a square pyramidal arrangement with sodium coordinated using the carbonyls of 2 residues in the unwound section of TM1, a carbonyl in TM8 and 2 hydroxyl side chains of either serine or threonine in TM8. The NCS1 family has conserved small side chains at the position of Ala39 and Gly29 in Mhp1 and CodB respectively and an Ala38Gly mutation in Mhp1 reduced transport activity of Mhp1 by 20% (Jackson, 2012). Maybe the size of the side chain in this position is optimised for the best packing of the protein core and doesn't have much effect on sodium binding as the interaction with sodium is because of the main chain interactions. The Ile41/Phe32 position has conserved hydrophobic residues across the NCS1, with PucI the exception with a histidine residue, and phenylalanine being conserved in a few of the yeast transporters. An Ile41Phe mutation for Mhp1was generated and was predicted sterically clash with Phe267/Phe305 preventing TM1 and TM8 to pack, this mutation reduced the ability of BH to bind in the presence and absence of sodium whilst having no effect on the transport ability of IMH. From this, it seems that the ability of Mhp1 to bind sodium has been disrupted as a result of the Ile41Phe (Jackson, 2012).

In CodB the carbonyl of Asn275 (corresponding to Ala309 in Mhp1) is in position to also be coordinating sodium, with the side chain forming hydrogen bonds with Ser34, which corresponds to Val43 in Mhp1. Asn275Ala demonstrated that this mutation reduced the ability of CodB to bind cytosine compared to wt CodB. Ala309 in Mhp1 is an interesting residue and coordinates sodium using its main chain carbonyl, alanine is well conserved in this position across the NCS1 family, with some of the yeast transporters having a serine residue at this position, whereas CodB has an asparagine. Ala309Asn Mhp1 exhibited high-affinity BH binding in the presence and absence of sodium and uptake of IMH was comparable to wt Mhp1 in the presence of sodium, suggesting that Mhp1 is no longer sodium-dependent.

Ser312 in Mhp1 is involved in sodium coordination via its hydroxyl side chain, with a Ser312Ala mutation reducing the binding of Mhp1 for substrate in a sodium dependent manner, whilst transport activity assays showed that this mutation could transport IMH at double at the rate of wt Mhp1 (Jackson, 2012). Thr278 is the corresponding residue in CodB and is interacting with sodium via its hydroxyl side chain. Thr278Ala behaved similarly to Asn275Ala suggesting that its ability to bind and be stabilised by cytosine is impaired. Thr313 in Mhp1/ Thr279 in CodB is the final residue implicated directly in sodium binding and uses its side-chain hydroxyl to interact directly with sodium. The mutation Thr313Ile to remove threonine functionality showed impairment of Mhp1 to binding BH in a sodium dependent manner, but had no impact on the rate of IMH uptake (Jackson, 2012). A limitation of the transport assay used for Mhp1 is the high concentrations of sodium used, which could counteract a reduction in sodium affinity. Thr279Ala also demonstrated cytosine impairment in GFP-TS assays, consistent with a loss in ability to coordinate cytosine. The position of the Thr279 side chains is shown to be coordinating the sodium ion using its hydroxyl whilst the methyl group is packing against the 5 position of cytosine, this Thr279 appears to be simultaneously coordinating sodium and selecting for cytosine and implicating itself as the residue linking the solute and cation sites. A Thr279Ser mutation would be interesting to investigate this residue, as serine would retain functionality of this residue for sodium coordination but might remove the hydrophobic packing against cytosine and potentially lose the link between the solute and sodium site. When an inward-open model of CodB was built,

the rotation of TM8 away from TM1 increased the volume of the sodium binding site by moving Asn275, Thr278, Thr279 away and would reduce the ability of CodB to coordinate sodium in this site. The movement of TM8 also means that the interaction holding Asn275 to Ser34 was impaired as well. The mutants Asn275Ala, Thr278Ala, Thr279Ala need to be investigated by transport assays to fully understand the role they play in sodium coordination and how this drives transport.

6.2.3. Interhelix Interactions

CodB differs from Mhp1 by having multiple residues on TM8 that are interacting with the other helices, this could be a way of coupling interactions between TM8 and sodium to the rest of CodB. Asn275 as discussed earlier is interacting with sodium via its main chain carbonyl but is also interacting with TM1 using its side chain. Mhp1 and other LeuT family members all have hydrophobic residues at this position and are not interacting with surrounding helices.

As discussed before, Asn280 is on TM8 just above Thr278 and Thr279 which are residues implicated in sodium binding and is found interacting with Met139 on TM4 at the position of TM4 that moves when the intracellular gate opens. In Mhp1, this interaction between TM4 and TM8 is disrupted in the inward-open state (Shimamura *et al.*, 2010) (PDB code:2X79) due to the opening of the intracellular gate. The inward-open model for CodB moves Asn280 away from the cytosine binding site so it is no longer in position to interact with water to coordinate cytosine, whilst the interaction with Met139 remains intact. This hydrogen bonding network could be more important than originally thought, TM4 is connected to TM5, the helix that acts as the intracellular thin gate in Mhp1; a change in TM4 could easily be transferred into TM5. Could this interaction between TM4 and TM8 be important for transport, and a way of coupling the sodium binding site to the intracellular gate? Sodium binding is thought to alter the position of TM8, sodium binding might bring TM8 into position to interact with TM4, movement of TM4 could be amplified into movement of TM5 to close the intracellular gate and how sodium binding appears to stabilise the outward-facing conformation. Both Mhp1 and CodB use the additional interactions of Asn318 and water respectively to potentially couple sodium binding, solute binding, and closure of the intracellular gate. Although it's more likely this

interaction would be a way of stabilisation the proteins and keeping protein fold intact. One caveat of this theory is that the interactions of methionine are weak, and that maybe this interaction wouldn't be strong enough to hold the helix in place. Asp281 in CodB can be seen interacting with Thr142 on TM4, at the position of TM4 that moves when the intracellular gate opens. Small hydrophobic side chains are generally conserved across the NCS1 at the Asp281 position. CodB and Mhp1 are the only members of the NCS1 to have Thr142/Thr156 at this position. The hydrogen bonding network between Asp281 and Thr142 is disrupted due to the bending of TM5 in the inward-open model that was built.

Ala317 in Mhp1 is next to Asn318 which was shown to be important in IMH and BH binding. Asparagine is conserved in this position across the NCS1 family, an Ala317Asn mutation was predicted to sterically hinder with Met39 in TM1 and change the packing of TM1 and TM8. The Ala317Asn mutation in Mhp1 reduced sodium-dependent BH binding as much as Ser312Ala and Thr313Ser, suggesting that this residue is important for sodium binding. Alignment of TM8 showed that either Asn282 or Ala283 was the corresponding residue in CodB, based on the packing of Asn282 onto TM1, where Ala283 doesn't, it seems more likely that Asn282 corresponding amino acid. In CodB, the side chain of Asn282 is in position to form interactions with the carbonyl of Val26 (Ile35 in Mhp1), the amide of Val26 is also coordinating with the side chain of Ser154 on TM5 whilst there is no hydrogen boding network in Mhp1. Investigation of Asn282 would be of interest. Ser154 is found at the intracellular end of TM5 which is the part of the helix that will bend. Could this hydrogen bonding network be way of coupling TM8 to TM5 directly? Changes in the position of TM8 are thought to be influenced by sodium binding, potentially sodium could position TM8, hydrogen bonds hold the position of TM1, 5, and 8 relative to each other and closes the intracellular thin gate in response to sodium binding. The inward-open model shows that this hydrogen bonding network is severely disturbed when CodB moves from an outward-facing conformation to an inward-facing conformation. Investigation of these hydrogen bonding networks and the role they play in transport would be interesting.

6.3. CodB and the LeuT-fold

Despite low sequence identity, CodB has the same 3D structure as other members of the LeuT-fold and all sodium symporters have the conserved sodium binding site termed Na2, found at the interface between the hash-motif and the 4-helix bundle. A structural based approach provides no evidence that CodB is able to coordinate sodium in the alternative sites due to the lack of observable density at above saturating concentrations of Na^+ and none of the other sodium sites having residues positioned to be capable of coordinating sodium. The Na1' site in BetP have been proposed as the secondary binding site for Na^+, the authors have discussed that in a number of homologs that a positive charge can be found in this site even for non-sodium binding transporters (Khafizov *et al.*, 2012). Mhp1 has Lys110 in this site, whilst CodB has only hydrophobic residues in this region. Of note, Lys110 interacts with Ser114 which corresponds to Gln105 in CodB, the importance of Lys110 in Mhp1 and its interaction with Ser114 is unclear, especially when CodB does not have a positive charge in the Na1' and Gln105 is involved with substrate binding. Lys110 is also interacting with Asp229 in Mhp1, this Asp229 is the most conserved residue in the NCS1 family and is found in every single characterised family member (Asp213 in CodB) (Appendix 2).

The polar residues found on TM8 interacting with surrounding helices are the most unique feature of CodB. Moving towards the c-terminal from the n-terminal end of the helix, Asn275 as it is interacting with TM1 using its side chain whilst coordinating with its main chain carbonyl; CodB is the only member of the LeuT family that has this unique interaction. The Asn280 interaction to hold TM4 and TM8 together is not preserved in any of the other homologs apart from in LeuT through Ser356 and Asn179; however, this interaction is still maintained in the inward-facing structures in contrast to Mhp1 where it is disrupted (Weyand *et al.*, 2008; Shimamura *et al.*, 2010; Krishnamurthy and Gouaux, 2012). TM5 does not change position in LeuT, unlike Mhp1 where it is the intracellular gate. None of the other LeuT family members have a polar residue at the Asp281to make interactions with TM4, apart from SiaT has serine residue which uses the side chain hydroxyl to coordinate sodium in the Na3 site (Wahlgren *et al.*, 2018). Moving finally to the Asn282 position, MhsT has a serine residue that is interacting with another serine

residue on TM1 in the inward-occluded state, and SiaT has a serine that is coordinating sodium in the Na3 site (Malinauskaite *et al.*, 2014; Wahlgren *et al.*, 2018).

The $GlyX_9Pro$ motif on the intracellular end of TM5 found in LeuT, MhsT, dDAT, SERT, and Mhp1 can also be found in CodB. In MhsT, this helix breaking motif is thought to enable unwinding of TM5 allow hydration of the Na2 site, resulting in Na^+ escaping into the cytoplasm (Malinauskaite *et al.*, 2014). However, more interestingly, this motif can be found on the structurally-symmetric TM10 in CodB. MhsT and Mhp1 have a $GlyX_{16}Pro$ and $GlyX_{14}Pro$ respectively, in the context of the gene duplication theory; then sequence divergence, potentially MhsT and Mhp1 added additional amino acids into the motif over the course of evolution. Investigation of the significance of this motif on TM5 and the role it plays in transport, why it is conserved across the NSS and NCS1 families despite variations in the mechanisms of transport would be of interest.

Structural investigation of CodB especially regarding residues found in TM8 has opened up many questions. CodB is unique amongst the NCS1 and the LeuT superfamily by having Asn275 in TM8 coordinating sodium via its carbonyl group and then interacting with TM1 using its side chain. Investigation of this residue would be of interest. Is this Asn275 interaction simply a way of stabilising TM1 and TM8 and allowing them to be held together more stably? But why would CodB need this interaction and not other members of the superfamily? CodB uses less residues for solute coordination than many of its structural homologs, many members of the LeuT superfamily use amino acids found on TM8 to help bind solute, CodB does not directly and instead coordinates through a water bridge, and BetP does use any residues on TM8 to interact with substrate (Ressl *et al.*, 2009). Asn280 is highly conserved within the NCS1 family and has had no obvious role. In Mhp1, this asparagine residue is found a hydrogen bonding network to connect TM4 and TM8. This interaction is preserved in CodB but is a weaker interaction due to methionine. Finally, Asn282 is forming hydrogen bonding network across TM1, 5, and 8 which cannot be seen in other homologs. The weaker interaction in CodB of Asn280 and Met139 compared to Asn314 and Gln153 in Mhp1 might have required an additional hydrogen bonding network from Asn282, Val26, and Ser154 to couple sodium

binding to the intracellular gate as seen in CodB. Understanding the role of all 3 of these asparagine residues will hopefully give insight into the role of sodium coordination and the opening of the intracellular gate.

6.4. Concluding Remarks

The work from this thesis has enabled the structural determination of CodB, a member of the NCS1 family that has not been structurally characterised. This structure has enabled identification of the substrate and cation binding site and confirmed the chemical and ion specificity of CodB. Detailed structural comparison with Mhp1 and other members of the LeuT-fold begins to explain the differences in ligand specificity and have allowed insight into conserved features of the NCS1 versus protein specific interactions. Complimentary functional studies have confirmed the sodium-dependence of CodB and the significance of some substrate binding residues and their role in transport. Further structures, binding assays, and transport experiments are required to fully understand transport mediated by CodB.

Bibliography

Abramson, J. *et al.* (2003) 'Structure and Mechanism of the Lactose Permease of Escherichia coli', *Science*, 301(5633), pp. 610–615.

Abramson, J. and Wright, E. M. (2009) 'Structure and function of Na+-symporters with inverted repeats', *Current Opinion in Structural Biology*, 19(4), pp. 425–432.

Adams, P. D. *et al.* (2010) 'PHENIX: A comprehensive Python-based system for macromolecular structure solution', *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp. 213–221.

Afonine, P. V. *et al.* (2012) 'Towards automated crystallographic structure refinement with phenix.refine', *Acta Crystallographica Section D: Biological Crystallography*, 68(4), pp. 352–367.

Ahmed, I. (2017) *Characterisation of membrane transport proteins of NCS1 and PACE families using biochemical and biophysical techniques*. PhD thesis, University of Leeds.

Alexandrov, A., Dutta, K. and Pascal, S. M. (2001) 'MBP fusion protein with a viral protease cleavage site: One-step cleavage/purification of insoluble proteins', *BioTechniques*, 30(6), pp. 1194–1198.

Alexandrov, A. I. *et al.* (2008) 'Microscale Fluorescent Thermal Stability Assay for Membrane Proteins', *Structure*, 16(3), pp. 351–359.

Anton, B. P. and Raleigh, E. A. (2016) 'Complete Genome Sequence of NEB 5-alpha , a Derivative of', *Genome Announcements*, 4(6), pp. 6–7.

Backmark, A. E. *et al.* (2013) 'Fluorescent probe for high-throughput screening of membrane protein expression', *Protein Science*, 22, pp. 1124–1132.

Bill, R. M. *et al.* (2011) 'Overcoming barriers to membrane protein structure determination', *Nature Biotechnology*, 29(4), pp. 335–340.

Boudker, O. *et al.* (2007) 'Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter', *Nature*, 445(7126), pp. 387–393.

Boudker, O. and Verdon, G. (2010) 'Structural perspectives on secondary active transporters', *Trends in pharmacological sciences*, 31(9), pp. 418–426.

Caffrey, M. (2009) 'Crystallizing membrane proteins for structure-function studies using lipidic mesophases', *Nature Protocol*, 4(5), pp. 709–731.

Calabrese, A. N. *et al.* (2017) 'Topological Dissection of the Membrane Transport Protein Mhp1 Derived from Cysteine Accessibility and Mass Spectrometry', *Analytical Chemistry*, 89(17), pp. 8844–8852.

Carpenter, E. P. *et al.* (2008) 'Overcoming the challenges of membrane protein crystallography', *Current Opinion in Structural Biology*, 18(5), pp. 581–586.

Cherezov, V., Abola, E. and Stevens, R. C. (2010) 'Toward drug design: recent progress in the structure determination of GPCRs, a membrane protein family with high potential as pharmaceutical targets', *Methods in Molecular Biology*, 654(1), pp. 141–168.

Coleman, J. A. *et al.* (2019) 'Serotonin transporter–ibogaine complexes illuminate mechanisms of inhibition and transport', *Nature*, 569, pp. 141–145.

Coleman, J. A., Green, E. M. and Gouaux, E. (2016) 'X-ray structures and mechanism of the human serotonin transporter', *Nature*, 532(7599), pp. 334–339.

Danielsen, S. *et al.* (1992) 'Characterization of the Escherichia coli codBA operon encoding cytosine permease and cytosine deaminase.', *Molecular microbiology*, 6(10), pp. 1335–1344.

Deng, D. *et al.* (2014) 'Crystal structure of the human glucose transporter GLUT1', *Nature*, 510(7503), pp. 121–125.

Dimaio, F. *et al.* (2013) 'Improved low-resolution crystallographic refinement with Phenix and Rosetta', *Nature Methods*, 10(11), pp. 1102–1106.

Drew, D. *et al.* (2008) 'GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae', *Nature Protocols*, 3(5), pp. 784–798.

Drew, D. and Boudker, O. (2016) 'Shared Molecular Mechanisms of Membrane Transporters', *Annual review of biochemistry*, 85, pp. 543–572.

Elbourne, L. D. H. *et al.* (2017) 'TransportDB 2.0: A database for exploring membrane transporters in sequenced genomes from all domains of life', *Nucleic Acids Research*, 45(D1), pp. D320–D324.

Emsley, P. *et al.* (2010) 'Features and development of Coot.', *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 4), pp. 486–501.

Evans, P. R. and Murshudov, G. N. (2013) 'How good are my data and what is the resolution?', *Acta Crystallographica Section D: Biological Crystallography*, 69(7), pp. 1204–1214.

Faham, S. *et al.* (2008) 'The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+ /sugar symport', *Science*, 23(1), pp. 810–814.

Fang, Y. *et al.* (2009) 'Structure of a prokaryotic virtual proton pump at 3.2 Å resolution', *Nature*, 460(7258), pp. 1040–1043.

Ferreira, T. *et al.* (1997) 'Functional analysis of mutated purine-cytosine permease from Saccharomyces cerevisiae. A possible role of the hydrophilic segment 371-377 in the active carrier conformation', *Journal of Biological Chemistry*, 272(15), pp. 9697–9702.

Forrest, L. R. *et al.* (2008) 'Mechanism for alternating access in neurotransmitter transporters', *Proceedings of the National Academy of Sciences of the United States of America*, 105(30), pp. 10338–10343.

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

Greenfield, N. J. (2007) 'Using circular dichroism spectra to estimate protein secondary structure', *Nature Protocols*, 1(6), pp. 2876–2890.

Hediger, M. A. *et al.* (2004) 'The ABCs of solute carriers: Physiological, pathological and therapeutic implications of human membrane transport proteins', *Pflugers Archiv European Journal of Physiology*, 447(5), pp. 465–468.

Hu, N.-J. *et al.* (2011) 'Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT', *Nature*, 478(7369), pp. 408–411.

Huang, X. and Miller, W. (1991) 'A time-efficient, linear-space local similarity algorithm', *Advances in Applied Mathematics*, 12(3), pp. 337–357.

Hughes, G. W. *et al.* (2019) 'Evidence for phospholipid export from the bacterial inner membrane by the Mla ABC transport system', *Nature Microbiology*, 4(10), pp. 1692–1705.

Hunte, C. *et al.* (2005) 'Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH', *Nature*, 435(7046), pp. 1197–1202.

Jackson, S. M. (2012) *Elucidating the molecular mechanisms of ligand binding and transport by the Na + -hydantoin transport protein , Mhp1*. PhD thesis, University of Leeds.

Jardetzky, O. (1966) 'Simple Allosteric Model for Membrane Pumps', *Nature*, 211(5052), pp. 969–970. doi: 10.1038/211969a0.

Jerabek-Willemsen, M. *et al.* (2011) 'Molecular Interaction Studies Using Microscale Thermophoresis', *ASSAY and Drug Development Technologies*, 9(4), pp. 342–353.

Jung, H. *et al.* (1998) 'Unidirectional reconstitution and characterization of purified Na+/proline transporter of Escherichia coli', *Biochemistry*, 37(31), pp. 11083–11088.

Jung, H., Hilger, D. and Raba, M. (2012) 'The Na+/L-proline transporter PutP', *Frontiers in Bioscience*, 17(2), pp. 745–759.

Jungnickel, K. E. J., Parker, J. L. and Newstead, S. (2018) 'Structural basis for amino acid transport by the CAT family of SLC7 transporters', *Nature Communications*, 9(1), pp. 1–12.

Kang, H. J., Lee, C. and Drew, D. (2013) 'Breaking the barriers in membrane protein crystallography', *International Journal of Biochemistry and Cell Biology*, 45(3), pp. 636–644.

Kappes, R. M., Kempf, B. and Bremer, E. (1996) 'Three transport systems for the osmoprotectant glycine betaine operate in Bacillus subtilis: Characterization of OpuD', *Journal of Bacteriology*, 178(17), pp. 5071–5079.

Kawate, T. and Gouaux, E. (2006) 'Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins', *Structure*, 14(4), pp. 673–681.

Khafizov, K. *et al.* (2012) 'Investigation of the sodium-binding sites in the sodium-coupled betaine transporter BetP', *Proceedings of the National Academy of Sciences of the United States of America*, 109(44), pp. 3035–3044.

Kleber-Janke, T. and Becker, W. M. (2000) 'Use of modified BL21(DE3) Escherichia coli cells for high-level expression of recombinant peanut allergens affected by poor codon usage', *Protein Expression and Purification*, 19(3), pp. 419–424.

Krishnamurthy, H. and Gouaux, E. (2012) 'X-ray structures of LeuT in substrate-free outward-open and apo inward-open states', *Nature*, 481(7382), pp. 469–474.

Krogh, A. *et al.* (2001) 'Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes', *Journal of Molecular Biology*, 305(3), pp. 567–580.

Lee, C. *et al.* (2013) 'A two-domain elevator mechanism for sodium/proton antiport', *Nature*, 501(7468), pp. 573–577.

Lee, C. *et al.* (2014) 'MemStar: A one-shot *Escherichia coli* -based approach for high-level bacterial membrane protein production', *FEBS Letters*, 588(20), pp. 3761–3769.

Lu, F. *et al.* (2011) 'Structure and mechanism of the uracil transporter UraA', *Nature*, 472(7342), pp. 243–247.

Ma, D. *et al.* (2012) 'Structure and mechanism of a glutamate-GABA antiporter', *Nature*, 483(7391), pp. 632–636.

Ma, P. (2010) *Structure-activity relationships of membrane proteins: the ncs1 family of transporters and sensor kinases of two-component systems*. PhD thesis, University of Leeds.

Ma, P. *et al.* (2016) 'Allantoin transport protein, Pucl, from Bacillus subtilis: Evolutionary relationships, amplified expression, activity and specificity', *Microbiology*, 162(5), pp. 823–836.

Madeira, F. *et al.* (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic Acids Research*, 47(W1), pp. W636–W641.

Malinauskaite, L. *et al.* (2014) 'A mechanism for intracellular release of Na+ by neurotransmitter/sodium symporters', *Nature Structural and Molecular Biology*, 21(11), pp. 1006–1012. d

Masson, J. *et al.* (1999) 'Neurotransmitter Transporters in the Central Nervous System', *Pharmacological Reviews*, 51(3), pp. 439 – 464.

McNicholas, S. *et al.* (2011) 'Presenting your structures: The CCP4mg molecular-graphics software', *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 386–394.

Nji, E. *et al.* (2018) 'An engineered thermal-shift screen reveals specific lipid preferences of eukaryotic and prokaryotic membrane proteins', *Nature Communications*, 9(4253), pp. 1–12.

Pao, S. S., Paulsen, I. T. and Saier, M. H. (1998) 'Major facilitator superfamily.', *Microbiology and molecular biology reviews : MMBR*, 62(1), pp. 1–34.

Pebay-Peyroula, E. *et al.* (1997) 'X-ray Structure of Bacteriorhodopsin at 2.5 Angstroms from Microcrystals Grown in Lipidic Cubic Phases', *Science*, 277(5332), pp. 1676–1681.

Pédelacq, J. D. *et al.* (2006) 'Engineering and characterization of a superfolder green fluorescent protein', *Nature Biotechnology*, 24(1), pp. 79–88. doi: 10.1038/nbt1172.

Penmatsa, A., Wang, K. H. and Gouaux, E. (2013) 'X-ray structure of the dopamine transporter in complex with tricyclic antidepressant', *Nature*, pp. 85–90.

Perez, C. *et al.* (2012) 'Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP', *Nature*, 490(7418), pp. 126–130.

Perez, C. *et al.* (2014) 'Substrate-bound outward-open state of the betaine transporter BetP provides insights into Na+ coupling', *Nature Communications*, 5(4231), pp. 1–11.

Quick, M. *et al.* (2009) 'Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation', *Proceedings of the National Academy of Sciences of the United States of America*, 106(14), pp. 5563–5568.

Quick, M. *et al.* (2018) 'The LeuT-fold neurotransmitter:Sodium symporter MhsT has two substrate sites', *Proceedings of the National Academy of Sciences of the United States of America*, 115(34), pp. E7924–E7931.

Quistgaard, E. M. *et al.* (2016) 'Major facilitator superfamily (MFS) transporters move a variety of small compounds across biological membranes', *Nature reviews. Molecular cell biology*, 17(134), pp. 123–132.

Rajarathnam, K. and Rösgen, J. (2014) 'Isothermal titration calorimetry of membrane proteins - Progress and challenges', *Biochimica et Biophysica Acta - Biomembranes*, 1838(1 PARTA), pp. 69–77.

Ressl, S. *et al.* (2009) 'Molecular basis of transport and regulation in the Na+/betaine symporter BetP', *Nature*, 458(7234), pp. 47–52.

Riehm, R. A. *et al.* (2005) 'A fluorescence-detection size-exclusion chromatography-based thermostability assay to identify membrane protein expression and crystallization conditions', *Structure*, 20(18), pp. 1293–1299.

Rodríguez, D. D. *et al.* (2009) 'Crystallographic ab initio protein structure solution below atomic resolution', *Nature Methods*, 6(9), pp. 651–653.

Saier, M. H. *et al.* (2015) 'The Transporter Classification Database (TCDB): Recent advances', *Nucleic Acids Research*, 44(D1), pp. D372–D379.

Schlegel, S. *et al.* (2012) 'Optimizing membrane protein overexpression in the Escherichia coli strain Lemo21(DE3)', *Journal of Molecular Biology*, 423(4), pp. 648–659. d

Schulze, S. *et al.* (2010) 'Structural basis of Na+-independent and cooperative substrate/product antiport in CaiT', *Nature*, 467(7312), pp. 233–236.

Sekiguchi, Y. (2014) 'Structural and functional studies of the Apical Sodium Dependent Bile Acid Transporter', pp. 1–182.
Shaffer, P. . *et al.* (2009) 'Structure and mechanism of a Na+independent amino acid transporter', *Science*, 325(5943), pp. 1010–1014.

Shi, Y. (2013) 'Common Folds and Transport Mechanisms of Secondary Active Transporters', *Annual Review of Biophysics*, 42(1), pp. 51–72.

Shimamura, T. *et al.* (2010) 'Molecular Basis of Alternating Access Membrane Transport by the Sodium-Hydantoin Transporter, Mhp1', *Science*, 328(5977), pp. 470–473.

Simmons, K. J. *et al.* (2014) 'Molecular mechanism of ligand recognition by membrane transport protein, Mhp1.', *The EMBO journal*, 33(16), pp. 1831–44.

Simpkin, A. J. *et al.* (2018) 'SIMBAD: A sequence-independent molecularreplacement pipeline', *Acta Crystallographica Section D: Structural Biology*, 74(7), pp. 595–605.

Singer, S. J. and Nicolson, G. L. (1972) 'The fluid mosaic model of the structure of cell membranes.', *Science (New York, N.Y.)*, 175(4023), pp. 720–31.

Singh, S. K. *et al.* (2008) 'A competitive inhibitor traps LeuT in an open-to-out conformation', *Science*, 322(5908), pp. 1655–1661.

Singh, S. K., Yamashita, A. and Gouaux, E. (2007) 'Antidepressant binding site in a bacterial homologue of neurotransmitter transporters', *Nature*, 448(7156), pp. 952–956.

Slotboom, D. J. (2014) 'Structural and mechanistic insights into prokaryotic energy-coupling factor transporters', *Nature Reviews Microbiology*, 12(2), pp. 79–87.

Sohlenkamp C, G. O. (2016) 'Bacterial membrane lipids: diversity in structures and pathways. - PubMed - NCBI', *FEMS Microbial Rev*, 40(1), pp. 133–59.

Sonoda, Y. *et al.* (2010) 'Tricks of the trade used to accelerate high-resolution structure determination of membrane proteins', *FEBS Letters*, 584(12), pp. 2539–2547.

Studier, F. W. (2005) 'Protein production by auto-induction in high-density shaking cultures', *Protein Expression and Purfication*, 41(1), pp. 207–234.

Suzuki, S. and Henderson, P. J. F. (2006) 'The hydantoin transport protein from Microbacterium liquefaciens', *Journal of Bacteriology*, 188(9), pp. 3329–3336.

Tanford, C. (1983) 'Translocation pathway in the catalysis of active transport.', *Proceedings of the National Academy of Sciences of the United States of America*, 80(12), pp. 3701–5.

Tang, L. *et al.* (2010) 'Crystal structure of the carnitine transporter and insights into the antiport mechanism', *Nature Structural and Molecular Biology*, 17(4), pp. 492–496.

Theobald, D. L. and Miller, C. (2010) 'Membrane transport proteins: Surprises in structural sameness', *Nature Structural and Molecular Biology*, 17(1), pp. 2–3.

Wahlgren, W. Y. *et al.* (2018) 'Substrate-bound outward-open structure of a Na+-coupled sialic acid symporter reveals a new Na+ site', *Nature Communications*, 9(1), pp. 1–14.

Waldo, G. S. *et al.* (1999) 'Rapid protein-folding assay using green fluorescent protein', *Nature Biotechnology*, 17(7), pp. 691–695.

Watanabe, A. *et al.* (2010) 'The mechanism of sodium and substrate release from the binding pocket of vSGLT', *Nature*, 468(7326), pp. 988–991.

Waterman, D. G. *et al.* (2016) 'Diffraction-geometry refinement in the DIALS framework', *Acta Crystallographica Section D: Structural Biology*, 72(4), pp. 558–575.

Weyand, S. *et al.* (2008) 'Structure and molecular mechanism of a nucleobase-cation- symport-1 family transporter', *Science.*, 322(October), pp. 709–713.

Weyand, S. *et al.* (2011) 'The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1', *Journal of Synchrotron Radiation*, 18(1), pp. 20–23.

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

Winter, G. (2010) 'Xia2: An Expert System for Macromolecular Crystallography Data Reduction', *Journal of applied crystallography*, 43, pp. 186–190.

Yamashita, A. *et al.* (2005) 'Crystal structure of a bacterial homologue of Na+/Cl -- dependent neurotransmitter transporters', *Nature*, 437(7056), pp. 215–223.

Yang, Z. R. *et al.* (2005) 'RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins', *Bioinformatics*, 21(16), pp. 3369–3376.

Yu, X. *et al.* (2017) 'Dimeric structure of the uracil:proton symporter UraA provides mechanistic insights into the SLC4/23/26 transporters', *Cell Research*, 27(8), pp. 1020–1033.

Zhang, X. *et al.* (2012) 'Crystal structure of an orthologue of the NaChBac voltage-gated sodium channel.', *Nature*, 486(7401), pp. 130–134.

Zhao, Y. *et al.* (2010) 'Single-molecule dynamics of gating in a neurotransmitter transporter homologue', *Nature*, 465(7295), pp. 188–193.

Zhao, Y. *et al.* (2011) 'Substrate-modulated gating dynamics in a Na+-coupled neurotransmitter transporter homologue', *Nature*, 474(7349), pp. 109–113.

Zhou, X. *et al.* (2014) 'Structural basis of the alternating-access mechanism in a bile acid transporter', *Nature*, 505(7484), pp. 569–573.

Zhou, Z. *et al.* (2007) 'LeuT-desipramine structure suggests how antidepressants inhibit human neurotransmitter transporters', *Science*, 317(5843), pp. 1390–1393.

Zhou, Z. *et al.* (2009) 'Antidepressant specificity of serotonin transporter suggested by three LeuT-SSRI structures', *Nature Structural and Molecular Biology*, 16(6), pp. 652–657.

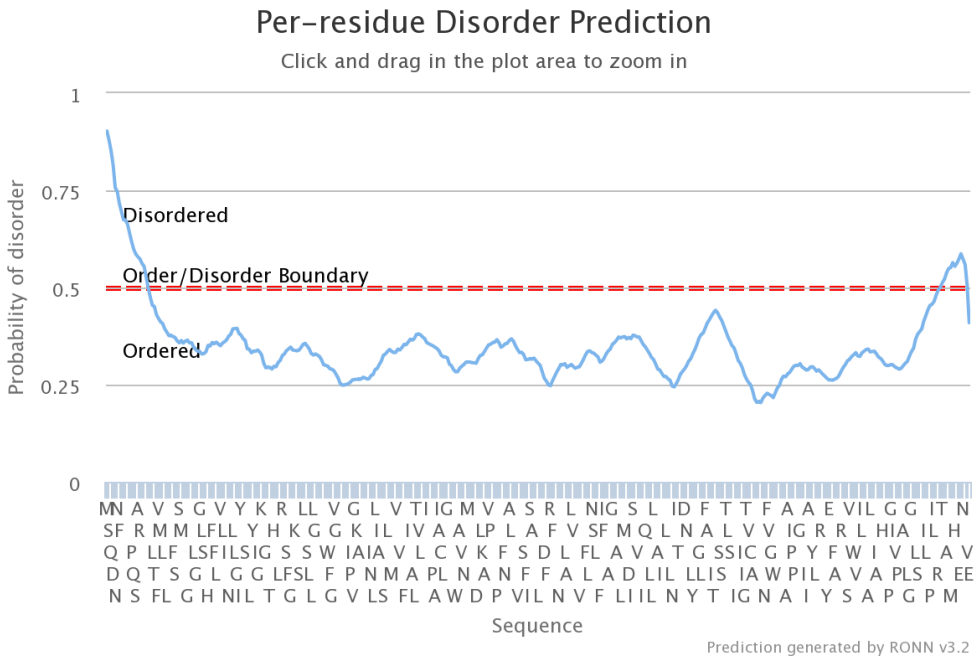
Ziegler, C., Bremer, E. and Krämer, R. (2010) 'The BCCT family of carriers: From physiology to crystal structure', *Molecular Microbiology*, 78(1), pp. 13–34.

Appendix 1

IDT Gene optimised DNA sequences of CodB and RONN disorder plots. Red writing denotes the NdeI restriction site, and blue the BamHI site.

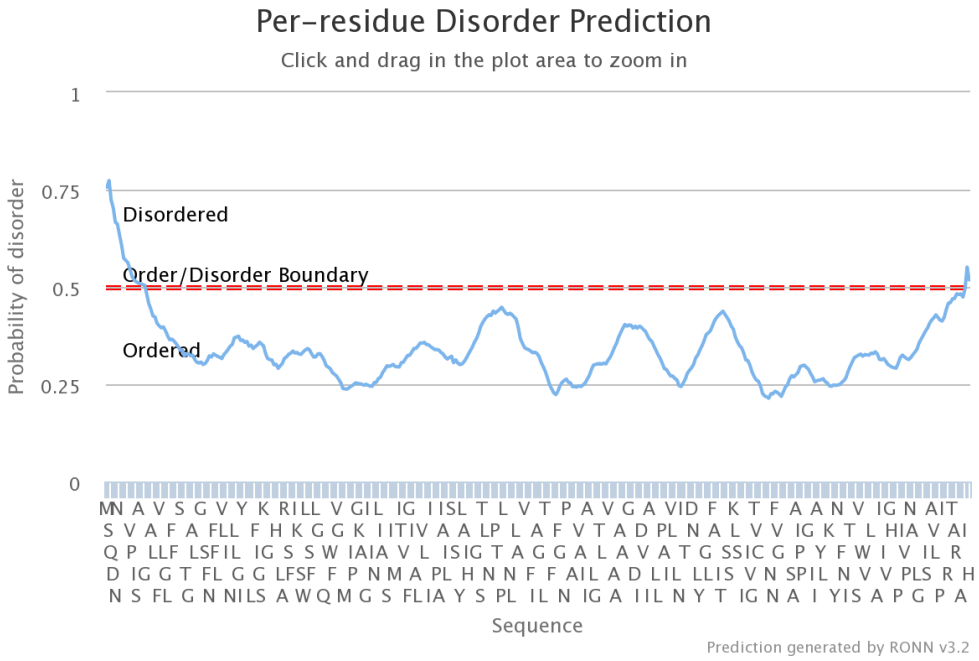
CodB *E. coli*

AGCAGCCATATGAGCCAGGACAACAATTTTAGTCAAGGTCCAGTCCCGC
AGTCTGCCCGCAAAGGTGTTTTGGCGCTTACCTTCGTGATGTTGGGTTTG
ACGTTTTTCTCCGCGTCTATGTGGACGGGAGGAACCTTGGGTACTGGTTT
ATCGTACCATGATTTCTTTTTAGCTGTTCTGATCGGCAACTTATTGCTGGG
CATCTATACCTCGTTTCTGGGGTATATCGGGGCAAAAACGGGTTTAACCA
CTCATCTTCTTGCCCGTTTCAGCTTCGGAGTTAAGGGGAGCTGGCTGCCC
AGTTTATTATTAGGCGGCACTCAAGTAGGGTGGTTCGGTGTAGGGGTTGC
GATGTTTGCAATTCCAGTGGGGAAGGCCACTGGATTAGATATCAATTTAT
TGATCGCTGTGTCCGGTCTTTTAATGACGGTCACCGTGTTTTTCGGAATTT
CCGCCCTTACGGTCCTGTCGGTCATTGCTGTCCCCGCCATCGCCTGCTTAG
GGGGCTACTCAGTGTGGTTAGCGGTGAACGGAATGGGGGGGCTTGACGC
GTTAAAAGCAGTGGTACCCGCACAACCATTAGATTTCAACGTAGCCCTTG
CTTTGGTCGTAGGGTCTTTTATTTCGGCAGGGACTCTTACTGCGGACTTCG
TGCGCTTTGGTCGCAATGCCAAACTTGCCGTATTAGTCGCTATGGTGGCT
TTCTTTCTGGGCAACTCTCTGATGTTCATTTTCGGCGCTGCGGGTGCCGCA
GCTTTGGGTATGGCCGACATTTCTGATGTTATGATTGCACAAGGATTGCT
TTTGCCCGCAATCGTAGTTCTTGGTTTGAATATTTGGACAACGAACGACA
ACGCGTTGTACGCCTCAGGACTTGGTTTTGCCAATATTACCGGGATGTCA
TCTAAAACTCTTTCTGTGATTAATGGCATCATTGGGACTGTGTGTGCTTTA
TGGTTATACAATAATTTTGTGGGATGGTTGACCTTCTTATCAGCAGCCAT
CCCGCCCGTAGGGGGTGTAATCATTGCGGACTACCTTATGAACCGTCGCC
GTTACGAACATTTCGCGACGACGCGCATGATGAGCGTCAATTGGGTTGC
AATCCTTGCTGTAGCCCTGGGTATTGCAGCAGGTCATTGGCTTCCTGGTA
TCGTCCCCGTGAACGCTGTGTTAGGTGGGGCGTTAAGTTACCTGATTCTG
AACCCCATTTTGAACCGCAAAACAACAGCGGCAATGACCCATGTTGAAG
CAAACTCAGTCGAAGGATCCATTATC



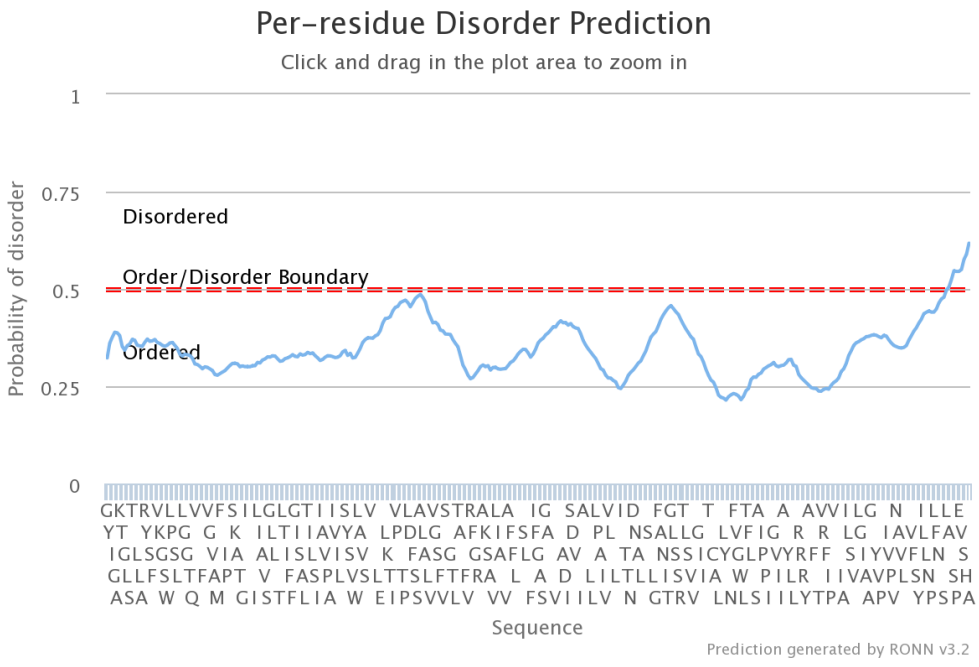
CodB *P. vulgaris*

TCAAGCCATATGAGCCAGGACAACAACTATAGTCAAGGGCCAGTCCCTA
TTTCCGCACGTAAGGGAGGACTTGCCCTTACCTTTGTCATGTTAGGTCTG
ACATTCTTTTCCGCATCTATGTGGACCGGAGGCGCTCTTGGGACAGGACT
GTCGTTCAACGACTTTTTCCTTGCGGTTCTGATCGGAAACCTGCTTCTGGG
TATCTACACGGCCTTCCTGGGTTTTATCGGGAGTAAGACTGGGTTAACTA
CTCACTTGCTTGCCCGTTACTCGTTCGGCATCAAAGGGTCCTGGTTACCCT
CATTTCTTCTGGGGGGGACTCAGGTTGGATGGTTTGGTGTAGGCGTCGCT
ATGTTTGCGATTCCGGTGGGTAAAGCCACGGGAATTGACATCAACTTACT
GATTGCTGTTAGTGGGATTCTGATGACCATTACTGTGTTCTTTGGTATCTC
TGCGCTGACCGTTTTATCCATCATCGCCGTTCCGGCTATCGCAATCCTTGG
CAGTTATAGCGTTTATCTGGCGATCCACGACATGGGCGGGCTGAGCACG
CTTATGAACGTGAAGCCCACACAACCATTAGACTTTAATTTAGCCCTTGC
GATGGTCGTGGGATCATTCATTAGCGCTGGTACACTTACAGCCGATTTCG
TCCGTTTCGGTCGTAACCCAAAAGTTGCAGTCGTTGTGGCAATCATCGCT
TTCTTTTTAGGCAATACGCTTATGTTTGTATTTGGCGCGGCTGGGGCCGC
GTCGTTGGGAATGGCCGACATCTCTGATGTCATGATCGCTCAGGGGTTAC
TTCTGCCGGCTATCGTGGTCCTGGGTTTGAATATTTGGACCACAAATGAC
AATGCCCTTTACGCCTCAGGATTAGGTTTTGCAAACATTACCGGGTTGAG
TAGCAAGAAGTTGAGCGTGATCAACGGCATCGTTGGAACGGTGTGTGCT
CTGTGGTTGTACAACAATTTCGTTGGTTGGCTTACATTCCTGTCAGCAGC
AATCCCGCCTGTCGGTGGGGTTATTATCGCCGACTATCTTATGAATAAGG
CACGCTATAACACTTTTAATATCGCAACCATGCAGTCCGTCAATTGGGTA
GCTTTGCTGGCCGTGGCTATTGGCATCGTAGCGGGCCATTGGTTGCCGGG
CATTGTGCCTGTTAACGCCGTTTTAGGCGGCGCCATTAGCTACGCAGTTT
TAAATCCGATCTTAAATCGCCGCACCGCACGCCAGGCTGAAATTTCTCAT
GCTGGATCCACTGTC



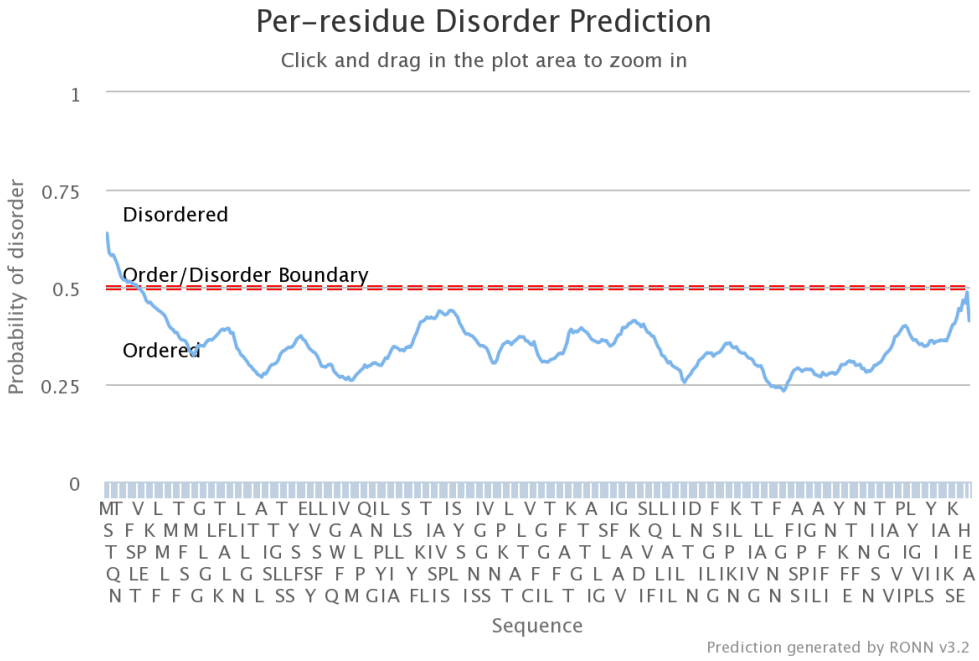
CodB *S. enterica*

GTGACACATATGTCCCAGGATAACAACTACTCGCAGGGTCCGGTACCAC
TGGCAGCACGCAAAGGAGTAATTCCTTTGACTTTCGTTATGTTGGGCCTT
ACCTTTTTTAGTGCTTCCATGTGGACTGGTGGTACATTAGGCACTGGGCT
GAGTTACAATGACTTCTTTTTAGCCGTGTTATTTGGAAATTTGTTACTGGG
TATTTACACAGCGTTTTTGGGTTACATCGGCGCCAAAACCGGTCTGTCTA
CGCACTTGCTGGCGCGTTACTCTTTTGGAGTCAAAGGATCTTGGCTTCCA
AGCCTTTTGCTGGGCGGTACGCAAGTAGGGTGGTTCGGCGTAGGCGTGG
CTATGTTCGCTATTCCGGTCTCCAAGGCTACTGGCATTGATGCTAACATC
TTGATTGCAGTCAGTGGTTTGCTGATGACCTTGACAATCTTTTTCGGTATT
TCTGCGTTAACCATCCTTAGCATCATTGCGGTCCCGGCCATCGTAATCCT
GGGTTCCTACTCAGTGTGGCTTGCAGTCAGCGGCGTAGGCGGACTTGAGC
ATTTGAAGACAATCGTGCCGCAGACTCCTCTGGACTTTTCCTCAGCATTA
GCACTTGTAGTAGGATCATTCGTTTCAGCTGGGACCCTTACCGCAGACTT
CGTTCGCTTTGGGCGTCATGCAAAATCTGCAGTTCTGATCGCAATGGTTG
CATTTTTCCTTGGAAATTCGCTTATGTTTATTTTTGGTGCTTCAGGTGCTG
CCGCAGTTGGACAGGCGGATATTTCCGACGTCATGATTGCTCAGGGCCTT
TTACTTCCCGCGATCGTAGTACTGGGCTTGAACATTTGGACAACGAATGA
TAACGCGCTGTACGCGTCGGGTTTGGGGTTCGCCAACATTACCGGATTAT
CTAGTCGCACGCTTAGTGTCGTCAACGGTATTATTGGCACGGTATGCGCA
CTGTGGTTATACAACAACTTCGTTGGTTGGCTGACGTTCCTTAGTAGTGC
TATCCCTCCCATCGGTGGTGTTATCATTGCTGACTATTTGTTAAATCGCCG
TCGCTATGCCGATTTCAACACAGTACGCTTCATCCCGGTCAACTGGATTG
CGATCTTGTCAGTGGCTTTAGGCATTGCAGCCGGACATTACGTCCCAGGC
ATCGTACCTGTGAATGCGGTACTGGGAGGGGTATTCAGTTACATCCTGCT
TAATCCCCTGTTCAATCGCTCCTTAGCGAAAAGCCCGGAGGTTTCCCACG
CTGAGCAGGGATCCATCAGC



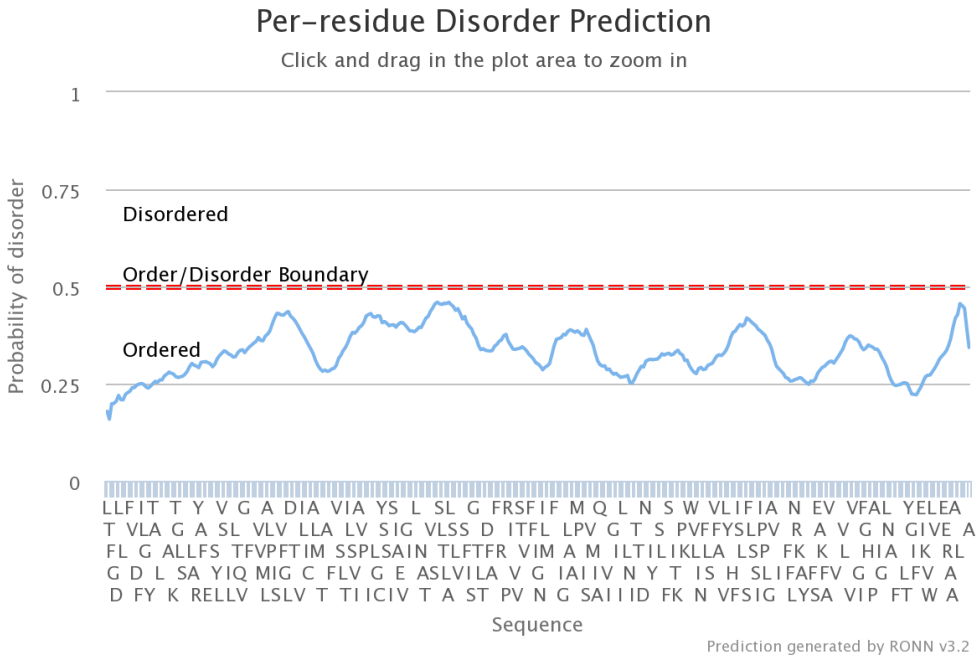
CodB *C. lundense*

AGTATGCATATGTCCACCCAGAACACGAACTATGATCACGACTTCAGTTT
GACCGTAGTACCCGAAGGAGAAAAAAAGGGGTTTTTATCAATGTTGGTC
GTTATGCTTGGATTCACTTTCTTCTCAGCTTCCATGTGGACAGGCGGCAA
GCTTGGGACAGGTCTTGATATGAAAACTTTTGCGCTTGCTGTACTTTCTG
GTAACCTGATTCTGGGGGGCCTATACGGGAGCGCTTGCATACATCTCTTGC
GAGACTGGATTGTCGACGCACCTTTTAAGCCGTTACAGTTTCGGCGAAAA
GGGCAGTTACCTGGTCAGCTTTTTGCTGGGTGGGACCCAGATCGGTTGGT
TCGGGGTGGGCGTGGCGATGTTTGCCTTGCCAGTACAAAAGGTAACCGG
GATCAATCCTTACATCTTAGTCCTGATTGCGGGGCTTTTGATGACGTCGT
CGGCATACTTCGGGATGAAGACTTTAACAATCATTAGTATCTTAGCAGTA
CCTTCTATTGCGGTTCTTGGTAGTTATTCCGCGATTAATGCAGTTAACAGT
ATCGGCGGCTTTTCCGTACTTATGAACTATCAGCCTAAAGAAACTTTAGC
GTTTGCCACGGCGCTTACAATGTGCGTCGGCTCCTTTATCAGCGGAGGGA
CCCTGACGCCTGACTTTACCCGCTTTGCTAAGACCAAAAAGGTTGGTGTC
TTGACTACCGTGATTGCATTTTTCTTGGGTAATTCCTTAATGTTTATCTTT
GGTGCCGTGGGGGCCGCAGCCACAGGCAAGTCCGATATTAGCGAGGTTA
TGTTTTTACAGGGGTTGATTCTGCCAGCGATTTTAATTCTTGGTTTGAACA
TCTGGACCACCAATGATAATGCTATTTATTCTTCGGGATTGGGCTTTTCA
AATATTACTAAGATTCCAAAGAACAAATTAGTCATCGTCAACGGAATTGT
GGGCACATTGGCCGCGATGTGGTTATACAACAATTTCGTAGGTTGGTTAA
CATTTCTGTCGAGTGCAATCCCGCCGATTGGGGGTGTAATCCTGGCGGAC
TTCTTTATTGTTAACCGTAAAATGTATGGAAAATTTGAAGAAACAAAGTT
CAAGAACGTCAACTGGAATGCGATTGTGTCTTGGACCATCGGCACCGTTG
CTGCCGAGGTTATCCCCGGAATCACCCCACTTTATGGGGTGCTTGGCGGA
GCGATCAGCTACATCATTATCGGCAAGGCTATGAAGTCCAAAGAGATTA
AGGAACGCCACGAAGCTGCAGCCGGATCCGTACAC



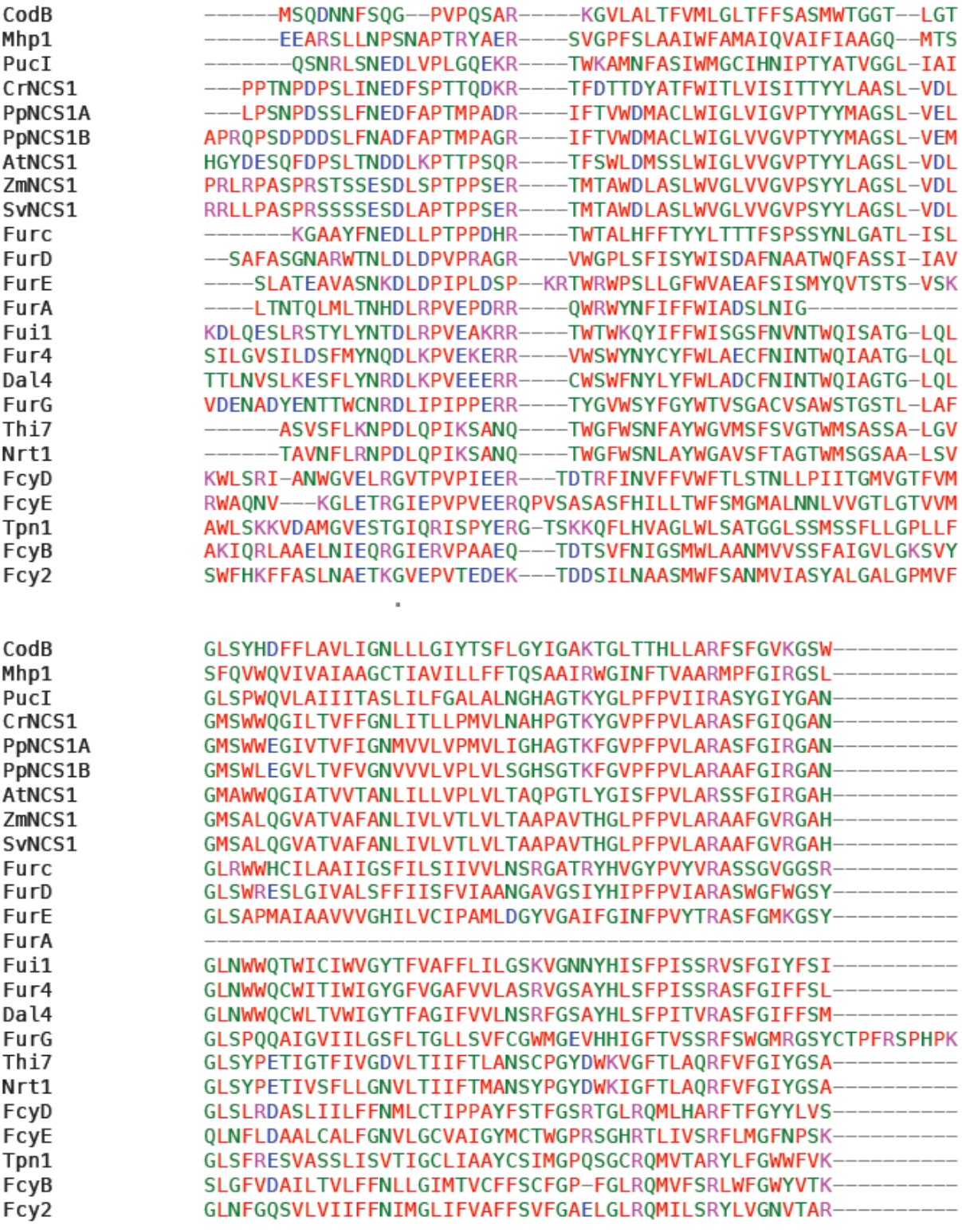
CodB *S. bovis*

TTCGTACATATGGAGAAGAAAATCGATAATGACTTCTCTCTTACACCAGT
ATCTAAAGAAGGCCGCCGTGGGTTAATTTCTATGATGGCCATCATGCTTG
GCTTTACTTTCTATACAGGAACTATGTTAACAGGAGGACGCTTGGGGACA
TCGTTAACATTCGGTGACTTAGCTTTAGTCCTTTTCGTCGGCGACTTTATC
TTGGGTGCGTATACGGCATTATTGGCCTACATGGCCGGAAAAACCGGGTT
ATCGACGCACTTATTGGCTCGCTACGCGTTTGGAGAAAAGGGTTCCTACC
TTGTCTCGGGGATCTTGGGATTGACACAAGTAGGATGGTTCGGAGTATCT
GTGGTAATGTTGGCACTTCCAATCTCAAAAGTTTTCGGATTGGATGTAAC
ACCCGTTATCCTGATCTGCGGAGCGTTAATGGTAACGACCGCCTACTTCG
GCGTAAAGTCGCTTACGATCTTGTCCGTGATTGCCGTACCTGCCATTGCT
GTACTTGGCTGTTACTCATCAAGCATTTCGATTGCAGAAGTTGGGGGAAT
CGGTGCACTTATGAATGCCACTGATGTGAGTAGCATGTCGCTTACTTTGG
CATTAAGTCTTGTAGTTGGTAGCTTTATTTCGGGAGGGACGCTTACGCCC
GACTTTGCACGTTTTTCCCGCACGCCACGTATCGCCGTAGTTTCGACGGT
GGCTGCGTTTTTCATCGGGAACATCTTAATGTTTGCTTTTGGAGCTATCGG
TGGCCTTGCTGCTGGTATGCCGGATATTTCCGACGTCATGATTGCTCAGG
GGTTGGTCATTAGCGGCATTGTTATCCTTGGACTGAATATTTGGACAACT
AACGATAACACGATTTATGCTGCGTCATTGGCTTTTAGCAATATTACCAA
GATGCCGAAAAAGCATTGGGTATTGATTAACGGATTTCTTAGCACCGTAT
TTGCGATGGTCCTTTACAACCACTTCATCTCATTACTTAGTTTTTTGTCCT
CCATCATCCCCCCCCTGGGGGCAGTAATGATTATGGATTACTTCTTCTTG
AATCGTAAGGCATACGCTGGGGCGTTCAGTGAGGCAAAGTTCGCCGTGG
TGAATGTTCCTGCTGTTCTTGCCGTGGTGGCAGGCGGAATCTTTGGGCAC
TTACCAGCGGGAATTGGCTGTCTTAATGCTGTTTTTGGGGCCATGTTGAC
CTACGGAATTTTTACTGAAATCAAGGTTTGGTTAGTCCGCCGTCGTGAGG
AGCGCGCTGCGGCCGGACTTCGTAAAGTTGCCGGATCCCTGTGA

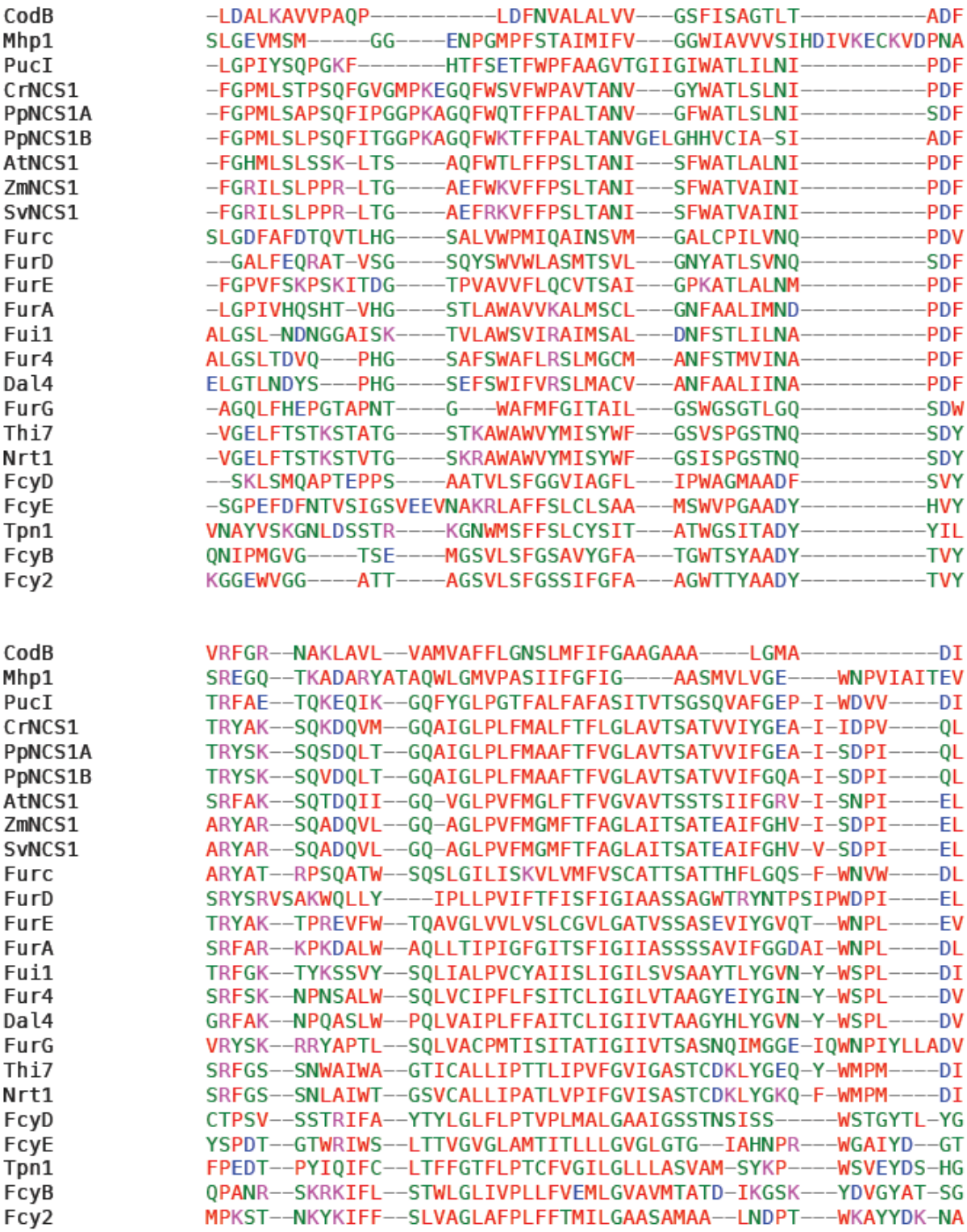


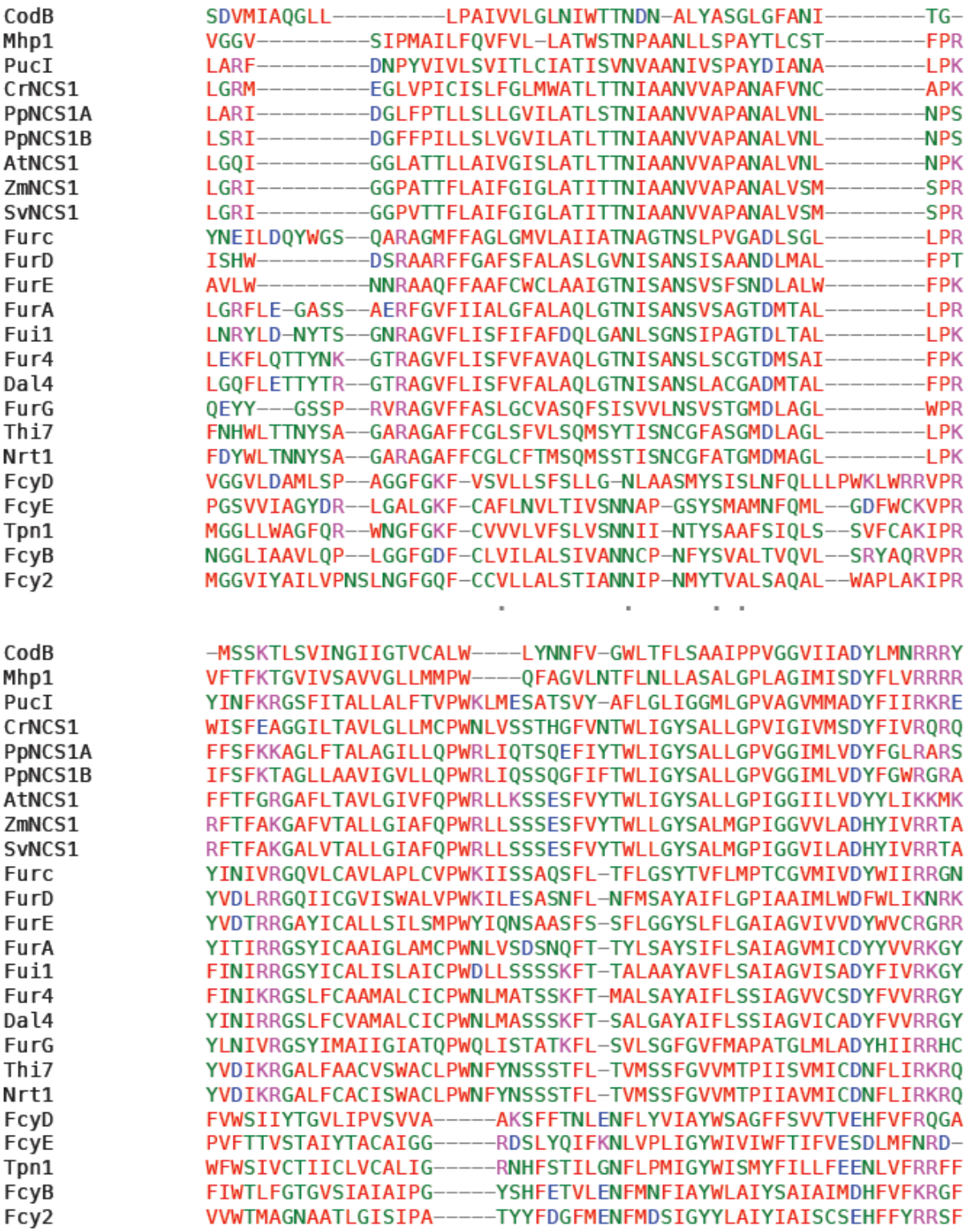
Appendix 2

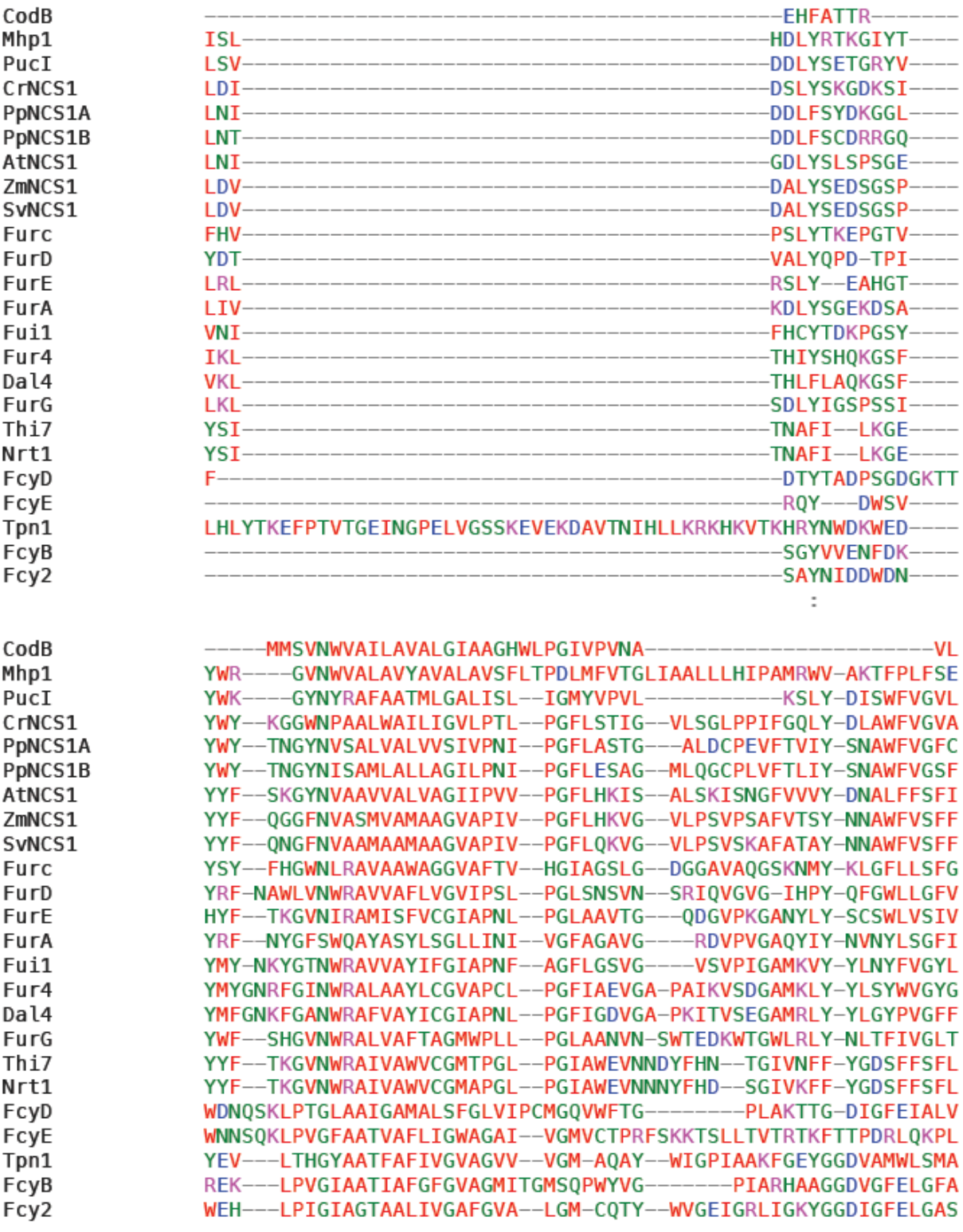
MUSCLE alignment of all characterised NCS1.

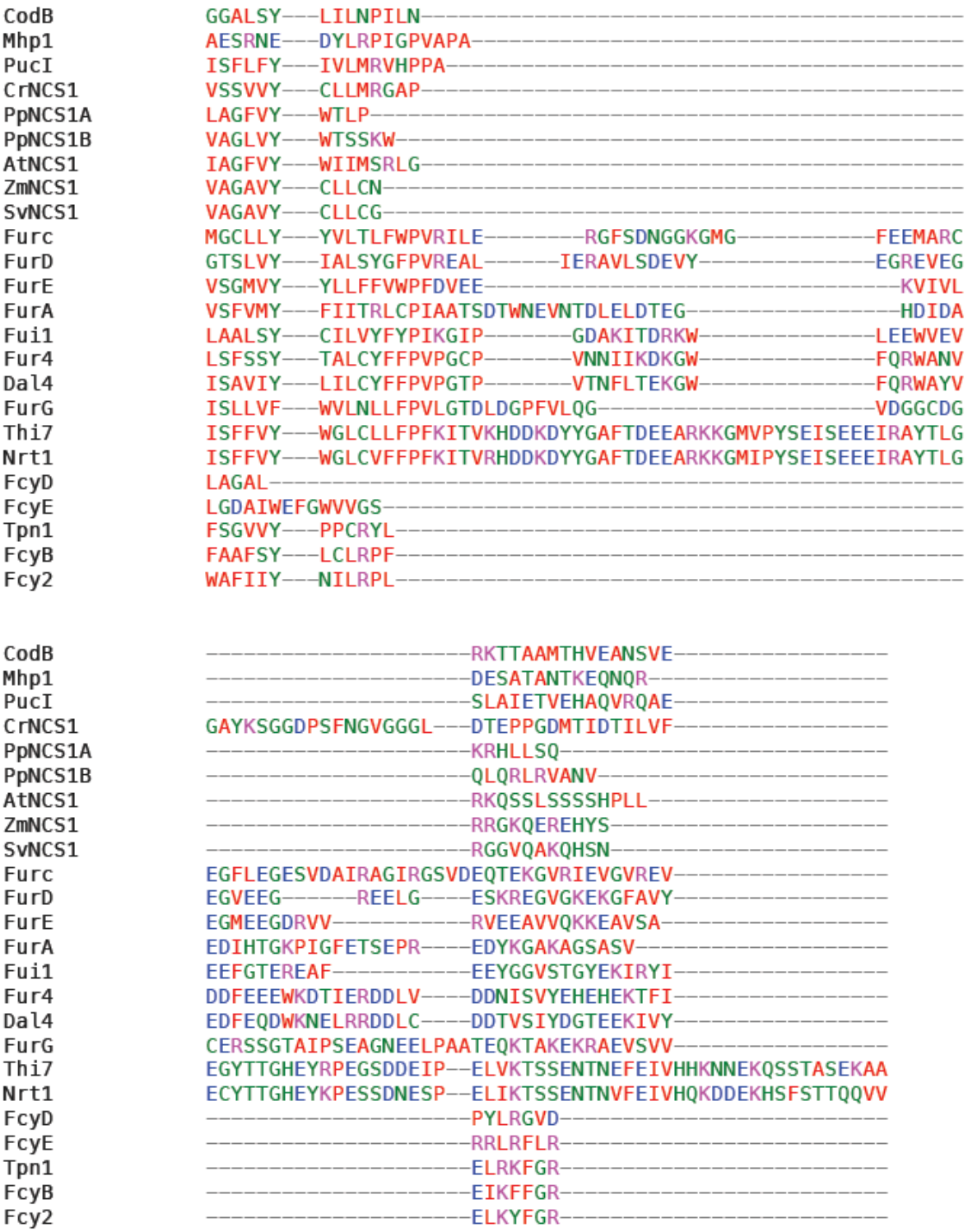

CodB -------MSQDNNFSQG--PVPQSAR------KGVLALTFVMLGLTFFSASMWTGGT--LGT
Mhp1 -------EEARSLLNPSNAPTRYAER-----SVGPFSLAAIWFAMAIQVAIFIAAGQ--MTS
PucI --------QSNRLSNEDLVPLGQEKR-----TWKAMNFASIWMGCIHNIPTYATVGGL-IAI
CrNCS1 ---PPTNPDPSLINEDFSPTTQDKR-----TFDTTDYATFWITLVISITTYYLAASL-VDL
PpNCS1A ---LPSNPDSSLFNEDFAPTMPADR-----IFTVWDMACLWIGLVIGVPTYYMAGSL-VEL
PpNCS1B APRQPSDPDDSLFNADFAPTMPAGR-----IFTVWDMACLWIGLVVGVPTYYMAGSL-VEM
AtNCS1 HGYDESQFDPSLTNDDLKPTTPSQR-----TFSWLDMSSLWIGLVVGVPTYYLAGSL-VDL
ZmNCS1 PRLRPASPRSTSSESDLSPTPPSER-----TMTAWDLASLWVGLVVGVPSYYLAGSL-VDL
SvNCS1 RRLLPASPRSSSSESDLAPTPPSER-----TMTAWDLASLWVGLVVGVPSYYLAGSL-VDL
Furc --------KGAAYFNEDLLPTPPDHR-----TWTALHFFTYYLTTTFSPSSYNLGATL-ISL
FurD --SAFASGNARWTNLDLDPVPRAGR-----VWGPLSFISYWISDAFNAATWQFASSI-IAV
FurE -----SLATEAVASNKDLDPIPLDSP--KRTWRWPSLLGFWVAEAFSISMYQVTSTS-VSK
FurA -----LTNTQLMLTNHDLRPVEPDRR-----QWRWYNFIFFWIADSLNIG----------
Fui1 KDLQESLRSTYLYNTDLRPVEAKRR-----TWTWKQYIFFWISGSFNVNTWQISATG-LQL
Fur4 SILGVSILDSFMYNQDLKPVEKERR-----VWSWYNYCYFWLAECFNINTWQIAATG-LQL
Dal4 TTLNVSLKESFLYNRDLKPVEEERR-----CWSWFNYLYFWLADCFNINTWQIAGTG-LQL
FurG VDENADYENTTWCNRDLIPIPPERR-----TYGVWSYFGYWTVSGACVSAWSTGSTL-LAF
Thi7 -------ASVSFLKNPDLQPIKSANQ-----TWGFWSNFAYWGVMSFSVGTWMSASSA-LGV
Nrt1 -------TAVNFLRNPDLQPIKSANQ-----TWGFWSNLAYWGAVSFTAGTWMSGSAA-LSV
FcyD KWLSRI-ANWGVELRGVTPVPIEER---TDTRFINVFFVWFTLSTNLLPIITGMVGTFVM
FcyE RWAQNV---KGLETRGIEPVPVEERQPVSASASFHILLTWFSMGMALNNLVVGTLGTVVM
Tpn1 AWLSKKVDAMGVESTGIQRISPYERG-TSKKQFLHVAGLWLSATGGLSSMSSFLLGPLLF
FcyB AKIQRLAAELNIEQRGIERVPAAEQ---TDTSVFNIGSMWLAANMVVSSFAIGVLGKSVY
Fcy2 SWFHKFFASLNAETKGVEPVTEDEK---TDDSILNAASMWFSANMVIASYALGALGPMVF
CodB GLSYHDFFLAVLIGNLLLGIYTSFLGYIGAKTGLTTHLLARFSFGVKGSW----------
Mhp1 SFQVWQVIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSL----------
PucI GLSPWQVLAIIITASLILFGALALNGHAGTKYGLPFPVIIRASYGIYGAN----------
CrNCS1 GMSWWQGILTVFFGNLITLLPMVLNAHPGTKYGVPFPVLARASFGIQGAN----------
PpNCS1A GMSWWEGIVTVFIGNMVVLVPMVLIGHAGTKFGVPFPVLARASFGIRGAN----------
PpNCS1B GMSWLEGVLTVFVGNVVVLVPLVLSGHSGTKFGVPFPVLARAAFGIRGAN----------
AtNCS1 GMAWWQGIATVVTANLILLVPLVLTAQPGTLYGISFPVLARSSFGIRGAH----------
ZmNCS1 GMSALQGVATVAFANLIVLVTLVLTAAPAVTHGLPFPVLARAAFGVRGAH----------
SvNCS1 GMSALQGVATVAFANLIVLVTLVLTAAPAVTHGLPFPVLARAAFGVRGAH----------
Furc GLRWWHCILAAIIGSFILSIIVVLNSRGATRYHVGYPVYVRASSGVGGSR----------
FurD GLSWRESLGIVALSFFIISFVIAANGAVGSIYHIPFPVIARASWGFWGSY----------
FurE GLSAPMAIAAVVVGHILVCIPAMLDGYVGAIFGINFPVYTRASFGMKGSY----------
FurA ----------
Fui1 GLNWWQTWICIWVGYTFVAFFLILGSKVGNNYHISFPISSRVSFGIYFSI----------
Fur4 GLNWWQCWITIWIGYGFVGAFVVLASRVGSAYHLSFPISSRASFGIFFSL----------
Dal4 GLNWWQCWLTVWIGYTFAGIFVVLNSRFGSAYHLSFPITVRASFGIFFSM----------
FurG GLSPQQAIGVIILGSFLTGLLSVFCGWMGEVHHIGFTVSSRFSWGMRGSYCTPFRSPHPK
Thi7 GLSYPETIGTFIVGDVLTIIFTLANSCPGYDWKVGFTLAQRFVFGIYGSA----------
Nrt1 GLSYPETIVSFLLGNVLTIIFTMANSYPGYDWKIGFTLAQRFVFGIYGSA----------
FcyD GLSLRDASLIILFFNMLCTIPPAYFSTFGSRTGLRQMLHARFTFGYYLVS----------
FcyE QLNFLDAALCALFGNVLGCVAIGYMCTWGPRSGHRTLIVSRFLMGFNPSK----------
Tpn1 GLSFRESVASSLISVTIGCLIAAYCSIMGPQSGCRQMVTARYLFGWWFVK----------
FcyB SLGFVDAILTVLFFNLLGIMTVCFFSCFGP-FGLRQMVFSRLWFGWYVTK----------
Fcy2 GLNFGQSVLVIIFFNIMGLIFVAFFSVFGAELGLRQMILSRYLVGNVTAR----------


CodB ----------LPSLLLGGTQVGWFGV----GVAMFAIPVGKATG----------
Mhp1 ----------IPITLKALLSLFWFGFQTWLGALALDEITRLLTG----------
PucI ----------IPALLRAFTAIMWLGIQTFAGSTALNILLLNMWP-GWGEIGGEWNILG-
CrNCS1 ----------LPSLSRAIVACGWFGIQTWIGGSSIFQMLMAVTGGAVAAAPIAWLG---
PpNCS1A ----------VPSLLRALVACGWFGIQTWIGGQAVFQLLNASLN---GSLAGSVVPWLG
PpNCS1B ----------VPSLLRAFVACGWFGIQTWIGGQAIYQLLNALFH---GRLASSVLPWLG
AtNCS1 ----------IPTLLRALVGCGWYGIETWIGGEAIFLLLPGHIK--KSALSHTLPWLG-
ZmNCS1 ----------VPAVIRALIGCGWFGIESWIGGRAIFLLLPSRLK-SYQPLLAPVPGLG-
SvNCS1 ----------VPAVIRALVGCGWFGIESWIGGRAIFLLLPSRLK-SYQPLLAPVPGLG-
Furc ----------LFVTVRASVAIIYFATQSYYGGMITSVCLRAIFGASWVNLPNRLPESAG
FurD ----------IAIISRVILAIFWFAIQNVNGANAVKAMISAIWP-SFLSMKNTIPQDQG
FurE ----------FAVFVRGIVAIIWFGTQTYQAGQCVSTMLSAIWP-SFNHFPNHLPSSGP
FurA ----------YIGGQCITLMIRAIWP-SYESLPNGIPESSG
Fui1 ----------WIVINRVVMACVWNSTLAYIGSQCVQLMLKAIFG---TNLNTRIKDTIK
Fur4 ----------WPVINRVVMAIVWYSVQAYIAATPVSLMLKSIFG---KDLQDKIPDHFG
Dal4 ----------WPIINRVVMAIVWYAVQAWLGATPVALMLKSIFG---KNLEDRIPNHFG
FurG GPNGRRLTRATVPVILRSFVGCMWFGMQAFWGGQATRVMIGAIIP-GFAHMKNYFSASSH
Thi7 ----------FGIIIRILMSIVNYGSNAWVGGLCINMILDS-WSHHYLHLPNTLSSKVA
Nrt1 ----------FGIIIRILMSIVNYGSNAWLGGLSINMILDS-WSHHYLHLPNTLSPSVA
FcyD ----------IIVALNLCTIAGFGVVSSVLGGSTLAAVSGG----------
FcyE ----------VCCFLNVLTNIGYGMMSSTVGGQILSKLSGG----------
Tpn1 ----------LVALASIIGVMGWSVVNSVVGGEMLAAISND----------
FcyB ----------GFAVLNILACLGWSAANAIVGAQMLHAVNSD----------
Fcy2 ----------IFSLINVIACVGWGIVNTSVSAQLLNMVNEGSGH----------
CodB ---LDINLLIAVS--GLLMTVTVFFGISALTVLSVIAVPAIACLGGYSVWLAVNGMGG--
Mhp1 ----FTNLPLWIVIFGAIQVVTTFYGITFIRWMNVFASPVL--LAMGVYMVYLMLDGADV
PucI ---IHLSGLLSFVFFWAIHLLVLHHGMESIKRFEVWAGPLVYLVFGGMVWWAVDIAGG--
CrNCS1 ---ISLPELLCFLGFWAAQVWIVVRGMESIRILEKYSAPILIGLSLALMGWAVTTAGG--
PpNCS1A ---ISAAEFGCFMAFWLLQVTIISHGIESIRKLENLSAPILIALSGALLAWAYVRAGG--
PpNCS1B ---ISASEFGCFMVFWLLQVGIVSNGIESIRELEKLSAPILIALSAALLAWAYVQAGG--
AtNCS1 ---TSPLEFSCFIVFWLAQLCIVWRGMDGIRKLEKYSAPILISLTSCLLAWSYLKAGG--
ZmNCS1 ---VAPLEFACFLAFWAAQLGVIMHGMEGIRKLEKLSAPVLIVLTSALLAWAYTSAGG--
SvNCS1 ---AAPLEFACFLAFWAAQLGVIMRGMEGIRKLEKFAAPVLFVLTSALLAWAYTSAGG--
Furc ---ITSRDLLAFFIFWILQFGVMFLHPTVLRHLFVIKAVYTTVALLGVLGWAVHMNGG--
FurD ---IETNTMIAYMIFWIVQMPFLCIHPNKVRWLFATKSVLVPAAWIAILIWAFVAEGK--
FurE ---ITSAELLCFFLAIILQAPLLWLKVSKLRYLFIVKTCIMPIFGIVLFAWAVKAANG--
FurA ---VDTKNFLSFFLFWLLSLPALWFPVHQIRHLFTVKSIYSPIAAIAFFAWAISRANG--
Fui1 NPNLTNFEFMCFMVFWVACLPFLWFPPDKLRHIFALKSAITPFAAFGFLIWTLCKAKGHL
Fur4 SPNATTYEFMCFFIFWAASLPFLLVPPHKIRHLFTVKAVLVPFASFGFLIWAIRRAHGRI
Dal4 SPNSTTFEFMCFFIFWVVSIPFVLVAPHKIRHLFTVKAALIPFAAFGFLIWALKKSHGKI
FurG ---LQTNDFIGLVIWMAGFIPAVLIKPEKLQIPFFSCFVLFCGTCFGLLIWSVSQAHG--
Thi7 ---MTTKELIGFIIFHVLTAFCYLMKPYHMNYILIWSCVATFFSMLGMVIYLAKQAHG--
Nrt1 ---MTTKQLVGFIIFHVLTALCYFMKPYHMNYLLIWSCVATCFAMLGIVIYLTKNAHG--
FcyD ----SIDDTAGIVVIGICGMVVSFGGYKFLHQYERYC--WLFALVAIVIATGV---GG--
FcyE ----AVSVVVGIIIVALLMVFCIL----------LGS-------
Tpn1 ----KVPLWVGIVIVTVCSFLVAIFGIKQVIKVETYL--SVPVLTAFLLLYISSSDKYSF
FcyB -----VPGFAAILIISICTLLVTFAGYKVVHLYEYWS--WIPTFIVFMIILGTFAHSGDF
Fcy2 ----VCPIWAGCLIIIGGTVLVTFFGYSVIHAYEKWS--WVPNFAVFLVIIAQLSRSGKF


CodB -LDALKAVVPAQP----------LDFNVALALVV---GSFISAGTLT----------ADF
Mhp1 SLGEVMSM-----GG----ENPGMPFSTAIMIFV---GGWIAVVVSIHDIVKECKVDPNA
PucI -LGPIYSQPGKF--------HTFSETFWPFAAGVTGIIGIWATLILNI----------PDF
CrNCS1 -FGPMLSTPSQFGVGMPKEGQFWSVFWPAVTANV---GYWATLSLNI----------PDF
PpNCS1A -FGPMLSAPSQFIPGGPKAGQFWQTFFPALTANV---GFWATLSLNI----------SDF
PpNCS1B -FGPMLSLPSQFITGGPKAGQFWKTFFPALTANVGELGHHVCIA-SI----------ADF
AtNCS1 -FGHMLSLSSK-LTS----AQFWTLFFPSLTANI---SFWATLALNI----------PDF
ZmNCS1 -FGRILSLPPR-LTG----AEFWKVFFPSLTANI---SFWATVAINI----------PDF
SvNCS1 -FGRILSLPPR-LTG----AEFRKVFFPSLTANI---SFWATVAINI----------PDF
Furc SLGDFAFDTQVTLHG----SALVWPMIQAINSVM---GALCPILVNQ----------PDV
FurD --GALFEQRAT-VSG----SQYSWVWLASMTSVL---GNYATLSVNQ----------SDF
FurE -FGPVFSKPSKITDG----TPVAVVFLQCVTSAI---GPKATLALNM----------PDF
FurA -LGPIVHQSHT-VHG----STLAWAVVKALMSCL---GNFAALIMND----------PDF
Fui1 ALGSL-NDNGGAISK----TVLAWSVIRAIMSAL---DNFSTLILNA----------PDF
Fur4 ALGSLTDVQ---PHG----SAFSWAFLRSLMGCM---ANFSTMVINA----------PDF
Dal4 ELGTLNDYS---PHG----SEFSWIFVRSLMACV---ANFAALIINA----------PDF
FurG -AGQLFHEPGTAPNT----G---WAFMFGITAIL---GSWGSGTLGQ----------SDW
Thi7 -VGELFTSTKSTATG----STKAWAWVYMISYWF---GSVSPGSTNQ----------SDY
Nrt1 -VGELFTSTKSTVTG----SKRAWAWVYMISYWF---GSISPGSTNQ----------SDY
FcyD --SKLSMQAPTEPPS----AATVLSFGGVIAGFL---IPWAGMAADF----------SVY
FcyE -SGPEFDFNTVSIGSVEEVNAKRLAFFSLCLSAA---MSWVPGAADY----------HVY
Tpn1 VNAYVSKGNLDSSTR----KGNWMSFFSLCYSIT---ATWGSITADY----------YIL
FcyB QNIPMGVG----TSE----MGSVLSFGSAVYGFA---TGWTSYAADY----------TVY
Fcy2 KGGEWVGG----ATT----AGSVLSFGSSIFGFA---AGWTTYAADY----------TVY
CodB VRFGR--NAKLAVL--VAMVAFFLGNSLMFIFGAAGAAA----LGMA----------DI
Mhp1 SREGQ--TKADARYATAQWLGMVPASIIFGFIG----AASMVLVGE----WNPVIAITEV
PucI TRFAE--TQKEQIK--GQFYGLPGTFALFAFASITVTSGSQVAFGEP-I-WDVV----DI
CrNCS1 TRYAK--SQKDQVM--GQAIGLPLFMALFTFLGLAVTSATVVIYGEA-I-IDPV----QL
PpNCS1A TRYSK--SQSDQLT--GQAIGLPLFMAAFTFVGLAVTSATVVIFGEA-I-SDPI----QL
PpNCS1B TRYSK--SQVDQLT--GQAIGLPLFMAAFTFVGLAVTSATVVIFGQA-I-SDPI----QL
AtNCS1 SRFAK--SQTDQII--GQ-VGLPVFMGLFTFVGVAVTSSTSIIFGRV-I-SNPI----EL
ZmNCS1 ARYAR--SQADQVL--GQ-AGLPVFMGMFTFAGLAITSATEAIFGHV-I-SDPI----EL
SvNCS1 ARYAR--SQADQVL--GQ-AGLPVFMGMFTFAGLAITSATEAIFGHV-V-SDPI----EL
Furc ARYAT--RPSQATW--SQSLGILISKVLVMFVSCATTSATTHFLGQS-F-WNVW----DL
FurD SRYSRVSAKWQLLY----IPLLPVIFTFISFIGIAASSAGWTRYNTPSIPWDPI----EL
FurE TRYAK--TPREVFW--TQAVGLVVLVSLCGVLGATVSSASEVIYGVQT--WNPL----EV
FurA SRFAR--KPKDALW--AQLLTIPIGFGITSFIGIIASSSSAVIFGGDAI-WNPL----DL
Fui1 TRFGK--TYKSSVY--SQLIALPVCYAIISLIGILSVSAAYTLYGVN-Y-WSPL----DI
Fur4 SRFSK--NPNSALW--SQLVCIPFLFSITCLIGILVTAAGYEIYGIN-Y-WSPL----DV
Dal4 GRFAK--NPQASLW--PQLVAIPLFFAITCLIGIIVTAAGYHLYGVN-Y-WSPL----DV
FurG VRYSK--RRYAPTL--SQLVACPMTISITATIGIIVTSASNQIMGGE-IQWNPIYLLADV
Thi7 SRFGS--SNWAIWA--GTICALLIPTTLIPVFGVIGASTCDKLYGEQ-Y-WMPM----DI
Nrt1 SRFGS--SNLAIWT--GSVCALLIPATLVPIFGVISASTCDKLYGKQ-F-WMPM----DI
FcyD CTPSV--SSTRIFA--YTYLGLFLPTVPLMALGAAIGSSTNSISS-----WSTGYTL-YG
FcyE YSPDT--GTWRIWS--LTTVGVGLAMTITLLLGVGLGTG--IAHNPR---WGAIYD--GT
Tpn1 FPEDT--PYIQIFC--LTFFGTFLPTCFVGILGLLLASVAM-SYKP----WSVEYDS-HG
FcyB QPANR--SKRKIFL--STWLGLIVPLLFVEMLGVAVMTATD-IKGSK---YDVGYAT-SG
Fcy2 MPKST--NKYKIFF--SLVAGLAFPLFFTMILGAASAMAA--LNDPT---WKAYYDK-NA


CodB SDVMIAQGLL----------LPAIVVLGLNIWTTNDN-ALYASGLGFANI---------TG-
Mhp1 VGGV----------SIPMAILFQVFVL-LATWSTNPAANLLSPAYTLCST---------FPR
PucI LARF----------DNPYVIVLSVITLCIATISVNVAANIVSPAYDIANA---------LPK
CrNCS1 LGRM----------EGLVPICISLFGLMWATLTTNIAANVVAPANAFVNC---------APK
PpNCS1A LARI----------DGLFPTLLSLLGVILATLSTNIAANVVAPANALVNL---------NPS
PpNCS1B LSRI----------DGFFPILLSLVGVILATLTTNIAANVVAPANALVNL---------NPS
AtNCS1 LGQI----------GGLATTLLAIVGISLATLTTNIAANVVAPANALVNL---------NPK
ZmNCS1 LGRI----------GGPATTFLAIFGIGLATITTNIAANVVAPANALVSM---------SPR
SvNCS1 LGRI----------GGPVTTFLAIFGIGLATITTNIAANVVAPANALVSM---------SPR
Furc YNEILDQYWGS--QARAGMFFAGLGMVLAIIATNAGTNSLPVGADLSGL---------LPR
FurD ISHW----------DSRAARFFGAFSFALASLGVNISANSISAANDLMAL---------FPT
FurE AVLW----------NNRAAQFFAAFCWCLAAIGTNISANSVSFSNDLALW---------FPK
FurA LGRFLE-GASS--AERFGVFIIALGFALAQLGTNISANSVSAGTDMTAL---------LPR
Fui1 LNRYLD-NYTS--GNRAGVFLISFIFAFDQLGANLSGNSIPAGTDLTAL---------LPK
Fur4 LEKFLQTTYNK--GTRAGVFLISFVFAVAQLGTNISANSLSCGTDMSAI---------FPK
Dal4 LGQFLETTYTR--GTRAGVFLISFVFALAQLGTNISANSLACGADMTAL---------FPR
FurG QEYY---GSSP--RVRAGVFFASLGCVASQFSISVVLNSVSTGMDLAGL---------WPR
Thi7 FNHWLTTNYSA--GARAGAFFCGLSFVLSQMSYTISNCGFASGMDLAGL---------LPK
Nrt1 FDYWLTNNYSA--GARAGAFFCGLCFTMSQMSSTISNCGFATGMDMAGL---------LPK
FcyD VGGVLDAMLSP--AGGFGKF-VSVLLSFSLLG-NLAASMYSISLNFQLLLPWKLWRRVPR
FcyE PGSVVIAGYDR--LGALGKF-CAFLNVLTIVSNNAP-GSYSMAMNFQML--GDFWCKVPR
Tpn1 MGGLLWAGFQR--WNGFGKF-CVVVLVFSLVSNNII-NTYSAAFSIQLS--SVFCAKIPR
FcyB NGGLIAAVLQP--LGGFGDF-CLVILALSIVANNCP-NFYSVALTVQVL--SRYAQRVPR
Fcy2 MGGVIYAILVPNSLNGFGQF-CCVLLALSTIANNIP-NMYTVALSAQAL--WAPLAKIPR
CodB -MSSKTLSVINGIIGTVCALW----LYNNFV-GWLTFLSAAIPPVGGVIIADYLMNRRRY
Mhp1 VFTFKTGVIVSAVVGLLMMPW----QFAGVLNTFLNLLASALGPLAGIMISDYFLVRRRR
PucI YINFKRGSFITALLALFTVPWKLMESATSVY-AFLGLIGGMLGPVAGVMMADYFIIRKRE
CrNCS1 WISFEAGGILTAVLGLLMCPWNLVSSTHGFVNTWLIGYSALLGPVIGIVMSDYFIVRQRQ
PpNCS1A FFSFKKAGLFTALAGILLQPWRLIQTSQEFIYTWLIGYSALLGPVGGIMLVDYFGLRARS
PpNCS1B IFSFKTAGLLAAVIGVLLQPWRLIQSSQGFIFTWLIGYSALLGPVGGIMLVDYFGWRGRA
AtNCS1 FFTFGRGAFLTAVLGIVFQPWRLLKSSESFVYTWLIGYSALLGPIGGIILVDYYLIKKMK
ZmNCS1 RFTFAKGAFVTALLGIAFQPWRLLSSSESFVYTWLLGYSALMGPIGGVVLADHYIVRRTA
SvNCS1 RFTFAKGALVTALLGIAFQPWRLLSSSESFVYTWLLGYSALMGPIGGVILADHYIVRRTA
Furc YINIVRGQVLCAVLAPLCVPWKIISSAQSFL-TFLGSYTVFLMPTCGVMIVDYWIIRRGN
FurD YVDLRRGQIICGVISWALVPWKILESASNFL-NFMSAYAIFLGPIAAIMLWDFWLIKNRK
FurE YVDTRRGAYICALLSILSMPWYIQNSAASFS-SFLGGYSLFLGAIAGVIVVDYWVCRGRR
FurA YITIRRGSYICAAIGLAMCPWNLVSDSNQFT-TYLSAYSIFLSAIAGVMICDYYVVRKGY
Fui1 FINIRRGSYICALISLAICPWDLLSSSSKFT-TALAAYAVFLSAIAGVISADYFIVRKGY
Fur4 FINIKRGSLFCAAMALCICPWNLMATSSKFT-MALSAYAIFLSSIAGVVCSDYFVVRRGY
Dal4 YINIRRGSLFCVAMALCICPWNLMASSSKFT-SALGAYAIFLSSIAGVICADYFVVRRGY
FurG YLNIVRGSYIMAIIGIATQPWQLISTATKFL-SVLSGFGVFMAPATGLMLADYHIIRRHC
Thi7 YVDIKRGALFAACVSWACLPWNFYNSSSTFL-TVMSSFGVVMTPIISVMICDNFLIRKRQ
Nrt1 YVDIKRGALFCACISWACLPWNFYNSSSTFL-TVMSSFGVVMTPIIAVMICDNFLIRKRQ
FcyD FVWSIIYTGVLIPVSVVA-----AKSFFTNLENFLYVIAYWSAGFFSVVTVEHFVFRQGA
FcyE PVFTTVSTAIYTACAIGG-----RDSLYQIFKNLVPLIGYWIVIWFTIFVESDLMFNRD-
Tpn1 WFWSIVCTIICLVCALIG-----RNHFSTILGNFLPMIGYWISMYFILLFEENLVFRRFF
FcyB FIWTLFGTGVSIAIAIPG-----YSHFETVLENFMNFIAYWLAIYSAIAIMDHFVFKRGF
Fcy2 VVWTMAGNAATLGISIPA-----TYYFDGFMENFMDSIGYYLAIYIAISCSEHFFYRRSF


CodB
Mhp1
PucI
CrNCS1
PpNCS1A
PpNCS1B
AtNCS1
ZmNCS1
SvNCS1
Furc
FurD
FurE
FurA
Fui1
Fur4
Dal4
FurG
Thi7
Nrt1
FcyD
FcyE
Tpn1
FcyB
Fcy2


CodB GGALSY---LILNPILN
Mhp1 AESRNE---DYLRPIGPVAPA
PucI ISFLFY---IVLMRVHPPA
CrNCS1 VSSVVY---CLLMRGAP
PpNCS1A LAGFVY---WTLP
PpNCS1B VAGLVY---WTSSKW
AtNCS1 IAGFVY---WIIMSRLG
ZmNCS1 VAGAVY---CLLCN
SvNCS1 VAGAVY---CLLCG
Furc MGCLLY---YVLTLFWPVRILE--------RGFSDNGGKGMG----------FEEMARC
FurD GTSLVY---IALSYGFPVREAL------IERAVLSDEVY--------------EGREVEG
FurE VSGMVY---YLLFFVWPFDVEE-----------------------------KVIVL
FurA VSFVMY---FIITRLCPIAATSDTWNEVNTDLELDTEG----------------HDIDA
Fui1 LAALSY---CILVYFYPIKGIP-------GDAKITDRKW--------------LEEWVEV
Fur4 LSFSSY---TALCYFFPVPGCP-------VNNIIKDKGW--------------FQRWANV
Dal4 ISAVIY---LILCYFFPVPGTP-------VTNFLTEKGW--------------FQRWAYV
FurG ISLLVF---WVLNLLFPVLGTDLDGPFVLQG---------------------VDGGCDG
Thi7 ISFFVY---WGLCLLFPFKITVKHDDKDYYGAFTDEEARKKGMVPYSEISEEEIRAYTLG
Nrt1 ISFFVY---WGLCVFFPFKITVRHDDKDYYGAFTDEEARKKGMIPYSEISEEEIRAYTLG
FcyD LAGAL--
FcyE LGDAIWEFGWVVGS--
Tpn1 FSGVVY---PPCRYL--
FcyB FAAFSY---LCLRPF--
Fcy2 WAFIIY---NILRPL--
CodB ---------------------RKTTAAMTHVEANSVE-----------------
Mhp1 ---------------------DESATANTKEQNQR-----------------
PucI ---------------------SLAIETVEHAQVRQAE-----------------
CrNCS1 GAYKSGGDPSFNGVGGGL---DTEPPGDMTIDTILVF-----------------
PpNCS1A ---------------------KRHLLSQ-----------------------
PpNCS1B ---------------------QLQRLRVANV---------------------
AtNCS1 ---------------------RKQSSLSSSSHPLL-----------------
ZmNCS1 ---------------------RRGKQEREHYS-----------------
SvNCS1 ---------------------RGGVQAKQHSN-----------------
Furc EGFLEGESVDAIRAGIRGSVDEQTEKGVRIEVGVREV-----------------
FurD EGVEEG------REELG-----ESKREGVGKEKGFAVY-----------------
FurE EGMEEGDRVV-----------RVEEAVVQKKEAVSA-----------------
FurA EDIHTGKPIGFETSEPR-----EDYKGAKAGSASV-----------------
Fui1 EEFGTEREAF-----------EEYGGVSTGYEKIRYI-----------------
Fur4 DDFEEEWKDTIERDDLV-----DDNISVYEHEHEKTFI-----------------
Dal4 EDFEQDWKNELRRDDLC-----DDTVSIYDGTEEKIVY-----------------
FurG CERSSGTAIPSEAGNEELPAATEQKTAKEKRAEVSVV-----------------
Thi7 EGYTTGHEYRPEGSDDEIP---ELVKTSSENTNEFEIVHHKNNEKQSSTASEKAA
Nrt1 ECYTTGHEYKPESSDNESP---ELIKTSSENTNVFEIVHQKDDEKHSFSTTQQVV
FcyD ---------------------PYLRGVD-----------------------
FcyE ---------------------RRLRFLR-----------------------
Tpn1 ---------------------ELRKFGR-----------------------
FcyB ---------------------EIKFFGR-----------------------
Fcy2 ---------------------ELKYFGR-----------------------

Appendix 3

Inhibition of ^{3}H-5-cytosine uptake in the presence of 0.5 mM inhibitor.