STRUCTURAL AND BIOCHEMICAL CHARACTERISATION OF ENZYMES REQUIRED FOR GRAM-NEGATIVE CAPSULE BIOSYNTHESIS

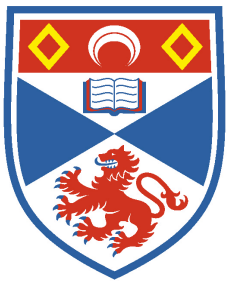
Laura Suzanne Woodward

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2018

Structural and Biochemical Characterisation of Enzymes Required for Gram-Negative Capsule Biosynthesis

Laura Suzanne Woodward



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of Doctor of Philosophy

at the

University of St Andrews

September 2017

Supervisor: Professor James H. Naismith

I. Abstract

Bacteria are coated with a wide-variety of complex polysaccharides, including lipopolysaccharide, crucial to their survival upon infection of a host organism. During lipopolysaccharide biosynthesis, the integral membrane protein Wzy polymerises undecaprenyl pyrophosphate-linked O-antigen repeat units. The mechanism of Wzy-catalysed polymerisation is unknown, as the difficulties associated with integral membrane proteins have precluded its study. Here we describe a protocol that takes membrane proteins from cloning to crystallisation trials, applied to 16 Wzy homologues, 14 of which could be expressed and purified, with 5 entered into crystallisation trials. The final O-antigen chain length is serotype-specific and is controlled by Wzz via an unknown mechanism. Proposed mechanisms are based on a direct interaction between Wzz and Wzy, however evidence for this was only provided recently. Here we describe the tandem affinity purification and preliminary results of the ongoing structural characterisation of Wzz with Wzy via cryo-EM. Not only the final O-antigen chain length is serotype-specific, but also the components of the O-antigen repeat unit. Modified heptoses present in the capsule of *C. jejuni* and *Y. pseudotuberculosis* have been shown to play a key role in virulence. Here we describe the structures of the C6 dehydratase DmhA, required for 6-deoxyheptose biosynthesis in *Y. pseudotuberculosis* O:2a, the C5/C3 epimerase MlghB and the C4 reductase MlghC, both required for *O*-methyl-L-*gluco*-heptose biosynthesis in *C. jejuni* strain NCTC 11168, and the C3 epimerase DdahB and the C4 reductase DdahC, both required for 6-deoxy-D-*altro*-heptose biosynthesis in *C. jejuni* strain 81-176. These structures and the mechanisms we propose will improve our understanding of the biosynthesis of these unique sugars.

II. Declarations

Candidate's declarations:

I, Laura Woodward, hereby certify that this thesis, which is approximately 40,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2013 and as a candidate for the degree of Doctor of Philosophy in September 2017; the higher study for which this is a record was carried out in the University of St Andrews between 2013 and 2017.

Date …….. Signature of candidate ……...

Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …….. Signature of supervisor ..……

3. Permission for publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Embargo on both the printed and electronic copy for a period of 2 years on the grounds that publication would preclude future publication.

Date …….. Signature of candidate …..……..

Date …..…….. Signature of supervisor ..……...

III. Contents

IV. List of figures

V. List of tables

VI. Abbreviations

ABC	ATP-binding cassette
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
Bis-tris	Bis(2-hydroxethyl)aminotris(hydroxymethyl)methane
BN	Blue native
BME	β-mercaptoethanol
BSA	Bovine serum albumin
CHES	2-(cylohexylamino)ethanesulfonic acid
CCD	Charge-coupled device
CCP4	Collaborative Computing Project Number 4
CFU	Colony forming units
CMC	Critical micelle concentration
CPS	Capsular polysaccharide
CSS	Complexation significance score
CV	Column volume
Da	Dalton
DCIA	7-diethyl-3-[4'-(iodoacetamido)phenyl]-4-methylcoumarin
DIALS	Diffraction integration for advanced light source
DID1	Dynein light chain-interacting domain 1
DME	Dimethyl ether
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphate
dRI	Differential refractive index
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
dTDP	Deoxythymidine diphosphate

dTMP	Deoxythymidine monophosphate
Dyn2	Yeast dynein light chain
ECA	Enterobacterial common antigen
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMDB	Electron microscopy data bank
FAB	Fragment antigen binding
FT	FT
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GlcNAc	*N*-acetylglucosamine
GST	Glutathione S-transferase
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
IPTG	Isopropyl-β-D-thiogalactopyranoside
IM	Inner membrane
ITC	Isothermal calorimetry
LB	Luria Bertani
LCP	Lipidic cubic phase
LDS	Lithium dodecyl sulfate
LLO	Lipid-linked oligosaccharide
LPS	Lipopolysaccharide
MALS	Multi-angle light scattering
MATE	Multidrug and toxic compound extrusion
MBP	Maltose binding protein
MES	4-morpholineethanesulfonic acid
MISTIC	Membrane-integrating sequence for translation of integral membrane constructs
MME	Monomethyl ether
Mn	Number average molecular weight

MOP	Multidrug/ oligosaccharyl-lipid/ polysaccharide transporter
MOPS	4-morpholinepropanesulfonic acid
MPD	2-methyl-2,4-pentandiol
MurNAc	*N*-acetylmuramic acid
MVF	Mouse virulence factor
Mw	Weight average molecular weight
NAD^+	β-nicotinamide adenine dinucleotide
NADH	β-nicotinamide dinucleotide reduced
$NADP^+$	β-nicotinamide adenine dinucleotide phosphate
NADPH	β-nicotinamide adenine dinucleotide phosphate reduced
NBD	Nucleotide binding domain
OD_{600}	Optical density measured at wavelength 600 nm
OECD	Organisation for Economic Co-operation and Development
OLF	Oligosaccharidyl-lipid flippase
OM	Outer membrane
OS	Oligosaccharide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline plus 5% (v/v) $Tween_{20}$
PCP	Polysaccharide co-polymerase
PCR	Polymerase chain reaction
PCT	Pre-crystallisation test
PDB	Protein data bank
PDC	Protein: detergent complex
PEG	Polyethylene glycol
PHPT	polyisoprenyl-phosphate hexose-1-phosphate transferase
PISA	Protein interfaces, surfaces and assemblies
PFO	Perfluorooctanoic acid
PNPT	polyisoprenyl-phosphate *N*-acetylaminosugar-1-phosphate transferase

PPI	Protein:protein interaction
PSI	Pounds per square inch
PVDF	Polyvinylidene difluoride
QPCR	Quantitative polymerase chain reaction
RT	Room temperature
RNA	Ribonucleic acid
SAXS	Small-angle X-ray scattering
SDM	Site-directed mutagenesis
SDR	Short chain dehydrogenase/reductase
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SPG	Succinic acid/ sodium Phosphate monobasic/ Glycine [ratio 2:7:7]
Strep	Streptactin
TBE	Tris-borate-ethylenediaminetetraacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TEM	Transmission electron microscope
TEV	Tobacco etch virus
TLR	Toll-like receptor
TLS	Translation, libration and screw-rotation
TLSMD	Translation, libration and screw-rotation motion determination
TMD	Transmembrane domain
TMH	Transmembrane helix
Tris	Tris(hydroxylmethyl)aminomethane
UDP	Uridine diphosphate
UV	Ultra-violet
WHO	World Health Organisation

Amino acids

Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartate/ aspartic acid
Cys	Cysteine
Glu	Glutamate/ glutamic acid
Gln	Glutamine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

Nucleic acids

A	Adenine
C	Cytosine
G	Guanine
T	Thymidine

Detergents and Lipids

C_8E_4	Octyl Tetraethylene Glycol Ether
$C_{12}E_8$	Polyoxyethylene(8)dodecyl Ether (ANAPOE-$C_{12}E_8$)
$C_{12}E_9$	Polyoxyethylene(9)dodecyl Ether (ANAPOE-$C_{12}E_9$)
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfate
Cym5	5-Cyclohexyl-1-Pentyl-β-D-Maltopyranoside (Cymal-5)
Cym5-NG	5-Cyclohexyl-1-Pentyl-β-D-Maltopyranoside Neopentyl Glycol
Cym6	6-Cyclohexyl-1-Hexyl-β-D-Maltopyranoside (Cymal-6)
Cym6-NG	6-Cyclohexyl-1-Hexyl-β-D-Maltopyranoside Neopentyl Glycol
DDG	n-Dodecyl-β-D-Glucopyranoside
DDM	n-Dodecyl-β-D-Maltopyranoside
DG	n-Decyl-β-D-Glucopyranoside
DM	n-Decyl-β-D-Maltopyranoside
DMNG	Decyl Maltose Neopentyl Glycol
DMPG	1,2-Dimyristoyl-sn-Glycero-3-[Phospho-rac-(1-Glycerol)]
DTG	n-Decyl-β-D-Thioglucopyranoside
DTM	n-Decyl-β-D-Thiomaltopyranoside
fos12	n-Dodecylphosphocholine (FOS-CHOLINE-12)
fos14	n-Tetradecylphosphocholine (FOS-CHOLINE-14)
Hega10	Decanoyl-N-Hydroxyethylglucamide
HTG	n-Heptyl-β-D-Thioglucopyranoside
LDAO	n-Dodecyl-N,N-Dimethylamine-N-Oxide
LMNG	Lauryl Maltose Neopentyl Glycol
LTM	n-Dodecyl-β-D-Thiomaltopyranoside
Mega10	Decanoyl-N-Methylglucamide
NDM	n-Nonyl-β-D-Maltopyranoside
NG	n-Nonyl-β-D-Glucopyranoside
NTG	n-Nonyl-β-D-Thioglucopyranoside
NTM	n-Nonyl-β-D-Thiomaltopyranoside
ODM	n-Octyl-β-D-Maltopyranoside
OG	n-Octyl-β-D-Glucopyranoside

OGNG	Octyl Glucose Neopentyl Glycol
OTG	n-Octyl-β-D-Thioglucopyranoside
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol
TDAO	n-Tetradecyl-N,N-Dimethylamine-N-Oxide
TDM	n-Tridecyl-β-D-Maltopyranoside
TetDM	n-Tetradecyl-β-D-Maltopyranoside
UDM	n-Undecyl-β-D-Maltopyranoside
UTM	n-Undecyl-β-D-Thiomaltopyranoside

VII. Acknowledgements

Firstly, I would like to thank Professor Jim Naismith for accepting me as a PhD candidate in his lab and for his expertise throughout. I would also like to thank the members of the "Sugar Team": Dr Stephen McMahon, Dr Audrey Le Bas, Dr Hui Liu and Dr Hexian Huang for a collaborative and supportive environment. Special thanks go to Dr Stephen McMahon for help during X-ray diffraction data collection, Dr Hui Liu for help during cloning and Dr Audrey Le Bas for constant discussion and advice. Special thanks also go to Dr Huanting Liu for support, advice and creating each vector I required. I would also like to thank Dr David Bhella, Centre for Virus Research MRC-University of Glasgow, for EM collaboration, Dr Carole Creuzenet, University of Western Ontario, for collaboration on the Heptose project and Professor Chris Whitfield, University of Guelph, for helpful advice and discussion. Thank you to Professor Martin Caffrey, Trinity College Dublin, for allowing me to train in his lab (training conducted by Dr Nicole Howe) and benefit from his LCP expertise. This training enabled us to implement this technique in our lab. I would also like to thank the remaining members of the Naismith Laboratory and BSRC annex level 2 for creating a collaborative environment, crucial to successful lab work.

My thanks go to the "Friday pub group": Dr Greg Mann, Dr Andrew Bent, Luana Ferrara, Miriam Weckener, Dr Emilia Oueis, Dr Rob Pengelly, Dr Laura Griffin, Dr C. David Owen, Dr Simon Dailey, Dr Joe Kenworthy and Dr Myles Garstang for their support and advice throughout.

Last but not least, my thanks go to my family; Sue, Peter, Rory and Ross, and my wonderful boyfriend Christoph, for their support and understanding throughout my PhD.

1 Introduction

Parts of this introduction, as well as the associated primary references, have been taken from the following review:

Woodward, L. & Naismith, J.H., 2016. Bacterial polysaccharide synthesis and export. *Current Opinion in Structural Biology*, 40, pp.81–88.

1.1 A comparison of the Gram-negative and Gram-positive cell envelopes

Bacteria have evolved numerous strategies to enable survival in a vast range of environments, the most challenging of which is arguably inside a host organism following infection. As the cell envelope provides the first point of contact between the bacteria and its environment, it provides a crucial barrier to protect the bacteria from the host environment (Raetz and Whitfield, 2002).

Bacteria fall into one of two classes, distinguished by Gram staining; first developed by the microbiologist Hans Christian Gram in 1884. This technique uses the dye crystal violet to differentiate bacteria based on the composition and arrangement of their cell envelope by detecting the presence of the polymer peptidoglycan (Wilhelm *et al*., 2015). Peptidoglycan consists of chains of *N*-acetylglucosamine-*N*-acetyl muramic acid (GlcNAc-MurNAc), cross-linked via peptide chains (Silhavy *et al*., 2010). Bacteria coated with peptidoglycan retain the dye in their peptidoglycan layer, and are therefore known as Gram-positive bacteria and those that do not retain the stain, due to the lack of peptidoglycan on the extreme cell surface, are known as Gram-negative bacteria (Wilhelm *et al*., 2015).

Gram-positive bacteria possess a cytoplasmic membrane, composed primarily of phospholipids, coated by a peptidoglycan layer, of 30-100 nm in thickness. This also contains lipoteichoic acid and teichoic acid polymers (Silhavy *et al*., 2010). Gram-negative bacteria possess a cytoplasmic membrane, known as the inner membrane (IM), again composed primarily of phospholipids, coated by a peptidoglycan layer, which, in contrast to Gram-positive bacteria, is only a few nm in thickness (Silhavy *et al*., 2010). Peptidoglycan-associated lipoproteins link the peptidoglycan layer to an additional asymmetric outer membrane, absent in Gram-positive bacteria (Silhavy *et al*., 2010). Like the Gram-negative IM, the inner leaflet of the outer membrane (OM) is composed of phospholipids, but the major component of the outer leaflet of the OM of Gram-negative bacteria is lipopolysaccharide (LPS) (Silhavy *et al*., 2010). Both Gram-positive and Gram-negative bacteria can also be enveloped by capsular polysaccharide (CPS). This consists of extremely long negatively-charged polysaccharide chains, forming a highly hydrated layer, anchored to the outer membrane by a phospholipid (Whitfield, 2006). Gram-negative enterobacteria are additionally coated with the enterobacterial common antigen (ECA) (Kuhn *et al*., 1988). These major differences in the components of the Gram-positive and Gram-negative cell envelopes are highlighted in figure 1.1.

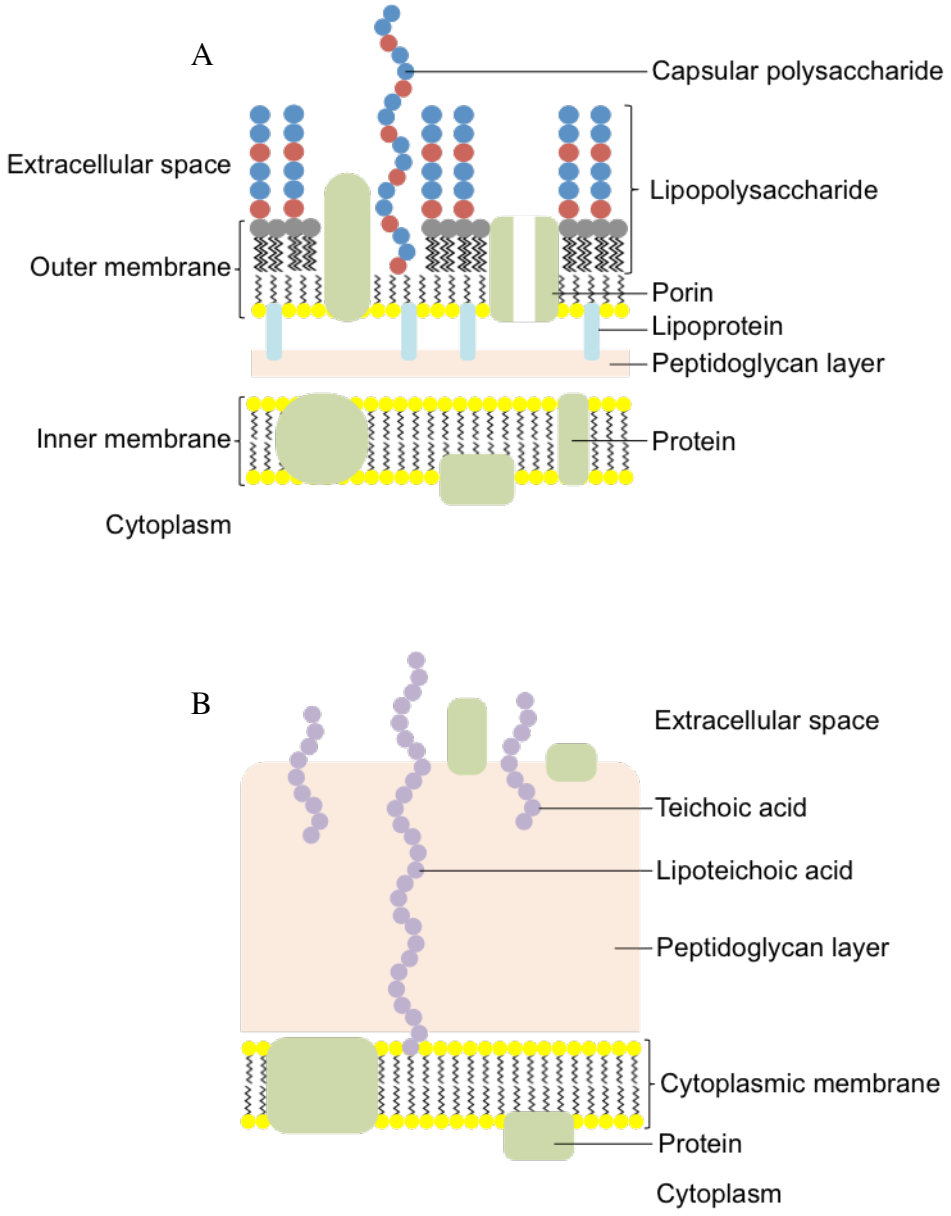


Figure 1.1 A simplified representation of the Gram-negative cell envelope and the Gram-positive cell envelope. The Gram-negative cell envelope is shown in "A" and the Gram-positive cell envelope is shown in "B".

1.2 LPS

1.2.1 LPS structure

As seen in figure 1.1, LPS is the major component of the outer leaflet of the OM of Gram-negative bacteria. LPS is composed of a lipid A anchor, and an inner and outer core, which together form the core oligosaccharide (core-OS), followed by the O-antigen (Raetz and Whitfield, 2002). Lipid A consists of a phosphorylated glucosamine disaccharide modified with acyl chains that function to anchor the molecule in the outer membrane (Raetz and Whitfield, 2002). The core-OS is non-repeating and relatively well conserved, due to the lack of exposure to environmental selective pressures (Raetz and Whitfield, 2002). The non-repeating core-OS consists of the inner core, which functions to attach the rest of the molecule to the lipid A anchor, and the outer core, which forms the attachment to the O-antigen chain (Raetz and Whitfield, 2002). The inner core contains two 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) and three heptoses, but the outer core is more variable (Raetz and Whitfield, 2002). As previously mentioned, the outer core is attached to the O-antigen chain. The O-antigen chain consists of polymerised chains of sugars, termed "repeat units", with each bacterium exhibiting a specific modal distribution of O-antigen chain lengths i.e. a specific modal distribution of polymerised repeat units (Whitfield, 2006). In contrast to the other components of LPS, the O-antigen is highly variable, with more than 170 different O-antigens found in different *Escherichia coli* (*E. coli*) alone (Whitfield, 2006). In addition, numerous K-antigens, which normally form CPS, can also be ligated to lipid A to form LPS (Whitfield, 2006). The structure of rough-form LPS (*E. coli*), which lacks the O-antigen and outer core present in full length LPS, taken from the crystal structure of rough-form LPS in complex with TLR4/MD-2, is shown in figure 1.2 (Park *et al*., 2009). The wide-variety of O-antigen structures is exemplified by figure 1.3.

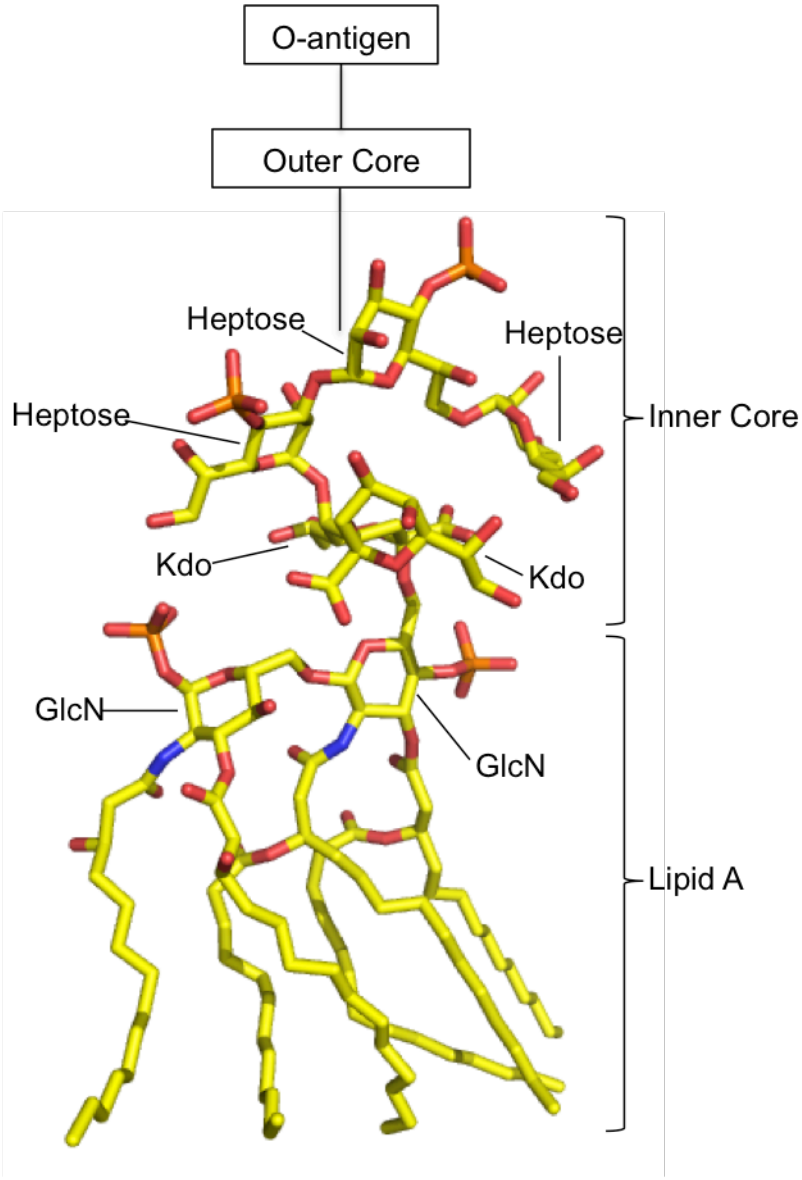


Figure 1.2. Crystal structure of *E. coli* EH100 rough-form LPS. Rough-form LPS is composed of lipid A and the inner core. Full-length LPS would have an O-antigen chain, composed of repeating units, linked to the inner core by an outer core-OS. This structure is taken from the TLR4/MD-2 rough-form LPS complex (PDB code: 3FXI) (Park *et al.*, 2009). This image was created using MacPyMOLEdu.

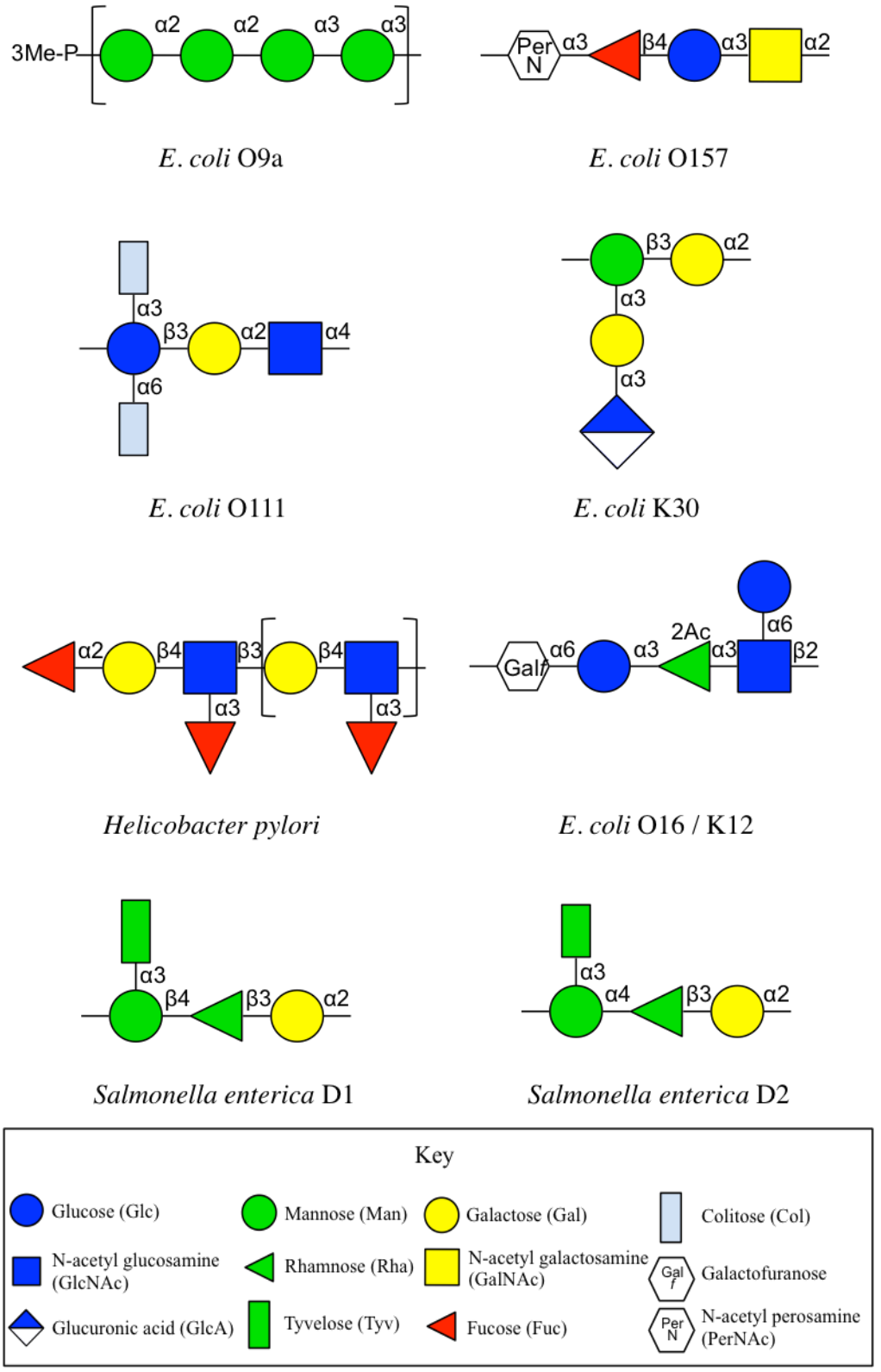


Figure 1.3. Gram-negative bacteria produce LPS with a wide-variety of O-antigen sugars (Hong and Reeves, 2014; King *et al*., 2014; Li *et al*., 2017) Carbohydrate structure database nomenclature used (Varki *et al*., 2015).

1.2.2 Role of LPS during infection of a host

Due to the surface-exposed nature of LPS, it is crucial to the evasion of the host immune response. The protective barrier function of LPS is highlighted by the finding that rough-form LPS mutants of the pathogen *Vibrio cholerae* (*V. cholerae*), so called because they produce a truncated LPS, possess a decreased ability to colonise the host intestine due to an increased susceptibility to components of the complement system and bile salts (Nesper *et al*., 2001). As well as functioning as a protective barrier, LPS can also function in the evasion of the host immune response via a mechanism known as "molecular mimicry" (Comstock and Kasper, 2006). An example of this is the O-antigen of the LPS of the gut pathogen *Helicobacter pylori* (*H. pylori*), seen in figure 1.3, which closely resembles the Lewis antigen, found on the surface of red blood cells. This similarity enables the establishment of a persistent infection as the *H. pylori* O-antigen is able to bind a C-type lectin found on the surface of dendritic cells, thereby preventing T-cell maturation (Bergman *et al*., 2004; Li *et al*., 2017).

1.3 Bacterial LLO biosynthesis

Given these crucial functions of just one component of the bacterial cell envelope during infection, it is not surprising that numerous antibiotics are known to target different stages of bacterial cell envelope biosynthesis, for example, penicillin and vancomycin are known to target transglycosylases and transpeptidases, respectively, required for the assembly of the peptidoglycan building blocks to form the completed peptidoglycan layer (Kohanski *et al*., 2010). Other antibiotics that target bacterial lipid-linked oligosaccharide (LLO) biosynthesis will be mentioned when discussing their particular targets. Increasing our understanding of the pathways for bacterial LLO biosynthesis is not only important for the development of novel antibiotics, but also important for the enzymatic synthesis of novel glycoconjugates, not accessible via chemical synthesis (Hug and Feldman, 2011).

As demonstrated in figure 1.3, bacteria produce a wide-range of LLOs. Despite this wide-range of possible structures, there are many commonalities between the biosynthesis pathways of these LLOs, indicating a shared evolution. Many of the proteins involved in these different pathways are homologous; therefore insights into each pathway can be gained by comparing with the other LLO biosynthesis pathways, allowing us to build a complete picture of LLO biosynthesis. Furthermore, these pathways share many parallels with human *N*-linked glycosylation (Hug and Feldman, 2011).

LLO biosynthesis can be reduced to the following sequence of events: initiation of biosynthesis on the inner leaflet of the IM via the attachment of the first sugar-1-phosphate onto a C_{55} undecaprenyl phosphate acceptor, flipping of the product to the periplasmic leaflet of the IM, repeat unit polymerisation/ chain extension, coupled with chain length regulation and transfer of the completed polymer to the appropriate acceptor, followed by its export to its final destination (Hug and Feldman, 2011). The various bacterial LLO biosynthesis pathways will be introduced briefly here, with steps that have been well characterised compared and contrasted in section 1.4.

1.3.1 Outline of Wzy-dependent LPS and CPS biosynthesis

There are two main mechanisms by which LPS and CPS are biosynthesised: the Wzy-dependent and ABC transporter-dependent mechanisms (Whitfield, 2006). The Wzy-dependent biosynthesis of both LPS and CPS begins on the inner leaflet of the IM with the transfer of a sugar-1-phosphate from its corresponding nucleotide-activated sugar onto undecaprenyl phosphate by a transferase enzyme (Whitfield, 2006). This initial sugar is then built upon by the sequential action of various membrane-associated glycosyltransferases to form the repeat unit (Whitfield, 2006). Wzx flips the undecaprenyl-pyrophosphate-linked repeat unit to the periplasmic face of the IM, where the repeat units are polymerised by the O-antigen polymerase Wzy (Whitfield, 2006). At this point, the biosynthesis of LPS and CPS diverges. During LPS biosynthesis, the final O-antigen chain length is regulated by Wzz, prior to the transfer of the final O-

antigen chain to lipid A by the O-antigen ligase WaaL, and its subsequent export to the outer membrane by the Lpt family of proteins (Whitfield, 2006). During Wzy-dependent CPS biosynthesis, chain length is regulated by Wzc and then the completed polysaccharide is exported to the outer membrane by Wza (Whitfield, 2006). Wzi is then believed to bind the polysaccharide as it emerges from Wza and is believed to play a role in capsule assembly (Bushell *et al*., 2013). Following the completion of both LPS and CPS biosynthesis, the remaining undecaprenyl pyrophosphate is dephosphorylated by the pyrophosphatase PgpB (Fan *et al*., 2014). Figures 1.4 and 1.5 show the Wzy-dependent LPS and CPS biosynthesis pathways, respectively.

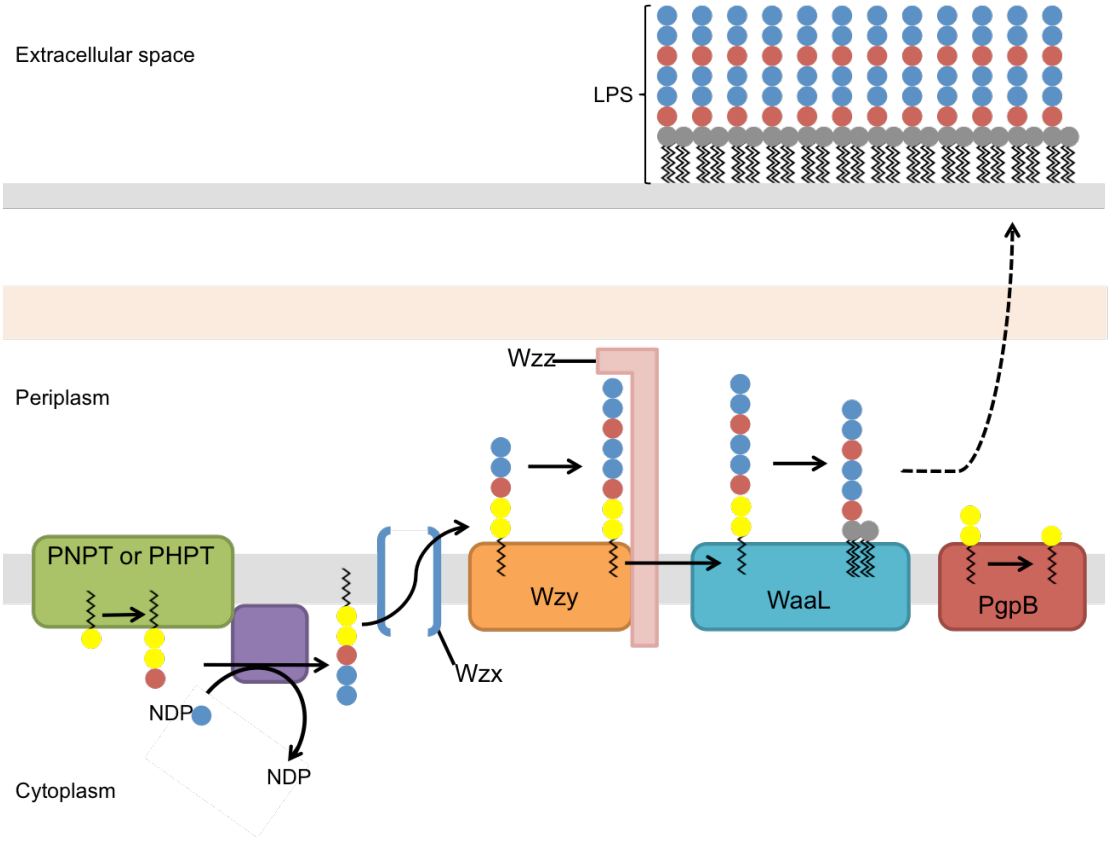


Figure 1.4. A simplified schematic of Wzy-dependent LPS biosynthesis. LPS biosynthesis occurs on the inner membrane. A sugar-1-phosphate or an *N*-acetylsugar-1-phosphate is transferred by a polyisoprenyl-phosphate hexose-1-phosphate transferase (PHPT) or a polyisoprenyl-phosphate N-acetylaminosugar-1-phosphate transferase (PNPT), respectively, onto undecaprenyl phosphate. This is then built upon by membrane-associated glycosyltransferases to form the repeat unit. Wzx flips this across the inner membrane to the periplasm, where Wzy polymerises the repeat units. This is under the control of Wzz. Once the O-antigen chain is completed, it is transferred onto lipid A by WaaL. The completed LPS is transported to the outer membrane and PgpB recycles undecaprenyl pyrophosphate to undecaprenyl phosphate (Hug and Feldman, 2011). The inner membrane and the periplasmic leaflet of the outer membrane are indicated with a grey line and the periplasmic peptidoglycan layer is indicated with a peach line. The sugars that form a representative O-antigen are shown with blue and red circles. Undecaprenyl is represented by a single zigzag and a phosphate group is represented by a yellow circle. Lipid A is represented by two grey circles and four zigzag lines.

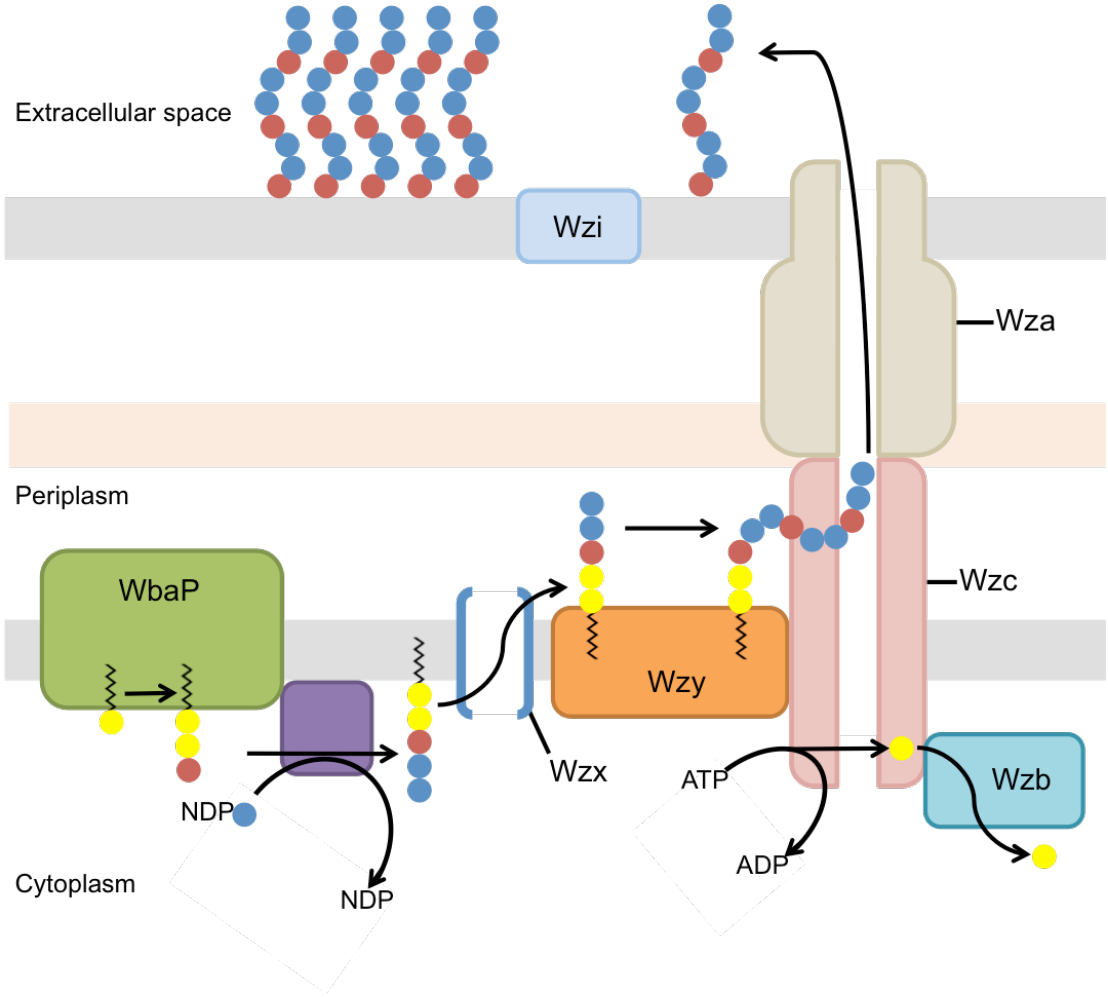


Figure 1.5. A simplified representation of Wzy-dependent CPS biosynthesis. WbaP transfers galactose from UDP-galactose (represented by a red circle) on to undecaprenyl phosphate in the inner leaflet of the inner membrane. Undecaprenyl is represented by a single zigzag and phosphate groups are represented by a yellow circle. Membrane-associated glycosyltransferases complete the repeat unit (represented by blue circles) and Wzx flips this in to the periplasmic leaflet of the inner membrane where Wzy polymerises these repeat units. The tyrosine autokinase Wzc regulates the final repeat unit chain length by an unknown mechanism, with Wzb functioning as a phosphatase to regulate Wzc function. Wzc also functions to couple CPS biosynthesis on the inner membrane to its export by Wza. Wzi is thought to play a role in the assembly of the final capsule (Whitfield, 2006). The inner and outer membranes are shown by a grey line and the periplasmic peptidoglycan layer is shown in peach.

1.3.2 Outline of ABC-transporter dependent LPS biosynthesis

During ABC transporter-dependent LLO biosynthesis, shown in figure 1.6, the complete O-antigen chain is biosynthesised on the inner leaflet of the IM. The archetype for the ABC-transporter dependent pathway is the O9a antigen from *E. coli*, the repeat unit of which is shown in figure 1.3. Like the Wzy-dependent biosynthesis of the O-antigen, the biosynthesis begins with the transfer of GlcNAc from UDP-GlcNAc to undecaprenyl-phosphate by WecA (Raetz and Whitfield, 2002), however this mechanism involves the stepwise addition of single sugars to the undecaprenyl phosphate carrier to form the complete O-antigen, rather than the polymerisation of repeating units. WbdC and WbdB catalyse the addition of three mannoses to the initial GlcNAc (Greenfield *et al*., 2012). WbdA then extends this polymannose chain via the stepwise addition of single mannoses, with the final chain length controlled by WbdD (Clarke *et al*., 2004; Greenfield *et al*., 2012; Hagelueken *et al*., 2015). This completed O-antigen is then transferred to lipid A by WaaL O-antigen ligase and transported to the outer membrane by the Lpt family of proteins (Whitfield, 2006).

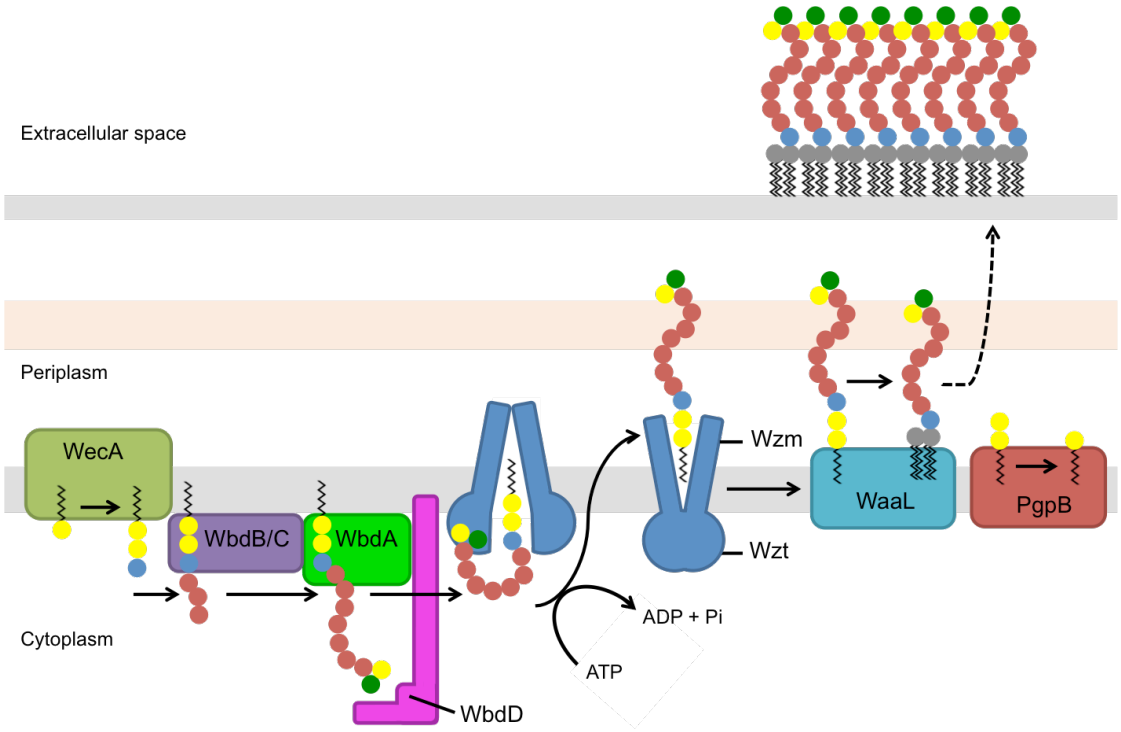


Figure 1.6. A simplified schematic of ABC transporter-dependent LPS biosynthesis. WecA transfers GlcNAc (represented by a blue circle) from UDP-GlcNAc onto undecaprenyl phosphate (undecaprenyl is represented by a single zigzag and phosphate groups are represented by a yellow circle). WbdB and WbdC then catalyse the addition of a total of three mannose residues (each represented by a red circle) to the initial GlcNAc. WbdA, shown in green, then sequentially adds mannose residues, with the length of the mannose chain controlled by WbdD, shown in pink. WbdD then phosphorylates and methylates (represented by a green circle) the terminal mannose to terminate WbdA-catalysed polymerisation. The completed polymer is then flipped to the periplasmic leaflet of the inner membrane by the ABC transporter composed of Wzt and Wzm. WaaL then transfers this completed polymer onto lipid A (represented by two grey circles and four zigzag lines). The completed LPS is exported to the outer membrane. PgpB recycles the undecaprenyl pyrophosphate to undecaprenyl phosphate (Fan *et al*., 2014; Hagelueken *et al*., 2012; King *et al*., 2014). The grey lines represent the inner membrane and periplasmic leaflet of the outer membrane and the periplasmic peptidoglycan layer is shown in peach.

1.3.3 Outline of peptidoglycan layer biosynthesis

During the biosynthesis of the peptidoglycan layer, the sequential action of the cytoplasmic enzymes MurA through to MurF forms the initial UDP-MurNAc pentapeptide (Laddomada *et al*., 2016). The integral membrane protein MraY then transfers the phosphoMurNAc-pentapeptide to undecaprenyl-phosphate; the same lipid carrier used for LPS and CPS biosynthesis (Chung *et al*., 2013). MurG then catalyses the addition of GlcNAc to the MurNAc of the undecaprenyl-phosphate-phosphoMurNAc-pentapeptide to form lipid II (Ha *et al*., 2000). MurJ flips lipid II across the inner membrane into the periplasm where the GlcNAc-MurNAc-pentapeptide is polymerised to form the final peptidoglycan layer (Kuk *et al*., 2016; Laddomada *et al*., 2016). A simplified view of the membrane-associated steps of peptidoglycan biosynthesis can be seen in figure 1.7.

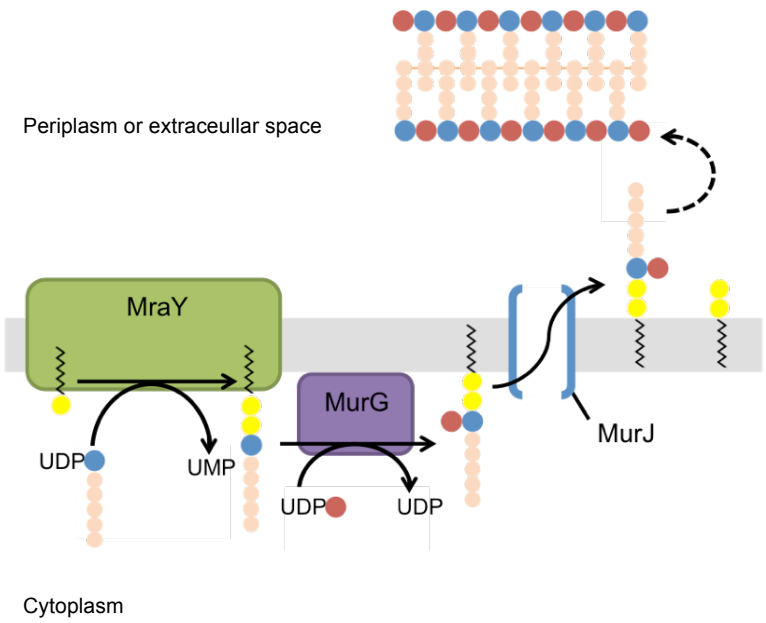


Figure 1.7. A simplified representation of peptidoglycan layer biosynthesis. On the inner leaflet of the inner membrane, MraY catalyses the addition of phosphoMurNAc pentapeptide to undecaprenyl phosphate from UDP-MurNAc pentapeptide. MurG then adds GlcNAc to the MurNAc residue to form lipid II, which MurJ flips across the inner membrane into the periplasm, where it is assembled to form the final peptidoglycan layer (Ha *et al*., 2000; Kuk *et al*., 2016; Laddomada *et al*., 2016). The inner membrane is shown in grey.

1.3.4 Outline of bacterial *N*-linked glycosylation

Previously believed to be confined to eukaryotes, *N*-linked glycosylation has been identified in prokaryotes (Hug and Feldman, 2011). The *N*-linked glycosylation pathway found in *Campylobacter jejuni* (*C. jejuni*) (shown in figure 1.8) not only shares many parallels with the biosynthesis of other LLOs, but also the eukaryotic *N*-linked glycosylation pathway present in the endoplasmic reticulum (Hug and Feldman, 2011). The function of *N*-linked glycosylation of bacterial proteins is often unclear, but protein *N*-linked glycosylation has been shown to be crucial for the colonisation of mice intestine by *C. jejuni* strain 81-176 (Szymanski *et al*., 2002). PglC catalyses the initial transfer of bacillosamine-1-phosphate from UDP-bacillosamine to the conventional undecaprenyl phosphate acceptor (Glover *et al*., 2006). This is then built upon by the sequential action of the glycosyltransferases PglA, PglJ, PglH and finally PglI to form the final undecaprenyl pyrophosphate-linked heptasaccharide (Glover *et al*., 2006). The ABC transporter PglK flips this completed LLO across the inner membrane and transfers it to the oligosaccharyltransferase PglB (Perez *et al*., 2015). The oligosaccharyltransferase PglB transfers the heptasaccharide to asparagine side chains of proteins (Lizak *et al*., 2011) with the following amino acid acceptor motif: D/E-Y-N-X-S/T, where X corresponds to any amino acid except P (Kowarik et al., 2006).

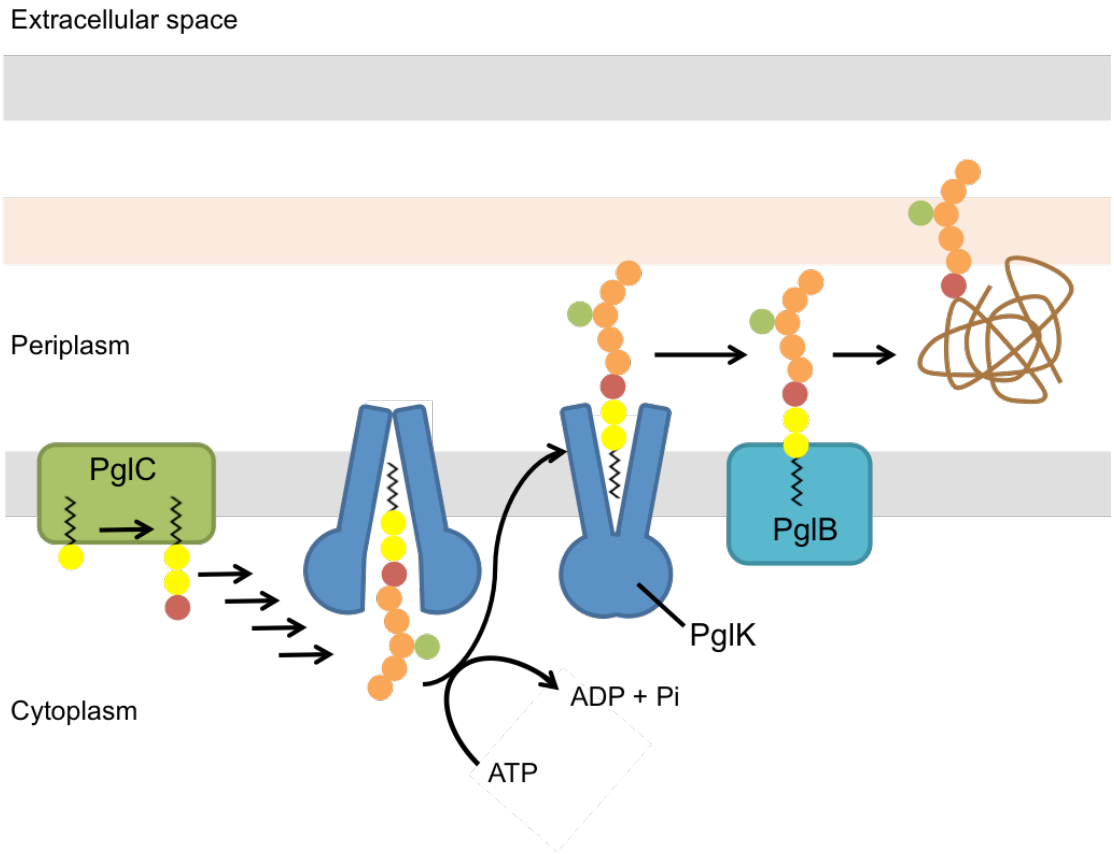


Figure 1.8. Simplified representation of bacterial *N*-linked glycosylation. PglC transfers bacillosamine-1-phosphate from UDP-bacillosamine (bacillosamine is represented by a red circle and phosphate groups are represented by a yellow circle) onto undecaprenyl phosphate (undecaprenyl is represented by a single zigzag). The ABC transporter PglK flips the completed undecaprenyl-pyrophophate-linked N-glycan precursor (shown in orange and green) across the inner membrane, where PglB transfers this on to the appropriate acceptor protein (Hug and Feldman, 2011; Lizak *et al*., 2011; Perez *et al*., 2015). The inner and outer membranes are shown in grey and the periplasmic peptidoglycan layer is shown in peach.

1.4 Stepwise comparison of the various bacterial LLO biosynthesis pathways

As highlighted by the brief introduction of each LLO biosynthetic pathway, these biosynthetic pathways follow a common theme and knowledge of one pathway can often be applied to aid our understanding of another. These LLO biosynthesis pathways will now be discussed in further detail by comparing and contrasting each step of their biosynthesis, focussing on those proteins that have been structurally characterised.

1.4.1 Initiation of LLO biosynthesis

The first step in lipid-linked oligosaccharide (LLO) biosynthesis is the formation of the sugar lipid conjugate that acts as the sugar carrier. This begins with the transfer of sugar-1-phosphate to undecaprenyl phosphate on the cytoplasmic face of the inner membrane from the corresponding nucleotide-activated sugar (Hug and Feldman, 2011). Gram-negative bacteria have two broad classes of initiating enzyme, the polyisoprenyl-phosphate hexose-1-phosphate transferase (PHPT) family and polyisoprenyl-phosphate *N*-acetylaminosugar-1-phosphate transferase (PNPT) family (Hug and Feldman, 2011). Humans only have PNPT-type enzymes. The nomenclature reflects an earlier understanding that PNPT enzymes used only UDP-GlcNAc and PHPT enzymes utilised UDP-galactose or other 'simple' sugars. A more useful, structural-based classification has since emerged; the PNPT class contains 10 or 11 transmembrane helices (TMHs), whilst the PHPT class has fewer TMHs and a cytoplasmic soluble domain containing the catalytic machinery (Hug and Feldman, 2011; Wang *et al*., 1996).

The first structure of a PNPT family member was solved by Chung *et al.* in 2013 and was that of the phosphoMurNAc pentapeptide translocase MraY, which initiates the process of peptidoglycan formation by transferring phosphoMurNAc pentapeptide onto undecaprenyl phosphate (Chung *et al*., 2013). Its crystal structure (from *Aquifex aeolicus*) reveals that it possesses 10 TMHs, with both N- and C- termini on the

periplasmic face of the inner membrane (Chung *et al*., 2013). The protein is found as a dimer in the crystal with interactions between helix 7 and 10 around a two-fold axis normal to the bilayer (Chung *et al*., 2013). The active site was identified, as expected, to be on the cytoplasmic surface with residues from helices 3, 4, 5 and as well as cytoplasmic loop E contributing to its architecture (Chung *et al*., 2013). TMH 9 is unusual in that it has a kink that breaks the helix into a short 9a and a longer 9b, which is almost parallel to the membrane (Chung *et al*., 2013). The unusual arrangement may reflect the complex donor substrate, which is predicted to bind here (Chung *et al*., 2013). The structure of the complex between MraY and its inhibitor muraymicin shows profound changes in the helix 9b and loop E as they bind to the peptidic component of muraymicin, shown in figure 1.9 (Chung *et al*., 2016). MraY is an atypical PNPT, in that it transfers a much larger and more complex initial substrate than other PNPTs, exemplified by WecA. WecA is predicted to have a periplasmic N-terminus and cytoplasmic C-terminus and consequently an odd number of helices. The PNPT family is the target of the antibiotic tunicamycin, which is a UDP-GlcNAc analogue (Chung *et al*., 2016; Hug and Feldman, 2011). The PHPT class of enzymes shows significant variation outside the conserved catalytically active cytoplasmic domain. WbaP, which transfers galactose-1-phosphate from UDP-galactose onto undecaprenyl phosphate, exemplifies the most common form of PHPT (Wang *et al*., 1996). WbaP-type PHPT topology predictions differ, with the number of predicted TMHs varying from 5 to 7 (Furlong *et al*., 2015). The topological study of the WbaP homologue WcaJ, the glucosyl-1-phosphate transferase required for the transfer of glucose from UDP-glucose to undecaprenyl during colanic acid biosynthesis, using a range of techniques found that WcaJ has a cytoplasmic N-terminus, four long TMHs followed by two shorter helices, giving six TMHs in total, with a large cytoplasmic loop located between TMHs 4 and 5 (Furlong *et al*., 2015). A large cytoplasmic domain is found at the C-terminus, which forms the catalytic domain (Furlong *et al*., 2015; Wang *et al*., 1996). The PHPT PglC, which initiates the biosynthesis of the glycan precursor required for *N*-linked glycosylation in *C. jejuni*, is predicted to have a periplasmic N-terminus and a single TMH (Hartley *et al*., 2011). The PHPT class of enzymes remains to be structurally characterised.

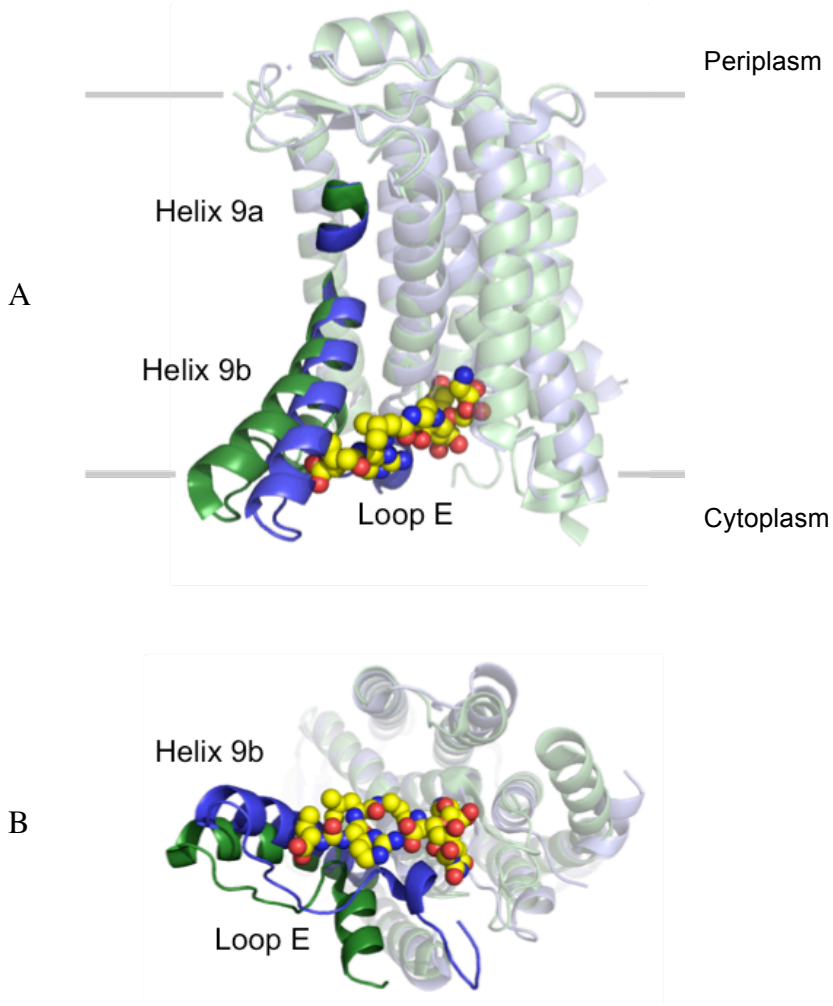


Figure 1.9. Overlay of the apo structure of MraY (green) with the muraymicin-bound structure (blue), with muraymicin shown in yellow. The inhibitor muraymicin binds MraY at the cytoplasmic face of the inner membrane and causes a shift in helix 9b and a conformational change in loop E upon binding. A sideview can be seen in "A" and a view looking from the cytoplasm is shown in "B". The grey lines indicate the approximate positions of the membrane boundary (apo structure PDB: 4J72; muraymicin-bound PDB: 5CKR) (Chung *et al*., 2013, 2016). This image was made using MacPyMOLEdu.

1.4.2 Transfer of LLOs across the inner membrane

In order to cross the membrane, the polar surface of the sugar must be masked from the lipid. There are believed to be two different mechanisms by which this is achieved, distinguished by the requirement of ATP (Whitfield, 2006).

1.4.2.1 ABC transporter-dependent LLO flipping

During the ABC transporter-dependent biosynthesis of CPS on the cytoplasmic face of the inner membrane, the lipid-linked oligosaccharide is flipped into the periplasmic face of the IM by the ABC transporter Wzt/Wzm, with Wzm forming the channel and Wzt forming the nucleotide binding domain (Whitfield, 2006). This protein complex has yet to be structurally characterised, however homologues have been structurally characterised.

The ABC transporter PglK is required for LLO flipping during bacterial *N*-linked glycosylation (Hug and Feldman, 2011). PglK is a homodimer with the typical ABC transporter fold and the structure of the *C. jejuni* homologue has been solved to a resolution of 2.9 Å for the apo structure, in an open inward-facing conformation and 5.9 Å for the outward occluded structure (Perez *et al*., 2015). These structures are shown in figure 1.10a and 1.10b, respectively. The ATP-dependent LLO flipping of PglK has been confirmed *in vitro* using wild type PglK, and various mutants, reconstituted in liposomes along with a radiolabelled LLO, with a shorter polysaccharide chain than the native substrate (Perez *et al*., 2015). The structural characterisation of numerous conformations of this protein has allowed the proposal of a mechanism for PglK-catalysed ATP-dependent LLO flipping, which is as follows: ADP-bound PglK interacts with the undecaprenyl tail of the LLO. ADP is then exchanged for ATP and the polysaccharide component of the LLO enters the arginine-lined cavity formed between the two monomers. ATP hydrolysis in the cytoplasmic NBDs then causes a "scissor-like" conformational change, closing the cavity formed between the two monomers, thereby forcing the polysaccharide chain into the periplasm (Perez *et al*., 2015). This conformational change to the outward-facing occluded form, shown in

figure 1.10b, would prevent the LLO returning to the cytoplasmic face of the inner membrane. This mechanism requires no specific interactions between PglK and the polysaccharide component of the LLO, which would go someway to explaining the observed substrate promiscuity of PglK (Perez *et al*., 2015). A simplified schematic of PglK-dependent LLO flipping can be seen in figure 1.10c. This mechanism remains unconfirmed, but a substrate-bound structure of PglK would help to confirm some aspects of the proposed mechanism (Perez *et al*., 2015).

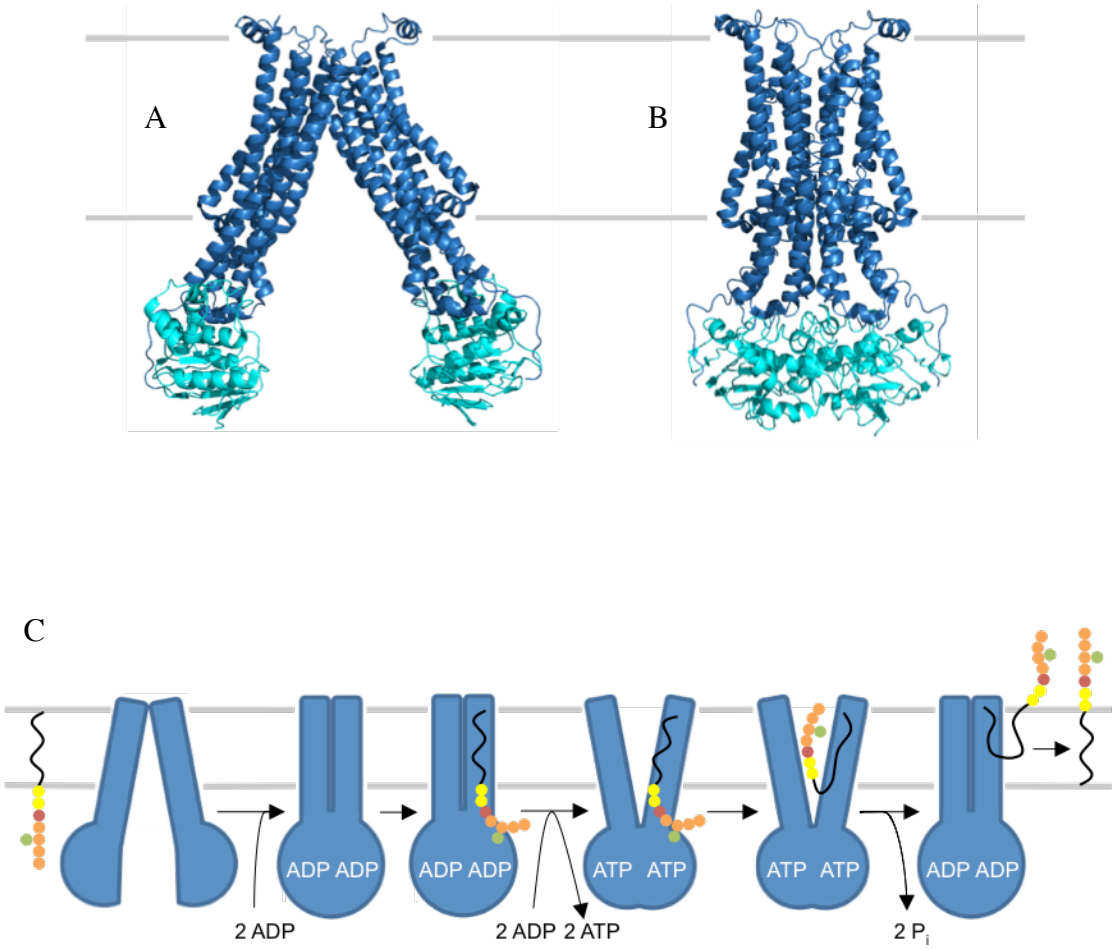


Figure 1.10. The structures of the ABC transporter PglK in different conformations and mechanism proposed based on these conformations. PglK in the open conformation (PDB: 5C73) is shown in "A" and PglK in an outward occluded conformation (PDB: 5C78) is shown in "B", with the TMHs shown in blue and the NBDs shown in cyan (Perez *et al*., 2015). The approximate boundaries of the membrane are shown in grey. The proposed mechanism for PglK-catalysed flipping is shown in "C". The images in "A" and "B" were created using MacPyMOLEdu.

1.4.2.2 ATP-independent LLO flipping

MurJ is the flippase required to flip lipid II from the cytoplasmic leaflet of the inner membrane to the periplasmic leaflet, during the biosynthesis of the peptidoglycan layer (Kuk *et al*., 2016). The inward-facing structure of this multidrug/ oligosaccharyl-lipid/ polysaccharide transporter (MOP) superfamily member, from the hyperthermophile *Thermosipho africanus*, was solved by X-ray crystallography from crystals grown in lipidic cubic phase (LCP), in the presence of substrate, to a resolution of 2.0 Å and is shown in figure 1.11a (Kuk *et al*., 2016). MurJ has 14 TMHs, with TMHs 1-5 forming the N-terminal lobe, TMHs 6-12 forming the C-terminal lobe, and the final two TMHs forming a hydrophobic groove (Kuk *et al*., 2016). A cationic cleft is located at the interface of the N- and C-terminal lobes (Kuk *et al*., 2016). In addition, a potential substrate binding site was identified using docking studies coupled with site-directed mutagenesis (SDM); this was found to be in agreement with the position of unassigned electron density in the structure, allowing the tentative modelling of the pentapeptide moiety (Kuk *et al*., 2016). Taken together these data enabled the proposal of a mechanism for lipid flipping by MurJ, in that the central cationic cavity traps the negatively charged substrate region. This binding would then, by an unknown mechanism, trigger a conformational change from inward to outward facing, thought to involve a hinge in TMH 7 at the interface between the two lobes, thereby releasing lipid II at the periplasmic side of the membrane (Kuk *et al*., 2016). The undecaprenyl moiety is proposed to remain shielded during this process in the hydrophobic cleft formed by the final two TMHs (Kuk *et al*., 2016). It is not known whether the action of this flippase requires the sodium- or proton- motive force (Kuk *et al*., 2016).

The MOP superfamily can be further subdivided into the mouse virulence factor (MVF), of which MurJ is a member, the multidrug and toxic compound extrusion (MATE), and the polysaccharide transporter and the oligosaccharidyl-lipid flippase (OLF) families (Hvorup *et al*., 2003). The flippase required during the biosynthesis of LPS is known as Wzx and this is a member of the MATE subclass of the MOP superfamily. No crystal structure of a Wzx homologue has been reported thus far, however it is predicted to be similar to that of the MATE family member NorM from *V.*

cholerae, which has been solved via X-ray crystallography to 3.65 Å resolution, in an outward-facing conformation, unlike the inward-facing MurJ structure (He *et al*., 2010; Kuk *et al*., 2016). The structure of NorM can be seen in figure 1.11b and it consists of 2 lobes, one of which is composed of TMHs 1-6 and the other TMHs 7-12 (He *et al*., 2010). These lobes are related by two-fold symmetry, with a cavity, lined with three conserved acidic residues, at the interface of the two lobes. This symmetry is in contrast to the lack of symmetry seen in the MurJ structure, seen in figure 1.11b (He *et al*., 2010; Kuk *et al*., 2016). Mutations in these conserved acidic residues have been found to disrupt the substrate/ Na^+ antiport relationship as they perturb cation binding (Jin *et al*., 2014). The outward-facing structure of NorM supports the following mechanism for compound export in the case of NorM, or LLO flipping in the case of Wzx: binding of cations to the outward-facing conformation triggers the switch to an inward-facing conformation, amenable to substrate binding. Substrate binding would then result in a conformational change to the outward-facing conformation, by an unknown mechanism, allowing substrate release (He *et al*., 2010). A schematic of this proposed mechanism can be seen in figure 1.11c. However, this proposed mechanism remains unproven and it remains unclear how Wzx flips such a large negatively charged substrate and which part of the LLO is recognised by Wzx. A high-resolution structure would go someway to answering these questions.

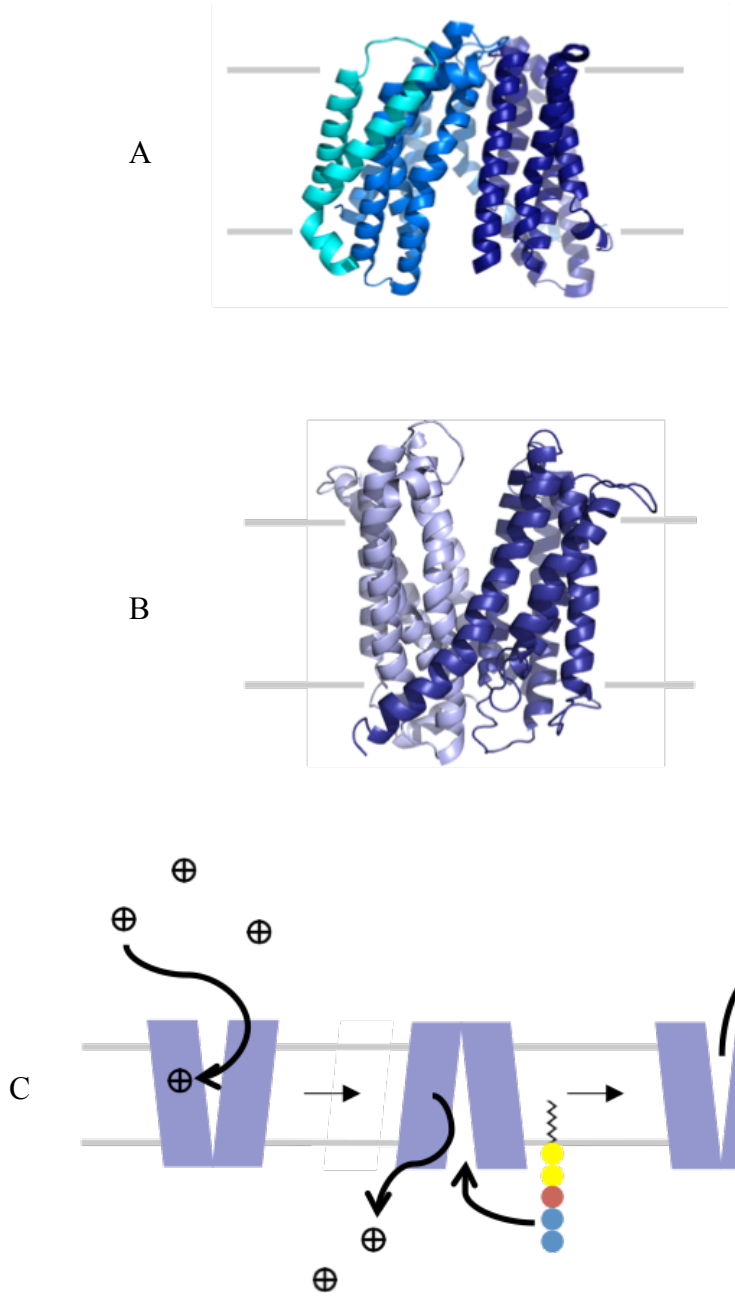


Figure 1.11. ATP-independent LLO flipping. The structure of MurJ, the ATP-independent flippase involved in peptidoglycan layer biosynthesis, is shown in "A" (PDB: 5T77) (Kuk *et al*., 2016). The N-terminal lobe, formed by TMHs 1-5 is shown in dark blue and the C-terminal lobe, formed by TMHs 6-12 is shown in mid blue and the hydrophobic groove, formed by TMHs 13 and 14, is shown in cyan. The structure of MATE family member NorM (PDB: 3MKT) is shown in "B". The proposed mechanism for Wzx-dependent LLO flipping, based on the structure of NorM is shown in "C" (He *et al*., 2010). The N-terminal lobe, consisting of TMHs 1.6, is shown in purple and the C-terminal lobe, consisting of TMHs 7-12, is shown in lilac. The approximate membrane boundaries are shown in grey. "A" and "B" were created using MacPyMOLEdu.

1.4.3 Repeat unit polymerisation and chain length regulation

The mechanism of sugar polymerisation, i.e. glycosidic bond formation, falls into one of two mutually exclusive categories. It can either be "processive" in that the chain extension occurs in one active site, with the product of the reaction released once the final chain length is reached or it can be "distributive" in that the polymerase possesses two active sites, between which the growing chain is transferred, with the product being released after each polymerisation reaction (Levengood *et al*., 2011). Two main models have been proposed to explain the regulation of polymer chain length, which is crucial in the absence of template. These are known as the "molecular clock" (Bastin *et al*., 1993) and "molecular ruler" models (Kintz and Goldberg, 2011). As the names suggest, a "molecular clock" would control the length of time during which polymerisation can take place (Bastin *et al*., 1993) and the "molecular ruler" provides a physical guide to polymerisation (Kintz and Goldberg, 2011). As the processes of polymerisation and chain length regulation are closely linked, they will be discussed in parallel.

1.4.3.1 *E. coli* O9a LPS polymerisation and chain length regulation

The first structurally characterised example of a "molecular ruler", WbdD, can be found in the ABC transporter-dependent biosynthesis of the *E. coli* O9a LPS, shown in figure 1.6. The O-antigen chain of the *E. coli* O9a LPS is shown in figure 1.3. The chain length regulator WbdD is composed of an N-terminal methyltransferase domain, followed by a eukaryotic kinase domain and a coiled-coil domain at the C-terminus, with the overall structure akin to an "umbrella" (Hagelueken *et al*., 2012, 2015). The termination of polymannose chain extension by the polymerase WbdA occurs by the addition of a phosphate group to the 3-OH of the terminal mannose by WbdD. The phosphate group is subsequently methylated (Clarke *et al*., 2004). This modification of the terminal mannose is essential for the export of the final O-antigen chain to the OM as it is recognised by the carbohydrate-binding module of the nucleotide binding protein Wzt, which forms part of the ABC transporter required for flipping of the LLO from the cytoplasmic face of the IM to the periplasmic face (Cuthbertson *et al*., 2005, 2007).

Despite knowing that this modification of the terminal mannose residue provides the "switch" from polymerisation to flipping, the mechanism of chain length regulation i.e. when this "switch" is flipped, resulting in a modal distribution of chain lengths, was discovered much more recently, with two methods found to influence chain length. WbdD has been found to perform the role of "molecular ruler" (Hagelueken *et al*., 2012, 2015) and additionally the role of a "molecular scaffold" (Clarke *et al*., 2009), introducing another possible chain length regulation mechanism. The coiled-coil domain of WbdD forms this "molecular ruler" (Hagelueken *et al*., 2012, 2015) and the base of the ruler is proposed to act as a "scaffold" to recruit the mannose polymerase WbdA to the cytoplasmic leaflet of the IM (Clarke *et al*., 2009). The modification of the length of this coiled-coil domain, modifies the length of the polymannose chain formed by WbdA (Hagelueken *et al*., 2015). Changes in the length of the coiled-coil domain result in changes in the distance between the chain extension reaction, catalysed by WbdA at the surface of the cytoplasmic leaflet of the IM, and the termination reaction catalysed by the kinase-methyltransferase domain of WbdD (Hagelueken *et al*., 2015). King *et al.* found that by varying the stoichiometry of this complex, by varying expression levels *in vivo*, they could vary the length of polymannose chain formed, with chain length found to increase with increasing WbdA expression and, conversely, decrease with increasing WbdD expression (King *et al*., 2014). In their model, King *et al.* propose that the more WbdA proteins that are associated with the WbdD scaffold, the more occluded the bifunctional kinase-methyltransferase domain becomes, allowing polymerisation of mannose moieties by WbdA to continue for longer, resulting in, on average, longer mannose chain lengths (King *et al*., 2014). Conversely, the fewer WbdA proteins that are associated with WbdD, the more accessible the kinase-methyltransferase domain becomes, allowing the polymannose chain to take the shortest possible route to termination (King *et al*., 2014). A proposed arrangement for the WbdA:WbdD complex can be seen in figure 1.12, however this is yet to be proven (King *et al*., 2014).

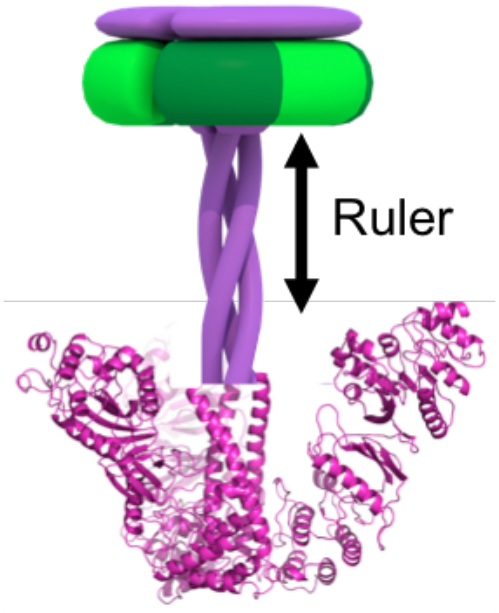


Figure 1.12. The proposed arrangement of the WbdA:WbdD complex. Membrane-associated WbdA is shown in green and proposed arrangement of the coiled-coil of WbdD in purple. The cytoplasmic domain of WbdD is shown in pink (taken from Woodward & Naismith 2016).

1.4.3.2 Polymerisation and chain length regulation during Wzy-dependent LPS and CPS biosynthesis

In Wzy-dependent LPS and CPS biosynthesis, undecaprenyl-linked sugar repeat units are polymerised at the periplasmic face of the IM by the O-antigen polymerase Wzy, with the number of repeat units polymerised controlled by the polysaccharide co-polymerase 1 (PCP-1) family member Wzz, resulting in an organism-specific modal distribution of polymer lengths (Morona *et al*., 2009). Using radiolabelled substrates *in vitro*, Woodward and colleagues were able to monitor purified *E. coli* O86 Wzy activity and found that O-antigen chain length modality could be restored purely upon the addition of Wzz to the reaction mixture (Woodward *et al*., 2010). Wzy is a predicted integral membrane protein with between 10-13 TMHs (Daniels *et al*., 1998), but it has has not been structurally characterised. Although the mechanism of Wzy-dependent repeat unit polymerisation remains unknown, it has been shown to be exquisitely

specific. *Salmonella enterica* D1 and D3 O-antigens are shown in figure 1.3. Despite these O-antigens only differing in the type of linkage between two sugars, different Wzy homologues are required to polymerise these O-antigen repeat units (Hong *et al*., 2015). Furthermore, one homologue cannot complement for the loss of the other *in vivo* (Hong *et al*., 2015). An interaction between Wzz and Wzy in *Shigella flexneri* has been detected via an *in vivo* cross-linking approach, although this remains to be further characterised (Nath and Morona, 2015a).

Despite its function in regulating O-antigen chain length, the presence of Wzz is not essential for O-antigen polymerisation (Woodward *et al*., 2010). Bacteria lacking in Wzz produce LPS with a much shorter average O-antigen chain length and a random distribution of O-antigen chain lengths (Murray *et al*., 2003). Wzz is a two-pass integral membrane protein, with the TMHs located at the termini and separated by an α-helical domain that stretches approximately 100 Å into the periplasm (Kalynych *et al*., 2015; Tocilj *et al*., 2008). Structures of the periplasmic domains of various Wzz homologues have been solved, with all sharing the same bell shape, despite only sharing 25 % sequence identity on average (Kalynych *et al*., 2015; Tocilj *et al*., 2008). Interestingly, Wzz homologues from different organisms show different oligomerisation states (Tocilj *et al*., 2008). The periplasmic domain of Wzz_{FepE} from *E. coli* O157:H7 has been solved to a resolution of 2.7 Å as a nonomer and that of WzzE from *E. coli* O157:H7 has been solved to a resolution 2.4 Å as an octamer, whereas the structure of the periplasmic domain of WzzB from *Salmonella enterica* subspecies *enterica* serovar Typhimurium ($WzzB_{ST}$) has been solved to a resolution of 3.1 Å as a pentamer (Tocilj *et al*., 2008). The structures of these periplasmic domains, in "sideward" and "downward" views, are shown in figure 1.13 to highlight the shared "bell" shape. The significance of the oligomerisation state of Wzz remains unclear, as $WzzB_{ST}$ confers a longer O-antigen chain length modality than WzzE, but forms a smaller oligomer (Tocilj *et al*., 2008). Furthermore different techniques appear to give rise to different observed oligomerisation states. Three full-length Wzz homologues have been characterised by cryo-electron microscopy (cryo-EM) in liposomes, including $WzzB_{ST}$, previously crystallised as a pentamer (Tocilj *et al*., 2008), with all homologues found in the same oligomerisation state, initially proposed to correspond to a hexamer (Larue *et al*., 2009),

but later corrected to an octamer (Kalynych *et al*., 2012a). This is in contrast to the different oligomeric states seen in the crystal structures of the periplasmic domains shown in figure 1.13. The inconsistency in the observed oligomeric states could be an artefact of the removal of the transmembrane and cytoplasmic domains or has also been suggested to be concentration-dependent (Chang *et al*., 2015). Full length $WzzB_{ST}$ has been recently studied via cryo-EM and was found to form a homododecamer (Collins et al., 2017). However the resolution of the transmembrane domain (TMD) was not sufficient to allow the modelling of the TMHs, therefore the periplasmic domain is shown in figure 1.13 (Collins *et al*., 2017). This dodecameric arrangement is in stark contrast to the pentameric crystal structure of the periplasmic domain of $WzzB_{ST}$, in absence of the TMHs (Tocilj *et al*., 2008). Taken together, the relevance of these different oligomerisation states to the physiological state remains unknown.

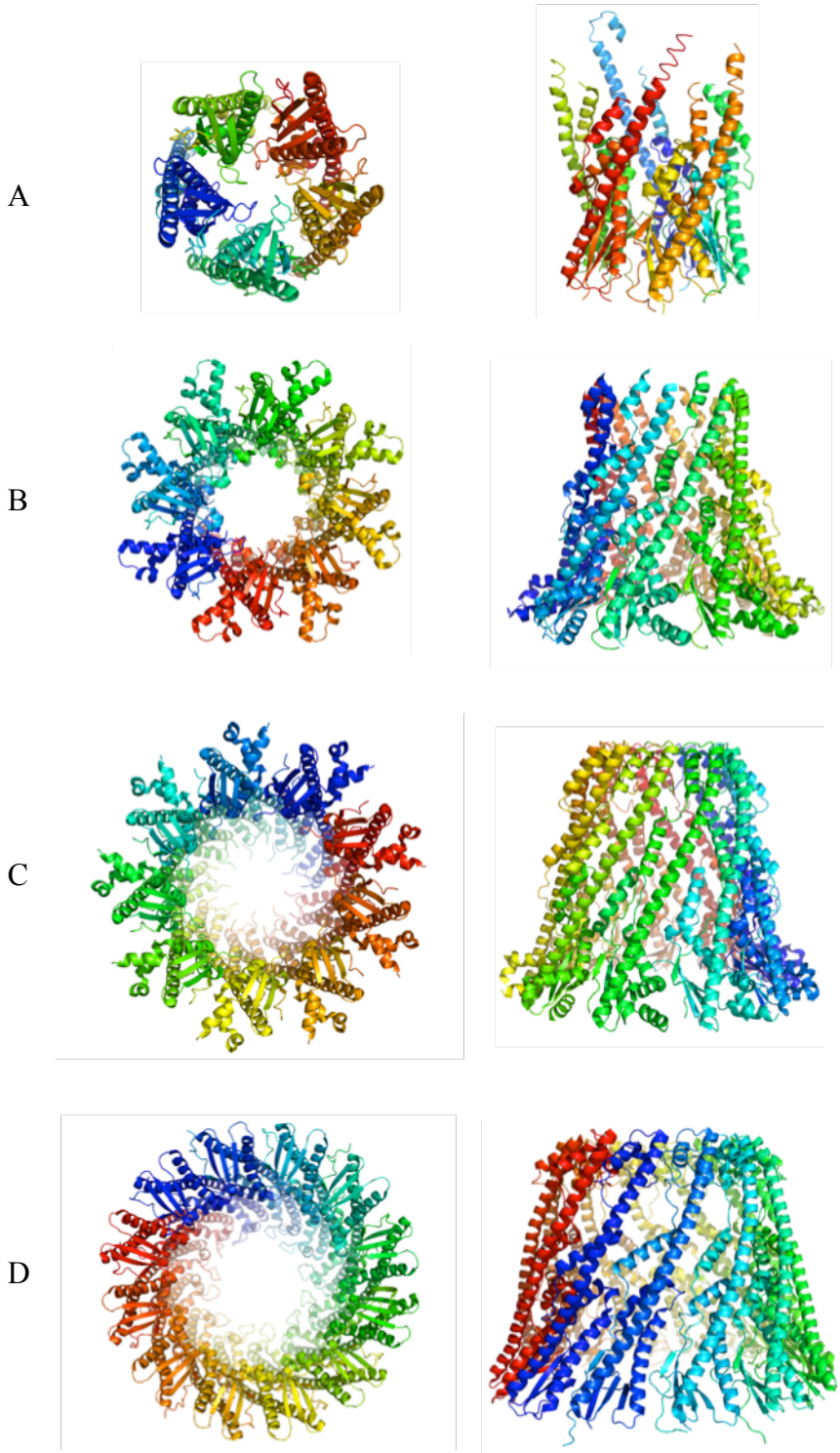


Figure 1.13. Different Wzz homologues have different oligomerisation states. The structures of the periplasmic domains of $WzzB_{ST}$ (A), Wzz_{FepE} (B) and WzzE (from *E. coli* O157:H7) (C) were solved by X-ray crystallography as a pentamer, octamer and nonamer, respectively (respective PDB codes: 3B8P, 3B8O and 3B8N) (Tocilj *et al*., 2008). The structure of the full length $WzzB_{ST}$ has been solved by cryo-EM as a dodecamer and its periplasmic domain is shown in "D" (PDB code: 5NBZ) (Collins *et al*., 2017). This image was created using MacPyMOLEdu.

A recent study reported the structure of the periplasmic domain of WzzB from *S. flexneri* ($WzzB_{SF}$) and identified a single point mutation (A107P) that affects the chain length of the synthesized carbohydrate, shown in figure 1.14a (Chang *et al*., 2015). This point mutation only results in slight conformational changes around the immediate area of the point mutation, with the overall structure, including the oligomerisation state, remaining unchanged (Chang *et al*., 2015). This lack of structural change resulting from this point mutation is highlighted by the alignment of the wild type structure with the mutant, with both forming an open trimer, shown in figure 1.14b. Despite this lack of structural change, the observed O-antigen modality decreases from 10-17 repeat units to 2-10 repeat units upon introduction of the point mutation (Chang *et al*., 2015). Furthermore, two strains of *Pseudomonas aeruginosa* (*P. aeruginosa*) O11, namely strains PA103 and 9882-80, exhibit different O-antigen modalities, conferred by two homologues of Wzz (Kintz and Goldberg, 2011). Kintz *et al.* found that when they mutated a single amino acid, predicted to be located in the internal coiled-coil domain that extends into the periplasm, in the 9882-80 homologue, to that found in the PA103 homologue, the resulting construct conferred the same O-antigen modality as the PA103 homologue, and vice versa (Kintz and Goldberg, 2011). Crucially the oligomeric state was not affected by the introduction of this point mutation, though it was proposed to be involved in contacts between adjacent monomers (Kintz and Goldberg, 2011). The structures of these Wzz homologues have not been determined, but the predicted position of this residue is highlighted on a single Wzz_{FepE}, shown in figure 1.14c and the position of this residue at the oligomerisation interface is shown in figure 1.14d (Kintz and Goldberg, 2011). A similar, but more drastic approach was taken by Kalynych *et al*. They constructed chimeras of $WzzBs_T$ and $WzzB_{SF}$, which share 72 % sequence identity, but confer different O-antigen modalities of approximately 26 and 14 repeat units, respectively (Kalynych *et al*., 2011). The structures of two chimeras, termed SF4 and SF5, which, despite sharing 96 % identity, confer different O-antigen modalities of approximately 17 and 15 repeat units, respectively, were determined (Kalynych *et al*., 2012a). Both structures were determined to be octamers, again indicating that the oligomeric state does not play a key role in the regulation of O-antigen modality, and most of the differences in amino acids could be found in the loop connecting the two α-helices that extend into the periplasm (Kalynych *et al*., 2012a). This correlates with the

finding that if this loop is replaced with a GSG linker in Wzz_{FepE}, the resulting O-antigen modality decreases, but the overall structure, and crucially oligomeric state, both remain unchanged from that of the wild type. This loop is visible in neither the wild type nor the mutant structures, but its approximate position is indicated in figure 1.14c (Kalynych *et al*., 2012a). These studies suggest that the Wzz amino acid sequence, rather than the oligomeric state, plays a role in the control of O-antigen chain length.

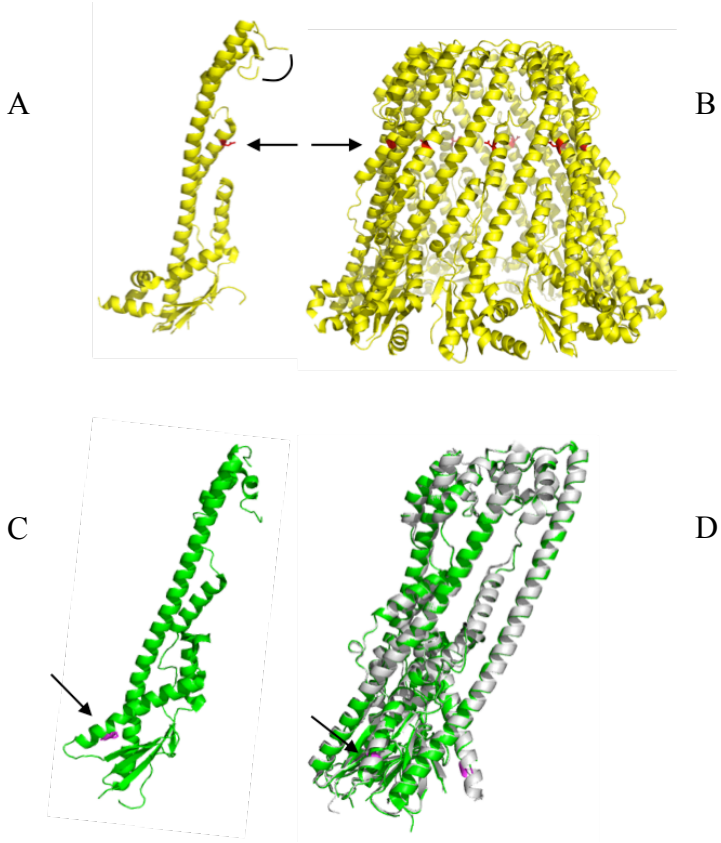


Figure 1.14. Identification of Wzz mutations that change conferred O-antigen modality. A point mutation that was identified in Wzz (from *P. aeruginosa* O11) as affecting O-antigen chain length modality is mapped on Wzz_{FepE}. This mutation is highlighted in red and with an arrow on a monomer (A) and the Wzz_{FepE} nonamer (B) (Kintz and Goldberg, 2011). The loop between the two helices that extend into the periplasm, shown to affect O-antigen chain length modality in Wzz_{FepE}, is not visible in the structure, so the approximate position of this loop is indicated with a semi-circle (A) (Kalynych *et al*., 2012a). An A107P point mutation in $WzzB_{SF}$ affects O-antigen chain length modality without changing the overall structure or oligomerisation state. The mutation is highlighted in pink and indicated with an arrow on a Wzz monomer (C) and on the alignment of the trimeric wild type structure (grey) and mutant structure (green) (D) (Chang *et al*., 2015). (Wzz_{FepE} PDB: 3B8N; $WzzB_{SF}$ wild type PDB: 4ZM1; $WzzB_{SF}$ A107P PDB: 4ZM5). This image was created using MacPyMOLEdu.

1.4.4 Chain length regulation and transport of capsular polysaccharide to the outer membrane are coupled

Regulation of CPS chain length is believed to be under the control of the PCP 2a family member Wzc (Whitfield, 2006). Wzc is predicted to share a similar "bell" shape to Wzz, but has an additional cytoplasmic domain at the C-terminus, consisting of Walker A and B ATP-binding motifs and a tyrosine-rich region, consistent with its identified tyrosine autokinase function, with Wzb acting as its cognate phosphatase (Wugeditsch *et al*., 2001). The exact function of this tyrosine autokinase activity in relation to capsular polysaccharide biosynthesis is unknown, with the phosphorylation of no particular residue required for function, but rather the overall level of phosphorylation is thought to play a role in Wzc function (Paiment *et al*., 2002). The requirement of the phosphatase Wzb suggests that the cycling between phosphorylated and non-phosphorylated Wzc plays an important role in Wzc function (Wugeditsch *et al*., 2001). The tyrosine autokinase activity has been shown to play no role in the Wzy-catalysed polymerisation of K30 repeat units in *E. coli* CPS biosynthesis, however it is required for high molecular weight K30 caspule production (Wugeditsch *et al*., 2001), suggesting a role for Wzc in the coordination of the events at the IM and OM during CPS biosynthesis (Collins *et al*., 2006, 2007).

The first carbohydrate translocase to be structurally characterised was the novel α-helical outer membrane protein Wza, required for the export of CPS to the outer membrane, shown in figure 1.15 (Dong *et al*., 2006). Although proposed from the structural study, it has only recently been experimentally established that the carbohydrate does indeed pass through the central pore (Nickerson *et al*., 2014). Blocking of Wza using specifically designed compounds has raised the possibility of targeting this event for novel drug therapies (Kong *et al*., 2013). Since the reported structure of Wza is of a closed vessel in the periplasm, further questions have yet to be answered regarding the function and regulation of Wza (Dong *et al*., 2006). The channel has been opened by applying voltage in a two droplet system with subsequent

translocation of carbohydrate detected in a partial recapitulation of the cell system (Kong *et al*., 2016).

The structure of the periplasm-spanning Wza-Wzc complex has been solved by cryo-EM, with the complex forming via interactions between the periplasmic domains of the two proteins (Collins *et al*., 2006, 2007). This complex is not a double-membrane spanning channel, but it has been proposed to coordinate the formation of the CPS with its export, with Wzc proposed to function in the regulation of the opening and closing of the channel formed by Wza (Collins *et al*., 2007).

Figure 1.15. Structure of the CPS transporter Wza. The approximate boundaries of the outer membrane are shown in grey, with the large periplasmic domain below (PDB code: 2J58) (Dong *et al*., 2006). This image was created using MacPyMOLEdu.

1.4.5 Transfer of completed polymer onto the appropriate acceptor

1.4.5.1 Ligation of the O-antigen onto Lipid A to form LPS

To form LPS, the O-antigen chain is ligated on to the lipid A core-OS by WaaL, mutations in which result in rough LPS, lacking the O-antigen, being translocated to the OM (Abeyrathne *et al*., 2005). Although the structure of WaaL is yet to be determined, WaaL homologues are predicted to contain a large periplasmic loop that is proposed to form two almost perpendicular alpha helices, with one helix contributing a conserved arginine and the other a conserved histidine (Pérez *et al*., 2008). Mutants in either of these residues are unable to rescue WaaL function in a knockout strain, although if the arginine is mutated to a lysine function can be restored, suggesting these residues are important for interaction with undecaprenyl pyrophosphate (Ruan *et al*., 2012). It was originally proposed that ligation of the O-antigen chain by WaaL onto lipid A core-OS requires ATP (Abeyrathne and Lam, 2007), but this now seems unlikely (Han *et al*., 2012; Ruan *et al*., 2012).

1.4.5.2 Glycosylation of target proteins

During bacterial *N*-linked glycosylation, the oligosaccharyltransferase PglB transfers the completed polysaccharide to asparagine side chains of proteins containing the following amino acid acceptor motif: D/E-Y-N-X-S/T, where X is any amino acid except P (Kowarik et al., 2006; Lizak et al., 2011). The structure of the *Campylobacter lari* homologue in complex with the acceptor peptide, DQNATF, has been solved via X-ray crystallography to 3.4 Å resolution (Lizak *et al*., 2011) and is shown in figure 1.16. The structure of the complete oligosaccharyltransferase shows why the isolated C-terminal periplasmic domain is not functional (Igura *et al*., 2008) as the acceptor protein binding pocket is formed by residues from both the transmembrane domain region and the C-terminal periplasmic domain, with the catalytic D-X-D motif located in the loop

between TMHs 3 and 4 (Lizak *et al*., 2011). The conserved "WWD" motif, found in all oligosaccharyltransferase homologues, is required for acceptor sequence binding (Lizak *et al*., 2011).

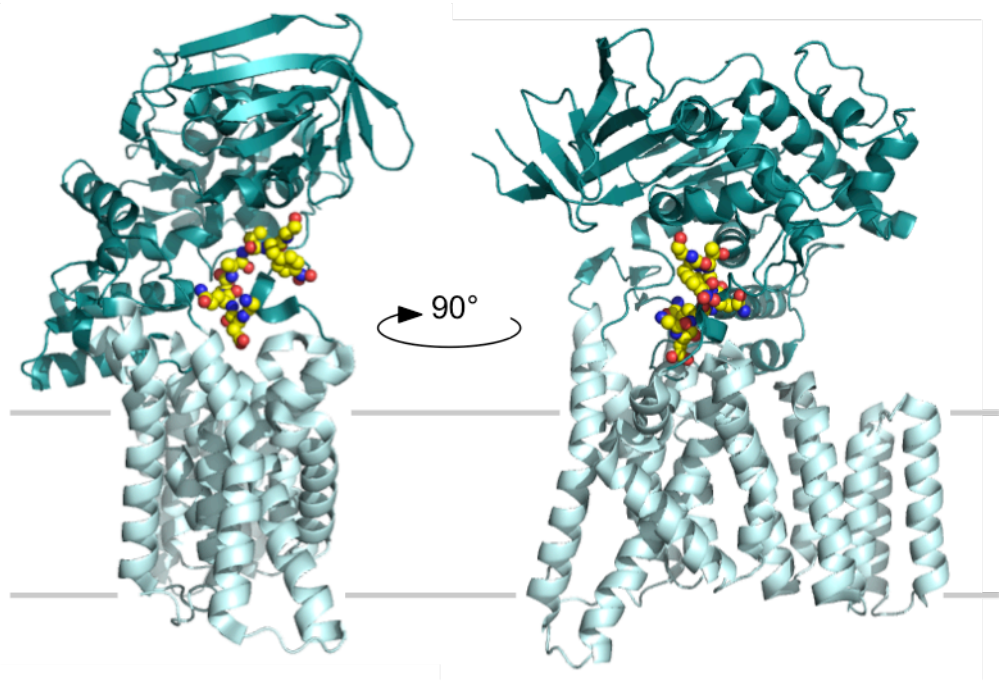


Figure 1.16. The structure of the oligosaccharyltransferase PglB in complex with acceptor peptide. The TMHs are shown in pale cyan and the approximate boundaries of the inner membrane in grey. The periplasmic domain is shown in dark teal and the acceptor peptide is shown in ball and stick representation (PDB code: 3RCE) (Lizak *et al*., 2011). This image was created using MacPyMOLEdu.

1.4.6 Transport of LPS to the outer membrane

In order to insert into the outer leaflet of the OM, the LPS molecule must first exit the IM and be transported from the IM across the periplasm and then translocated across the OM. Seven genes constitute the pathway that accomplishes this transport, termed *lptA* through to *lptG*. This process begins with the extraction of the completed LPS from the periplasmic leaflet of the inner membrane by the ABC transporter complex $LptB_2FG$ (Luo *et al*., 2017). Recently the structure of the nucleotide-free $LptB_2FG$ complex from *P. aeruginosa* PAO1 has been solved by X-ray crystallography to 3.46 Å resolution (Luo *et al*., 2017) and is shown in figure 1.17. It consists of one LptG and one LptF,

each of which contain six TMHs and a β-jellyroll-like domain in the periplasm, and two LptB proteins, which form the NBD (Luo *et al*., 2017). A large V-shaped cleft is located between the TMHs of LptG and LptF (Luo *et al*., 2017). A coupling helix located between TMHs 2 and 3 of both LptG and LptF forms the LptG-LptB and LptF-LptB interface, with LptG and LptF each interacting with one LptB of the cytoplasmic dimer (Luo *et al*., 2017). The structure of the LptB dimer has been solved in absence of the LptF-G proteins, and undergoes a large conformational change upon ATP hydrolysis (Luo *et al*., 2017; Sherman *et al*., 2014; Wang *et al*., 2014). In the proposed model for LPS extraction from the IM, ATP binding to LptB causes a conformational change in LptB, which is then transmitted to LptF-G via the interaction with the coupling helices. This conformational change in LptF-G would allow lateral LPS entry. LptB-catalysed ATP hydrolysis would then be expected to cause the expulsion of LPS into the β-jellyroll-like periplasmic domain of LptF-G (Luo *et al*., 2017). Following release of LPS from the IM, it is then transferred on to LptC in an ATP-dependent manner (Okuda *et al*., 2012). The structure of the periplasmic domain of LptC, shown in figure 1.17, shows it too possesses an exposed β-sheet, suggesting a model in which the LptF-G periplasmic domain docks to LptC via β strand-strand interactions (Tran *et al*., 2010), with one LptC binding each $LptB_2FG$ (Narita and Tokuda, 2009). LPS is then transferred to LptA (Tran *et al*., 2010) via interactions between the periplasmic domain of LptC and the N-terminus of LptA (Freinkman *et al*., 2012). The transfer of LPS from LptC to LptA has been found to be ATP-dependent (Okuda *et al*., 2012). LptA forms an extended polymer via β-sheet interactions between monomers, with each monomer consisting of 16 anti-parallel β-strands (Suits *et al*., 2008), and can be seen in figure 1.17. LptA completes the bridge across the periplasm from the IM to the OM via interactions between the C-terminus of LptA and the N-terminal periplasmic domain of the OM protein LptD (Freinkman *et al*., 2012).

Two groups determined the structure of the LptE and LptD heterodimeric complex, required for the insertion of LPS into the OM, at the same time (Dong *et al*., 2014; Qiao *et al*., 2014), with the structures of the LptDE complex from three other organisms solved at a later date (Botos *et al*., 2016). LptD is a 26 anti-parallel stranded β-barrel in which LptE, which consists of a β-sheet and an α-helix sits; akin to a plug (Dong *et al*.,

2014; Qiao *et al*., 2014), shown in figure 1.17. LptD has a large N-terminal periplasmic domain, which is only visualised in one of these structures (Qiao *et al*., 2014). This N-terminal domain forms a β-sheet jellyroll, like LptA and LptC, with an exposed strand (Qiao *et al*., 2014) and is thought to bind to the C-terminus of LptA through β-strand-strand interactions, thereby completing the double-membrane spanning complex (Freinkman *et al*., 2012; Qiao *et al*., 2014). During the transfer of LPS from the N-terminus of LptD into the outer leaflet of the OM, the O-antigen chain is believed to enter the negatively charged lumen (Botos *et al*., 2016) of the LptD barrel via an opening between the first and final β-strands of the barrel, whilst lipid A is proposed to remain shielded in an "intramembrane hole" (Gu *et al*., 2015; Li *et al*., 2015). Repulsion between the negatively charged LptD barrel lumen and the O-antigen is believed to play a role in the transfer of LPS into the outer leaflet of the OM as the mutation of negatively charged residues found in the LptD lumen to positively charged residues results in decreased cell viability (Botos *et al*., 2016). A simplified image of this proposed double-membrane spanning complex composed of LptA-G, plus the currently known structures, can be seen in figure 1.17.

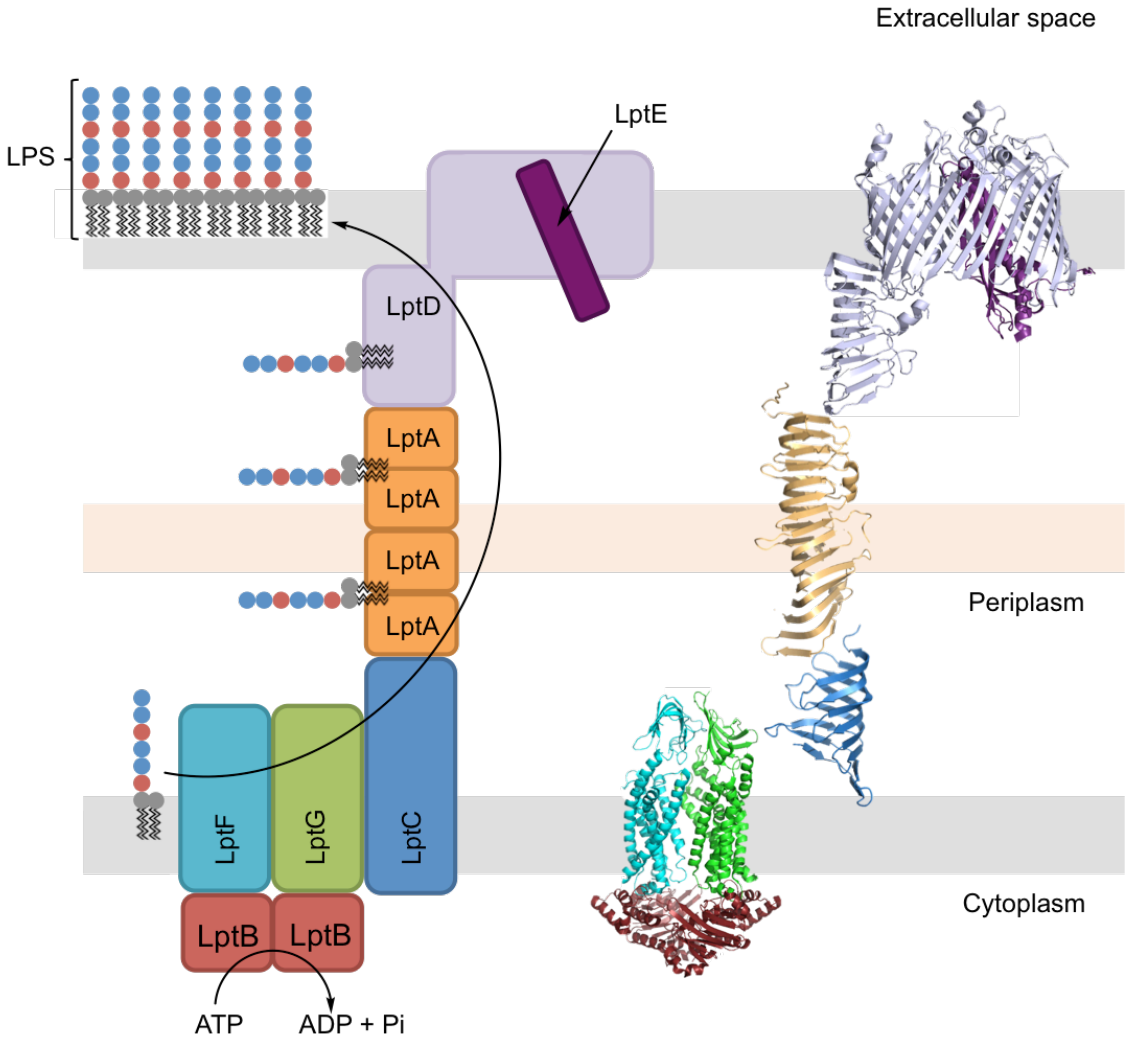


Figure 1.17. A cartoon representation of the export of the completed LPS to the outer membrane by the proposed double-membrane spanning complex formed by the Lpt family, along with the known Lpt family member structures. The inner and outer membranes are shown with a grey line and the peptidoglycan layer with a peach line. The lipid A component of LPS is represented by the grey circles and 4 lines and the sugars forming the O-antigen component of LPS are represented by the blue and red circles. The same colour code is used for the structures as in the cartoon. (LptDE PDB: 4Q35; LptA PDB: 2R19; LptC PDB: 3MY2; $LptB_2FG$ PDB: 5X5Y) (Luo *et al*., 2017; Qiao *et al*., 2014; Suits *et al*., 2008; Tran *et al*., 2010). The cartoon representation is adapted from Woodward and Naismith, 2016.

1.4.7 Recycling of undecaprenyl-pyrophosphate

Once the completed polysaccharide chain has been transferred to the appropriate acceptor, the leftover undecaprenyl pyrophosphate needs to be recycled to form the initial undecaprenyl phosphate, which can then be used as a carrier lipid during the formation of a new LLO. This is achieved by a membrane-integrated member of the type II phosphatidic acid phosphatase family, exemplified by phosphatidylglycerol-phosphate phosphatase B, known as PgpB (Fan *et al*., 2014; El Ghachi *et al*., 2005). This, along with the other phosphatases responsible for this final step in the LLO biosynthesis pathways, known as UppP (previously BacA), YbjG and LptX (previously YeiU) is the target of the antiobiotic Bacitracin (El Ghachi *et al*., 2004, 2005). The structure of the *E. coli* homologue of PgpB has been solved, showing a six pass transmembrane topology, with both the N- and C-termini located in the cytoplasm (Fan *et al*., 2014). A large α-helical periplasmic domain is located between TMHs 3 and 4 and one α-helix forms the cytoplasmic domain at the C-terminus (Fan *et al*., 2014). The protein shares the same overall fold and active site as soluble members of this family. The active site residues are located at the interface between the membrane and periplasm (Fan *et al*., 2014). The structure of PgpB in complex with the phospholipid phosphatidyl ethanolamine has been solved, with phosphatidyl ethanolamine bound in a V-shaped cleft formed by TMHs 3, 4 and 6 (Tong *et al*., 2016). This can be seen in figure 1.18. Mutations at the proposed lateral substrate entrance of this cleft result in a reduced ability to dephosphorylate a range of substrates compared to the wild type (Fan *et al*., 2014).

Figure 1.18. The undecaprenyl pyrophosphatase PgpB in complex with phosphatidyl ethanolamine. The approximate boundaries of the inner membrane are shown in grey and phosphatidyl ethanolamine is shown using a ball and stick representation (PDB: 5JWY) (Tong *et al*., 2016). This image was created using MacPyMOLEdu.

1.5 *C. jejuni*

According to the World Health Organisation, there are an estimated 550 million cases diarrhoeal food-borne of diseases per year, approximately half of which are in children below the age of 5 years, and the Gram-negative pathogen *C. jejuni* is one of the 4 main causative agents of this (WHO, 2016) [Accessed 19/06/17]. *C. jejuni* is a commensal bacterium in chicken and therefore most cases of *C. jejuni* are transmitted through incorrect handling or cooking of raw chicken, although *C. jejuni* can also be found in water supplies (Young *et al*., 2007). Numerous complications can arise post *C. jejuni* infection. These include the autoimmune diseases Reactive Arthritis, affecting 1-5 % of patients post-infection (WHO, 2012) [Accessed 19/06/17], and Guillain-Barré syndrome, affecting approximately 0.1 % of patients post-infection (Young *et al*., 2007). Despite this low percentage, *C. jejuni* is estimated to cause approximately one third of Guillain-Barré cases (WHO, 2012) [Accessed 19/06/17]. Guillain-Barré syndrome results in damage to the myelin sheaths of the nerves of the peripheral

nervous system, resulting in muscle weakness and even paralysis (WHO, 2012; Young *et al*., 2007) [Accessed 19/06/17]. As *C. jejuni* infections are usually self-limiting, maintaining proper hydration is usually the recommended course of treatment. This, along with ensuring that food, especially chicken, is prepared appropriately as a preventative measure, is often sufficient, however a course of antibiotics is required in extreme cases (Allos, 2001). Antibiotic resistance is on the increase, with many *C. jejuni* strains no longer responding to the broad-spectrum class of antibiotics, the fluoroquinolones. This increase in fluoroquinolone-resistant *C. jejuni* infections corresponded with an increase in the use of fluoroquinolones in poultry farming (Allos, 2001).

1.5.1 Roles of *C. jejuni* LOS and CPS during infection

The complete *C. jejuni* genome sequence has proven key in characterising some of its key pathogenic traits. The genomes of *C. jejuni* isolates can vary enormously, with much of this variation occurring in the genes required for the biosynthesis of various carbohydrate components of the capsule and cell envelope, including lipooligosaccharide (LOS) biosynthesis (Young *et al*., 2007). LOS is a truncated form of LPS, which lacks the O-antigen. Much of this observed sequence variation results from a process known as "phase variation", which allows *C. jejuni* to switch expression of particular genes on and off as required. This results in a hypervariable LOS and capsule structure (Young *et al*., 2007). As previously described, LPS is crucial to the survival of bacteria upon infection of a host and can enable the evasion of the host immune response via "molecular mimicry" (Li *et al*., 2017; Young *et al*., 2007). The same is true for LOS in the case of *C. jejuni* (Yuki *et al*., 1993). Unlike most prokaryotes, *C. jejuni* LOS contains *N*-acetylneuraminic acid, which enables it to mimic the carbohydrate moiety of the human neuronal gangloside structure (Prendergast *et al*., 1998; Yuki *et al*., 1993). This molecular mimicry is in agreement with the finding that approximately 30 % of Guillain-Barré patients produce antibodies against neuronal gangliosides (Gregson *et al*., 1993). The presence of these antibodies is also associated with a poorer prognosis when compared to patients who do not produce these antibodies

(Gregson *et al*., 1993), confirming the crucial role played by LOS in the development of Guillain-Barré syndrome post *C. jejuni* infection (Young *et al*., 2007). *C. jejuni* LOS has also been demonstrated to play a role in adherence to and invasion of cells during infection as strains with a mutated *galE* gene, which encodes a UDP-glucose 4-epimerase required for LOS biosynthesis, exhibited attenuated ability to adhere to and invade INT407 epithelial cells *in vitro* (Fry *et al*., 2000). However, this mutant strain could still colonise the guts of chicks, suggesting a multi-factorial approach to infection (Fry *et al*., 2000).

In addition to the role played by LOS, CPS has also been shown to play a role in *C. jejuni* infection (Bacon *et al*., 2001; Wong *et al*., 2015). Acapsular *C. jejuni* strain 81-176 mutants, generated via the mutation of the *kpsM* gene encoding the ABC transporter required for undecaprenyl pyrophosphate-linked repeat unit flipping across the IM during ABC transporter-dependent CPS biosynthesis, show reduced ability to invade INT407 epithelial cells *in vitro* (Bacon *et al*., 2001).

1.5.2 Modified heptose sugars in *C. jejuni* CPS

Modified heptoses have been found in the CPS of various *C. jejuni* strains. *C. jejuni* strain 81-176 has been found to produce CPS containing various forms of D-*altro*-heptose, including 6-deoxy and 3-*O*-methyl forms (Papp-Szabo *et al*., 2005). *C. jejuni* strain NCTC 11168 produces CPS containing 6-*O*-methyl-L-*gluco*-heptose and 3,6-*O*-methyl-L-*gluco*-heptose (St. Michael *et al*., 2002; Wong *et al*., 2015). These sugars, in their nucleotide-activated forms, and their positions in their corresponding repeat units can be seen in figures 1.19 and 1.20, respectively.

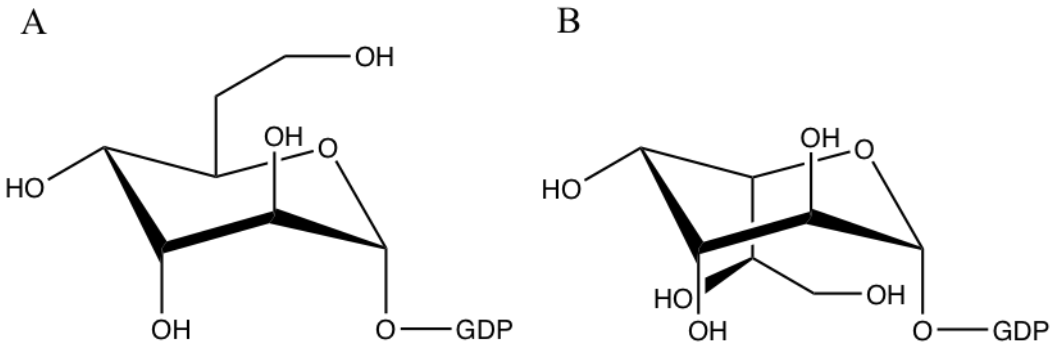


Figure 1.19. GDP-6-deoxy-D-*altro*-heptose and GDP-L-*gluco*-heptose. GDP-6-deoxy-D-*altro*-heptose is shown in "A" and GDP-L-*gluco*-heptose is shown in "B" (McCallum *et al*., 2012, 2013).

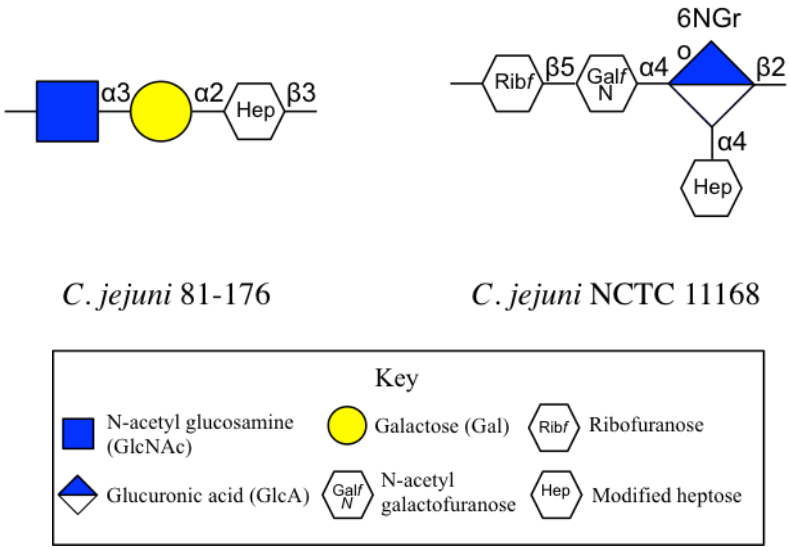


Figure 1.20. Modified heptoses are found in the CPS repeat units of *C. jejuni* strains 81-176 and NCTC 11168. In the case of strain 81-176, "Hep" corresponds to various forms of D-*altro*-heptose, including 6-deoxy and 3-*O*-methyl forms (Papp-Szabo *et al*., 2005). In the case of strain NCTC 11168, "Hep" corresponds to 6-*O*-methyl-L-*gluco*-heptose or 3,6-*O*-methyl-L-*gluco*-heptose. NGr corresponds to 2-amino-2-deoxyglycerol (St. Michael *et al*., 2002; Wong *et al*., 2015). Carbohydrate Structure Database nomenclature used (Varki *et al*., 2015).

C. jejuni strain 81-176 was first isolated from a 1981 outbreak caused by the ingestion of raw milk at a farm in Minnesota, USA (Korlath *et al*., 1985) and has been found to be between 250 and 2500 fold more efficient at invading INT407 cells than other *C. jejuni* strains and up to 25-fold more efficient at invading INT407 cells than the negative control, a non-invasive *E. coli* HB101 strain (Oelschlaeger *et al*., 1993). Furthermore, *C. jejuni* 81-176 is most efficient at invading INT407 cells when exposed to them at an extremely low multiplicity of infection of just 0.02 (Hu and Kopecko, 1999).

C. jejuni strain NCTC 11168 was first isolated from a patient in 1977 (Gaynor *et al*., 2004) and it became the first strain of *C. jejuni* to have its entire genome sequenced (Parkhill *et al*., 2000). The *C. jejuni* NCTC 11168 genome-sequenced strain (termed NCTC 11168-GS) has been found to be unable to colonise 1 day old chicks, however the original *C. jejuni* NCTC 11168 (termed NCTC 11168-O) is able to colonise chicks, suggesting that the genome-sequenced strain has evolved to survive in a laboratory environment (Gaynor *et al*., 2004). Doses of 10^6 colony-forming units (CFU) of *C. jejuni* strain NCTC 11168 have been found to be sufficient to cause symptoms of campylobacteriosis in an infant rabbit model, in comparison to *C. jejuni* strain ATCC 3350 which needed to be used at 10^{10} CFU to have a similar effect, confirming that *C. jejuni* NCTC 11168 is more virulent than other strains, in this model (Shang *et al*., 2016). Despite *C. jejuni* NCTC 11168-GS being determined to be less able to colonise chicks, it is still capable of infecting human hosts (Thomas *et al*., 2014).

1.5.3 Roles of modified heptoses during *C. jejuni* infection

The role of capsular modified heptoses in the virulence of other Gram-negative pathogens has been determined. *Yersinia pseudotuberculosis* strain O:2a produces LPS containing the modified heptose 6-deoxy-D-*manno*-heptose (Pacinelli *et al*., 2002). Mutations in the genes encoding 6-deoxy-D-*manno*-heptose biosynthetic enzymes result in the production of LPS containing D-*glycero*-D-*manno*-heptose, instead of 6-deoxy-D-*manno*-heptose (Ho *et al*., 2008). These mutants display an increased susceptibility to

both bile salts and the antibiotic polymyxin B, indicating that the modified heptose plays a specific role in the barrier function of LPS (Ho *et al*., 2008).

C. jejuni strain 81-176 mutated in the *waaF* gene, encoding the heptosyltransferase required for the transfer of L-*glycero*-D-*manno*-heptose onto the outer core of lipid A during LOS biosynthesis, has been found to be unable to invade INT407 cells (Kanipes *et al*., 2004). This highlights a specific role for this heptose during invasion, as a previous study found that mutations causing the production of LOS without *N*-acetylneuramic acid had no effect on the ability of this strain to invade INT407 cells (Guerry *et al*., 2002). Although a function of L-*glycero*-D-*manno* heptose in the *C. jejuni* LOS has been determined, the function of D-*altro*-heptose, and its derivatives found in the CPS of *C. jejuni* 81-176 remains unknown (McCallum *et al*., 2012).

Whilst the role of capsular D-*altro*-heptose, and its derivatives, in the virulence of *C. jejuni* strain 81-176 remains unknown, the role of capsular modified heptose in *C. jejuni* strain NCTC 11168 has been studied (Wong *et al*., 2015). Inactivating mutations in the genes required for GDP-6-*O*-methyl-L-*gluco*-heptose biosynthesis in *C. jejuni* strain NCTC 11168, namely *mlghB* and *mlghC*, resulted in the production of CPS lacking this modified heptose (Wong *et al*., 2015). These mutants lacking capsular modified heptose possessed a decreased ability to invade Caco-2 epithelial cells compared to the wild type (Wong *et al*., 2015). Furthermore these mutants also possessed a decreased ability to colonise and persist in the chicken gut (Wong *et al*., 2015). These findings highlight a role of these modified heptoses in the CPS in *C. jejuni* virulence. Despite their demonstrated role in *C. jejuni* strain NCTC 11168 virulence (Wong *et al*., 2015), and their potential role in *C. jejuni* 81-176 virulence, their respective biosynthetic pathways remain relatively uncharacterised. Furthermore the enzymes required for the biosynthesis of modified heptoses in *C. jejuni* represent potential drug targets, but their structures are unknown (McCallum *et al*., 2012, 2013).

1.5.4 Biosynthesis of modified heptoses in *C. jejuni*

The sugars involved in the biosynthesis of LOS and CPS are biosynthesised as nucleotide-activated sugars. The biosynthesis of GDP-6-deoxy-D-*altro*-heptose, in *C. jejuni* strain 81-176, is shown in figure 1.21 and begins with the dehydration of GDP-D-*manno*-heptose at the C4, C6 position, catalysed by DdahA, also known as WcbK, to form GDP-4-keto-6-deoxy-D-*lyxo*-heptose (McCallum *et al*., 2011). DdahB then catalyses the epimerisation of GDP-4-keto-6-deoxy-D-*lyxo*-heptose at the C3 position to form GDP-4-keto-6-deoxy-D-*arabino*-heptose. Finally this is reduced by DdahC at the C4 position to form GDP-6-deoxy-D-*altro*-heptose (McCallum *et al*., 2012).

The biosynthesis of GDP-L-*gluco*-heptose, in C. jejuni strain NCTC 11168, shown in figure 1.22, follows the same scheme as the biosynthesis of GDP-6-deoxy-D-*altro*-heptose, however it is less well characterised and somewhat more complex. The biosynthesis of GDP-6-*O*-methyl-L-*gluco*-heptose begins with the oxidation of the same starting sugar, GDP-D-*manno*-heptose, at the C4 position to form GDP-4-keto-D-*lyxo*-heptose, catalysed by MlghA. The presence of MlghA and its oxidase, rather than dehydratase, activity are inferred from the absence of a dehydrated final product and the requirement of a 4-keto group for the epimerisation activity of MlghB, however the existence of MlghA remains to be confirmed (McCallum *et al*., 2013). The methyltransferase MlghD is then proposed to methylate GDP-4-keto-D-*lyxo*-heptose at the 6-*O* position to form GDP-6-*O*-methyl-4-keto-D-*lyxo*-heptose as the second step of GDP-6-*O*-methyl-L-*gluco*-heptose biosynthesis (McCallum *et al*., 2013). However, this is based on mutagenesis studies and remains to be biochemically confirmed (Sternberg *et al*., 2013). Due to this uncertainty, coupled with the finding that this methyl group is not required for the activity of the remaining enzymes in the GDP-6-*O*-methyl-L-*gluco*-heptose biosynthetic pathway, this step has been omitted from figure 1.22 (McCallum *et al*., 2013). This is followed by the MlghB-catalysed epimerisation at both the C3 and C5 positions to form GDP-(6-*O*-methyl)-4-keto-L-*xylo*-heptose (McCallum *et al*., 2013). This is in contrast to the GDP-6-deoxy-D-*altro*-heptose biosynthesis, which only involves the epimerisation at the C3 position (McCallum *et al*., 2012). This is then

followed by a reduction at the C4 position, catalysed by MlghC, to give the final product GDP-(6-*O*-methyl)-L-*gluco*-heptose (McCallum *et al*., 2013).

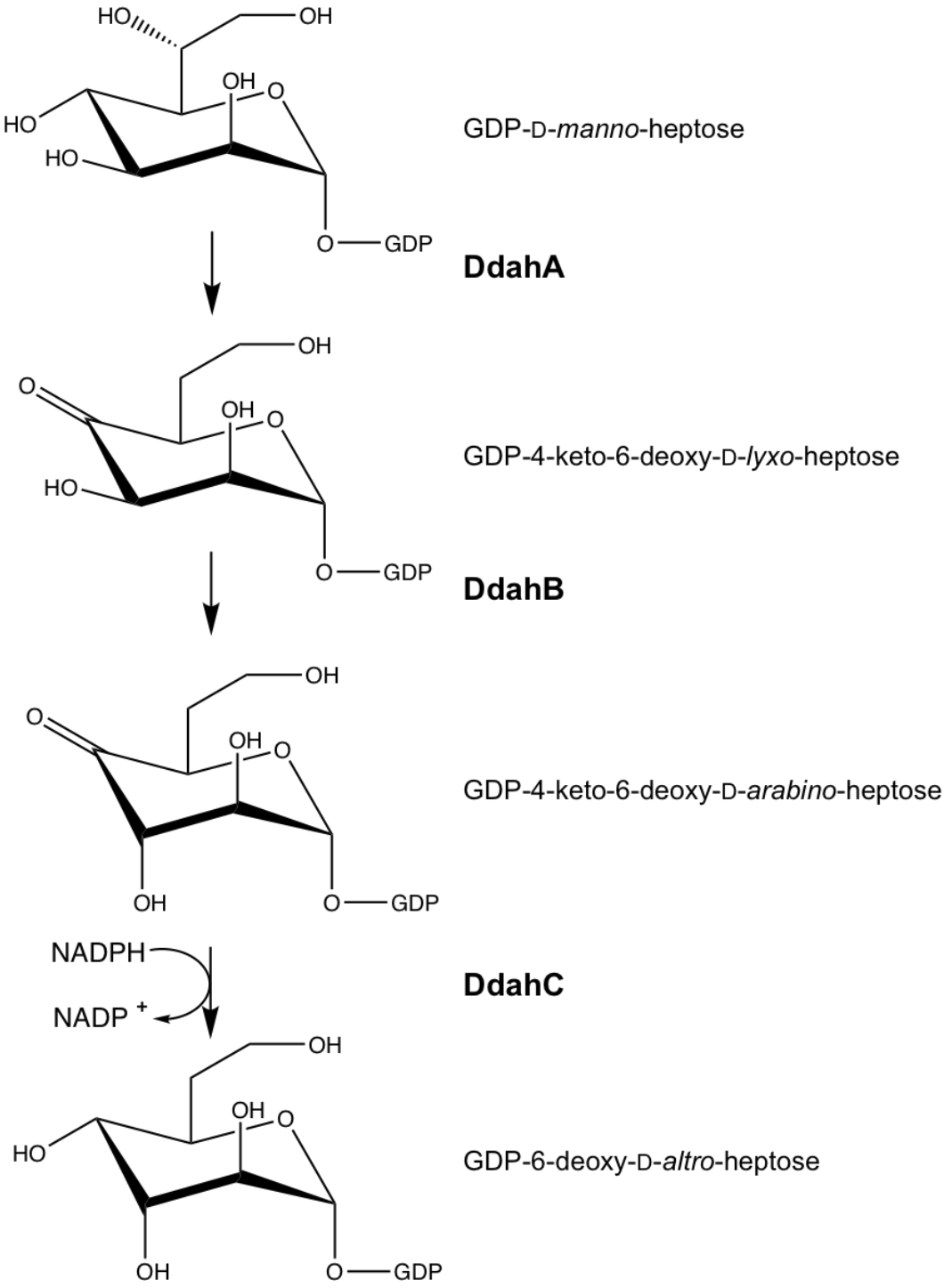


Figure 1.21. The biosynthesis of GDP-6-deoxy-D-*altro*-heptose in *C. jejuni* strain 81-176. (McCallum *et al*., 2012). The names of the enzymes proposed to catalyse each step are shown in bold. This figure was created using ChemDraw 16.0.

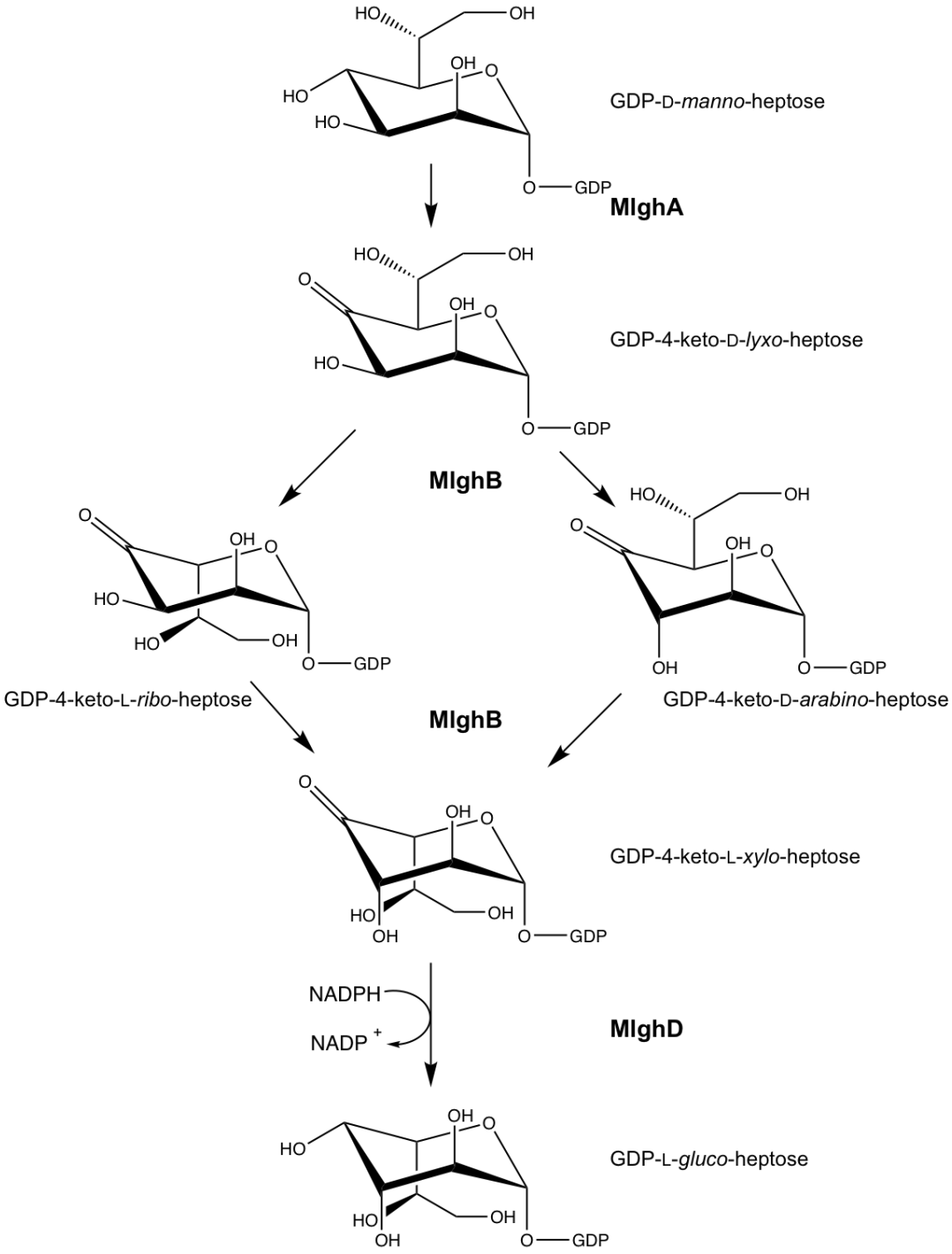


Figure 1.22. The biosynthesis of GDP-L-*gluco*-heptose in *C. jejuni* strain NCTC 11168. The names of the enzymes proposed to catalyse each step are shown in bold (McCallum *et al*., 2013). This figure was created using ChemDraw 16.0.

The modified heptose biosynthesis pathways described here share many parallels with the more characterised dTDP-L-rhamnose biosynthesis pathway, shown in figure 1.23. Rhamnose is an important component of the cell envelope both Gram-negative and Gram-positive pathogens, for example *P. aeruginosa* and *Streptococcus suis*, and like the modified heptose biosynthetic enzymes, its biosynthesis also offers multiple potential drug targets (Dong *et al*., 2003a). Rhamnose biosynthesis begins with the activation of glucose-1-phosphate via RmlA-catalysed transfer of dTMP from dTTP to form dTDP-glucose. The structure of RmlA from *P. aeruginosa* has been solved to 1.66 Å (Blankenfeldt *et al*., 2000). The dTDP-D-glucose 4, 6-dehydratase RmlB then catalyses the dehydration of dTDP-glucose at the C4, C6 position to form dTDP-6-deoxy-D-*xylo*-hexulose. The structure of RmlB from *S.* Typhimurium has been solved to 2.47 Å (Allard *et al*., 2001). This C4, C6 dehydration step also occurs in GDP-6-deoxy-D-*altro*-heptose biosynthesis in *C. jejuni* strain 81-176 and is catalysed by DdahA (McCallum *et al*., 2011, 2012). The dTDP-6-deoxy-D-*xylo*-4-hexulose 3,5-epimerase RmlC then catalyses the epimerisation at the C3 and C5 positions to form dTDP-6-deoxy-L-*lyxo*-4-hexulose. The structure of RmlC from *S.* Typhimurium has been solved to 2.17 Å and the structure of this homologue in complex with the substrate analogue dTDP-phenol allowed the identification of the active site (Giraud *et al*., 2000). This step also occurs in GDP-(6-*O*-methyl)-L-*gluco*-heptose biosynthesis in *C. jejuni* strain NCTC 11168 and is catalysed by MlghB (McCallum *et al*., 2013). In contrast, during GDP-6-deoxy-D-*altro*-heptose biosynthesis in *C. jejuni* strain 81-176, DdahB catalyses the epimerisation of its substrate only at the C3 position (McCallum *et al*., 2012). Finally the dTDP-6-deoxy-L-*lyxo*-4-hexulose reductase RmlD reduces the keto group at the C4 position to a hydroxyl group, a process requiring NADPH, to form dTDP-L-rhamnose (Graninger *et al*., 1999). The structure of RmlD from *S.* Typhimurium in complex with NADPH has been solved to 2.7 Å and the structure of RmlD in complex with dTDP-L-rhamnose allowed the identification of the active site residues (Blankenfeldt *et al*., 2002). MlghC and DdahC catalyse the equivalent steps in GDP-(6-*O*-methyl)-L-*gluco*-heptose biosynthesis in *C. jejuni* strain NCTC 11168 and GDP-6-deoxy-D-*altro*-heptose biosynthesis in *C. jejuni* strain 81-176, respectively (McCallum *et al*., 2012, 2013).

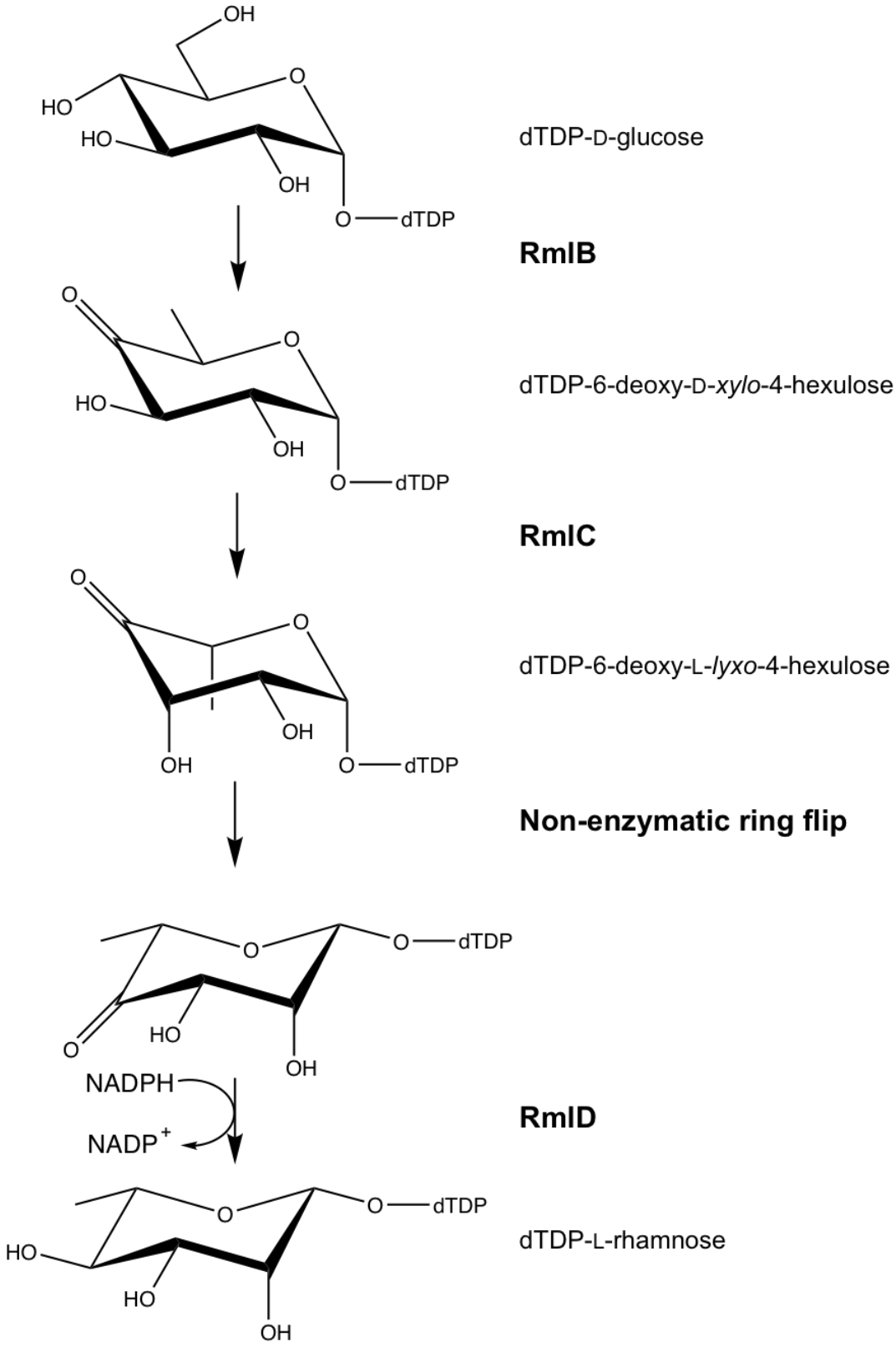


Figure 1.23. The biosynthesis of dTDP-L-rhamnose. (Naismith, 2004) The names of the enzymes responsible for each step are shown in bold. This figure was created using ChemDraw 16.0.

1.6 Aims

This thesis concerns the structural characterisation of proteins required for the biosynthesis of LPS and CPS.

Multiple homologues of the O-antigen polymerase Wzy were cloned, expressed, and purified for crystallisation trials, with multiple methods for the optimisation of expression and purification of membrane proteins explored and optimised.

The hypothesis that the O-antigen polymerase Wzy and its polysaccharide co-polymerase Wzz interact directly to form a complex will be investigated, using both biochemical and structural methods, including tandem affinity purification and cryo-electron microscopy.

Modified heptose sugars have been found in the capsule of the pathogen *C. jejuni* and although their biosynthetic enzymes represent potential drug targets, they remain structurally uncharacterised. In collaboration with Dr Carole Creuzenet (University of Western Ontario) we will use X-ray crystallography to determine the structures of the enzymes required for GDP-6-deoxy-D-*altro*-heptose in *C. jejuni* strain 81-176 and GDP-6-deoxy-L-*gluco*-heptose in *C. jejuni* strain NCTC 11168, both in the absence and presence of substrate.

2 Materials and Methods

2.1 Transformation

Approximately 100 ng of plasmid DNA was added to 50 µl competent cells and left on ice for 15 minutes. These were heat-shocked at 42 °C for one minute and then returned to the ice for a further 15 minutes. 600 µl of LB medium was added and cells were grown at 37 °C, at 200 rpm, for one hour. Transformations were centrifuged at 20,817 x *g* for one minute and excess supernatant was discarded. The pellet was resuspended in the remaining supernatant (approximately 100 µl) and then plated on the appropriate antibiotic-selective LB-agar plate and left overnight at 37 °C. *E. coli* C43 (DE3) and C41 (DE3) (Miroux and Walker, 1996) competent cells were prepared in-house by Dr Hui Liu, University of St Andrews. α-select bronze competent *E. coli* DH5 α cells were obtained from Bioline.

2.2 Small-scale expression trials

A single colony from an LB-agar plate, or a glycerol stock, was used to inoculate 10 ml LB-medium, supplemented with the appropriate antibiotic (ampicillin at 100 µg/ml; kanamycin at 50 µg/ml; chloramphenicol at 25 µg/ml), cells were grown overnight at 37 °C, shaking at 200 rpm. 10 ml LB-medium, supplemented with the appropriate antibiotic at the aforementioned concentration, was inoculated with overnight culture at a 1:50 dilution. Cells were grown at 37 °C, shaking at 200 rpm, until the culture reached the required OD_{600} (either OD_{600} 0.4-0.6 or OD_{600} > 1). Expression of genes in expression vectors with an *araC* promoter was induced with either 0.02 % (w/v) or 0.2 % (w/v) arabinose and expression of genes in expression vectors with a T7 promoter was induced with 0.1 mM, 0.5 mM or 1.0 mM IPTG. Following the induction of expression, cells were grown at the following temperatures: 37 °C for 3 hours, 25 °C for

4 hours or overnight or 15 °C overnight (overnight corresponds to approximately 20 hours). To allow comparison of expression levels, cell cultures were typically normalised to an OD_{600} of 1 via dilution with LB medium. 1 ml of cells at an OD_{600} of 1 were harvested via centrifugation at 20,817 x *g* for one minute. Samples were prepared for SDS-PAGE and western blotting as follows: an *E. coli* pellet was resuspended in 50 µl 4 x LDS sample buffer (Novex) with 50 µl PBS, 5 µl 10 % (w/v) SDS and 5 µl 1 M DTT, sonicated for one minute in a sonicator bath, boiled for five minutes and centrifuged for 10 minutes at 20,817 x *g*. Samples were either used directly or stored at -20 °C until required.

2.3 SDS-PAGE

Whole cell expression trial SDS-PAGE samples were prepared as described in section 2.2. Samples for SDS-PAGE analysis of the purification progress were prepared using 2 x Laemmli sample buffer (Biorad). Samples for analysis via SDS-PAGE followed by western blotting or Coomassie Brilliant Blue staining analysis were loaded onto a 4-12 % Bis-tris NuPAGE gel (Novex), along with Sharp Pre-stained Protein Standard (Novex) and run at 200 V for 35 minutes in 1 x MES SDS Running buffer (NuPAGE). Samples for SDS-PAGE analysis were loaded onto an AnykD Mini-PROTEAN TGX Stain-Free Protein Gel (Biorad), along with Precision Plus Unstained Marker (Biorad) and run at 400 V for 20 minutes in 1 x Tris/Glycine/SDS buffer (Biorad). AnykD Mini-PROTEAN TGX Stain-Free Protein Gels (Biorad) were visualised using the ChemiDoc MP system (Biorad).

2.4 Western blotting and Dot blotting

SDS-PAGE was conducted using a 4-12 % Bis-tris NuPAGE gel (Novex) as described in section 2.3. Four pieces of Whatman filter paper, cut to approximately the size of an SDS-PAGE gel, were soaked in 1 x Tris/Glycine/SDS buffer (Biorad) supplemented with 20 % (v/v) ethanol, along with a piece of nitrocellulose membrane (Amersham

Protran 0.45 μm pore size, GE Healthcare), also cut to approximately the size of an SDS-PAGE gel. The transfer stack was assembled on the Semi-Dry Transfer Unit (Biorad) as follows: two pieces of Whatman filter paper, followed by the nitrocellulose membrane then the gel, pre-washed in in 1 x Tris/Glycine/SDS buffer (Biorad) supplemented with 20 % (v/v) ethanol, and then the remaining two pieces of Whatman filter paper. Proteins were transferred onto the membrane at 5 W for one hour. Following transfer, the membrane was blocked with 5 % (w/v) instant dried skimmed milk powder (Tesco) in Phosphate buffered saline with 0.05 % (v/v) TWEEN 20 (Sigma Aldrich) (PBST) for one hour and then incubated with the appropriate antibody at 4 °C overnight. The following antibody dilutions were used: horseradish peroxidase (HRP)-conjugated anti-His tag antibody (Sigma Aldrich) at a 1:10,000 dilution; mouse anti-GST tag antibody (Generon) at a 1:20,000 dilution; HRP-conjugated anti-GFP (Proteintech) at a 1:20,000 dilution and HRP-conjugated anti-Strep tag antibody (IBA) at a 1:10,000 dilution. All antibody dilutions were prepared in 20 ml PBST. For the mouse anti-GST tag (Generon) primary antibody, the membranes were washed three times with PBST, with each wash lasting five minutes, prior to the incubation with anti-mouse secondary antibody (Generon) at a 1:20,000 dilution for one hour at room temperature (RT). The membrane was then washed three times in PBST, with each wash lasting 5 minutes. SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific) was applied to the membrane, according to the manufacturer's protocol, and visualised using the Biorad ChemiDoc MP unit.

During dot blotting, the sample was pipetted directly onto the nitrocellulose membrane (Amersham Protran 0.45 μm pore size, GE Healthcare) using the Minifold 96-well Dot-blot system (GE Healthcare), according to the manufacturer's protocol. All steps subsequent to transfer are the same as the standard western blotting protocol previously described.

2.5 Membrane preparation

An *E. coli* cell pellet, harvested following expression of the target protein, was resuspended in PBS plus 2.5 mg DNAse (Sigma Aldrich), and lysed by passing the resuspended cells through a cell disrupter once at 30 kpsi and then once at 20 kpsi (Constant Systems). The lysate was centrifuged at 15,000 x g for 15 minutes at 4 °C using a JA-25.50 rotor (Beckman Coulter) in a J-26XP centrifuge (Beckman Coulter) and the supernatant then ultracentrifuged at 40,000 rpm for one hour at 4 °C using a 50.2 Ti rotor (Beckman Coulter). The supernatant was discarded and the pellet washed in PBS. The pellet was flash frozen in liquid N_2 and stored at -80 °C until required.

2.6 Size exclusion chromatography with multi-angle light scattering (SEC-MALS)

A Dawn Heleos II light scattering detector (Wyatt) and an Optilab T-rEX dRI detector (Wyatt) were connected downstream to a Biorad NCG Chromatography system for measurement following SEC. In order to equilibrate the Dawn Heleos II light scattering detector and Optilab T-rEX dRI detector prior to measurement, the SEC buffer for the protein of interest was flowed through the machine at a constant flow rate, until the baselines of the Dawn Heleos II light scattering detector and the Optilab T-rEX dRI detector were stable. A method was created using the Astra software package, ensuring that the flow rate recorded in the Astra software package matched that of the Biorad NCG Chromatography system and that the data collection duration was sufficient to record the full dataset. Sample injection into the Biorad NCG Chromatography system was accompanied by a sharp increase in light scattering, measured by the Dawn Heleos II light scattering detector. The data collection was initiated once the light scattering levels returned to approximately pre-injection levels. Data were processed using the ASTRA software package (Wyatt).

2.7 Mass spectrometry

Mass spectrometry to confirm the identity of the purified protein was conducted by the BSRC Mass Spectrometry and Proteomics facility at the University of St Andrews. The protein sample of interest was run on an SDS-PAGE gel gel, as described in section 2.3 and the SDS-PAGE gel gel was Coomassie Brilliant Blue stained to allow the visualisation of the protein/ proteins present in the sample. The band of interest was removed from the SDS-PAGE gel gel using a scalpel. Gel bands were reduced, alkylated and trypsin digested and then typically analysed by nano-high throughput liquid chromatography electrospray ionization mass spectrometry and tandem mass spectrometry. The resulting spectra were then compared to an internal database and the National Center for Biotechnology Information (NCBI) database.

2.8 Pre-crystallisation test (PCT)

Two 0.5 µl drops of protein were pipetted onto a glass coverslip and 0.5 µl of two different pre-crystallisation test (PCT) reagents (Hampton Research) were added, one per protein drop. The presence of precipitate was determined using a microscope and the protein concentration adjusted following the manufacturer's protocol.

2.9 Sitting-drop vapour diffusion crystallisation trials

Sitting-drop vapour diffusion plates were set up using 96-well Intelli-plates using the Gryphon Crystallisation Robot (Art Robbins Instruments), at 4 °C and 20 °C, with the following protein: mother liquor ratios: 1:1, 2:1 and 1:2 (1 corresponds to 0.15 µl here). In-house crystallisation screens along with the following commercially available crystallisation screens were used: MemGold (Molecular Dimensions), MemGold2 (Molecular Dimensions), JCSG-plus (Molecular Dimensions), Wizard 1&2 Classic (Molecular Dimensions), The PEGs Suite (Qiagen) and The PEGs II suite (Qiagen). In-house crystallisation screens were prepared using the custom screens service provided

by Molecular Dimensions. The components of the in-house screens discussed here can be found in appendix 8.4. Plates were stored at 4 °C or 20 °C, as required, in the Minstrel HT-UV imaging system (Rigaku). Samples were centrifuged at at 20,817 x *g* for 10 minutes at 4 °C to pellet any insoluble material prior to crystallisation trials.

2.10 In-house crystal screening

Crystals were cryoprotected in mother liquor supplemented with 20 % (v/v) glycerol and frozen in liquid N_2. These were tested using a Rigaku 007HFM rotating copper anode X-ray generator and Saturn 944 charge-coupled device (CCD) detector. This was conducted in collaboration with Dr Stephen McMahon, University of St Andrews.

2.11 Lipidic cubic phase (LCP) crystal plate preparation

LCP methods followed here (described in sections 2.11 and 2.12) were learnt during training with Dr Nicole Howe at the laboratory of Professor Martin Caffrey, Trinity College Dublin.

The 127.8 x 85.5 x 1 mm glass base plates (Marienfeld) were coated in Rain-X water repellent, left to dry and wiped to remove excess Rain-X. The plates were then rinsed in milliQ water and buffed with tissue. The 96-well spacer was then applied to the base plate and the plates were stored until required.

2.12 LCP crystallisation trials

Protein at between 12-15 mg/ml was applied to one LCP syringe (Art Robbins Instruments) and monoolein applied to a second pre-warmed LCP syringe. The required volume ratio of protein: monoolein was 2:3. The two syringes were connected using a

LCP coupling ferrule (Art Robbins Instruments) and the monoolein pushed towards the protein. The protein and monoolein were mixed approximately 100 times until the mixture appeared clear. If the mixture remained cloudy after mixing, the syringe was cooled on ice and the mixing step was repeated. If the mixture remained cloudy after cooling on ice, 2-3 µl of monoolein was added to the mixture and the mixing step repeated.

The clear mixture was transferred into one syringe, an short dispense needle (Art Robbins instruments) was connected to the syringe, and crystallisation trials were set up using the Crystal Gryphon LCP robot at RT (Art Robbins Instruments) according to the manufacturer's protocol, with 50 nl of LCP mixture and 800 nl of mother liquor per well. The LCP mixture was applied to the plate first followed by the mother liquor. The coverslip was then placed over the wells and the plate stored and imaged using the Minstrel HT-UV imaging system at 20 °C (Rigaku). In addition to the crystallisation screens described in section 2.9, the following commercially available crystallisation screens were used: MemMeso LCP screen (Molecular Dimensions) and JBScreen LCP (Jena Bioscience). Additional in-house screens were also used with the following components (Dr Nicole Howe and Professor Martin Caffrey, Trinity College Dublin, personal communication): each 96 well block contained either 8 % (v/v) MPD or 40 % (v/v) PEG 400, StockOptions Salts (Hampton Research) at 0.1 M and 0.4 M and each block contained a different buffer at 0.1 M. The following buffers were used: sodium citrate at pH 3.0 and 3.5, sodium acetate at pH 4.0, 4.5, 5.0 and 5.6, MES at pH 6.0 and 6.5, HEPES at pH 7.0 and 7.5 and Tris at pH 8.0 and 8.5. These additional screens were prepared using a Scorpion (Art Robbins Instruments) by Dr Audrey Le Bas. The components of StockOptions Salts (Hampton Research) are shown in appendix 8.5.

3 The expression and purification of the O-antigen polymerase Wzy for crystallisation trials

3.1 Introduction

The O-antigen polymerase Wzy catalyses the polymerisation of O-antigen repeat units on an undecaprenyl-phosphate acceptor molecule once it has been flipped into the periplasmic leaflet of the IM by the flippase Wzx, during both Wzy-dependent LPS and CPS biosynthesis (Hug and Feldman, 2011). Mutations in the *wzy* gene in multiple organisms have been shown to result in LPS with only one O-antigen repeat unit, supporting the role of Wzy as O-antigen polymerase (Raetz and Whitfield, 2002; Whitfield, 2006). Despite the overexpression of the *S. flexneri* homologue of Wzy first being described almost 20 years ago, little is known about the mechanism of Wzy (Daniels *et al*., 1998). Furthermore, until recently the majority of evidence supporting the aforementioned role of Wzy in O-antigen polymerisation came from genetics-based and *in vivo* studies (Raetz and Whitfield, 2002; Whitfield, 2006). The first *in vitro* evidence supporting the role of Wzy as the O-antigen polymerase came in 2010 from Woodward *et al.*, who showed that purified Wzy was the only component of the LPS biosynthesis pathway required for successful O-antigen polymerisation by reconstituting elements of the *E. coli* serotype O86 LPS biosynthetic pathway *in vitro* (Woodward *et al*., 2010). A major contributing factor to this lack of characterisation is that Wzy is an integral inner membrane protein consisting of between 10 and 13 TMHs, with a large periplasmic loop (Daniels *et al*., 1998). A membrane topology prediction of Wzy from *Chromobacterium violaceum* (referred to as Wzy_{CV}) is shown in figure 3.1

and this homologue will be the focus of the results presented here. This particular homologue is predicted to consist of 12 TMHs, with both the N- and C-termini located in the cytoplasm, and a large periplasmic loop is predicted to be located between TMHs 9 and 10 (Omasits *et al*., 2014). Furthermore, Wzy homologues from different organisms share a very low percentage sequence similarity, complicating the identification of residues essential to Wzy function (Raetz and Whitfield, 2002).

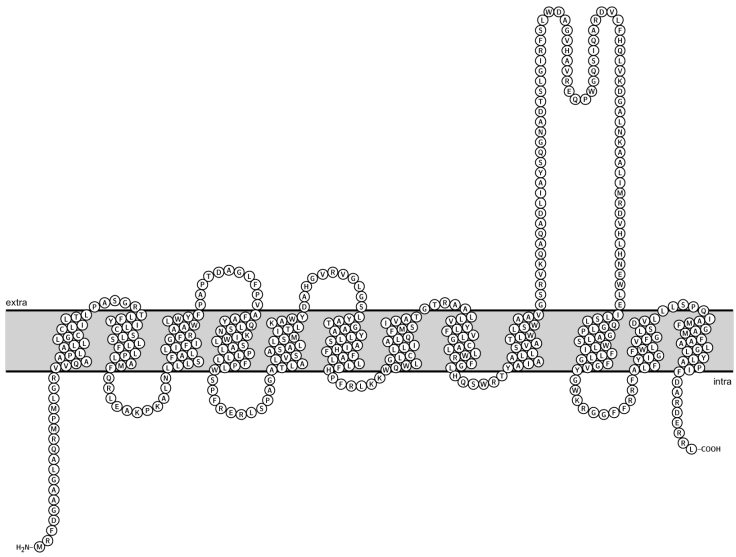


Figure 3.1. Predicted membrane topology diagram of Wzy_{CV}. This homologue consists of 429 amino acids and is 47.6 kDa in molecular weight. It is predicted to consist of 12 TMHs with both the N- and C-termini located in the cytoplasm and a large periplasmic loop located between TMHs 9 and 10. Membrane topology prediction created using Protter (Omasits *et al*., 2014).

Despite accounting for a predicted 30 % of the eukaryotic genome and for approximately 60 % of drug targets, unique structures of membrane proteins currently represent only 0.5 % of the total structures deposited in the PDB (http://blanco.biomol.uci.edu/mpstruc/ Stephen White lab, UC Irvine. Accessed: 24/6/17) (Moraes *et al*., 2014). A major bottleneck in the structural characterisation of membrane proteins is obtaining pure, monodisperse protein in sufficient quantities for crystallisation trials. This is due to numerous factors, found at different stages of the process, for example, low levels of expression or aggregation of the target protein during expression, both of which result in subsequent low yields. Another major factor that can contribute to low yields is the poor stability of membrane proteins once

extracted from their native membrane environment using detergents. Careful screening and optimisation of expression conditions, followed by the identification of suitable buffers and detergents for purification, is the key to successful membrane protein purification. Furthermore, membrane proteins are notoriously difficult to crystallise. Membrane proteins can crystallise in multiple different forms. Type I membrane protein crystals consist of 2D arrays of crystalline membrane proteins, with crystal contacts formed between both the hydrophobic transmembrane regions and the hydrophilic loops (Krauss *et al*., 2013). Type II crystals result from polar crystal contacts between the surface-exposed loops (Krauss *et al*., 2013). A cartoon representation of the protein arrangement in type I and type II crystals can be seen in figure 3.2, with the type I arrangement shown in 3.2a and the type II arrangement shown in figure 3.2b.

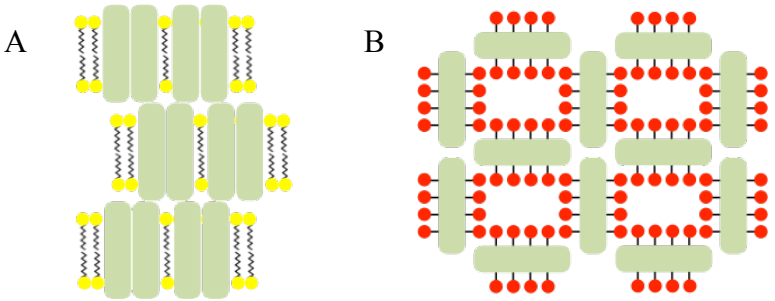


Figure 3.2. A comparison of type I and type II membrane protein crystals. Type I crystals (shown in A) contain contacts between both the hydrophobic TMHs and the hydrophilic loops of the membrane proteins, whereas type II crystals (shown in B) contain contacts between only the hydrophilic loops (Krauss *et al*., 2013). Type I crystals are formed usually in LCP and type II crystals are formed from membrane proteins in detergent micelles (Caffrey, 2015; Privé, 2007). Proteins are shown in green, lipids in yellow and black and detergents in red and black.

Most membrane proteins in detergent micelles form type II crystals, as the TMHs are shielded by the detergent micelle (Krauss *et al*., 2013). Type II crystals usually diffract poorly due to the low amount of possible crystal contacts (Krauss *et al*., 2013). Various crystallisation techniques have been developed with the aim of obtaining type I crystals, for example LCP (Caffrey, 2015). Approximately 16% of the membrane protein structures deposited in the PDB were solved using crystals grown in LCP (http://www.rcsb.org/pdb/home/home.do Accessed: 25/6/17). Briefly, LCP consists of a continuous lipid bilayer with aqueous channels and upon incorporation of a protein:

detergent complex (PDC) into LCP, the lipids displace the detergent from the target protein, surrounding the protein in a native-like lipid bilayer environment. The addition of a precipitant to this mixture during crystallisation trials results in phase separation between the LCP and the target protein. The target protein diffuses laterally in the bilayer, leading to areas of increased local protein concentration, which act as sites for nucleation of crystal growth (Caffrey, 2015).

In expressing and purifying multiple different homologues of Wzy, we aimed to obtain sufficient quantities of monodisperse protein for crystallisation trials, in the hope of obtaining crystals of sufficient quality to solve the structure of this protein. In solving the structure of this protein, we aimed to determine the mechanism of Wzy-dependent O-antigen polymerisation. In order to achieve this, multiple methods for the optimisation of buffer and detergent choice for purification were explored, with crystallisation trials conducted using both conventional sitting-drop vapour diffusion and LCP.

3.2 Materials and Methods

The recipes for buffers described here can be found in appendix 8.1.1.

3.2.1 Cloning

The gene encoding Wzy from *E. coli* serotype O9a:K30 was amplified via PCR from a plasmid received from Prof. Whitfield (University of Guelph, Canada) using the following primer sequences: Wzy K30 Reverse AAA-TTT-CTC-GAG-ATC-ATT-TAC-ATT-CCT-TCT-TCC-GA and Wzy K30 Forward AAA-TTT-CCA-TGG-TGA-GGA-CAC-TTT-CGA-AAG-CGA. 100 ng of plasmid, 2.5 µl dNTPs, 1 µl *Pfu* polymerase, 1 x *Pfu* buffer, and 2.5 µl of the appropriate primers were mixed in a total volume of 50 µl with milli Q water. The following PCR programme was used: 1 cycle of 4 minutes at 98 °C, then 30 cycles of 30 seconds at 98 °C, 30 seconds at 55 °C and 2 minutes at 72 °C, followed by 1 7 minute cycle at 72 °C. PCR products were then run on a 1 % agarose gel, prepared using 0.5 x Tris-borate-ethylenediaminetetraacetic acid

buffer (TBE) and 1 µl of SYBR safe stain (Biorad), at 90 V for 30 minutes. Bands were visualised using a Biorad ChemiDoc MP and purified from the gel using a QIAquick Gel Extraction Kit (Qiagen). Approximately 100 ng of PCR product and 200 ng of expression vector were then digested using 0.5 µl *Nco*I and 0.5 µl *Xho*I with 1 x Buffer D (Promega) (made up to a total volume of 20 µl with milliQ water), for two hours at 37 °C. 1 µl of alkaline phosphatase (Promega) was added to the vector digest and this was left at 37 °C for a further 30 minutes. The *wzy* K30 gene was ligated into the following digested in-house expression vectors using T4 DNA ligase, with 1 x T4 DNA ligase buffer (Promega), at RT overnight: pEBSRCTEVC10HIS and pBADTEVC10His. Both vectors encode a C-terminal Tobacco etch virus (TEV) protease-cleavable His_{10} tag, but possess a T7 and *araC* promoter, respectively. Both of these vectors confer ampicillin resistance and were constructed by Dr Huanting Liu (University of St Andrews). Vector maps of the pEBSRCTEVC10HIS and pBADTEVC10HIS vectors can be seen in figure 3.3.

Genes encoding Wzy homologues from *Chromobacterium violaceum* and *Nitrosospira multiformis* were provided by the Joint Centre for Structural Genomics (JSCG, USA; http://www.jcsg.org), as part of the JCSG high-throughput structural biology pipeline, in the pSpeedET vector, which encodes an N-terminal TEV protease-cleavable His_6 tag, and both an *araC* and T7 promoter. These genes were subcloned into the pBADTEVC10HIS vector, using *BspH*I and *Xho*I restriction sites, by Dr Hui Liu (University of St Andrews). The gene encoding Wzy from *Pectobacterium atrosepticum* SCRI 1043 was amplified from genomic DNA and cloned into the pEBSRCTEVC10HIS and pBADTEVC10HIS vectors, using *Nco*I and *Xho*I restriction sites, by Dr Hui Liu (University of St. Andrews).

The genes encoding the remaining 10 Wzy homologues studied here were identified from the following database: http://www-archbac.u-psud.fr/genomics/phylotaxBlast.html (Altschul *et al*., 1997). These were ordered from Eurofins Genomics, including either an *Nco*I or *BspH*I site at the 5' end and an *Xho*I site at the 3' end. 1 mg of DNA was digested with the appropriate enzyme combination and

ligated into the pEBSRCTEVC10HIS and pBADTEVC10HIS vectors at RT overnight using T4 DNA ligase (Promega), as previously described.

The genes encoding Wzy from *E. coli* serotype O9a:K30 and *Clostridium thermocellum* were also ligated, as described previously, into the pBADMISTICTEV10HIS vector. This vector encodes an N-terminal Membrane-integrating sequence for translation of integral membrane constructs (MISTIC) tag (Kefala et al., 2007) and an N-terminal His tag, both of which can be removed from the target protein by TEV protease cleavage, and a C-terminal TEV-cleavable His_{10} tag. This vector was constructed by Dr Huanting Liu (University of St Andrews). The vector map for the pBADMISTICTEV10HIS vector can be seen in figure 3.3.

Ligation mixtures were transformed into *E. coli* DH5 α cells (Bioline), as described in section 2.1. Subsequent olonies were grown in 10 ml LB-medium at 37 °C overnight in the presence of the appropriate antibiotic and DNA was purified from the resulting cell pellet using a QIAprep Spin Miniprep kit (Qiagen).

The presence of insert was determined using a control digest using *Nco*1 and *Xho*I at 37 °C for two hours, followed by agarose gel analysis. Sequencing was performed using primers against the vector backbone (Eurofins; GATC).

A list of all Wzy homologues studied here can be found in table 3.1.

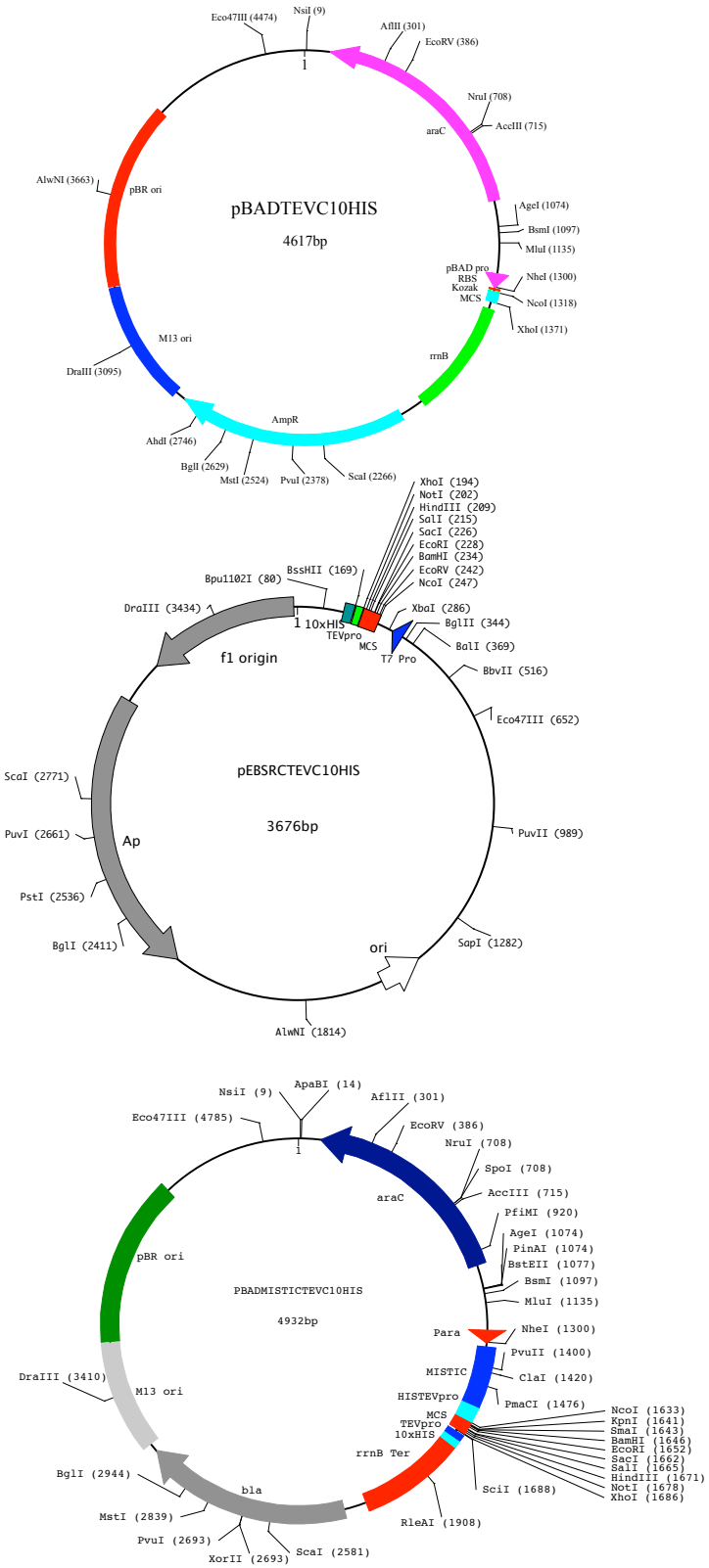


Figure 3.3. pBADTEVC10His, pEBSRCTEV10His and pBADMISTICTEV10HIS vector maps. Vectors used here were constructed by Dr Huanting Liu (University of St Andrews) and vector maps were provided by Dr Huanting Liu (unpublished data).

Homologue	Vector	Cell line	Growth medium	OD_{600} at induction	Level of inducer (% (w/v) arabinose for pBAD vector or mM for pET vector)	Growth temperature after induction (°C)	Duration of growth after induction
E. coli O9:K30	pET	C43	TB	0.6	0.5	37	3 hours
Chromobacterium violaceum	pBAD	C43	LB	0.6	0.02	37	3 hours
Nitrosospira multiformis	pBAD	C43	LB	1.0	0.2	25	4 hours
Clostridium thermocellum	pBAD	C43	TB	1.0	0.2	15	Overnight
Pectobacterium atrosepticum	pBAD	C43	LB	1.0	0.02	25	4 hours
Actinobacillus pleuropneumoniae	pET	C43	LB	0.6	0.5	25	Overnight
Thermomicrobium roseum	pBAD	C41	LB	0.6	0.2	25	Overnight
Thermodesulfo-bacterium geofontis	pBAD	C41	LB	1.0	0.2	15	Overnight
Rhodothermus marinus	pET	C41	LB	1.0	0.1	15	Overnight
Sulfobacillus acidophilus	pBAD	C43	LB	0.6	0.2	15	Overnight
Thermocrinus albus	pBAD	C43	LB	0.6	0.2	25	Overnight
Marinotoga piezophila	pBAD	C43	LB	1.0	0.02	25	Overnight
Marinomonas mediterranae	pBAD	C43	LB	0.6	0.2	37	3 hours
Streptococcus thermophilium	pBAD	C41	LB	0.6	0.2	37	4 hours

Table 3.1. Large-scale expression conditions of each Wzy homologue tested.

3.2.2 Large-scale expression

Plasmids were transformed into *E. coli* C41 (DE3) and C43 (DE3) expression cell lines as described in section 2.1. The appropriate expression conditions were determined using the expression trials protocol described in section 2.2, and expression was confirmed via SDS-PAGE and western blotting against the affinity tag, detailed in sections 2.3 and 2.4, respectively. Large-scale expression conditions identified using this approach and subsequently used for all homologues tested here can be seen in table 3.1. Following expression, cells were harvested at 6,000 rpm for 10 minutes at 4 °C

using a JLA-8.1000 rotor (Beckman Coulter) in a J-26XP centrifuge (Beckman Coulter).

3.2.3 Membrane solubilisation detergent screening

An *E. coli* pellet, obtained from 1 L *E. coli* cell culture, was resuspended in PBS, lysed using a cell disrupter (Constant Systems) at 30 kpsi and centrifuged at 15,000 x g for 15 minutes at 4 °C using a JA 25.50 rotor (Beckman Coulter) in a J-26XP centrifuge (Beckman Coulter). The supernatant was ultracentrifuged at 80,000 rpm for 10 minutes at 4 °C using a TLA-100 rotor (Beckman Coulter) in an Optimax Max-XP centrifuge (Beckman Coulter). Each pellet was resuspended in PBS plus 1 % (w/v) detergent (or 1 % v/v in the case of Elugent and $C_{12}E_8$) and solubilised at 4 °C for two hours and the ultracentrifugation step was repeated. The detergents tested are shown in table 3.2. Samples of the supernatant and pellet were taken for analysis via SDS-PAGE and western blotting against the affinity tag, as described in sections 2.3 and 2.4, respectively.

	Detergent (1 %)	Detergent Abbreviation	CMC (%)	Micelle size (kDa)
1	n-Dodecyl-β-D-Maltopyranoside	DDM	0.0087	72
2	n-Decyl-β-D-Maltopyranoside	DM	0.087	33
3	n-Octyl-β-D-Glucopyranoside	OG	0.53	8-27
4	Lauryl Maltose Neopentyl Glycol	LMNG	0.001	393
5	n-Dodecyl-N,N-Dimethylamine-N-Oxide	LDAO	0.023	17
6	Anapoe $C_{12}E_8$	$C_{12}E_8$	0.0048	66
7	5-Cyclohexyl-1-Pentyl-β-D-Maltoside (Cymal-5)	Cym5	0.12	23
8	Elugent	-	-	-
9	Decanoyl-N-Methylglucamide	MEGA-10	0.21	-
10	3-[(cholamidopropyl)dimethylammonio]-1-propanesulfonate	CHAPS	0.49	6
11	Fos Choline-12	Fos12	0.047	19

Table 3.2. Membrane solubilisation detergent screen. Detergents used here for the membrane solubilisation detergent screen described in section 3.2.3. All detergents were used at 1% (w/v), apart from $C_{12}E_8$ and Elugent that were used at 1% (v/v). Critical micelle concentrations (CMC) and micelle sizes were obtained from www.anatrace.com. The CMC of LMNG was obtained from Chaptal *et al*. (Chaptal *et al*., 2017).

3.2.4 Nickel affinity column buffer screening

An *E. coli* pellet, obtained from 1 L of *E. coli* cell culture, was resuspended in PBS, lysed using a cell disrupter (Constant systems) at 30 kpsi and centrifuged at 15,000 x g for 15 minutes at 4 °C using a JA 25.50 rotor (Beckman Coulter) in a J-26XP centrifuge (Beckman Coulter). The supernatant was ultracentrifuged at 50,000 rpm for 30 minutes at 4 °C using a TLA-110 rotor (Beckman Coulter) in an Optimax Max-XP centrifuge (Beckman Coulter). Each pellet was resuspended in a chosen buffer, shown in table 3.3, plus 1% (w/v) dodecyl maltoside (DDM) and solubilised at 4 °C two hours. Following solubilisation in the test buffers, the ultracentrifugation step was repeated and 500 µl of the supernatant was taken and run on a Biosprint 15 (Qiagen) with 50 µl MagneHis beads (Promega) per sample (both added to well 1). Beads were washed with 2 x 500 µl of the test buffer plus 0.024 % (w/v) DDM (in wells 2 and 3) and bound protein was eluted in 2 x 50 µl of test buffer plus 0.024 % (w/v) DDM), with 250 mM imidazole (in wells 4 and 5). A sample of the eluate (well 4) was taken for analysis via SDS-PAGE in combination with western blotting against the His tag, as described in sections 2.3 and 2.4, respectively.

	Original Buffer Screen
1	20 mM Tris pH 7.5, 300 mM NaCl, 20 mM imidazole
2	20 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole
3	20 mM Tris pH 7.5, 750 mM NaCl, 20 mM imidazole
4	20 mM Tris pH 7.5, 1 M NaCl, 20 mM imidazole
5	20 mM Tris pH 6.8, 500 mM NaCl, 20 mM imidazole
6	20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole
7	20 mM Tris pH 8.5, 500 mM NaCl, 20 mM imidazole
8	20 mM Tris pH 7.5, 500 mM NaCl, 20% glycerol, 20 mM imidazole
9	20 mM Tris pH 7.5, 500 mM NaCl, 50 mM imidazole
10	20 mM Tris pH 7.5, 500 mM NaCl, 100 mM imidazole
11	20 mM Tris pH 7.5, 500 mM NaCl, 5% (v/v) ethanol, 20 mM imidazole
12	20 mM Tris pH 7.5, 500 mM NaCl, 20 mM BME, 20 mM imidazole
13	20 mM Tris pH 7.5, 500 mM NaCl, 25 mM glucose, 20 mM imidazole
14	20 mM Tris pH 7.5, 500 mM NaCl, 100 mM glycine, 20 mM imidazole
15	20 mM Tris pH 7.5, 500 mM NaCl, 100 mM arginine, 20 mM imidazole

Table 3.3. Buffer screen. Buffers used for the buffer screen described in section 3.2.4.

3.2.5 Differential filtration detergent screening

This protocol is based on the Analytic Selector Kit (Anatrace) (Vergis *et al*., 2015). Approximately 1 mg of decyl maltoside (DM)-purified His-tagged protein, in a volume of 5 ml of chosen buffer without imidazole, was applied to 2.4 ml nickel sepharose Fast Flow resin (GE Healthcare) overnight. 50 µl of protein-bound nickel resin was added to each well of a 0.22 µm 96-well filter plate. Contaminant proteins were removed by washing each well with 150 µl of the chosen buffer plus DM at 3 x CMC (wash and elution buffers were prepared by mixing two-times concentrated buffer and two-times concentrated detergent) six times via centrifugation at 218 x *g* for five minutes. The nickel-bound protein was exchanged into a new detergent; one detergent or detergent mixture per well. Well A1 was kept in DM as a positive control. Well A2 was exchanged into buffer without detergent as a negative control. Each well was washed seven times with 30 µl of chosen buffer plus new detergent, or detergent mixture, at 3 x CMC via centrifugation at 218 x *g* for five minutes. The protein was then eluted in 70 µl of the chosen buffer plus 400 mM imidazole plus the new detergent at 3 x CMC via centrifugation at 218 x *g* for five minutes. 30 µl of the eluate was applied to a filter plate with a 100 kDa Mw cut-off and 30 µl was applied to a filter plate with a 300 kDa Mw cut-off (Agilent Technologies). These were centrifuged at 218 x *g* for five minutes. The samples passed through the filter plates were analysed via dot-blotting against the His-tag, as described in section 2.4. A list of the detergents and detergent mixtures used in this assay can be found in appendix 8.2.

3.2.6 His-tagged inner membrane protein purification

The components of buffers described here can be found in appendix 8.1.1. Membranes from *E. coli* expressing Wzy_{CV}, prepared as described in section 2.5, was homogenised and solubilised in 100 ml of Wzy_{CV} solubilisation buffer plus 1 % (w/v) detergent, for two hours at 4 °C. Suitable buffer and detergents were identified as described in sections 3.2.4 and 3.2.5, respectively. Following solubilisation, the sample was ultracentrifuged at 40,000 rpm for one hour at 4 °C using a 50.2 Ti rotor (Beckman

Coulter) and the supernatant applied to 2 ml high density nickel agarose resin (ABT) and incubated for one hour at 4 °C. Following this incubation, the mixture was applied to an empty column and the beads were washed with 50 column volumes (CV) of Wzy_{CV} wash buffer 1, followed by 50 CV of the Wzy_{CV} wash buffer 2. The protein was eluted in 10 CV of Wzy_{CV} elution buffer (all buffers contain the chosen detergent at 3 x CMC). The presence of the target protein in the eluate fraction was confirmed via SDS-PAGE, as described in section 2.3. The eluate fraction was concentrated at 3,488 x *g* at 4 °C using a Vivaspin 20 concentrator with a 50 kDa Mw cut-off (GE Healthcare). The protein was resuspended every 10 minutes, until a volume of 10 ml was reached. This was then applied to a CentriPure P100 Desalting column (EMP Biotech), pre-equilibrated in Wzy_{CV} desalt buffer, and the protein eluted using 14 ml Wzy_{CV} desalt buffer. 1 mg of TEV protease was added to the sample and incubated overnight at 4 °C. Following TEV protease cleavage, the protein was applied to 2 ml nickel Sepharose Fast Flow resin (GE Healthcare) and incubated for one hour at 4 °C. The mixture was applied to an empty column and the beads washed with 20 ml desalting buffer and the non-cleaved protein eluted using 10 ml Wzy_{CV} elution buffer. The presence of TEV protease-cleaved target protein in the FT (FT) and wash was confirmed via SDS-PAGE, as described in section 2.3. The TEV protease-cleaved protein was pooled and concentrated to a volume of 1 ml using a Vivaspin 20 concentrator with a 50 kDa Mw cut-off (GE Healthcare) at 3,488 x *g* at 4 °C, resuspending every 10 minutes during the course of the concentration process. Once the target volume had been reached, the sample was centrifuged at 20,817 x *g* for 10 minutes at 4 °C (Eppendorf 5417R centrifuge) and applied to either a superdex 200 10/300 GL SEC column or a Hiload 16/60 superdex 200 pg size SEC column (GE Healthcare), pre-equilibrated in Wzy_{CV} SEC buffer, using a Biorad NCG Chromatography System. Fractions containing target protein were identified and the purity assessed via SDS-PAGE, as described in section 2.3. Protein was pooled and concentrated using a Vivaspin 20 concentrator with a 50 kDa Mw cut-off (GE Healthcare) at 3,488 x *g* at 4 °C. The sample was resuspended every 10 minutes until the desired concentration was reached.

If different detergents were used for solubilisation and purification, detergent exchange was conducted on the first nickel affinity column, with the resin was first washed with

50 CV Wzy_{CV} wash buffer 1 with the solubilisation detergent at 3 x CMC, then with 25 CV Wzy_{CV} wash buffer 1 with the new detergent at 3 x CMC and then the remaining purification steps were conducted with the new detergent at 3 x CMC.

The same purification protocol was also used for MISTIC-tagged constructs and both the MISTIC-tag and the His-tag were removed from the target protein via TEV protease-cleavage. The His-tagged MISTIC tag was separated from the target protein during the second nickel affinity column (Kefala *et al*., 2007).

SEC-MALS was conducted as described in section 2.6.

Purified protein was used for sitting-drop vapour diffusion and LCP crystallisation trials, detailed in sections 2.9 and 2.12, respectively.

3.2.7 Differential scanning fluorimetry (DSF) buffer screening

Protein at a final concentration of 3.5 µM was incubated with 35 µM 7-diethyl-3-[4'-(iodoacetamido)phenyl]-4-methylcoumarin (DCIA) (Sigma Aldrich) (Branigan *et al*., 2013) with the RUBIC buffer screen (Molecular Dimensions) (Boivin *et al*., 2013; Newman, 2004) in a final volume of 50 µl. The components of the RUBIC buffer screen can be found in appendix 8.3. The samples were heated for one minute at 25 °C. The temperature was increased by 1 °C increments, each of one minute, up to a final temperature of 95 °C. The samples were heated and DCIA fluorescence was measured using a Stratagene Mx 3005-P QPCR machine using the MxPro software package. Data was interpreted using the Prism software in collaboration with Dr Clarissa Melo Czekster (University of St. Andrews).

3.3 Results

3.3.1 Expression of C-terminal TEV protease-cleavable His_{10} tagged Wzy homologues

A total of 16 different Wzy homologues were cloned into our in-house pBADBSRCCTEVHis10 and pEBSRCCTEVHis10 vectors, shown in figure 3.3. These vectors both encode a TEV protease-cleavable His_{10} affinity tag at the C-terminus and were transformed into *E. coli* C43 (DE3) and C41 (DE3) cells. These were chosen as they were developed for the overexpression of membrane proteins and other proteins normally toxic to *E. coli* (Miroux and Walker, 1996). Small-scale expression trials were conducted to identify the appropriate expression conditions for each construct, with samples induced at two different OD_{600} values with a low and high level of inducer, either 0.02 % or 0.2 % (w/v) of arabinose for the *araC* promoter-based vector or 0.1 mM or 0.5 mM IPTG for the T7 promoter-based vector. Following induction, small-scale cultures were grown for either 3 hours at 37 °C or 4 hours at 25 °C or overnight at either 15 °C or 25 °C. A flow chart showing the expression trial strategy used can be seen in figure 3.4. A western blot, against the His_{10}-affinity tag, of the expression trials for Wzy_{CV} in the pBADBSRCCTEVHis10 vector can be seen in figure 3.5. Highest expression could be obtained following induction at OD_{600} of 0.6 with 0.2 % (w/v) arabinose, with subsequent growth at 25 °C for 4 hours, with induction with 0.02 % (w/v) resulting in no detectable target protein expression. The same approach was followed for each construct, with the optimal conditions identified during expression trials and subsequently used for large-scale expression of each homologue shown in table 3.1.

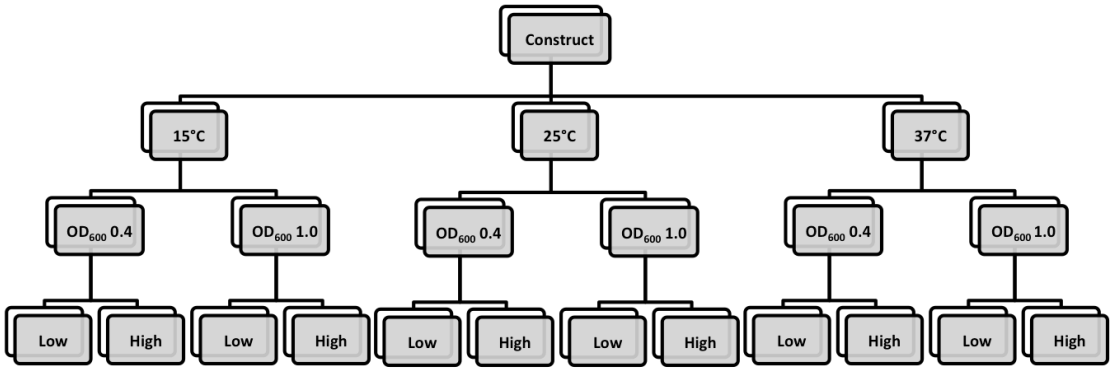


Figure 3.4. Expression trials flow chart. Once a construct had been cloned (level 1), different growth temperatures after induction (level 2), different OD_{600} levels at induction (level 3) and different levels of inducer (level 4) were screened. "Low" corresponds to 0.02% (w/v) arabinose for the pBADBSRCCTEVHis10 vector and 0.1 mM IPTG for the pEBSRCECTEVHis10 vector. "High" corresponds to 0.2% (w/v) arabinose for the pBADBSRCCTEVHis10 and 0.5 mM IPTG for the pEBSRCCTEVHis10 vector.

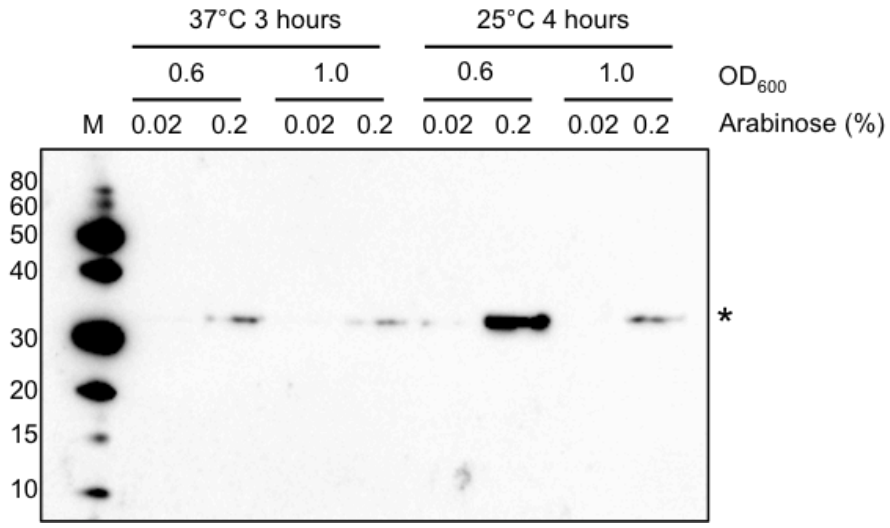


Figure 3.5. Western blot probed with an anti-His tag antibody of whole cells following Wzy_{CV} expression trials. The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*". Cell cultures were normalised to an OD_{600} of 1 following expression trials in order to allow the comparison of different expression conditions. The highest level of expression resulted from growth at 25 °C for 4 hours following induction of expression with 0.2% (w/v) arabinose at an OD_{600} of 0.6.

3.3.2 Membrane solubilisation detergent screening

Once suitable expression conditions had been identified, detergents that are able to extract the target protein from the membrane, without denaturing the target protein, need to be identified. Different detergents were screened on their ability to extract the target protein from the membrane, as confirmed via SDS-PAGE of the soluble fraction, remaining after centrifugation of the solubilisation mixture, and the insoluble pellet. The use of a detergent suitable for the extraction of the target protein from the membrane would result in the majority of the target protein being found in the detergent-soluble fraction. Table 3.2 shows the detergents tested for the ability to extract Wzy_{CV} from the membrane. An SDS-PAGE gel of the soluble and insoluble fractions following extraction of the target protein from the membrane using different detergents can be seen in figure 3.6, with the band corresponding to Wzy_{CV} at approximately 35 kDa indicated with a "*". Most detergents tested here, for example, DDM (detergent 1) and cym5 (detergent 7), could extract Wzy_{CV} from the membrane, as determined by the presence of the majority of the protein in the detergent-soluble fraction following centrifugation of the membrane:detergent mixture after solubilisation. However, the use of OG (detergent 3) and CHAPS (detergent 10) was not successful in extracting the target protein from the membrane as the majority of the target protein could be found in the detergent-insoluble pellet.

3.3.3 Nickel affinity column buffer screening

Different buffers were screened on their resulting target protein yield following a small-scale nickel affinity column, as well as the presence of contaminating proteins. Table 3.3 shows the buffers screened. DDM was used during the buffer screening process as this detergent performed well in the detergent screen and is one of the most commonly used detergents for IM protein purification, as it is non-denaturing (Moraes *et al.*, 2014). Figure 3.7 shows an SDS-PAGE gel of the eluates from the nickel affinity columns conducted during buffer screening, with each lane corresponding to the eluate of the target protein in a different buffer. The target protein can be seen at

approximately 35 kDa, indicated with a "*". All buffers tested enabled the purification of Wzy_{CV}, with the inclusion of the additive glycerol at 20% (v/v) resulting in the highest yield. The increase of the pH from 7.5 to 8.0 also resulted in an increase in yield. The inclusion of arginine or glycine at 100 mM resulted in a decreased level of a contaminant protein, seen at approximately 45 kDa, in the eluate of the nickel affinity column, without resulting in a decreased yield of target protein.

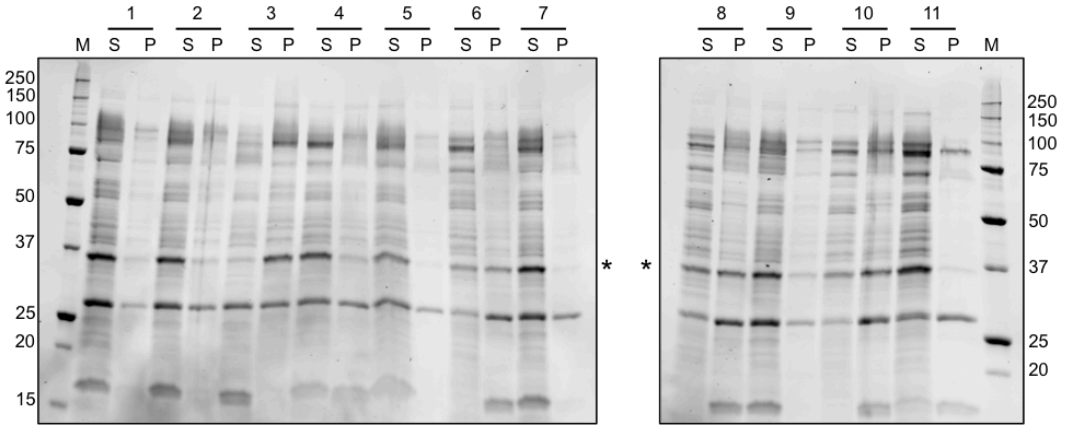


Figure 3.6. SDS-PAGE gel of Wzy_{CV} membrane solubilisation detergent screen. The band corresponding to Wzy_{CV} at 35 kDa is indicated with a "*". "M" corresponds to marker; "S" corresponds to the detergent-soluble fraction; "P" corresponds to the detergent-insoluble fraction; numbers 1-11 indicate the detergents used, which corresponds to those in table 3.2.

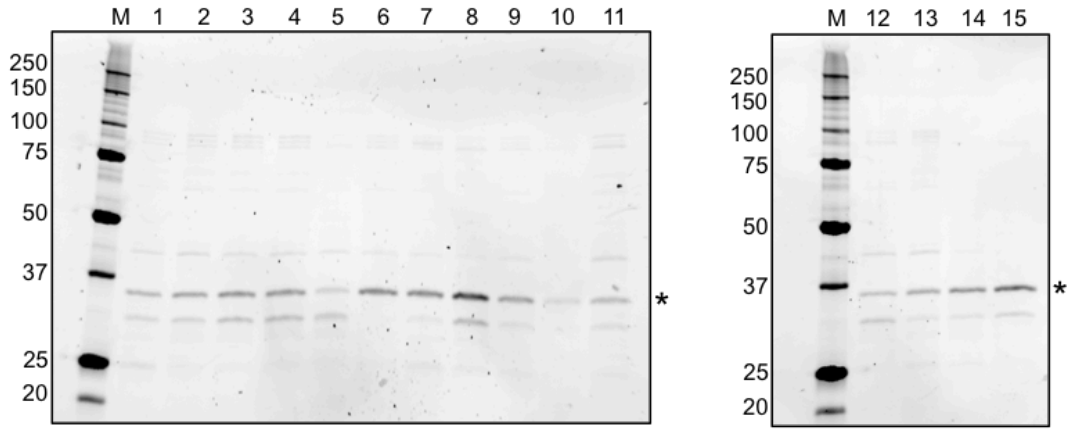
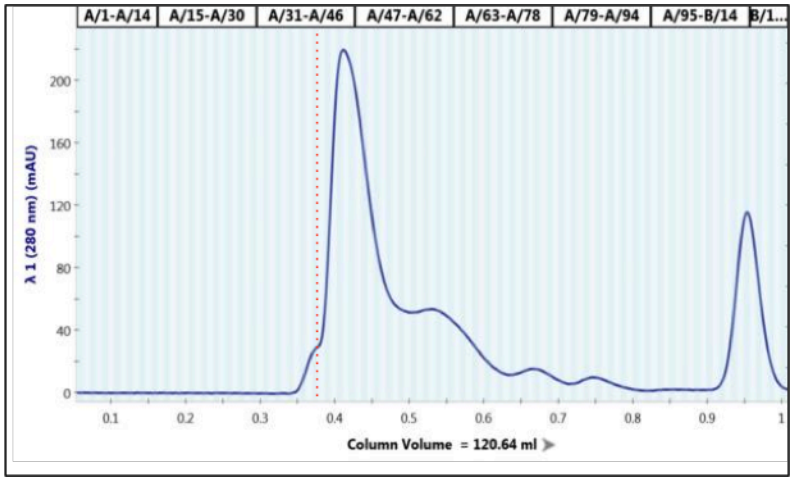


Figure 3.7. SDS-PAGE gel of Wzy_{CV} buffer screen. The band corresponding to Wzy_{CV}, purified in the presence of different buffers on a small-scale, is indicated with a "*". The numbers for the buffers correspond to those used in table 3.3. The addition of glycerol at 20 % (v/v) (lane 8) resulted in the highest yield and the increase of the pH from 7.5 (lane 2) to 8 (lane 6) also resulted in an increase in yield. The addition of arginine or glycine at 100 mM (lanes 14 and 15, respectively) decreased the level of a contaminant at 45 kDa.

3.3.4 Large-scale purification of Wzy_{CV}

Once a suitable combination of detergent and buffer for purification had been identified on a small-scale, purification was conducted on a larger scale. Large-scale purification of the C-terminal TEV protease-cleavable His_{10} affinity tagged proteins expressed followed the following scheme: extraction of the target protein from the membrane using the detergent and buffer combination identified during screening, followed by the purification of the target protein via a nickel affinity column, with the protein eluted from the affinity resin using imidazole. The target protein was dialysed to remove the excess imidazole and cleaved with TEV protease overnight. The sample was reapplied to a nickel affinity column to remove the contaminant proteins and the His-tagged TEV protease. The TEV protease-cleaved target protein, present in the FT from the second nickel affinity column, was then concentrated for SEC, with the aim of obtaining the elution of the protein as a single peak, not in the void volume of the column, indicating that the purified protein is not aggregated and monodisperse. The most suitable buffer for purification of Wzy_{CV} was identified as 20 mM Tris pH 7.5, 500 mM NaCl, 100 mM arginine, shown in figure 3.7, with DDM identified as a suitable detergent, shown in figure 3.6. This combination was therefore chosen for the initial large-scale purification, with the pH increased to pH 8.0, see figure 3.7. The target protein yield following the first nickel affinity column was 0.74 mg/L of *E. coli* culture, this decreased to 0.6 mg/L following TEV protease-cleavage of the His_{10}-tag and the second nickel affinity column. However, the purified protein eluted in the void volume of a hiload 16/60 superdex 200 pg column, as seen in figure 3.8a, with the position of the void volume indicated with a dotted red line. It was therefore deemed to be aggregated. The SDS-PAGE gel of the SEC fractions is shown in figure 3.8b, with higher order SDS-stable oligomers visible, also indicative of aggregation.

A



B

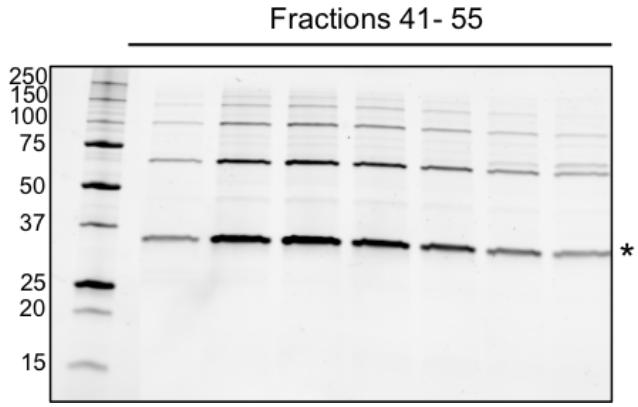


Figure 3.8. Wzy_{CV} purified as an aggregate in DDM. Following the upscale of the conditions identified in sections 3.3.2 and 3.3.3 for Wzy_{CV}, Wzy_{CV} eluted from the SEC column as an aggregate, as indicated by its presence in the void volume (A). The position of the void is indicated with a red dotted line and the fraction numbers are indicated. The fractions corresponding to the SEC peak were run on SDS-PAGE (B). The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*". In addition, SDS-stable oligomers can also be seen, again indicating aggregation of the target protein.

Following lengthy optimisation of the buffer conditions for purification to include the additional additives TCEP at 2.5 mM and glycerol at 10 % (v/v), the modification of the expression conditions from growth for 4 hours at 25 °C with induction of expression at OD_{600} of 0.6 with 0.2 % (w/v) arabinose to growth for 3 hours at 37 °C with induction of expression at OD_{600} of 0.6 with 0.02 % (w/v) arabinose, and the transfer of the purification to 4 °C from RT, the protein could be purified in a non-aggregated form. The overlay of the SEC profiles obtained at RT and at 4 °C, using 20 mM Tris pH 8.0, 500 mM NaCl, 100 mM arginine, 2.5 mM TCEP, 3 x CMC DDM, can be seen in figure

3.9, with the purification at 4 °C resulting in a peak with a decreased shoulder at approximately 0.5 CV. Figure 3.10 shows the SDS-PAGE gel of the final optimised first and second nickel affinity columns, as well as the SDS-PAGE gel of the SEC fractions and the corresponding SEC profile. The addition of glycerol at 10 % (v/v) could further reduce the shoulder at approximately 0.5 CV, shown by comparing figures 3.9 and 3.10.

Purified protein from the final optimised purification, shown in figure 3.10, was used to set up crystallisation trials using conventional sitting-drop vapour diffusion at both RT and 4 °C, and using LCP, however no crystals could be obtained for this homologue in this buffer and detergent combination.

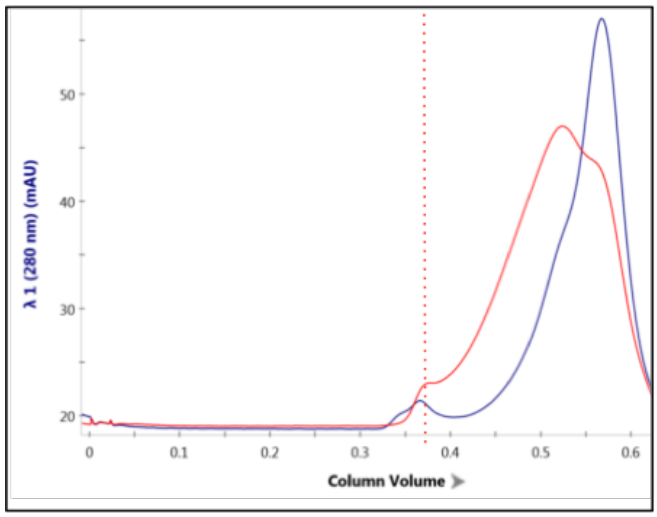


Figure 3.9. An overlay of the SEC profiles of Wzy_{CV} purified at RT and at 4 °C. The profile obtained at RT is shown in red and that obtained at 4 °C in blue. The position of the void volume of the SEC column is indicated with a dotted red line.

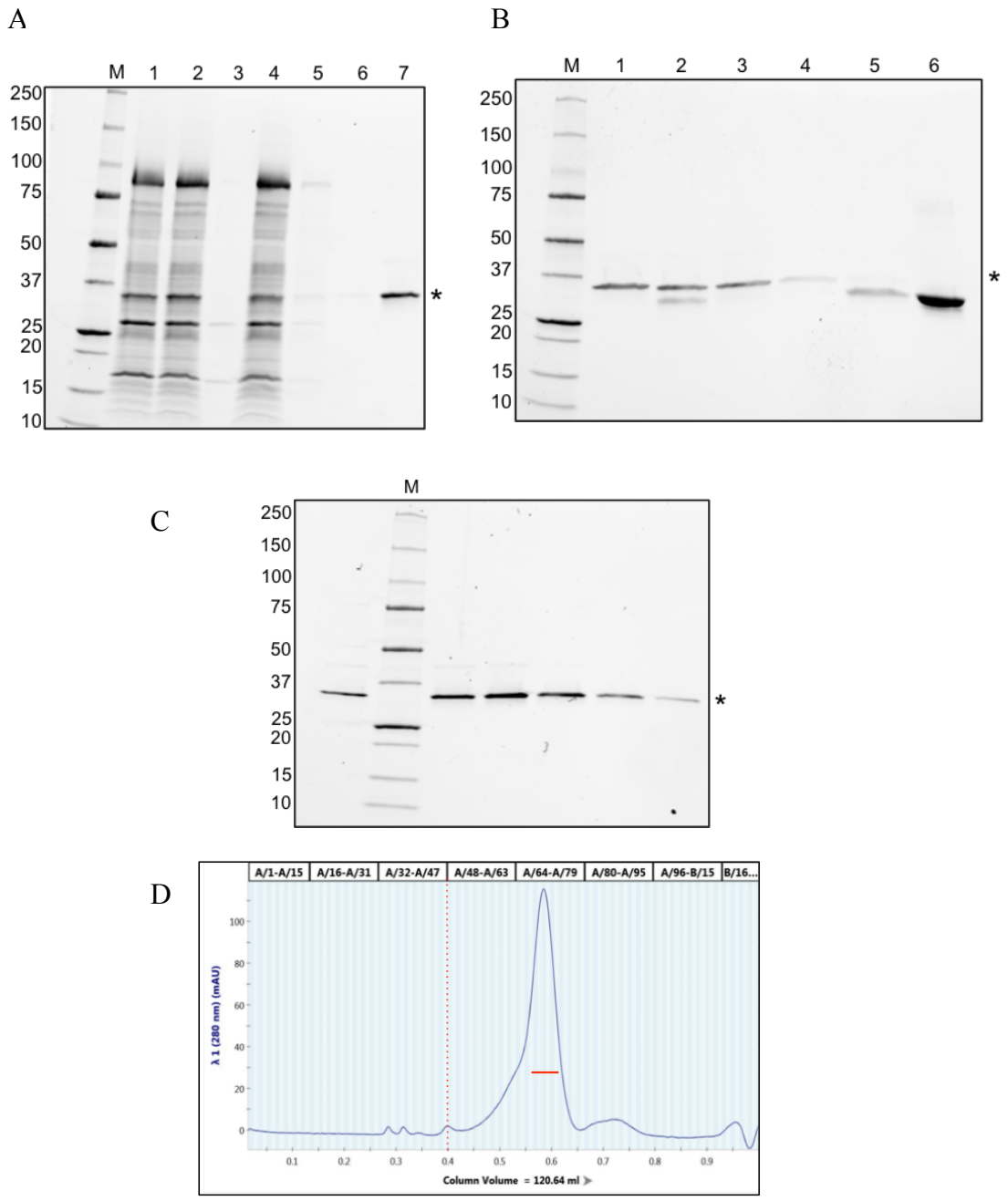


Figure 3.10. Optimised purification of Wzy_{CV} in DDM. "A" shows the SDS-PAGE gel of the first nickel affinity column (M: Marker; 1: membrane solubilisation; 2: DDM-soluble fraction; 3: DDM-insoluble fraction; 4: nickel affinity FT; 5: nickel affinity wash 1; 6: nickel affinity wash 2; 7: nickel affinity eluate). "B" shows the SDS-PAGE gel of second nickel affinity column (M: Marker; 1: first nickel affinity eluate; 2: TEV protease-cleaved sample; 3: second nickel affinity FT; 4: second nickel affinity wash; 5: second nickel affinity eluate; 6: TEV protease). "C" shows the SDS-PAGE gel of the SEC peak shown in "D", with the position of the void volume indicated with a dotted red line. The approximate position of the fractions in "D" shown in "C" is shown by the solid red line. The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*".

3.3.5 MISTIC-tag to improve expression levels

As a major improvement to the purification of Wzy_{CV} resulted from decreasing the expression levels, we decided to explore different tagging strategies in order to improve the expression levels and solubility of Wzy. Here we use the example of Wzy from *Clostridium thermocellum*, referred to as Wzy_{CT}. MISTIC, is a 13 kDa membrane-associated protein from *Bacillus subtilis*, which is believed to aid in targeting of membrane proteins to the membrane (Roosild *et al*., 2005). MISTIC-tagging has previously been shown to result in an increased level of overexpression of members of the histidine kinase receptor family in *E. coli* (Kefala *et al*., 2007). A vector map of the pBADMISTICTEV10HIS vector constructed by Dr Huanting Liu, University of St. Andrews, can be seen in figure 3.3. A whole cell western blot against the His_{10}-affinity tag of the expression trials of MISTIC-tagged Wzy_{CT} can be seen in figure 3.11, with lane 1 (indicated with "WT") showing the expression level obtained without the addition of the MISTIC tag. The addition of the MISTIC tag results in much higher expression levels in the pBAD vector than the pET vector, with both vectors resulting in higher expression levels than the non-MISTIC tagged construct. However, the increased expression levels resulting from the MISTIC fusion did not correspond with an increased final yield of soluble protein per litre of *E. coli*. A comparison of the yields of pure protein obtained prior to SEC showed that the use of the MISTIC-tagged construct resulted in only a slightly higher yield of non-aggregated protein than that of the non-MISTIC tagged construct, with yields of 0.4 and 0.3 mg/L of *E. coli* obtained, respectively. This is due to an increased level of aggregated protein, rather than soluble protein, as can be seen in the overlay of the respective SEC profiles in figure 3.12. The MISTIC tag was removed by TEV protease-cleavage prior to SEC.

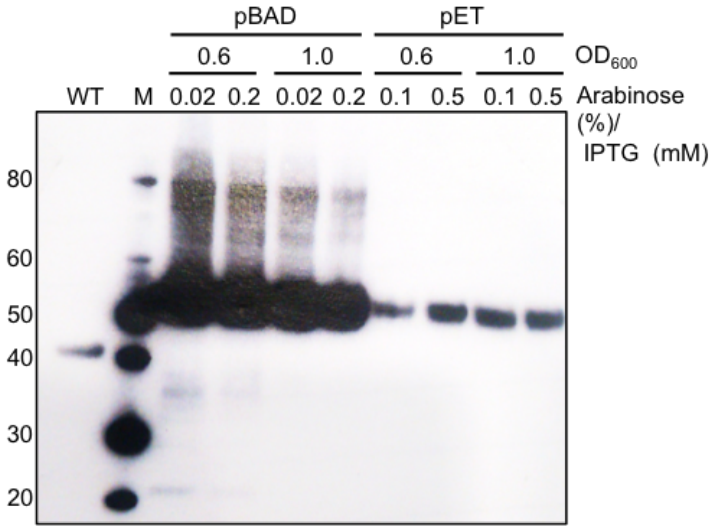


Figure 3.11. The fusion of MISTIC to the N-terminus of Wzy_{CT} increases expression levels. The western blot, probed with an anti-His tag antibody, of whole cells following expression trials of MISTIC-tagged Wzy_{CT}, with cells grown overnight at 15 °C following induction of expression, shows that the MISTIC fusion resulted in a higher level of expression than that obtained without the MISTIC tag. The expression condition giving the highest level of expression for the non-MISTIC tagged protein (15 °C overnight following induction at an OD_{600} of 1.0 with 0.2 % (w/v) arabinose) is shown in lane 1, indicated with "WT", to allow the comparison of expression levels. The expression of MISTIC-tagged Wzy_{CT} from the pBAD vector was much higher than that from the pET vector. (Cell cultures obtained during these expression trials were normalised to an OD_{600} of 1 prior to western blot analysis).

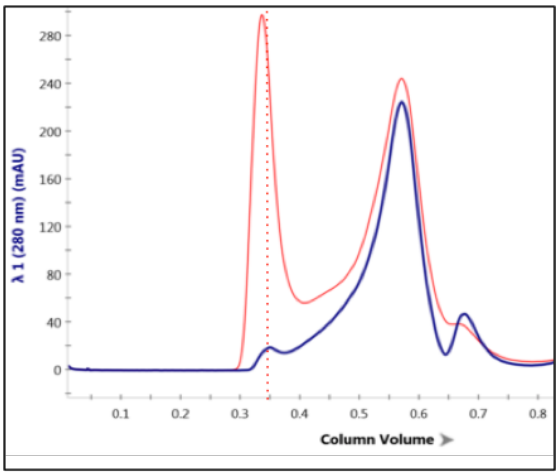


Figure 3.12. The effect of the MISTIC tag on Wzy_{CT} purification yield. The SEC profile following purification of the non-MISTIC-tagged and MISTIC-tagged Wzy_{CT} are overlaid here and shown in blue and red, respectively. The use of the MISTIC tag did not result in an increased purification yield, with more protein found in the void volume of the SEC column, the position of which is indicated with a dotted red line.

3.3.6 Differential filtration detergent screening

As no crystals were obtained using the detergent and optimised buffer combination identified using the membrane solubilisation detergent screening and nickel affinity column buffer screening, described in sections 3.3.2 and 3.3.3, respectively, additional, higher throughput, buffer and detergent screening approaches were explored. These additional experiments were also introduced in order to minimise the extent to which the "trial and error" approach to final purification conditions, discussed in section 3.3.4, is required. A differential filtration-based detergent screen was used to identify the smallest stable PDC, proposed to be ideal for crystallisation as the purification of target proteins in detergents with a smaller micelle size results in a larger exposed protein surface area, increasing the surface area available for potential crystal contacts (Privé, 2007; Vergis *et al*., 2015). The differential filtration-based detergent screen required purifying Wzy_{CV} in a detergent with a low CMC, so that it could be readily exchanged into a new detergent or detergent mixture. DM was used as the initial exchangeable detergent for the purification of the Wzy_{CV} for the differential filtration-based detergent screen. As seen in figure 3.13 the use of DM during the purification of Wzy_{CV}, using the optimised buffer and purification conditions previously described in section 3.3.4, resulted in pure, non-aggregated protein suitable for use in this screen. The purified protein was batch bound onto nickel affinity resin, which was aliquoted across a 96 well filter plate. Each well was exchanged into a new detergent or detergent mixture, with the detergent used in well A1 corresponding to DM, the detergent used during the purification, as a positive control and A2 was exchanged into Wzy_{CV} SEC buffer with no detergent, as a negative control. The target protein was eluted from the nickel affinity resin in the new detergent or detergent mixture and passed through either a 100 kDa cut-off filter plate or a 300 kDa cut-off filter plate, with the samples analysed by dot-blotting against the His_{10} affinity tag, to determine the approximate size of the PDC.

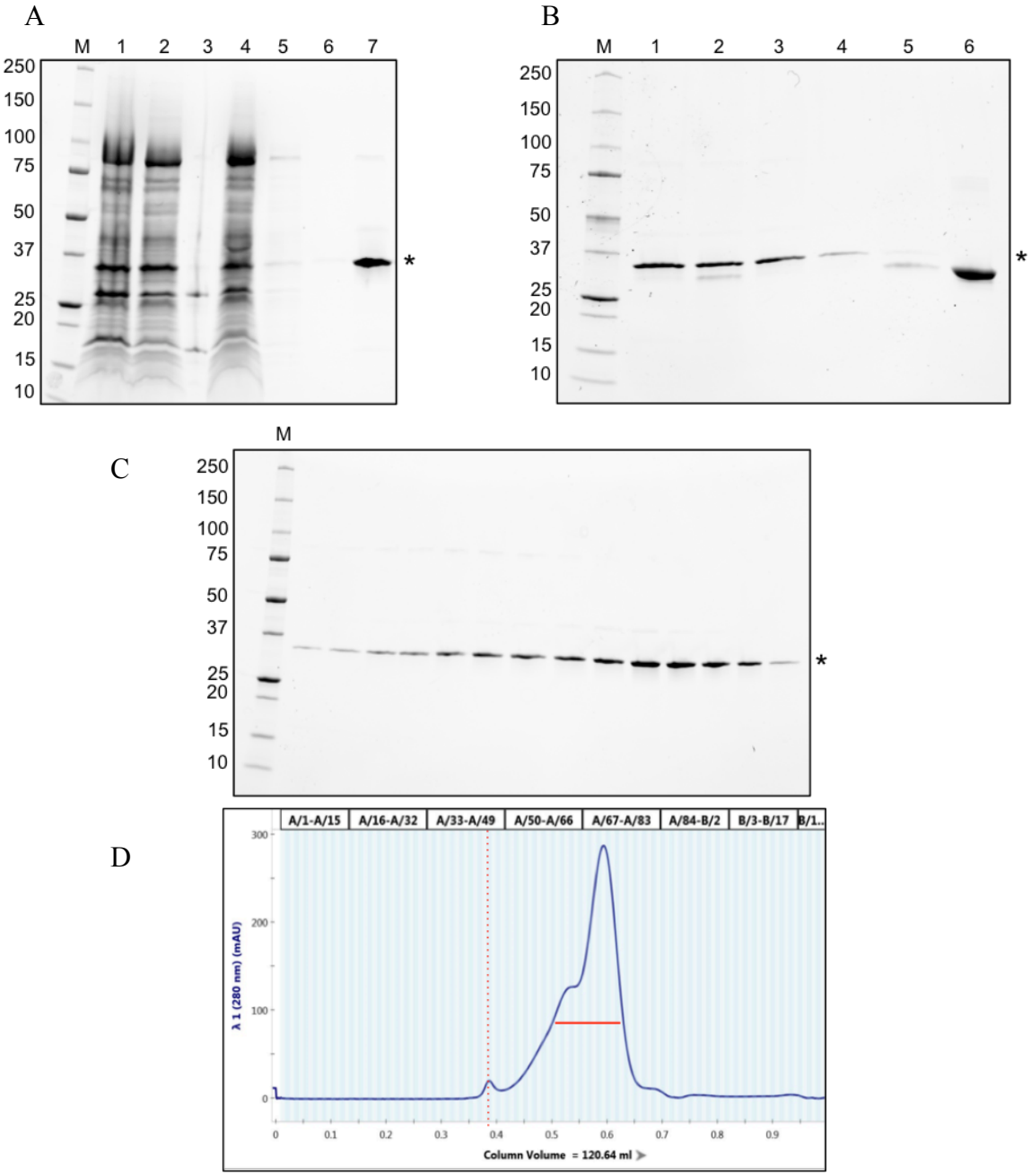


Figure 3.13 Wzy_{CV} purification in DM for the differential filtration detergent screen. "A" shows the SDS-PAGE gel of the first nickel affinity column (lane 1: membrane solubilisation; 2: DM-soluble fraction; 3: DM-insoluble fraction; 4: nickel affinity FT; 5: nickel affinity wash 1; 6: nickel affinity wash 2; 7: nickel affinity eluate). "B" shows the SDS-PAGE gel of second nickel affinity column (lane 1: first nickel affinity eluate; 2: TEV protease-cleaved sample; 3: second nickel affinity FT; 4: second nickel affinity wash; 5: second nickel affinity eluate; 6: TEV protease). "C" shows the SDS-PAGE gel of the central fractions of the SEC peak shown in "D", with the position of the void volume indicated with a dotted red line. The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*".The approximate position of the fractions in "D" shown in "C" is shown by the solid red line.

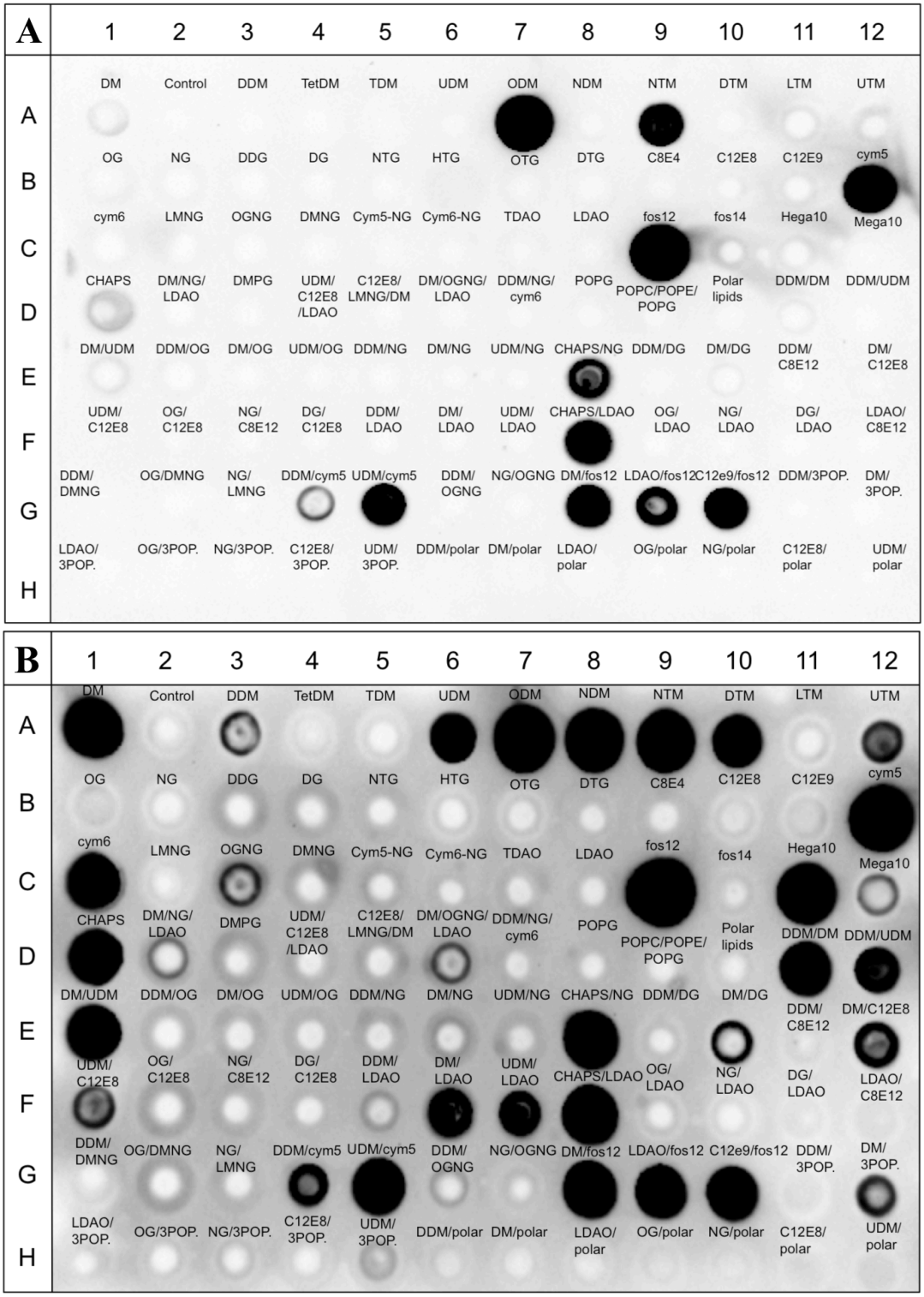


Figure 3.14. Differential filtration screen of Wzy_{CV}. Western dot blot, probed with an anti-His tag antibody, of the Wzy_{CV} that passed through the 100 kDa cut-off filter plate (A) and the 300 kDa cut-off filter plate (B). Each dot corresponds to Wzy_{CV} in a different detergent or detergent mixture, with selected detergents named. A list of all of the detergents used in this screen can be found in appendix 8.2.

The results of the differential filtration-based detergent screen can be seen in figure 3.14. Figure 3.14a shows the dot-blot against the His_{10} affinity tag for the nickel affinity eluate that was passed through the 100 kDa cut-off filter plate and figure 3.14b shows the dot-blot against the His_{10} affinity tag for the nickel affinity eluate that was passed through the 300 kDa cut-off filter plate. A signal can be seen in well A1 on both western blots, corresponding to the DM positive control, and no signal can be seen in well A2, corresponding to the no detergent negative control. Detergents that gave a signal on the western blot for the 100 kDa cut-off filter, i.e. those that produce a PDC smaller than 100 kDa, include cymal-5 (cym5; well B12), octyl maltoside (ODM; well A7), nonyl thiomaltoside (NTM; well A9) and fos-choline 12 (fos12; well C9). These detergents were chosen for large-scale purification for crystallisation trials. The micelle sizes and CMCs of the detergents chosen for large-scale purification can be seen in table 3.4.

Detergent	Detergent Abbreviation	CMC (%)	Micelle size (kDa)
n-Decyl-β-D-Maltopyranoside	DM	0.087	33
n-Nonyl-β-D-Thiomaltopyranoside	NTM	0.15	-
n-Octyl-β-D-Maltopyranoside	ODM	0.89	21
5-Cyclohexyl-1-Pentyl-β-D-Maltoside (Cymal-5)	Cym5	0.12	23
Fos-Choline 12	Fos12	0.047	19

Table 3.4. Detergents identified via differential filtration screening chosen for large-scale purification.

3.3.7 Large-scale purifications in various detergents

The results of the large-scale purification of Wzy_{CV} in cym5, ODM, NTM and fos12 are shown in figures 3.15, 3.16, 3.17 and 3.18, respectively. During the purification of Wzy_{CV} in the detergents identified in the differential filtration-based detergent screen, shown in figure 3.14, the target protein was extracted from the membrane using DM and this was exchanged for the new detergent during the first nickel affinity purification step. This approach was chosen in order to mimic that of the differential filtration-based detergent screen and resulted in pure, non-aggregated protein in all of the detergents tested.

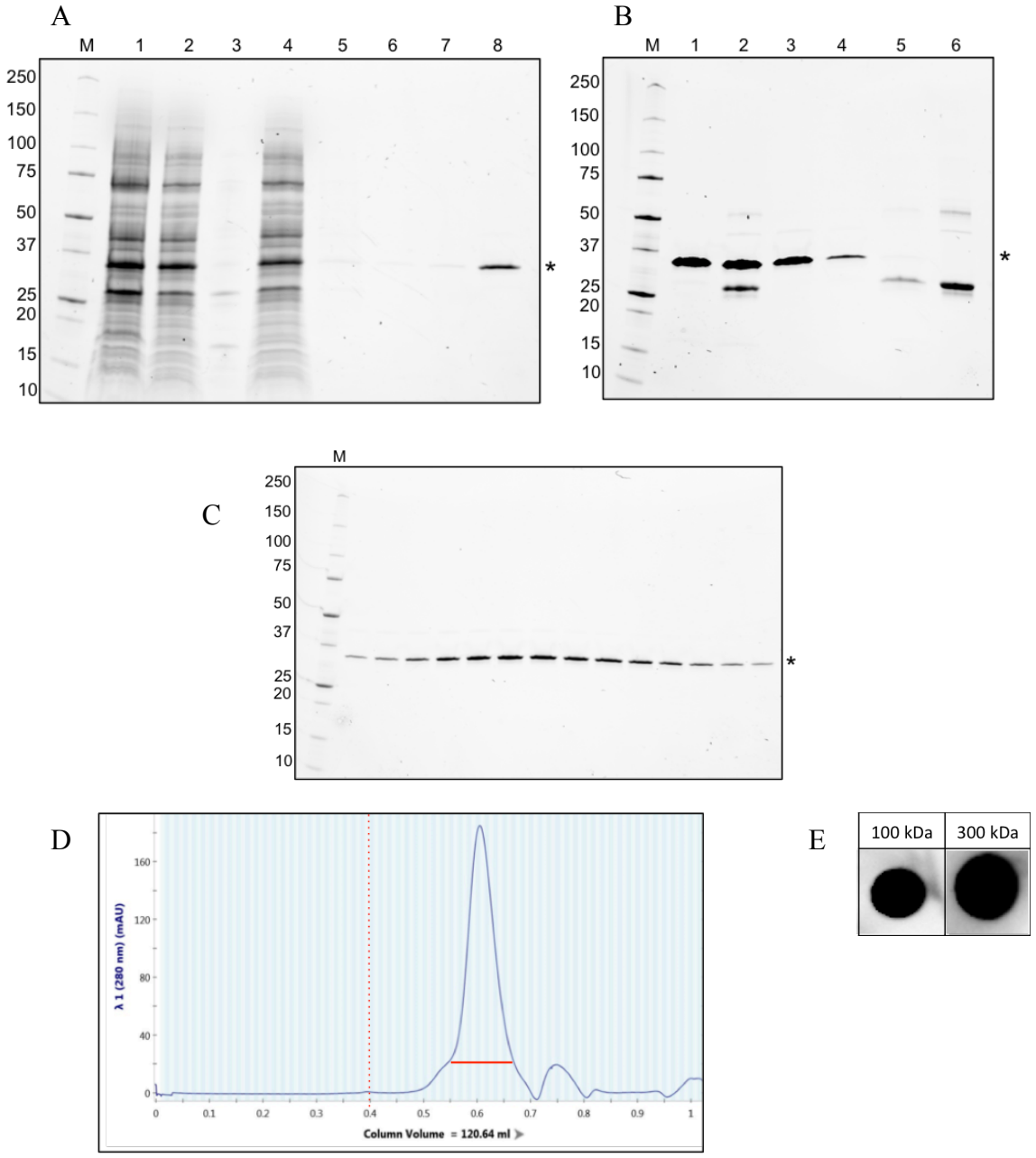


Figure 3.15. Wzy_{CV} purification in cym5. "A" shows the SDS-PAGE gel of the first nickel affinity column (M: marker; 1: membrane solubilisation; 2: DM-soluble fraction; 3: DM-insoluble fraction; 4: nickel affinity FT; 5: nickel affinity wash 1; 6: nickel affinity wash 2; 7: nickel affinity wash 3; 8: nickel affinity eluate). "B" shows the SDS-PAGE gel of second nickel affinity column (M: marker; 1: first nickel affinity eluate; 2: TEV protease-cleaved sample; 3: second nickel affinity FT; 4: second nickel affinity wash; 5: second nickel affinity eluate; 6: TEV protease). "C" shows the SDS-PAGE gel of the central fractions of the SEC peak shown in "D", with the position of the void volume indicated with a dotted red line. The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*".The approximate position of the fractions in "D" shown in "C" is shown by the solid red line. For reference, the results of the differential filtration screen for cym5 are shown in "E".

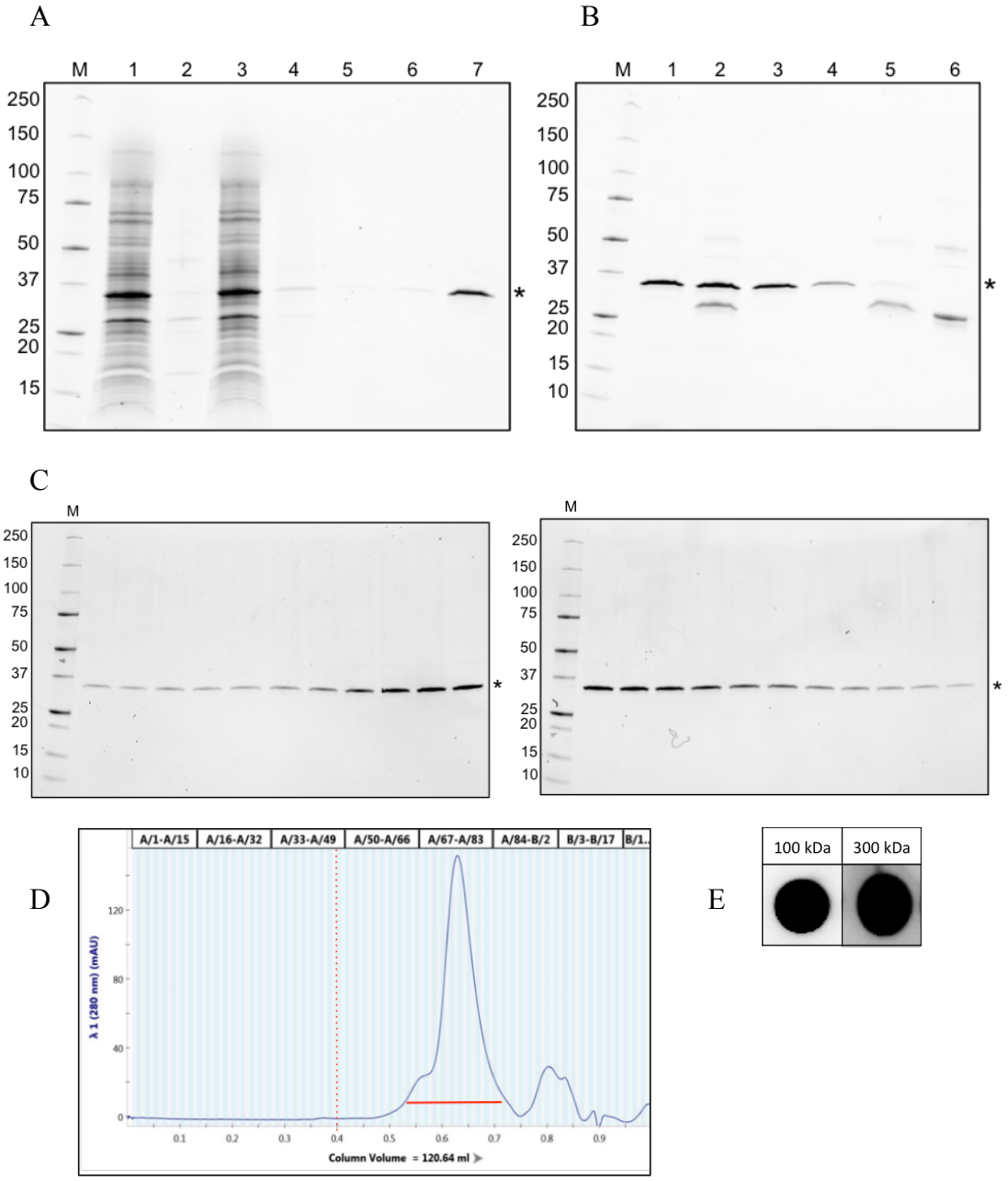


Figure 3.16. Wzy_{CV} purification in ODM. "A" shows the SDS-PAGE gel of the first nickel affinity column (M: marker; 1: DM-soluble fraction; 2: DM-insoluble fraction; 3: nickel affinity FT; 4: nickel affinity wash 1; 5: nickel affinity wash 2; 6: nickel affinity wash 3; 7: nickel affinity eluate). "B" shows the SDS-PAGE gel of second nickel affinity column (M: marker; 1: first nickel affinity eluate; 2: TEV protease-cleaved sample; 3: second nickel affinity FT; 4: second nickel affinity wash; 5: second nickel affinity eluate; 6: TEV protease). "C" shows the SDS-PAGE gel of the central fractions of the SEC peak shown in "D", with the position of the void volume indicated with a dotted red line. The approximate position of the fractions in "D" shown in "C" is shown by the solid red line. For reference, the results of the differential filtration screen for ODM are shown in "E". The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*".

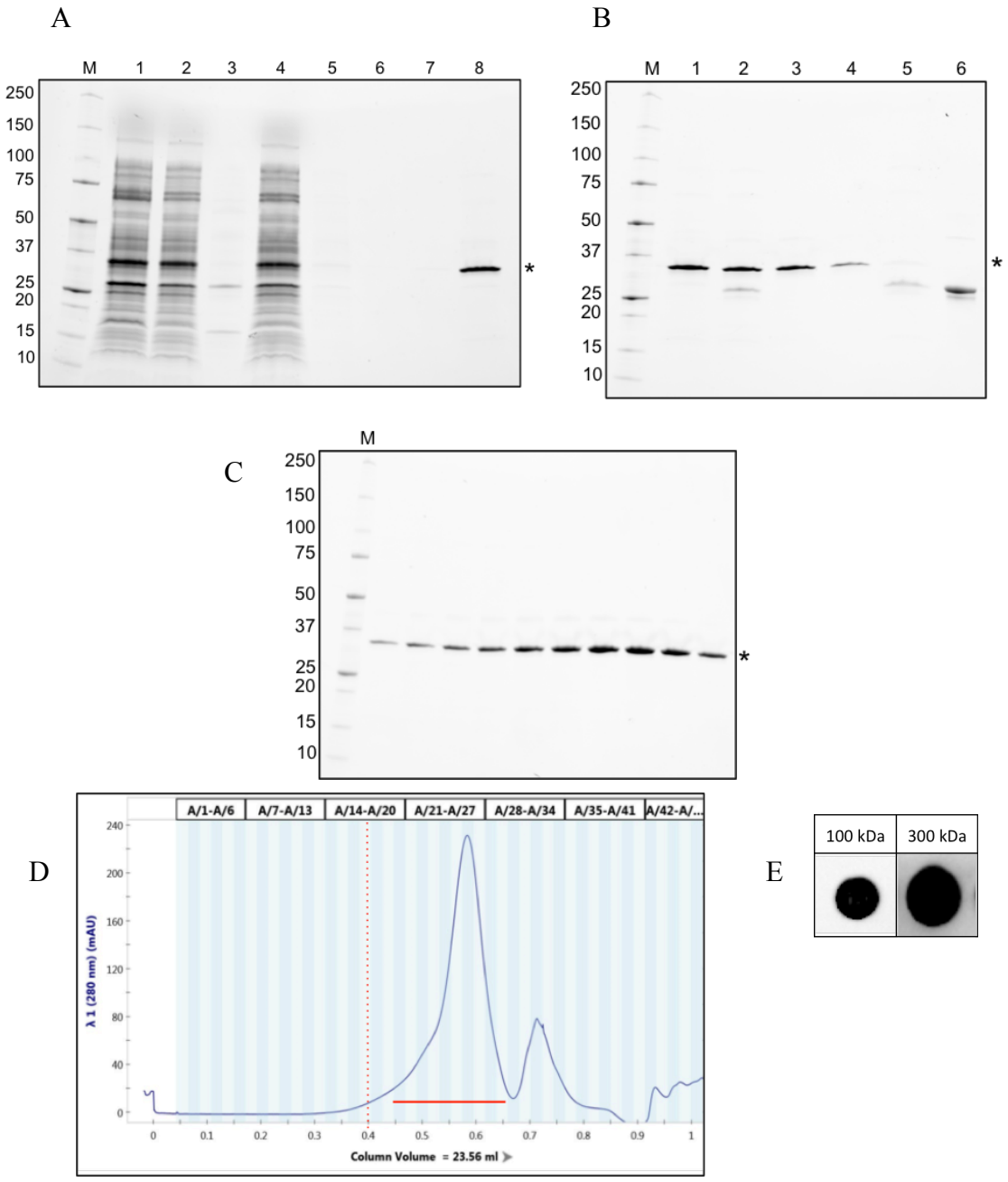


Figure 3.17. Wzy_{CV} purification in NTM. "A" shows the SDS-PAGE gel of the first nickel affinity column (M: marker; 1: membrane solubilisation; 2: DM-soluble fraction; 3: DM-insoluble fraction; 4: nickel affinity FT; 5: nickel affinity wash 1; 6: nickel affinity wash 2; 7: nickel affinity wash 3; 8: nickel affinity eluate). "B" shows the SDS-PAGE gel of second nickel affinity column (M: marker; 1: first nickel affinity eluate; 2: TEV protease-cleaved sample; 3: second nickel affinity FT; 4: second nickel affinity wash; 5: second nickel affinity eluate; 6: TEV protease). "C" shows the SDS-PAGE gel of the central fractions of the SEC peak shown in "D", with the position of the void volume indicated with a dotted red line. The approximate position of the fractions in "D" shown in "C" is shown by the solid red line. For reference, the results of the differential filtration screen for NTM are shown in "E". The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*".

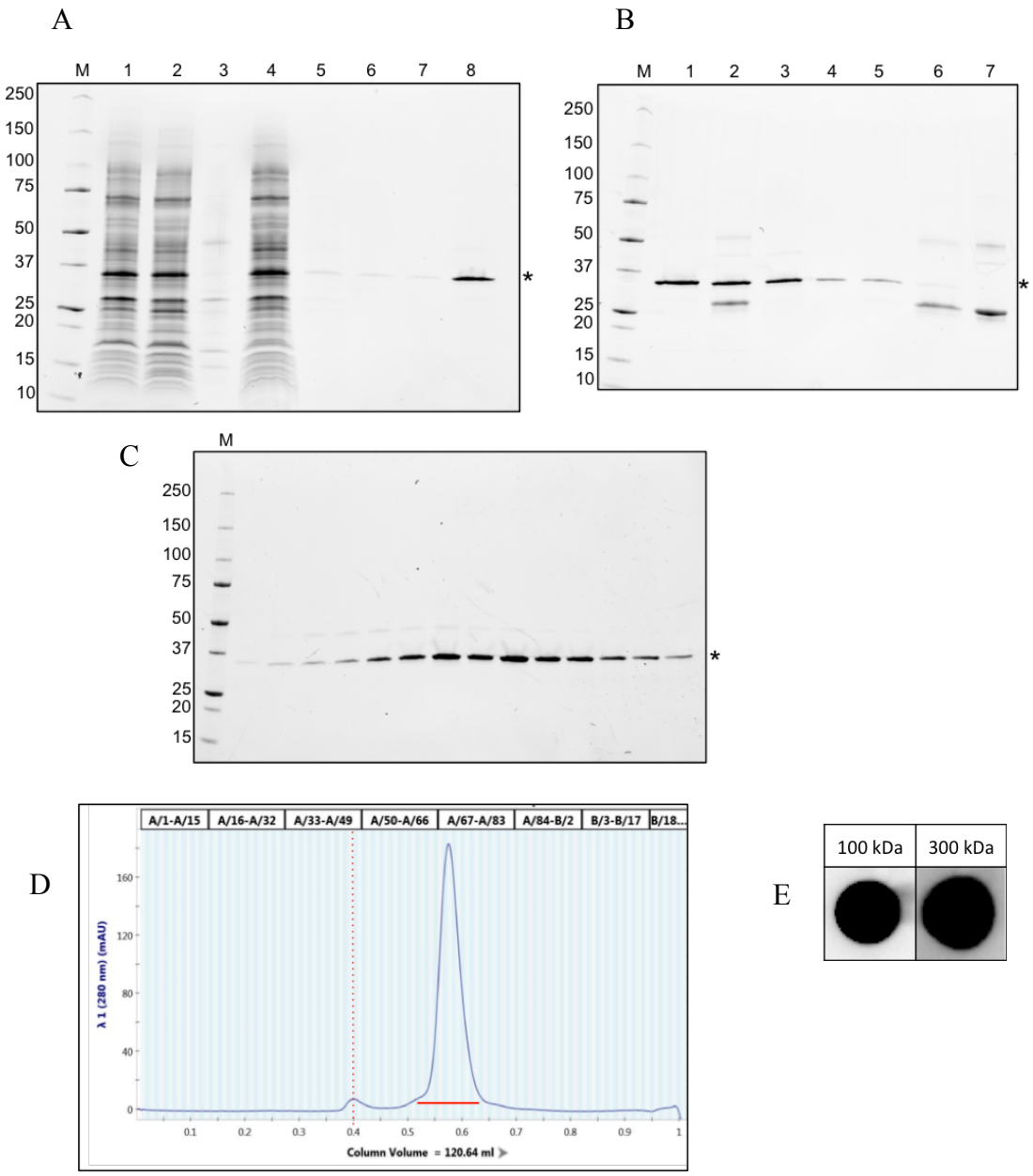


Figure 3.18. Wzy_{CV} purification in fos12. "A" shows the SDS-PAGE gel of the first nickel affinity column (M: marker; 1: DM-soluble fraction; 2: DM-insoluble fraction; 3: nickel affinity FT; 4: nickel affinity wash 1; 5: nickel affinity wash 2; 6: nickel affinity wash 3; 7: nickel affinity eluate). "B" shows the SDS-PAGE gel of second nickel affinity column (M: marker; 1: first nickel affinity eluate; 2: TEV protease-cleaved sample; 3: second nickel affinity FT; 4: second nickel affinity wash fraction 1; 5: second nickel affinity wash fraction 2; 6: second nickel affinity eluate; 7: TEV protease). "C" shows the SDS-PAGE gel of the central fractions of the SEC peak shown in "D", with the position of the void volume indicated with a dotted red line. The approximate position of the fractions in "D" shown in "C" is shown by the solid red line. For reference, the results of the differential filtration screen for fos12 are shown in "E". The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*".

3.3.8 Sitting-drop vapour diffusion crystallisation trials of Wzy

Sitting-drop vapour diffusion crystallisation trials, on a 96 well plate scale, were implemented in an attempt to obtain crystals of Wzy. Crystallisation trials were conducted at both RT and 4 °C, mainly using a drop volume ratio of 1 protein: 1 mother liquor, using both in-house stochastic screens and the commercially available screens MemGold (Newstead *et al*., 2008) and MemGold 2 (Newstead *et al*., 2008; Parker and Newstead, 2012). Both of these screens were developed using data mining of the components of crystallisation conditions that have lead to membrane protein structure deposition in the PDB, MemGold is based on conditions used to crystallise both α-helical and β-barrel membrane proteins (Newstead *et al*., 2008), whereas MemGold 2 is based on α-helical membrane protein crystallisation conditions (Newstead *et al*., 2008; Parker and Newstead, 2012). Sitting drop vapour diffusion plates were set up using protein purified in a variety of detergents, as described in section 2.9, with the main focus being on detergents with a low micelle size as these result in a small PDC (purifications shown in figures 3.15-3.18). Despite small PDCs being proposed to facilitate an increased amount crystal contacts when compared to larger PDCs, resulting from detergents with a larger micelle size, crystals could not be obtained when using the majority of these detergents. Crystals could be obtained following the incubation of Wzy_{CV}, purified in NTM (figure 3.17), at a concentration of 4.5 mg/ml, in MemGold well A10 (1.1 M sodium citrate tribasic dihydrate, 0.0665 M HEPES pH 7.5) (Newstead et al., 2008) in a sitting-drop vapour diffusion plate (drop volume ratio 1 protein: 1 mother liquor) stored at 20 °C following the purification of Wzy_{CV} in NTM. These crystals and their progression over time can be seen in figure 3.19. They appeared after four days incubation at 20 °C and reached full size and number after 17 days. However, these crystals proved difficult to reproduce and did not diffract when screened on Beamline I04-1 at Diamond Light Source. This could be due to the precipitation of the protein at RT in the time during which the crystallisation trials were set up. This, in addition to the reduced aggregation observed upon purification at 4 °C (discussed in section 3.3.4), provided further indication that this protein is not thermally stable. Therefore further optimisation of buffer conditions was implemented in order to

increase the thermal stability of this protein prior to further optimisation of this crystallisation condition.

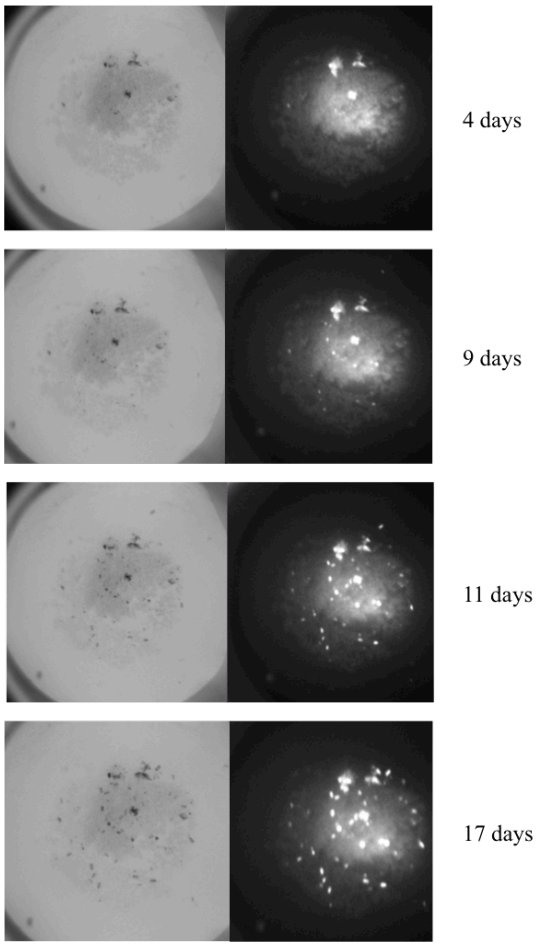


Figure 3.19. Crystals obtained in a sitting-drop vapour diffusion crystallisation trial following purification of Wzy_{CV} in NTM. Crystals were obtained in MemGold well A10 (Newstead *et al*., 2008) and reached full size and number after 17 days incubation at 20 °C in a Minstrel HT-UV system. The normal image can be seen on the left and the UV image can be seen on the right. The images were taken using the Minstrel HT-UV system (Rigaku) and viewed using the CrystalTrak software.

3.3.9 LCP crystallisation trials of Wzy

During LCP crystallisation trials, the target protein was mixed with the lipid monoolein, in a 2:3 volume ratio, in syringes to form the LCP mixture as shown in figure 3.20 (Caffrey, 2015). This mixture was used to set up crystallisation trials using a glass base plate, with the LCP mixture applied first followed by the mother liquor. The wells were sealed with a glass coverslip. An example of a completed 96 well plate can be seen in figure 3.21. In-house LCP crystallisation screens are based around the precipitants PEG 400 at 40 % (v/v) or MPD at 8 % (v/v) with the StockOptions Salts (Hampton Research) at 0.1 M or 0.4 M and a pH range of 3.5 to 8.5 (Dr Nicole Howe and Professor Martin Caffrey, Trinity College Dublin, personal communication). An example LCP well can be seen in figure 3.22. LCP is believed to promote crystal contacts not only between the hydrophilic loops, but also between the TMHs, (Caffrey, 2015; Krauss *et al*., 2013), however no crystals could be obtained during LCP crystallisation trials of Wzy_{CV}.

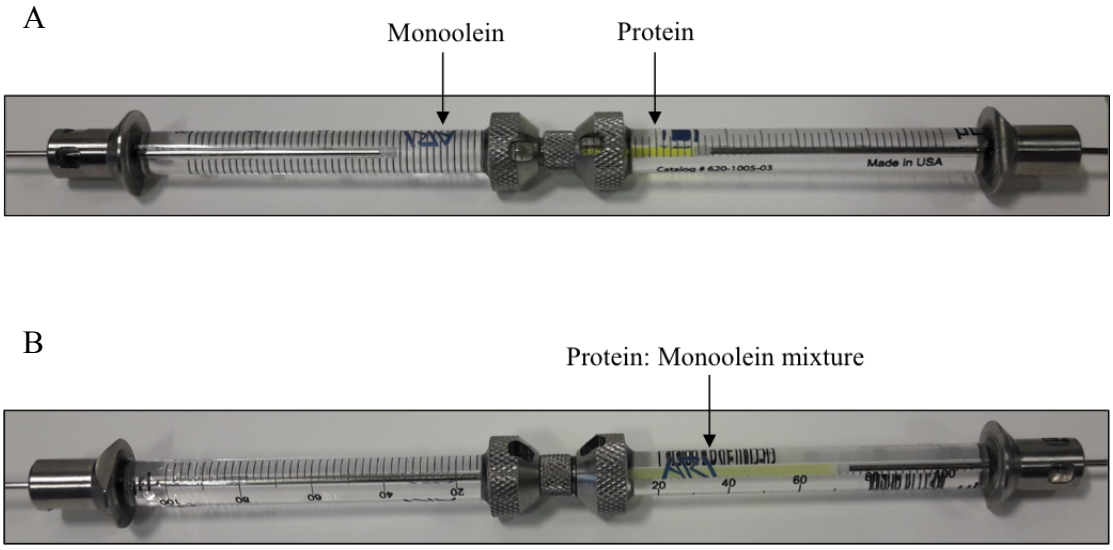


Figure 3.20 Preparation of LCP mixture. Syringes, with one containing protein and the other containing monoolein, prior to mixing are shown in "A" and after mixing to form the LCP mixture in "B". Here GFP was used for clarity.

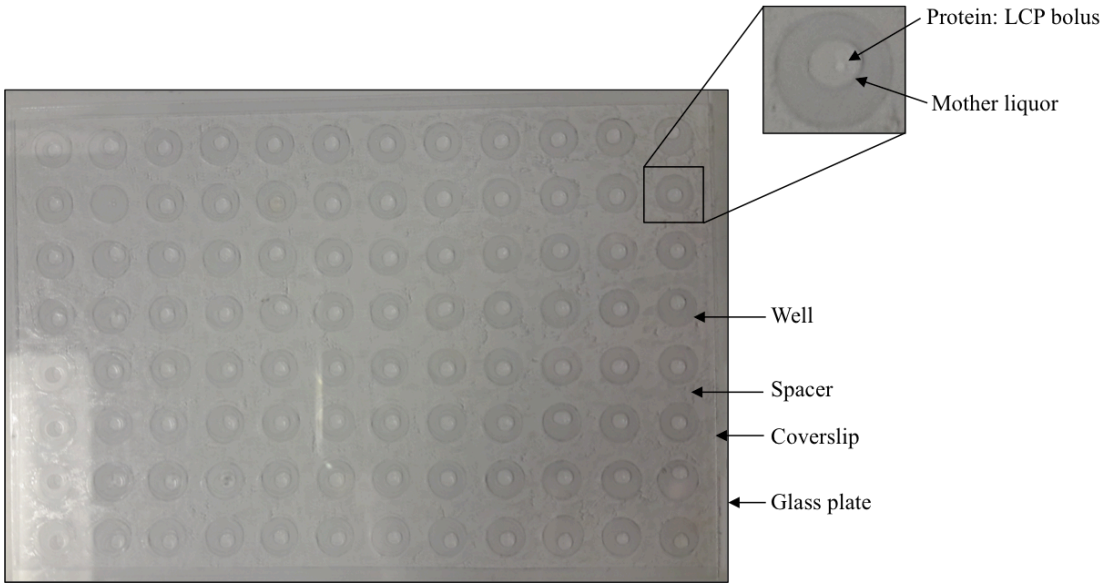


Figure 3.21. A completed LCP plate

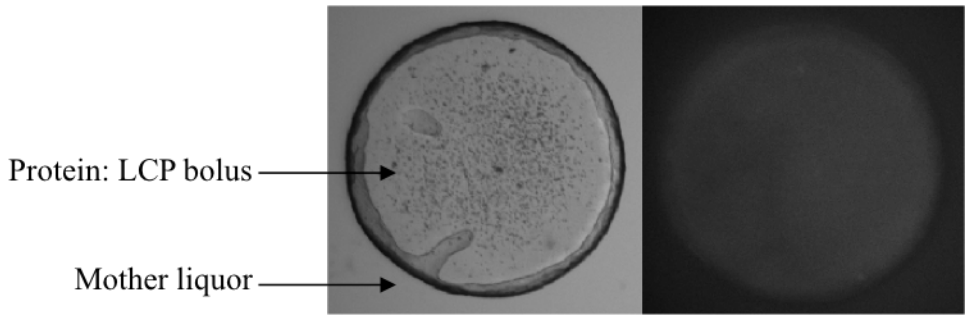


Figure 3.22. An example LCP well. A normal image can be seen on the left and a UV image can be seen on the right. The images were taken using the Minstrel HT-UV system (Rigaku) and viewed using the CrystalTrak software.

3.3.10 DSF buffer screening

As discussed in section 3.3.4, an improvement in the purification of Wzy_{CV} was observed when purifying the target protein at 4 °C in comparison to purifying the target protein at RT (figure 3.9). Furthermore, Wzy_{CV}, purified in NTM, precipitated during the time it took to set up crystallisation trials at RT, as discussed in section 3.3.8. These results indicate that the target protein is not thermally stable. Therefore a DSF screen was implemented in order to improve the thermal stability of this protein by testing the effect of different buffers, pH values, buffer concentrations and salt concentrations on the melting temperature (T_m) of the target protein (Boivin *et al*., 2013; Newman, 2004). This is the temperature at which 50 % of the target protein is unfolded (Boivin *et al*., 2013). The conventional DSF approach uses Sypro Orange, which binds to the hydrophobic residues of soluble proteins exposed upon unfolding of the protein as the temperature is increased, resulting in an increase in fluorescence (Lo *et al*., 2004). Due to their high hydrophobicity, this approach is not suitable for membrane proteins. Furthermore Sypro Orange will also bind to detergents required for membrane protein purification, leading to a high background signal (Kohlstaedt *et al*., 2015; Lo *et al*., 2004). We therefore used a newly characterised dye DCIA that binds to free cysteines exposed upon unfolding (Branigan *et al*., 2013). This method is therefore unsuitable for target proteins without cysteine residues (Branigan *et al*., 2013). Once bound to the thiol side chain of a free cysteine, DCIA emits at 440 nm, upon excitation at 350 nm, as shown in figure 3.23 (Branigan *et al*, 2013).

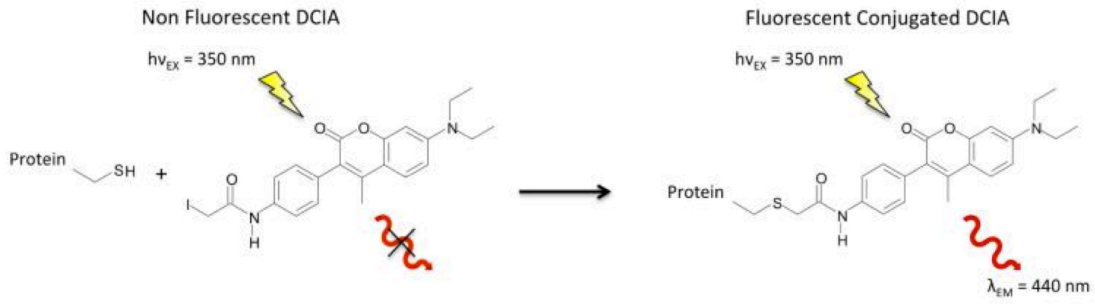


Figure 3.23. DCIA fluorescence. DCIA dye fluoresces at 440 nm upon excitation at 350 nm, following conjugation to a free cysteine (taken from Branigan *et al*. 2013)

An example of a melting curve obtained here for Wzy_{CV}, which was used to derive the T_m, can be seen in figure 3.24, with the temperature plotted on the X-axis and the fluorescence response plotted on the Y-axis.

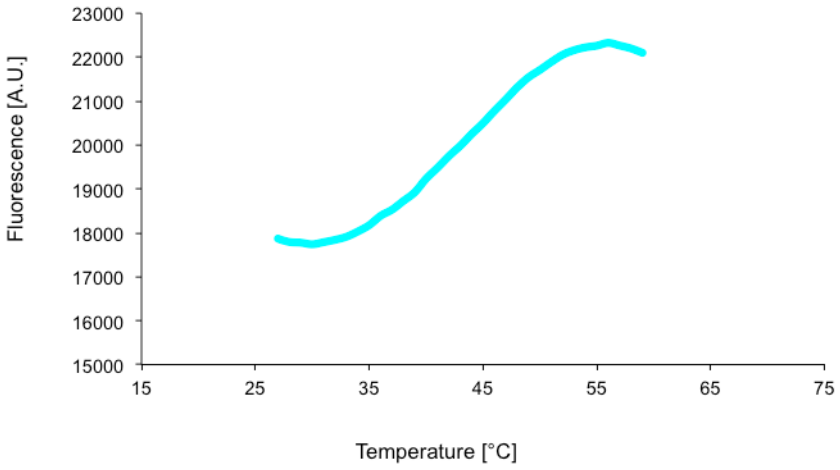


Figure 3.24. Typical melting curve of Wzy_{CV} obtained during the DSF studies.

A graph comparing the derived T_m for each of the buffers and additives tested can be seen in figure 3.25, with each line corresponding to the T_m in a different condition. T_m values for selected conditions that will be discussed in more detail can be seen in table 3.5. Well D7 most closely resembles the optimised purification buffer and had a T_m of 38 °C. This could be increased to a maximum of 85 °C in 0.119 M citrate pH 4.0 (well A2). The melting curves for wells D7 and A2 can be compared in figure 3.26. Some general trends can be observed in the T_m. As can be seen in table 3.5, the target protein is more thermally stable at lower pH values, with the T_m dropping from 85 °C in 0.119 M citrate pH 4.0 to just 34 °C in 0.119 M CHES pH 9.0. The use of the broad range buffer SPG (succinic acid/ sodium phosphate monobasic/ glycine in a 2:7:7 ratio) over a pH range of 4.0 to 10.0 (wells E1-E12, figure 3.25) enabled the deconvolution of the effect of changes in pH from that of the buffering agent itself (Newman, 2004). Although T_m curves could not be reliably plotted for wells E1 to E5, corresponding to pH 4.0 to 6.0, a slight pH dependence could still be observed, with the T_m decreasing from 53 °C at pH 6.5 to 35 °C at pH 10.0 (wells E6 to E12, figure 3.25 and table 3.5). This shows that the previously described decrease in T_m with increasing pH is not due to changes in buffering agent, but due to the change in pH. The effect of the addition of

NaCl at 0.298 M to the buffer depended on the buffer used. No change in the T_m was observed in the presence of 0.119 M MES pH 6.5 or 0.119 M Tris-HCl pH 8.0 upon addition of 0.298 M NaCl, whereas the addition of 0.298 M NaCl in the presence of 0.1 M Bis-tris pH 6.5 caused the T_m to decrease by approximately 25 °C (table 3.5). There appeared to be a slight dependence of T_m on the ionic strength of the buffer, with the T_m increasing as the concentration of HEPES pH 7.5 and sodium phosphate pH 7.5 increased, with a 10-fold increase in concentration resulting in a 5 °C and 8 °C increase in T_m, respectively (table 3.5). The buffer that results in the highest level of thermal stability of the target protein, 0.119 M citrate pH 4.0 (figure 3.25 and table 3.5) is not suitable for the nickel affinity purification process, therefore the following condition was chosen for large-scale purification: 0.119 M MES pH 6.5, 0.298 M NaCl, which gave a T_m for the target protein of 67 °C (figure 3.25 and table 3.5).

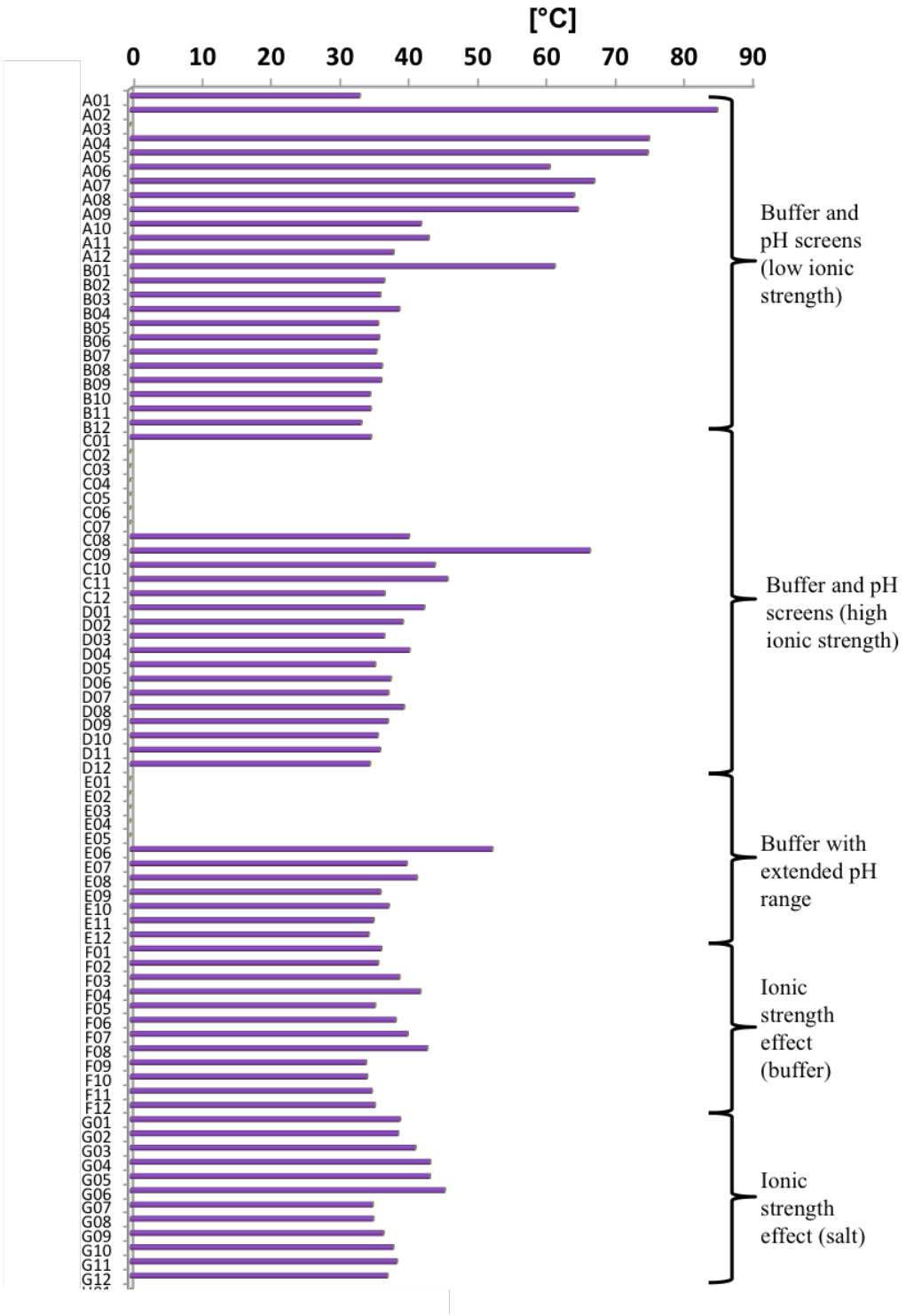


Figure 3.25. Graph comparing the derived T_m for each buffer tested during the DSF studies. The RUBIC buffer screen (Molecular Dimensions) was used and the components of this screen can be found in appendix 8.3.

	Condition	T_m (°C)
A2	0.119 M citrate pH 4.0	85
A4	0.119 M citrate pH 5.0	75
A5	0.119 M MES pH 6.0	75
A6	0.119 M potassium phosphate pH 6.0	61
A7	0.119 M citrate pH 6.0	67
A8	0.119 M Bis-tris pH 6.5	64
A9	0.119 M MES pH 6.5	65
B1	0.119 M MOPS pH 7.0	62
B7	0.119 M Tris-HCl pH 8.0	36
B12	0.119 M CHES pH 9.0	34
C8	0.119 M Bis-tris pH 6.5, 0.298 M NaCl	40
C9	0.119 M MES pH 6.5, 0.298 M NaCl	67
D7	0.119 M Tris-HCl pH 8.0, 0.298 M NaCl	37
E6	0.119 M SPG pH 6.5	53
E7	0.119 M SPG pH 7.0	40
E8	0.119 M SPG pH 7.5	42
E9	0.119 M SPG pH 8.0	36
E10	0.119 M SPG pH 8.5	38
E11	0.119 M SPG pH 9.0	35
E12	0.119 M SPG pH 10.0	35
F1	0.024 M HEPES pH 7.5	36
F2	0.06 M HEPES pH 7.5	36
F3	0.149 M HEPES pH 7.5	39
F4	0.298 M HEPES pH 7.5	42
F5	0.024 M sodium phosphate pH 7.5	36
F6	0.06 M sodium phosphate pH 7.5	39
F7	0.149 M sodium phosphate pH 7.5	40
F8	0.298 M sodium phosphate pH 7.5	43

Table 3.5. T_m values obtained for a selected range of buffers.

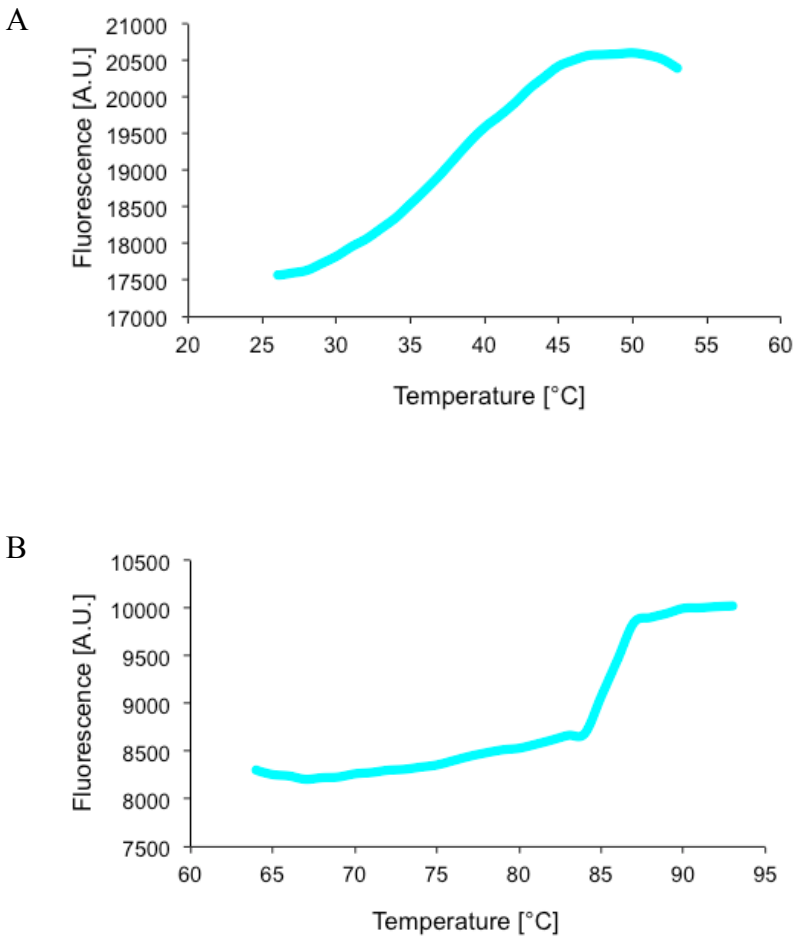


Figure 3.26. A comparison of the melting curve for Wzy_{CV} obtained in the optimised SEC buffer and the melting curve of Wzy_{CV} in the buffer that gave the highest T_m. The melting curve obtained in the condition that most closely resembled the optimised SEC buffer is shown in "A", resulting in a T_m of 37 °C. The use of 0.119 M citrate pH 4.0 resulted in a T_m of 85 °C, the melting curve for this buffer is shown in "B".

3.3.11 Large scale purification using buffer identified via DSF

Using an adaptation of the buffer identified via DSF, namely 50 mM MES pH 6.5, 200 mM NaCl, and DDM as the detergent, the final SEC step of the purification of Wzy_{CV} could be conducted at RT without resulting in aggregation, confirming the increased

thermal stability in this buffer. The SEC profile of the target protein in this buffer, as well as the SDS-PAGE gels of the purification, can be seen in figure 3.27. However, the SDS-PAGE gels show that the protein is much more contaminated than that obtained previously (figure 3.10, section 2.3.3). Therefore further optimisation of the purification of Wzy_{CV} is required as this buffer is optimised for the stability of the protein rather than the purification.

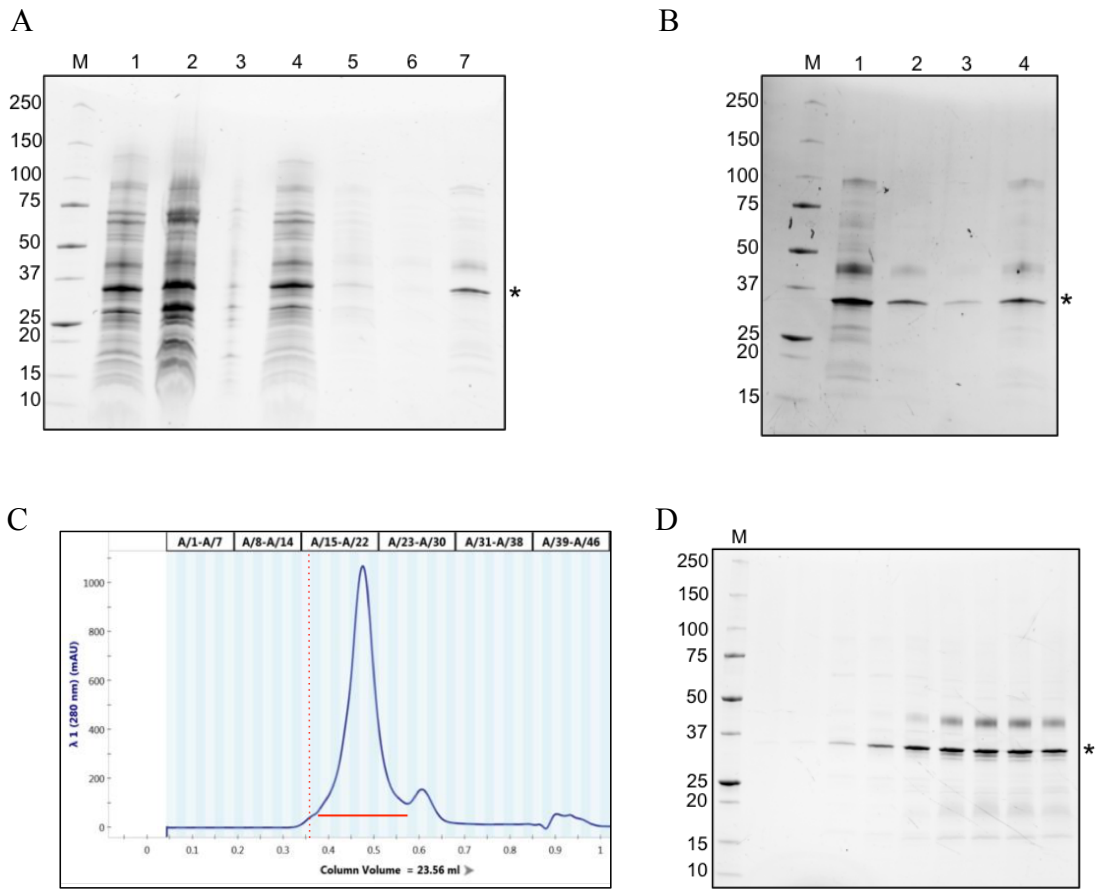


Figure 3.27. Wzy_{CV} purified in 50 mM MES pH 6.5, 200 mM NaCl, 3 x CMC DDM. "A" shows the SDS-PAGE gel of the first nickel affinity column (M: marker; 1: membrane solubilisation; 2: DDM-soluble fraction; 3: DDM-insoluble fraction; 4: nickel affinity FT; 5: nickel affinity wash 1; 6: nickel affinity wash 2; 7: nickel affinity eluate). "B" shows the SDS-PAGE gel of second nickel affinity column (M: marker; 1: TEV protease-cleaved sample; 2: second nickel affinity FT; 3: second nickel affinity wash; 4: second nickel affinity eluate). "C" shows the SDS-PAGE gel of the central fractions of the SEC peak shown in "D", with the position of the void volume indicated with a dotted red line. The approximate position of the fractions in "D" shown in "C" is indicated by the solid red line. The band at approximately 35 kDa corresponding to Wzy_{CV} is indicated with a "*".

3.3.12 SEC-MALS of various Wzy homologues

SEC-MALS was used in order to determine whether the various Wzy homologues purified share a common oligomerisation state, when purified in DDM. However, of the Wzy homologues tested here, Wzy_{CV} and Wzy (from *Pectobacterium atrosepticum*) purified as trimers, and Wzy (from *Clostridium thermocellum*) purified as a monomer. The SEC-MALS profiles for Wzy_{CV}, Wzy (from *Pectobacterium atrosepticum*) and Wzy (from *Clostridium thermocellum*) can be seen in figures 3.28, 3.28 and 3.30, respectively, with the total molar mass and the protein molar mass of each PDC of each homologue seen in table 3.6.

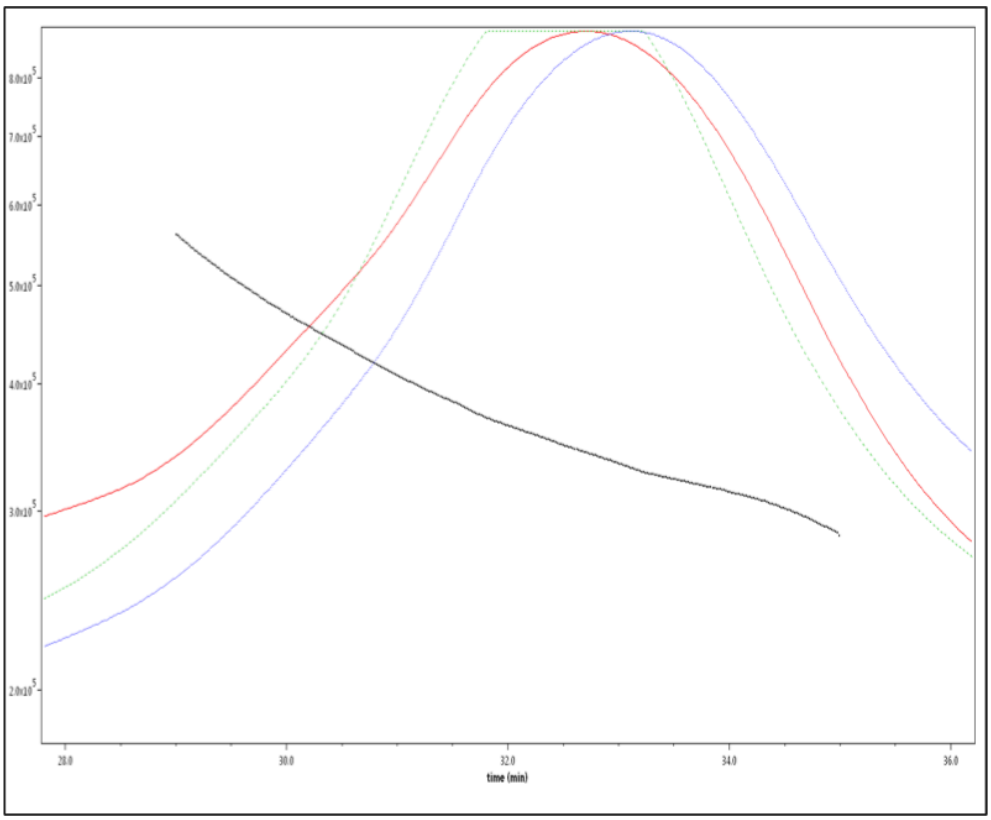


Figure 3.28. SEC-MALS profile of Wzy_{CV}. Change in UV signal over time is shown in green, change in light scattering over time is shown in red, change in refractive index over time is shown in blue and the change in molar mass (g/ml) over time is shown in black.

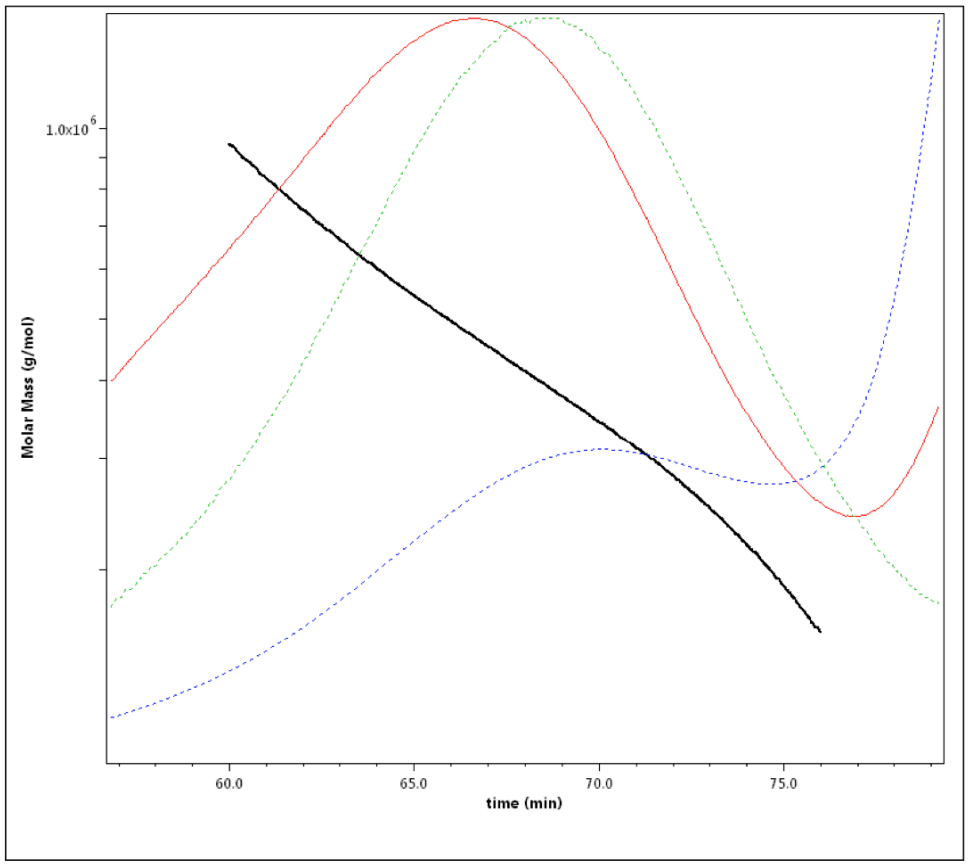


Figure 3.29. SEC-MALS of Wzy from *Pectobacterium atrosepticum*. Change in UV signal over time is shown in green, change in light scattering over time is shown in red, change in refractive index over time is shown in blue and the change in molar mass (g/ml) over time is shown in black.

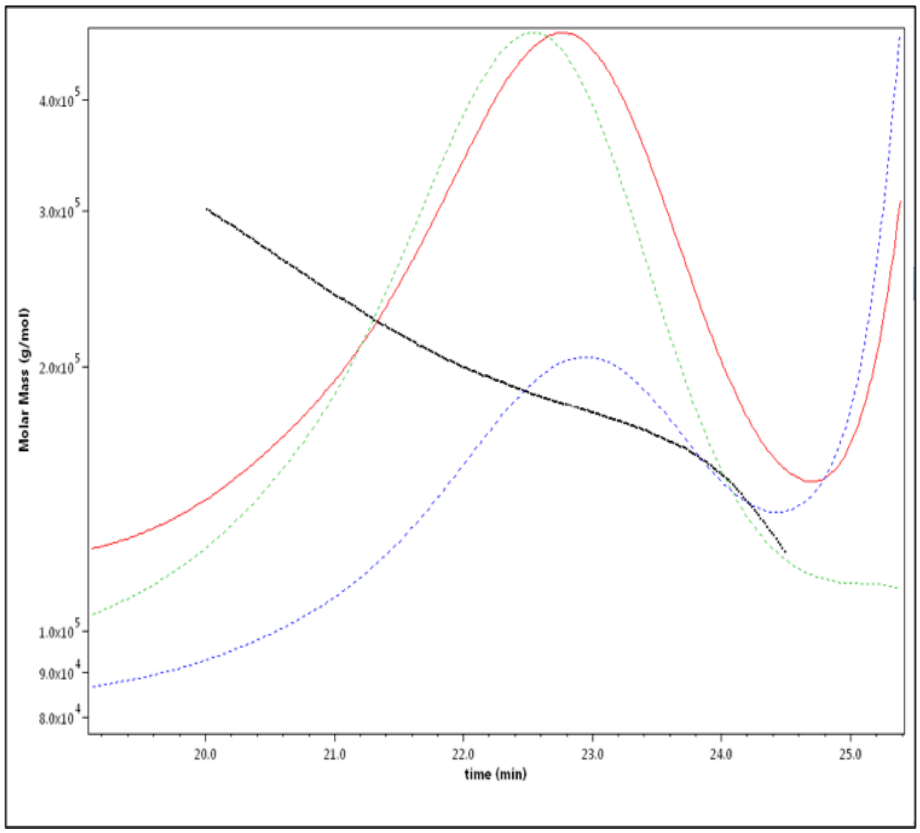


Figure 3.30. SEC-MALS of Wzy from *Clostridium thermocellum*. Change in UV signal over time is shown in green, change in light scattering over time is shown in red, change in refractive index over time is shown in blue and the change in molar mass (g/ml) over time is shown in black.

Homologue	Monomeric molar mass (kDa)	SEC-MALS PDC molar mass (kDa)	SEC-MALS protein molar mass (kDa)	Oligomeric state
Chromobacterium violaceum	47.6	359.5	146.5	Trimer
Pectobacterium atrosepticum	53.0	409.5	196.8	Trimer/Tetramer
Clostridium thermocellum	61.1	186.7	87.8	Monomer/Dimer

Table 3.6. SEC-MALS comparison the oligomeric states of different Wzy homologues.

Homologue	Expressed	Detergent extraction screen	Buffer screen	Purified	Yield (mg/ L of *E. coli cell culture*)	Aggregated	Differential filtration detergent screen	Thermofluor	Vapour diffusion crystallisation trials	LCP crystallisation trials	Crystals
***E. coli* O9:K30**	Yes	Yes	Yes	Yes	0.4	Yes	No	No	Yes	No	No
***E. coli* K12**	No	-	-	-	-	-	-	-	-	-	-
***E. coli* K12 ECA**	No	-	-	-	-	-	-	-	-	-	-
Chromobacterium violaceum	Yes	Yes	Yes	Yes	0.4	No	Yes	Yes	Yes	Yes	Yes
Nitrosospira multiformis	Yes	Yes	Yes	Yes	0.5	No	No	No	Yes	No	No
Clostridium thermocellum	Yes	Yes	Yes	Yes	0.7	No	No	No	Yes	Yes	No
Pectobacterium atrosepticum	Yes	Yes	Yes	Yes	0.2	No	No	No	Yes	No	No
Actinobacillus pleuropneumoniae	Yes	Yes	Yes	Yes	0.2	No	No	No	No, impure	No	No
Thermomicrobium roseum	Yes	Yes	Yes	Yes	-	No	No	No	Yield too low	No	No
Thermodesulfo-bacterium geofontis	Yes	Yes	Yes	Yes	-	Yes	No	No	No	No	No
Rhodothermus marinus	Yes	Yes	Yes	Yes	0.02	Yes	No	No	No	No	No
Sulfobacillus acidophilus	Yes	Yes	Yes	Yes	0.06	No	No	No	Yield too low	No	No
Thermocrinus albus	Yes	Yes	Yes	Yes	-	No	No	No	Yield too low	No	No
Marinotoga piezophila	Yes	Yes	Yes	Yes	0.2	No	No	No	Yield too low	No	No
Marinomonas mediterranae	Yes	Yes	Yes	Yes	0.6	Yes	No	No	No	No	No
Streptococcus thermophilium	Yes	Yes	Yes	Yes	-	No	No	No	Yield too low	No	No

Table 3.7. A summary of the techniques used and results obtained for each Wzy homologue investigated.

3.4 Discussion

A major bottleneck in the structural characterisation of integral membrane proteins is obtaining the target protein in sufficient quantities for crystallisation trials. Numerous strategies have been employed here in an attempt to obtain pure, monodisperse Wzy, in sufficient quantities for crystallisation trials. The structure of the O-antigen polymerase Wzy would shed light on the mechanism of polymerisation, which remains uncharacterised. To the best of our knowledge the overexpression and purification of a Wzy homologue has only been described once previous to this report. This previous study used the coexpression of the chaperone GroEL/GroES to facilate overexpression of one Wzy homologue in *E. coli* (Woodward *et al.*, 2010). The coexpression of chaperones was not required for Wzy overexpression in *E. coli* C43 (DE3) cells here.

Table 3.7 summarises the techniques used and results obtained for each Wzy homologue tested, including whether it could be expressed, the level of buffer and detergent screening techniques explored for each homologue, if it could be purified, its aggregation status and the purification yield, expressed in mg of protein per L of *E. coli* cell culture, and whether it could be purified in sufficient yields for crystallisation trials. Out of the 16 different homologues tested here, 14 could be expressed. Detergents for extraction of the target protein from the IM and buffers for nickel affinity chromatography were tested for all of the expressed homologues. Following larger scale purification using the detergent and buffer combination identified during the initial screening, four homologues were purified in an aggregated state and five homologues were entered into crystallisation trials. The remaining homologues could not be purified in sufficient yields, or to a sufficient level of purity. The Wzy_{CV} was used to trial the differential filtration-based detergent screen and the DSF buffer screen. Despite the large number of homologues screened, as well as the extent of buffer and detergent screening conducted, crystals of Wzy_{CV} could be obtained in one condition following the purification of this homologue in one detergent, namely NTM. However, these crystals resulted from protein that precipitated during the time it took to set up the crystallisation trials, therefore future optimisation focussed on improving the thermal stability of this protein, rather than the optimisation of this crystallisation condition.

Furthermore, the identity of the crystallised protein remains unconfirmed, as these crystals were not analysed via SDS-PAGE gel.

Important factors to take into consideration during membrane protein expression are: ensuring that the expressed protein is intact, active and that the expression level is sufficient for your chosen technique. These factors can be influenced by initial construct design, expression conditions and expression systems. Here we describe conventional expression trials, testing level of inducer, OD_{600} at induction, and growth temperature following induction, using *E. coli* C41 (DE3) and C43 (DE3) cells, which are optimised for the overexpression of membrane proteins (Miroux and Walker, 1996). Despite demonstrated success with other target proteins and the increased overexpression level observed here, the use of the MISTIC tag did not result in an increased yield of soluble target protein (Kefala *et al*., 2007).

Key factors to take into consideration during membrane protein purification are as follows: ensuring that the purified protein is not denatured or aggregated, but monodisperse, pure and thermally stable. These factors can be influenced by purification method of choice i.e. the chosen affinity tagging system, the detergent chosen for extraction of the target protein from the membrane and for subsequent purification, and the buffer chosen for purification, along with the use of additives. We used two methods for buffer screening: a nickel affinity chromatography-based approach and a DSF approach, using the recently characterised dye DCIA (Branigan *et al*., 2013). A major advantage of the nickel affinity chromatography-based approach is the direct comparison of purification yields and levels of contaminant proteins obtained in different buffers on a small-scale, allowing the easy comparison of the effects of different ionic strengths, buffering agents, pH and various additives. A disadvantage of this screen is that only 15 buffers can be tested at once. Furthermore, the SDS-PAGE gel, or western-blotting output, of this screen does not show whether the purified protein is aggregated or not. In contrast, the DSF screen is higher throughput, as up to 96 conditions can be screened at a time. Furthermore, a wider range of buffering agents can be investigated using DSF, as not all pHs are suitable for nickel affinity purification. Disadvantages of DSF include that it requires purified protein, the

purification of which must be optimised prior to the DSF screening. Furthermore, the purified target protein must contain cysteine residues.

Two methods were also used for detergent screening: a solubilisation-based approach to determine which detergents can extract the target protein from the membrane and a differential filtration-based approach to determine which detergent results in the smallest PDC, proposed to be ideal for type II crystal formation (Krauss *et al*., 2013). The advantage of the solubilisation-based detergent screen is that it allows the easy comparison of the level of target protein extracted from the membrane over a specific time by each detergent tested, as well as the levels of contaminant proteins extracted. The disadvantage of this screen is that no information regarding the aggregation status of the target protein is obtained. Also this screen is not high-throughput. Furthermore, the SDS-PAGE gel and western blotting analysis required for this screen does not enable the determination of the aggregation state of the extracted protein. The differential filtration-based detergent screen is high-throughput as 96 detergents or detergent mixtures can be tested at once. The purification of membrane proteins using only five different detergents has resulted in more than 50 % of the total membrane protein structures solved (Vergis *et al*., 2015). These detergents are the maltosides, DDM and DM, and the glucosides, OG and NG, and finally the zwitterionic detergent, LDAO (Vergis *et al*., 2015). Therefore the purification of membrane proteins in detergents other than the five most popular enabled the determination of the remaining 40 % of membrane protein structures, highlighting the importance of high-throughput screening of different detergents (Vergis *et al*. 2015). Furthermore, this method of screening requires only small amounts of each detergent and is therefore a cost-effective method of screening. Disadvantages of the differential filtration-based detergent screen include that the target protein must be purified in an exchangeable detergent, i.e. a detergent with a low CMC, to allow its displacement from the target protein by other detergents during screening. The identification of a suitable exchangeable detergent can require lengthy optimisation. Furthermore, a relatively large amount of purified target protein is required for this screen. It must be noted that we do not know to what extent each new detergent displaces the detergent used for purification; therefore it is possible that exchanging into some of the detergents results in a mixture of both the original and

new detergent. This is not necessarily a disadvantage as the addition of a second detergent during purification enabled the determination of almost 20 % of the protein structures solved in 2012 (Moraes *et al*., 2014). Therefore it is also important to screen for appropriate detergent mixtures for purification.

As can be seen when comparing figure 3.6 and 3.14 some detergents, for example DM and cym5, performed well in both the solubilisation screen and the differential filtration screen, indicating that they are suitable for both the extraction of the target protein from the membrane and also result in a small PDC, proposed to be ideal for type II crystal contacts (Krauss *et al*., 2013). However, some detergents, for example LDAO, could extract the majority of the target protein from the membrane during the solubilisation screen, but once the target protein was exchanged from DM into LDAO during the differential filtration screen, no signal could be obtained via dot-blotting indicating that the target protein aggregated in this detergent. This aggregation could be prevented by the addition of a second detergent, for example UDM. Furthermore, some detergents, for example CHAPS, could not extract the target protein from the membrane during the solubilisation screen, but the use of this CHAPS during the differential filtration-based filtration screen resulted in a small PDC of below 100 kDa, as indicated by the signal on the dot-blot of the protein that passed through the 100 kDa cut-off filter plate. Taken together, the results indicate that different detergents are suitable for different stages of the purification process and both screens should be used in future.

SEC-MALS was used to determine whether the O-antigen polymerase Wzy homologues studied share a common oligomerisation state. Despite purifying each homologue in DDM, the homologues tested did not share the same oligomerisation state (figure 3.28 and table 3.6). This could be due to multiple reasons including the non-native environment of the detergent micelle, which has been proposed to influence the oligomeric state of the PCP Wzz (Larue *et al*., 2009; Tocilj *et al*., 2008). Furthermore, the oligomerisation state of Wzy could also depend on the presence of other proteins in the LPS or CPS biosynthesis pathway (Nath and Morona, 2015a). This is something that should be explored further in the future, perhaps using the same proteoliposome-based approach used with the PCP Wzz (Larue *et al*., 2009).

3.5 Conclusions and future work

Obtaining pure membrane proteins in sufficient quantities is a major bottleneck in their structural study. Here we describe a protocol for the purification of membrane proteins at the levels required for crystallisation trials, covering expression to purification. The example studied is the O-antigen polymerase Wzy, although the methods discussed here are applicable to the other membrane proteins. The use of this protocol has taken multiple homologues of Wzy from cloning and expression trials to large-scale purification and crystallisation trials, providing a foundation to build upon.

Despite multiple homologues of Wzy entering crystallisation trials, in both conventional sitting-drop vapour diffusion plates and, more recently, LCP, crystals could be obtained only in one condition for one homologue in one detergent. However, these crystals resulted from protein that precipitated during the time taken to set up crystallisation trials at RT. This precipitation could have removed unstable or misfolded protein and should be explored further. Both the differential filtration-based detergent screen and the DSF screen were only tested using the Wzy_{CV}. Although this protein was purified using a range of different detergents, identified during the differential filtration-based screen, this protein was only purified using the MES pH 6.5 and MOPS pH 7.0 conditions identified during DSF (data for MOPS pH 7.0 not shown), therefore more buffers should be tested on a larger scale. Particular focus should be on exchanging into the lower pH buffers, shown to result in the highest T_m, during SEC, but these buffers are not suitable for nickel affinity purification. As these screens could identify suitable buffers and detergents for the purification of Wzy_{CV}, they should also be implemented in the optimisation of the purification conditions of other Wzy homologues tested. The different methods described for buffer and detergent screening aid in the identification of the buffers and detergents most suitable for a particular stage in the process, and the future approach should continue to include both methods of screening, as for example, the most suitable detergent and buffer combination for purification of the protein, might not be the most suitable for downstream applications.

Factors that need to be taken into consideration during future membrane protein crystallisation trials are the detergent micelle size and ensuring that the target protein does not occupy multiple conformational states. The paradigm of type II crystals, shown in figure 3.2, is that shorter chain detergents leave a greater surface area of target protein exposed for potential crystal contacts during crystallisation trials (Krauss *et al*., 2013). An alternative approach that could both increase the level of protein surface area exposed for crystal contacts and ensure that the target protein entering into crystallisation trials occupies a single conformational state, is the use of nanobodies. Nanobodies, produced by camelids, are single-domain antibodies of approximately 15 kDa, consisting only of the variable domain of an antibody heavy chain (Pardon *et al*., 2014). Once injected with a structurally intact target protein, camelids produce a range of nanobodies to recognise conformational epitopes of this target protein, rather than the primary sequence, with high affinity (Pardon *et al*., 2014). The purified nanobodies can then be used during crystallisation trials of the target protein both to increase the exposed surface area, of particular relevance to membrane protein crystallisation, and to lock the target protein in a particular conformation, producing a more homogeneous and less flexible sample for crystallisation trials and allowing the study of functionally relevant conformational changes (Pardon *et al*., 2014). A nanobody-based approach has been used to capture and determine the structure of a novel conformational state of the ABC-transporter flippase PglK (Perez *et al*., 2017), which flips the LLO across the IM during bacterial *N*-glycan biosynthesis, to those previously structurally characterised (Perez *et al*., 2015). As Wzy is an integral membrane protein, possible type II crystal contacts are limited. Furthermore, Wzy homologues have a large, potentially flexible, periplasmic loop. A membrane topology prediction of the Wzy_{CV} homologue, the focus of this chapter, is shown in figure 3.1. Both of these factors demonstrate that Wzy would be a suitable candidate for nanobody-assisted crystallisation and this should be explored in the future.

Ensuring that the target protein is expressed and purified in a functional form is crucial, i.e. the protein in a functionally relevant oligomerisation state and not denatured. We currently lack an assay to monitor Wzy O-antigen polymerase function. This is largely due to the requirement of a highly complex substrate, consisting of undecaprenyl

pyrophosphate-linked sugar-repeat units, with a different undecaprenyl pyrophosphate-linked repeat unit required for each Wzy homologue (Hong *et al*., 2015). Furthermore, Woodward *et al.* observed that the level of Wzy (from *E. coli* serotype O86)-catalysed O-antigen polymerisation *in vitro* reduced as the length of the lipid anchor used decreased (Woodward *et al*., 2010). The number of *cis* double bonds in the lipid anchor was also observed to play a role in the level of observed Wzy-catalysed O-antigen polymerisation (Woodward *et al*., 2010). As well as ensuring that the target protein is purified intact, a functional assay would allow the identification of suitable substrates for co-crystallisation. Co-crystallisation in the presence of substrate could aid in the reduction of conformational flexibility, thereby promoting crystallisation, and also allow the identification of the active site residues of Wzy. The ATP-independent flippase MurJ, responsible for the flipping of lipid II across the IM during the peptidoglycan layer biosynthesis, was crystallised in the presence of lipid II (Kuk *et al*., 2016). A functional assay would also allow the screening of mutations to identify the catalytic residues of Wzy. An inactive mutant construct could be more suitable for crystallisation due to a potentially reduced level of conformational change. Possible functional assays could include radiation-based assays to monitor the polymerisation of a radiolabelled substrate (Woodward *et al*., 2010) or a malachite green-based colorimetric assay (Geladopoulos *et al*., 1991) to detect free phosphate release, as the by-product of the polymerisation reaction is undecaprenyl-pyrophosphate, which is then dephosphorylated by PgpB to yield undecaprenyl-phosphate (Fan *et al*., 2014).

Proteins with inherent flexibility and multiple functionally relevant conformations are not suitable for crystallography. Whilst nanobodies have been developed in order to lock the target protein in a particular conformation to aid crystallisation (Pardon *et al*., 2014), this prevention of flexibility and conformational change is not required for all methods of structural characterisation (Bai, *et al*., 2015). Once reserved for the structural characterisation of icosahedral viruses, technological advances in the field of cryo-electron microscopy (cryo-EM), including the development of a direct electron detector, which reduces the signal to noise ratio in the images produced, and improvements in motion correction during data processing, have vastly increased its applicability to and usage for the structural characterisation of proteins (Bai *et al*.,

2015). This is demonstrated by the fact that the number of maps deposited per year in the Electron Microscopy Data Bank (EMDB) is currently increasing exponentially (https://www.ebi.ac.uk/pdbe/emdb/statistics_main.html/ Accessed: 30/6/17). Cryo-EM is being used to determine the structures of proteins and protein complexes below 200 kDa in molecular weight, and is even breaking the 100 kDa barrier, with the structure of isocitrate dehydrogenase (93 kDa) solved to 3.8 Å resolution (Merk *et al*., 2016). Furthermore, cryo-EM can now be used to solve structures at atomic resolution, with the structure of glutamate dehydrogenase (334 kDa) solved to 1.8 Å resolution (Merk *et al*., 2016). As well as determining apo protein structures, cryo-EM is now being used to determine the mode of action of small molecule binding to proteins, for example the structure of β-galactosidase (from *E. coli*) in complex with the inhibitor phenylethyl β-D-thiogalactopyranoside has been solved to a resolution of 2.2 Å allowing the determination of the mode of inhibition (Bartesaghi *et al*. 2015). As demonstrated using SEC-MALS, the Wzy_{CV} homologue studied here is an oligomer, therefore cryo-EM would be a viable alternative to X-ray crystallography and could not only be used to solve the structure of Wzy, but also be used to determine the mode of substrate binding and therefore the mechanism of Wzy-catalysed O-antigen polymerisation.

4 Do the O-antigen polymerase Wzy and the polysaccharide co-polymerase Wzz form a complex?

4.1 Introduction

Chain length regulation is an essential component of Wzy-catalysed O-antigen polymerisation during LPS and CPS biosynthesis as this, unlike DNA polymerisation, occurs in the absence of a template (Whitfield, 2006). Like the composition of the O-antigen repeat unit, O-antigen chain length modality is serotype-specific (Tocilj *et al*., 2008; Whitfield, 2006; Woodward *et al*., 2010). Different O-antigen chain lengths have been shown to play different roles during infection and regulation of O-antigen chain length has been demonstrated to be crucial to the survival of bacteria in different environments (Hoa Tran *et al*., 2014; Hong and Payne, 1997). *S. flexneri*, for example, produces a bimodal distribution of O-antigen chain lengths, resulting from the expression of two Wzz homologues, one of which confers an O-antigen chain length modality of 11-17 repeat units and the other a very long modality of 90-100 repeat units (Hong and Payne, 1997). Deletion of the gene encoding the Wzz homologue responsible for the very long O-antigen chain length results in a decreased resistance to complement-mediated killing when compared to the wild type (Hong and Payne, 1997). In contrast, deletion of the gene encoding the Wzz homologue responsible for the O-antigen chain modality of 11-17 repeat units had no effect on resistance to complement-mediated killing (Hong and Payne, 1997). Mutations in the Wzz homologue that confers the very long O-antigen chain length modality had no effect on the ability of *S. flexneri* to invade HeLa cells, multiply and subsequently spread to neighbouring cells, whereas mutations in the Wzz conferring the shorter O-antigen modality resulted in a reduced

ability to invade, survive within and spread to neighbouring HeLa cells (Hong and Payne, 1997). These findings demonstrate that O-antigen chain length regulation is a highly tuned process crucial to bacterial survival, with different O-antigen chain lengths fulfilling different functions during the bacterial life cycle.

In the absence of template, chain length regulation in biology relies either on a "molecular ruler" or a chemical clock not only to control the final O-antigen chain length, but also to provide, or recognise, a signal that switches the equilibrium from being in favour of O-antigen chain polymerisation to being in favour of ligating the O-antigen chain on to the appropriate acceptor or exporting the O-antigen chain to the OM. The first example of a "molecular ruler" to be characterised at a molecular level was WbdD (Clarke *et al*., 2004; Hagelueken *et al*., 2012, 2015). WbdD regulates O-antigen chain length during *E. coli* O9 antigen biosynthesis, with its coiled-coil domain forming the ruler (Hagelueken *et al*., 2015) and the modification of the terminal mannose of the O9 antigen by the bifunctional methyltransferase-kinase domain acting as the export signal (Clarke *et al*., 2004), which is then recognised by the nucleotide-binding protein Wzt of the ABC transporter Wzt/Wzm (Cuthbertson *et al*., 2005, 2007). Further to its function as a "molecular ruler", WbdD also acts as a "molecular scaffold" for the assembly of the polymannosyl transferase WbdA (Clarke *et al*., 2009) and together they form a "chain extension complex", the stoichiometry of which can affect the final O-antigen chain length (King *et al*., 2014).

During Wzy-dependent O-antigen biosynthesis, the final O-antigen chain length is controlled by the PCP1a family member, Wzz (Daniels and Morona, 1999). Woodward *et al.* showed that the O-antigen polymerase Wzy is the only component of the LPS biosynthesis pathway required to polymerise the O-antigen repeat units (Woodward *et al*., 2010). Furthermore, they found that O-antigen chain length modality could be restored purely upon addition of the chain length regulator Wzz (Woodward *et al*., 2010). Despite the importance of O-antigen chain length regulation to pathogenicity, the mechanism of chain length regulation by Wzz remains unknown. Various studies have found that mutations throughout Wzz can affect O-antigen modality, with no specific "hotspot" region identified, suggesting that Wzz does not possess catalytic activity

(Daniels and Morona, 1999; Kintz and Goldberg, 2011). This lack of catalytic activity is in contrast to WbdD (Clarke *et al*., 2004; Cuthbertson *et al*., 2005).

4.1.1 Proposed mechanisms of Wzz-mediated O-antigen chain length regulation

Multiple, not necessarily mutually exclusive, mechanisms have been proposed to explain how Wzz acts to regulate the O-antigen chain length, in the absence of both catalytic activity and template (Islam and Lam, 2014; Kintz and Goldberg, 2011; Tocilj *et al*., 2008), and each will be briefly introduced here.

4.1.1.1 Molecular clock

Wzz and Wzy have been proposed to form a complex that exists in two states, one of which favours the continued polymerisation of O-antigen repeat units by Wzy, the other favours the transfer of the O-antigen chain to the O-antigen ligase WaaL (Bastin *et al*., 1993). In this mechanism, Wzz is proposed to act as a "molecular clock", represented by figure 4.1, controlling the time during which each conformation of Wzy is occupied (Bastin *et al*., 1993). However, it is unsure how Wzz would function as a "molecular clock" as little evidence has been put forward to support this proposal (Islam and Lam, 2014).

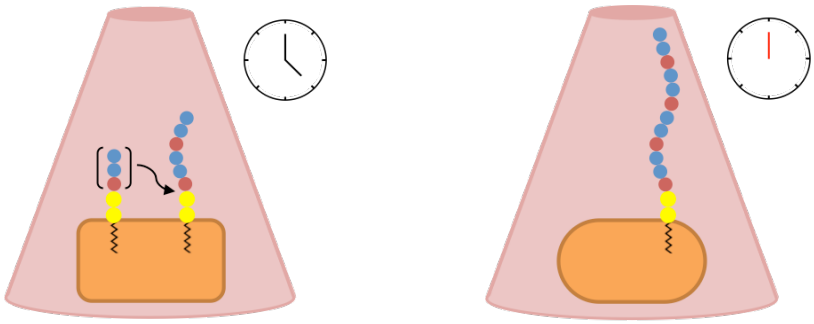


Figure 4.1. Proposed mechanism for Wzz acting as a molecular clock. Wzy (shown in orange) is proposed to exist in two conformational states, one of which favours polymerisation, shown in "A", and the other favours ligation of the O-antigen chain onto lipid A, shown in "B". Wzz (shown in pink) is proposed to control the time during which each conformational state is occupied by Wzy (Bastin *et al*., 1993).

4.1.1.2 Chaperone

Wzz has been hypothesised to act as a chaperone that mediates and controls the interaction between the O-antigen polymerase Wzy and the O-antigen ligase WaaL, represented in figure 4.2 (Islam and Lam, 2014; Morona *et al.*, 1995). However, no evidence in favour of an interaction between Wzy and WaaL has been provided. Furthermore, Daniels *et al.* found that the final O-antigen chain length is reached prior to its ligation onto lipid A by WaaL (Daniels *et al.*, 2002). Woodward *et al.* found that O-antigen chains with a modal distribution could be produced when Wzz was incubated with Wzy in the presence of substrate *in vitro*, indicating that O-antigen modality is conferred without the requirement of additional components of the LPS biosynthesis pathway (Woodward *et al.*, 2010). Feldman *et al.* also found that O-antigen chain length modality is still conferred in a WaaL knockout *E. coli* K-12 O16 cell line and that the completed O-antigens can function as donor substrates during *N*-linked glycosylation catalysed by PgpB (from *C. jejuni*), suggesting that the transfer of the O-antigen on to the acceptor lipid A plays no role in chain length determination (Feldman *et al.*, 2005). Taken together, the data indicate that this proposed mode of O-antigen chain length regulation is unlikely.

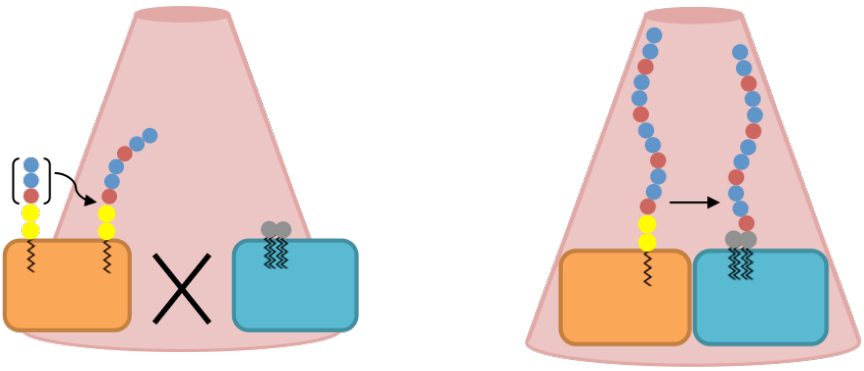


Figure 4.2. Proposed mechanism for Wzz acting as a chaperone. Wzz (shown in pink) has been proposed to act as a chaperone to control the interaction between Wzy (shown in orange) and WaaL (shown in turquoise). By blocking the interaction between Wzy and WaaL allows O-antigen repeat unit polymerisation, and Wzz would allow polymerisation to continue. The facilitation of the interaction between Wzy and WaaL by Wzz would result in the termination of O-antigen repeat unit polymerisation (Morona *et al.*, 1995). However, the final O-antigen chain length is reached prior to WaaL-catalysed ligation onto lipid A core-OS (Daniels *et al.*, 2002) and O-antigen modality is conferred in absence of WaaL (Woodward *et al.*, 2010), making this mechanism unlikely.

4.1.1.3 Ruler

Akin to WbdD, Wzz has also been proposed to act as both a ruler (Kintz and Goldberg, 2011) and also a scaffold (Tocilj *et al*., 2008). Both of these proposals are based on the finding that different Wzz homologues crystallise in different oligomeric states (Chang *et al*., 2015; Kalynych *et al*., 2015; Tocilj *et al*., 2008). The ability of Wzz to oligomerise has been demonstrated to play a role in O-antigen chain length modulation as mutants that are unable to oligomerise are also unable to regulate chain length (Kintz and Goldberg, 2011). A "molecular ruler" would allow polymerisation to continue until a maximum length is reached, shown in figure 4.3. The periplasmic "bell" is believed to function as this molecular ruler, with the size of the barrel imposing constraints on O-antigen chain length (Kintz and Goldberg, 2011). This proposal is supported by the finding that the periplasmic domains of different Wzz homologues crystallise in different oligomeric states, however no obvious correlation between the oligomeric state of the Wzz homologue and the final O-antigen chain length modality has yet been identified (Tocilj *et al*., 2008). Furthermore, more recent studies using cryo-EM to determine the oligomeric states of various homologues of Wzz in a lipid bilayer environment have found that different Wzz homologues all form octamers in a native-like environment, including WzzB (from *S.* Typhimurium), which previously crystallised as a pentamer (Kalynych *et al*., 2012a; Larue *et al*., 2009). SDM of Wzz (*P. aeruginosa*) enabled the identification of a single point mutation that could change O-antigen modality without altering the oligomerisation state of Wzz, as judged by comparing cross-linking profiles (Kintz and Goldberg, 2011). This mutation was predicted, using homology modelling, to be located at the monomer:monomer interface, suggesting that the packing of the monomers into the characteristic bell shape, and the stability of the bell, rather than the oligomeric state, might play a role in O-antigen chain length modulation (Kintz and Goldberg, 2011). On the basis of this SDM study, Kintz *et al* proposed that a more compact periplasmic bell would result in a shorter O-antigen chain length modality and vice versa (Kintz and Goldberg, 2011).

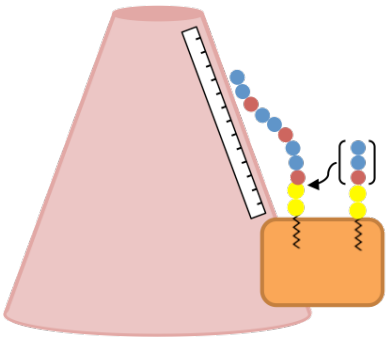


Figure 4.3. Proposed mechanism for Wzz acting as a molecular ruler. The periplasmic bell of Wzz (shown in pink) is proposed to form the ruler that determines the length of the O-antigen chain polymerised by Wzy (shown in orange) (Kintz and Goldberg, 2011). This proposal is supported by the finding that different Wzz homologues crystallise in different oligomeric states (Tocilj *et al*., 2008). However, Kintz *et al*. proposed, on the basis of mutagenesis studies, that the compactness of the Wzz oligomer might play a role in O-antigen chain length regulation (Kintz and Goldberg, 2011).

4.1.1.4 Scaffold

It is also thought that Wzz could act as a scaffold for the assembly of Wzy, with different oligomer sizes allowing different numbers of Wzy to assemble around Wzz, with the growing O-antigen chain proposed to be transferred from one Wzy to the neighbouring Wzy during polymerisation, resulting in different O-antigen chain lengths (Tocilj *et al*., 2008), shown in figure 4.4. However, the transfer of the growing O-antigen chain from one Wzy to the neighbouring Wzy would be unfavourable (Islam and Lam, 2014). This proposal is also contrary to the finding that Wzy is not only expressed at very low levels in *S. flexneri* (Daniels *et al*., 1998), but also the Wzy expression level is much lower than that of Wzz (Carter *et al*., 2009). However the level of Wzy expression has been shown to play a role in O-antigen chain length. Multiple studies have shown that the overexpression of Wzy in *S. flexneri* resulted in an increased O-antigen modal distribution (Carter *et al*., 2007, 2009; Daniels *et al*., 1998). As previously mentioned, *S. flexneri* 2a expresses two different Wzz homologues, one of which confers a short O-antigen modality of approximately 17 repeat units, the other confers a very long modality of approximately 90 O-antigen repeat units (Carter *et al*., 2009). Carter *et al.* proposed that the previously described effect of Wzy overexpression

on O-antigen chain length in *S. flexneri* 2a was due to a preferential interaction between Wzy and the Wzz homologue that confers the short O-antigen modality as Wzy only produces the very long antigen when it is expressed at higher levels (Carter *et al.*, 2009). Furthermore, they found increasing the expression of one Wzz homologue resulted in an increase in the level of O-antigen produced with the corresponding chain length and a coordinated decrease in the alternative O-antigen chain length (Carter *et al.*, 2009). This suggests that the Wzz homologues compete for the interaction with Wzy, however no evidence of a direct interaction between Wzy and either of the Wzz homologues was reported in this study (Carter *et al.*, 2009).

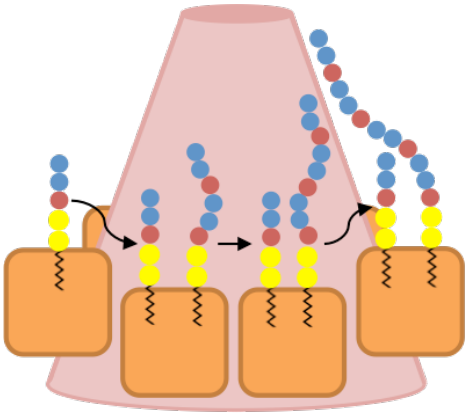


Figure 4.4. Proposed mechanism for Wzz acting as a scaffold for Wzy assembly. Wzy (orange) has been proposed to assemble around a Wzz (pink) scaffold (Tocilj *et al.*, 2008). This proposal does not agree with the finding that native Wzy expression levels in *S. flexneri* are lower than those of Wzz (Carter *et al.*, 2009), but modification of the Wzy expression level have been found to result in changes in O-antigen modality (Carter *et al.*, 2007, 2009; Daniels *et al.*, 1998).

4.1.1.5 Chain feedback

In the chain feedback model, it is proposed that Wzz maintains the growing O-antigen chain in a suitable conformation to allow the continued O-antigen polymerisation by Wzy (Kalynych *et al.*, 2012b). As the O-antigen chain length increases, the adoption of a higher order structure by the O-antigen can no longer be prevented by the interaction of the growing O-antigen chain with Wzz (Kalynych *et al.*, 2012b). This adoption of a higher order structure results in the dissociation of the undecaprenyl pyrophosphate-

linked O-antigen chain from Wzy, preventing further polymerisation and facilitating ligation of the O-antigen chain onto lipid A core-OS by WaaL (Kalynych *et al*., 2012b). This is represented by figure 4.5. This proposal is based on the finding that Wzz co-immunoprecipitates with O-antigen (Daniels *et al*., 2002). Furthermore changes in the Wzz small-angle X-ray scattering (SAXS) (Tang *et al*., 2007) and circular dichroism (CD) spectra (Guo *et al*., 2006) following incubation with O-antigen also suggest an interaction between Wzz and the O-antigen chain, supporting this mechanism of O-antigen chain regulation.

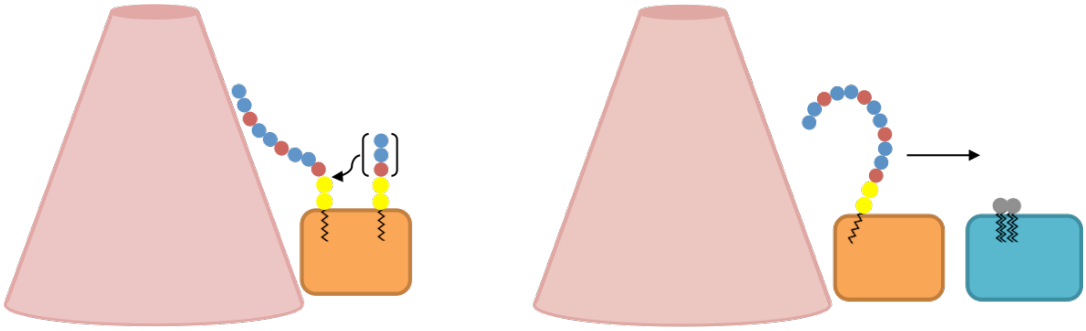


Figure 4.5. Chain feedback model. Wzz (pink) has been proposed to regulate the formation of higher order structures by the growing O-antigen chain, allowing O-antigen polymerisation to occur. The adoption of a higher order structure by the growing O-antigen chain is proposed to cause the dissociation of undecaprenyl pyrophosphate from the Wzy (orange) active site, allowing the release of undecaprenyl pyrophosphate-linked O-antigen for ligation on to lipid A by WaaL (turquoise) (Kalynych *et al*., 2012b). This model is supported by evidence of an interaction between Wzz and the O-antigen chain (Daniels *et al*., 2002; Guo *et al*., 2006; Tang *et al*., 2007).

4.1.1.6 Hybrid mechanism: molecular clock and ruler

The cryo-EM structure of full length WzzB (from *S.* Typhimurium) provided the first insight into the effect of the TMD on the structure of Wzz (Collins *et al*., 2017). This structure, along with modelling, enabled the hypothesis that Wzy associates with the outside of the Wzz oligomer and then begins to polymerise O-antigen repeat units, as shown in figure 4.6. The growing chain would associate either with the outside or the inside of the periplasmic bell of Wzz. In this model, the final O-antigen chain length

would be determined by either the duration of time Wzy remains associated with Wzz, the dissociation of Wzy would become less likely as the O-antigen chain length increases, or the constraints of the periplasmic bell of Wzz, combining both the molecular timer and ruler models (Collins *et al*., 2017). However, Wzy remains structurally uncharacterised.

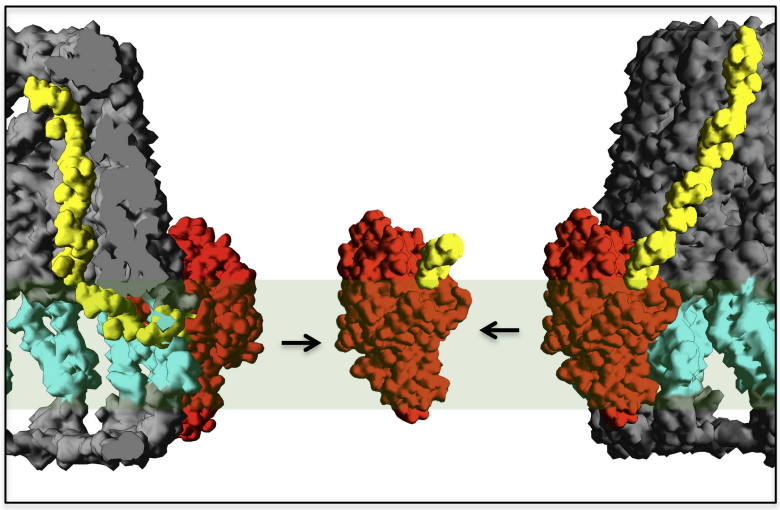


Figure 4.6. Proposed hybrid model for Wzz acting as molecular clock and a ruler. Taken from Collins *et al*. (Collins *et al*., 2007). A model of Wzy, based on the structure of acyl-CoA dehydrogenase (PDB: 1EGC), is shown in red. The growing O-antigen chain is shown in yellow. The periplasmic and cytoplasmic domains of Wzz are shown in grey and the TMHs of Wzz are shown in cyan (Collins *et al*., 2017).

4.1.1.7 Hybrid mechanism: Chain feedback and molecular ruler

This mechanism, proposed by Islam *et al.* (Islam and Lam, 2014), combines the previously discussed ruler (Kintz and Goldberg, 2011) and chain feedback (Kalynych *et al*., 2012b) models. They propose that the association of the growing O-antigen chain with Wzz prevents the formation of higher order structures by the growing O-antigen chain (Kalynych *et al*., 2012b). The tip of the periplasmic domain of Wzy is proposed to act as the ruler. Once the growing O-antigen chain reaches the top of the periplasmic domain, the adoption of a higher order structure would cause the dissociation of the undecaprenyl pyrophosphate-linked O-antigen chain from Wzz and Wzy, allowing the

WaaL-catalysed ligation of the O-antigen onto the lipid A core-OS, thereby completing LPS biosynthesis (Islam and Lam, 2014).

4.1.2 A direct interaction between Wzz and Wzy

The molecular clock, scaffold and chaperone models for Wzz-mediated chain length regulation infer a direct interaction between Wzz and Wzy, whereas the chain feedback model infers an indirect interaction between Wzz and Wzy mediated by the growing O-antigen chain. In contrast, a ruler-based mechanism could function via a direct or indirect interaction. A study from Marczak *et al.* set a precedent for a direct interaction between Wzz and Wzy (Marczak *et al*., 2013). Using a bacterial two-hybrid screen, Marczak *et al.* detected an *in vivo* interaction between the polymerase PssT and its PCP-2a PssP, both of which are required for exopolysaccharide biosynthesis, which follows the same scheme as that of Wzy-dependent LPS and CPS biosynthesis (Marczak *et al*., 2013). D3 bacteriophage produce an inhibitor of α-polymerase (Iap) peptide, which inhibits Wzy_{α} from *P. aeruginosa* PAO1 (Taylor *et al*., 2013). The level of Iap-induced inhibition of Wzy_{α} is increased in Wzz_1 and Wzz_2 mutant strains, suggesting that Iap competes with these Wzz homologues for interaction with Wzy. The finding that this peptide shares a high level of sequence similarity with the first TMD of both Wzz_1 and Wzz_2 supports this idea (Taylor *et al*., 2013). These results are suggestive of a direct interaction between Wzz and Wzy. The first evidence that Wzz and Wzy interact directly was obtained by Nath and Morona (Nath and Morona, 2015a). Using an *in vivo* crosslinking approach, they were able to detect the co-purification of Wzz when coexpressed with a GFP-His-tagged Wzy (from *S. flexneri*). (Nath and Morona, 2015a).

With the aim of gaining structural insight into the proposed interaction between Wzz and Wzy, Wzy (from *P. atrosepticum*) and Wzz (from *P. atrosepticum*) were coexpressed in *E. coli* C43 (DE3) cells and tandem affinity purified. The tandem affinity purified sample was then further analysed by Blue Native (BN)-PAGE, quantitative western blotting, negative stain and cryo-EM.

4.2 Materials and Methods

The recipes for buffers described here can be found in appendix 8.1.2.

4.2.1 Cloning

The gene encoding Wzy (from *P. atrosepticum*) was cloned into the pBADTEVC10His vector (ampicillin resistant), following amplication from *P. atrosepticum* SCRI 1043 genomic DNA, by Dr Hui Liu. The *Nco*I/*Xho*I digested fragment, provided by Dr Hui Liu, was ligated into the pEBSRCTEVDISTREP vector (ampicillin resistant), which a C-terminal TEV protease-cleavable Strep tag, shown in figure 4.7. This vector was constructed in-house by Dr Huanting Liu, University of St Andrews.

Wzz (from *P. atrosepticum*) was amplified from *P. atrosepticum* SCRI 1043 genomic DNA using the following primers: Wzz forward primer: AAA-TTT-CCA-TGG-TGA-AAT-CAG-AGA-ACT-TG; Wzz reverse primer: AAA-TTT-CTC-GAG-ACG-CAG-ACG-GCG-AGC-CAG; and the following PCR protocol: 95 °C for four minutes, followed by 30 cycles of 95 °C for 30 seconds, 50 °C for 30 seconds, 72 °C for 2.5 minutes, followed by seven minutes at 72 °C. Following PCR amplification, the PCR product was run on a 1 % agarose gel, prepared using 0.5 x TBE and 0.5 µl SYBR safe stain (Biorad), at 90 V for 30 minutes. Bands were visualised using Biorad ChemiDoc MP and purified from the gel using a QIAquick Gel Extraction Kit (Qiagen). The samples were digested using *Nco*I and *Xho*I for two hours at 37 °C and ligated into the pACYCGSTTEVDuet vector (chloramphenicol resistant), which encodes an N-terminal TEV protease-cleavable GST tag and a C-terminal S-tag, using T4 ligase overnight at RT. The pACTCGSTTEVDuet vector map can be seen in figure 4.7. The following primers were used for the insertion of a stop codon before the S-tag coding sequence in the pACYC construct of N-terminally GST-tagged and C-terminally S-tagged Wzz to produce the final N-terminally GST-tagged Wzz via SDM (Liu and Naismith, 2008): Wzz reverse primer for insertion of stop codon by SDM: ACT-CGA-GCT-AAC-GCA-GAC-GGC-GAG-CCA-GGG-CAA-C; Wzz forward primer for insertion of stop codon by SDM: TCT-GCG-TTA-GCT-CGA-GTC-TGG-TAA-AGA-AAC-CGC-TGC-TG. The following PCR protocol was used: 94 °C for seven minutes, followed by 12 cycles

of 94 °C for one minute, 55 °C for one minute and 72 °C for 10 minutes, followed by 3 cycles of 95 °C for one minute, 50 °C for one minute, 72 °C for 10 minutes. This was followed by 72 °C for 20 minutes. Following PCR amplification, the sample was treated with *Dpn*I for one hour at 37 °C and a PCR cleanup (Qiagen) was conducted. This final construct was digested with *Nco*I and *Xho*I and the insert was ligated into the pEHisTEV vector (kanamycin resistant), which encodes a TEV protease-cleavable N-terminal His_6-tag (Liu and Naismith, 2009) using T4 ligase (Promega) overnight at RT.

Ligation mixtures were transformed into *E. coli* DH5 α cells (Bioline), as described in section 2.1. Colonies were grown on a 10 ml scale in LB-medium at 37 °C overnight with the appropriate antibiotic. DNA was purified, from the resulting cell pellets, using a QIAprep Spin Miniprep kit (Qiagen). The presence of insert was detected using a control digest using *Nco*I and *Xho*1 at 37 °C for two hours, followed by agarose gel analysis. Sequencing was performed using primers against the vector backbone (Eurofins; GATC).

All constructs were transformed into *E. coli* C43 (DE3) cells, as described in section 2.1. Expression trials were conducted as described in section 2.2. Successful expression was confirmed via a combination of SDS-PAGE and western blotting against the affinity tags, as described in sections 2.3 and 2.4, respectively.

4.2.2 Co-transformation of target plasmids

A 10 ml LB culture, supplemented with the appropriate antibiotic, was grown from a glycerol stock or a single colony (*E. coli* C43 (DE3) cells), containing one of the plasmids of interest, overnight. This was used to inoculate 10 ml LB, supplemented with the appropriate antibiotic, at a 1:10 dilution. Following two hours growth at 37 °C, cells were harvested via centrifugation at 1,962 x *g* for five minutes. The supernatant was discarded and the pellet carefully resuspended in 1.5 ml ice-cold 100 mM $CaCl_2$ and 1.5 ml ice-cold 40 mM $MgSO_4$. This was incubated on ice for 30 minutes, the centrifugation step repeated, and the pellet resuspended in 100 µl ice-cold 100 mM

$CaCl_2$ and 100 µl ice-cold 40 mM $MgSO_4$. Transformation with the second plasmid of interest was then conducted as described in section 2.1.

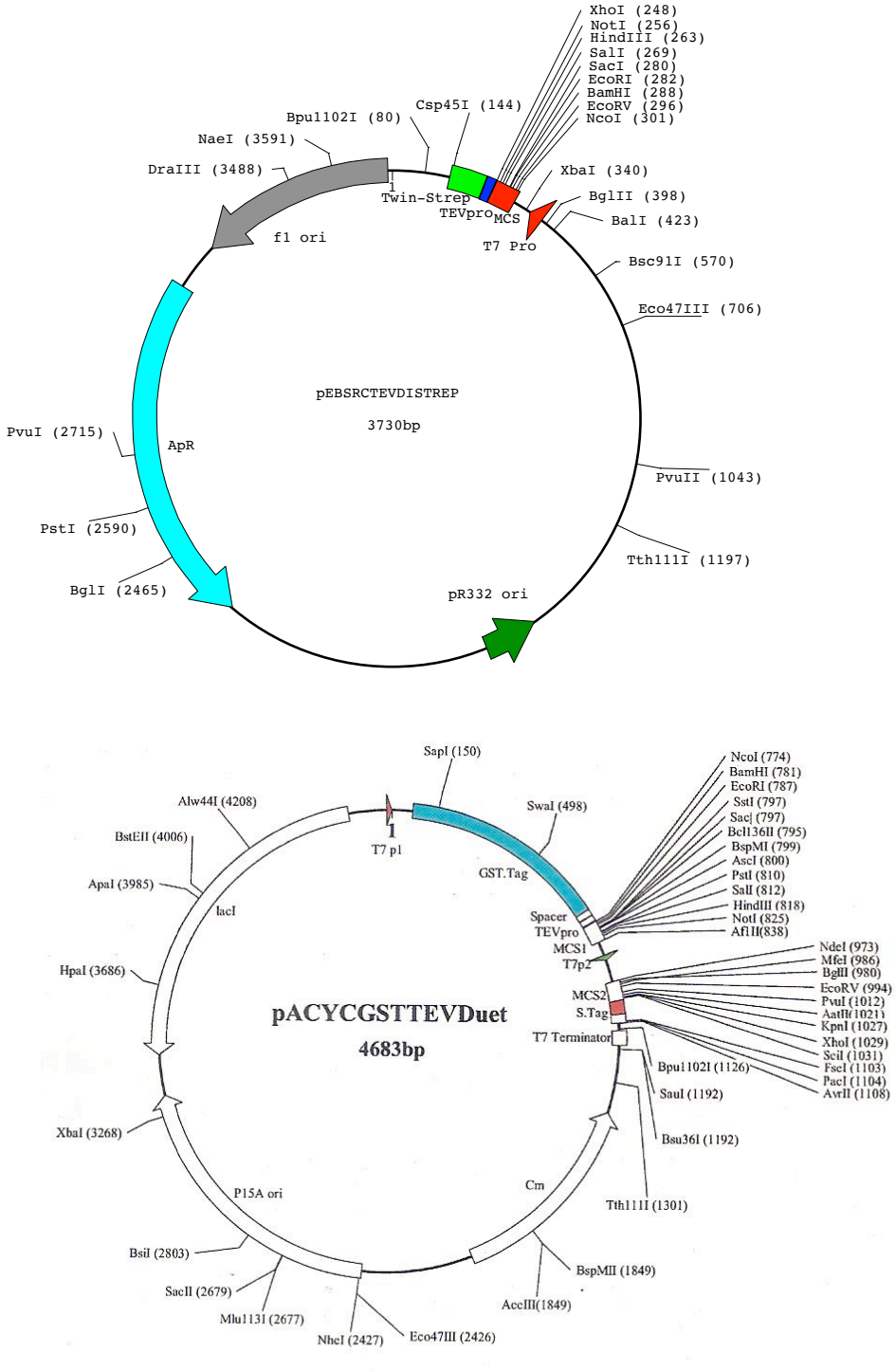


Figure 4.7. pEBSRCTEVDISTREP and pACYCGSTTEVDuet vector maps. These vectors were constructed in-house by Dr Huanting Liu, Univeristy of St Andrews. These vector maps were also provided by Dr Huanting Liu, University of St Andrews.

4.2.3 Large-scale cell cultures

All cultures were grown on a 12 L scale using LB medium. Wzz, expressed from the pACYCGSTTEVDuet vector (chloramphenicol resistant) with an N-terminal TEV protease-cleavable GST tag and C-terminal S-tag, was expressed at 25 °C overnight, following induction at an OD_{600} of 0.4 with 0.1 mM IPTG. Wzz, expressed from the pACYCGSTTEVDuet vector (chloramphenicol resistant) with an N-terminal TEV protease-cleavable GST tag, was expressed at 25 °C overnight, following induction at an OD_{600} of 0.4 with 0.1 mM IPTG. Wzz, expressed from the pEHisTEV vector (kanamycin resistant) (Liu and Naismith, 2009) with an N-terminal TEV-cleavable His_6 tag, was expressed at 15 °C overnight, following induction at an OD_{600} of 0.4 with 0.5 mM IPTG. Wzy, expressed from the pBADTEVC10His vector (ampicillin resistant) with a TEV protease-cleavable His_{10} tag at the C-terminus, was grown at 25 °C for 4 hours, following induction at an OD_{600} of 1.0 with 0.2 % (w/v) arabinose. Wzy, expressed from the pEBSRCTEVDISTREP vector (ampicillin resistant) with a TEV protease-cleavable C-terminal Strep tag, was expressed at 25 °C for 3 hours following induction at an OD_{600} of 0.4 with 0.5 mM IPTG. His-tagged Wzy was co-expressed with GST-tagged Wzz at 25 °C following induction of expression with 0.1 mM IPTG and 0.2 % arabinose at an OD_{600} of 1.0. Strep-tagged Wzy was co-expressed with His-tagged Wzz at 15 °C overnight, following induction at an OD_{600} of 0.4 with 0.5 mM IPTG. All cultures were harvested at 6,000 rpm for 10 minutes using a JLA 8.1000 rotor (Beckman Coulter) in a J-26XP centrifuge (Beckman Coulter) and stored at -80 °C until required.

4.2.4 Purification of GST-tagged Wzz

A membrane pellet prepared, using a cell pellet obtained from 6 L *E. coli* cell culture, as described in section 2.5, was homogenised in solubilisation buffer and left solubilising at 4 °C for two hours. Insoluble material was removed via ultracentrifugation for one hour at 40,000 rpm at 4 °C in a 50.2 Ti rotor (Beckman Coulter). The supernatant was incubated with 3 ml of glutathione sepharose 4B beads (GE Healthcare), equilibrated

with glutathione sepharose 4B wash buffer, for approximately two hours at 4 °C. The resin was applied to an empty column and the FT was discarded. The resin was washed with 100 ml glutathione sepharose 4B wash buffer, resuspended in 30 ml co-purification buffer, and incubated with 2 mg of TEV protease on a roller at 4 °C overnight. The FT was collected and the resin washed with 20 ml co-purification buffer and the non-cleaved protein was eluted in 30 ml glutathione sepharose 4B elution buffer. The FT and wash fractions were pooled and concentrated to 1 ml in a 50 kDa Mw cut-off Vivaspin 20 concentrator (GE Healthcare) via centrifugation at 3,488 x *g* at 4 °C. This sample was centrifuged at 20,817 x *g* for 10 minutes at 4 °C prior to loading on a superose 6 SEC column using a Biorad NCG chromatography system in co-purification buffer at RT.

To ensure that GST-tagged Wzz does not bind non-specifically to HIS-select cobalt affinity resin (Sigma Aldrich), 5 ml of the supernatant obtained following ultracentrifugation to remove the insoluble material after membrane solubilisation was incubated with 100 µl of HIS-select cobalt affinity resin (Sigma Aldrich) for two hours at 4 °C. This was centrifuged at 3,488 x *g* for 10 minutes at 4 °C to pellet the resin and the supernatant was discarded. The wash step was repeated twice. The resin was incubated with 50 CV PBS plus 0.024 % (w/v) DDM for 10 minutes and the centrifugation step was repeated to pellet the resin, and the supernatant was discarded. This washing was repeated twice and the bound protein was eluted, using the same method, in 30 CV co-purification nickel elution buffer.

The presence of the target protein was determined using SDS-PAGE, as described in section 2.3.

4.2.5 Purification of His-tagged Wzy

A membrane pellet, prepared as described in section 2.5, from an *E. coli* cell pellet obtained from 6 L of *E. coli* cell culture, was homogenised in 100 ml solubilisation buffer and solubilised at 4 °C for two hours. The solubilisation mixture was

ultracentrifuged at 4 °C for one hour at 40,000 rpm in a 50.2 Ti rotor (Beckman Coulter). The supernatant was incubated with 3 ml HIS-Select cobalt resin (Sigma Aldrich) at 4 °C for two hours. The mixture was then packed into a column and the FT was run through the column twice. The resin was then washed with 100 ml co-purification nickel wash 1, 100 ml co-purification nickel wash 2 buffer and bound protein was eluted in 30 ml co-purification nickel elution buffer. The protein was then dialysed at 4 °C for one hour in co-purification buffer, the dialysis buffer was then replaced and 2 mg of TEV protease was added to the sample and dialysis continued overnight at 4 °C. The sample was then incubated with 3 ml of nickel sepharose 6 fast flow resin (GE Healthcare) for two hours at 4 °C. The mixture was applied to an empty column, the FT was collected and the resin washed with 20 ml co-purification buffer. Bound protein was eluted in 30 ml co-purification nickel elution buffer. The FT and wash were pooled and concentrated at 3,488 x *g* at 4 °C using a Vivaspin 20 concentrator with a 50 kDa Mw cut-off (GE Healthcare). The sample was resuspended every 10 minutes until a volume of 1 ml was reached. The sample was centrifuged at 20,817 x *g* at 4 °C for 10 minutes and then applied to a superose 6 10/300 GL column in co-purification buffer using a Biorad NCG Chromatography system.

To ensure that His-tagged Wzy does not bind non-specifically to glutathione sepharose 4B resin, 5 ml of the supernatant obtained following ultracentrifugation to remove the insoluble material after membrane solubilisation was incubated with 100 µl glutathione sepharose 4B resin (GE Healthcare) for two hours at 4 °C. This was centrifuged at 3,488 x *g* for 10 minutes at 4 °C to pellet the resin and the supernatant was discarded. The resin was incubated with 50 CV PBS supplemented with 0.024 % (w/v) DDM for 10 minutes and the centrifugation step was repeated again to pellet the resin, and the supernatant was discarded. This washing was repeated four times and the bound protein was eluted, using the same method, in 30 CV glutathione sepharose 4B elution buffer.

The presence of target protein was determined using SDS-PAGE, as described in section 2.3.

4.2.6 Purification of His-tagged Wzz

A membrane pellet, prepared as described in section 2.5, from an *E. coli* cell pellet obtained from 6 L of *E. coli* cell culture, was solubilised in 100 ml solubilisation buffer for two hours at 4 °C and this was ultracentrifuged at 40,000 rpm for one hour at 4 °C in a 50.2 Ti rotor (Beckman Coulter). The supernatant was incubated with 2 ml HIS-select cobalt affinity resin (Sigma Aldrich) for two hours at 4 °C. The FT was discarded and the resin washed with 50 CV co-purification nickel wash 1 buffer, followed by 50 CV co-purification nickel wash 2 buffer. Bound protein was eluted in 15 CV co-purification nickel elution buffer. The eluate was dialysed for one hour at 4 °C in 1 L co-purification buffer. The dialysis buffer was replaced and dialysis continued overnight at 4 °C.

If TEV protease cleavage was required, 2 mg of TEV protease was added to the dialysis tubing and incubated with the sample overnight at 4 °C. The sample was incubated with 2 ml nickel sepharose 6 fast flow resin for two hours at 4 °C. The sample was applied to an empty column, the resin washed with 10 CV co-purification buffer and bound protein eluted in 15 CV co-purification nickel elution buffer.

The sample was concentrated to 1 ml using a Vivaspin 20 concentrator with a 50 kDa Mw cut-off (GE Healthcare) at 3,488 x *g* at 4 °C, centrifuged at 20,817 x *g* for 10 minutes and run on a superose 6 10/300 SEC column in co-purification buffer using a Biorad NCG chromatography system at RT.

To ensure that His-tagged Wzz does not exhibit any non-specific binding to streptactin resin (IBA), 5 ml of the supernatant obtained following ultracentrifugation to remove the insoluble material after membrane solubilisation was incubated with 100 µl streptactin resin (IBA) for two hours at 4 °C. This was centrifuged at 3,488 x *g* for 10 minutes at 4 °C to pellet the resin and the supernatant was discarded. The resin was incubated with 50 CV PBS supplemented with 0.024 % (w/v) DDM for 10 minutes and the centrifugation step was repeated to pellet the resin, and the supernatant was discarded. This washing was repeated twice and the bound protein was eluted, using the same method, in 30 CV PBS, 0.024 % (w/v) DDM, 2.5 mM desthiobiotin, pH 8.0.

The presence of target protein was determined via SDS-PAGE, as described in section 2.3.

4.2.7 Purification of Strep-tagged Wzy

A membrane pellet, prepared as described in section 2.5 from an *E. coli* cell pellet obtained from 6 L of *E. coli* cell culture, was solubilised in co-purification buffer, using 1 % (w/v) DDM rather than 0.024 % (w/v) DDM, for two hours at 4 °C and then ultracentrifuged for one hour at 40,000 rpm at 4 °C, using a 50.2 Ti (Beckman Coulter) to pellet any insoluble material. The supernatant was batch-bound for one hour on 1 ml streptactin resin (IBA) at 4 °C. The FT was discarded, the resin washed with 100 CV co-purification buffer and bound protein eluted in 30 CV co-purification streptactin elution buffer. The eluate was concentrated to 1 ml at 3,488 x *g* at 4 °C using a Vivaspin 20 concentrator with a 50 kDa Mw cut-off (GE Healthcare), centrifuged at 20,817 x *g* for 10 minutes at 4 °C and run on a superose 6 10/300 GL SEC column in co-purification buffer, at RT, using a Biorad NCG chromatography system.

To ensure that Strep-tagged Wzy does not show non-specific binding to HIS-Select cobalt affinity resin (Sigma Aldrich), 5 ml of the supernatant obtained following ultracentrifugation to remove the insoluble material after membrane solubilisation was incubated with 100 µl HIS-Select cobalt affinity resin (Sigma Aldrich) for two hours at 4 °C. This was centrifuged at 3,488 x *g* for 10 minutes at 4 °C to pellet the resin and the supernatant was discarded. The resin was incubated with 50 CV PBS supplemented with 0.024 % (w/v) DDM for 10 minutes and the centrifugation step was repeated to pellet the resin, and the supernatant was discarded. This washing was repeated twice and the bound protein was eluted, using the same method, in 30 CV co-purification nickel elution buffer.

The presence of the target protein was determined using SDS-PAGE, as described in section 2.3.

4.2.8 Tandem Affinity Purification: GST-tagged Wzz and His-tagged Wzy

The membrane pellet was solubilised in solubilisation buffer for two hours at 4 °C and then ultracentrifuged for one hour at 40,000 rpm at 4 °C in a 50.2 Ti rotor (Beckman Coulter) to pellet any insoluble material. If purifying first via the GST-tag of Wzz and then via the His-tag of Wzy, the supernatant was batch-bound for two hours at 4 °C to glutathione sepharose 4B resin and washed with 100 ml glutathione sepharose 4B wash buffer. Bound protein was eluted in 30 ml glutathione sepharose 4B elution buffer. The sample was concentrated at 3,488 x *g* at 4 °C, resuspending every 10 minutes until a final volume of 10 ml was reached. The sample was desalted into co-purification buffer using a CentriPure P100 desalting column (Generon), according to the manufacturer's protocol. Protein was bound onto 2 ml HIS-select cobalt affinity gel (Sigma Aldrich) overnight at 4 °C, washed with 50 ml co-purification buffer, and then eluted in 30 ml co-purification nickel elution buffer. If required, the desalting step was repeated and the protein was cleaved with TEV protease for 5 hours at 4 °C.

If purifying first via the His-tag of Wzy and then via the GST-tag of Wzz, the supernatant was batch-bound to 2 ml nickel sepharose 6 fast flow resin (GE Healthcare) for two hours at 4 °C. The FT was applied to the column twice and the resin was washed with 50 CV co-purification nickel wash 1, followed by 50 CV co-purification nickel wash 2 and the bound protein was eluted in 15 CV co-purification nickel elution buffer. The sample was dialysed for one hour at 4 °C in co-purification buffer. The dialysis buffer was replaced and dialysis continued overnight at 4 °C. The sample was bound to 3 ml glutathione sepharose 4B for two hours at 4 °C and the resin was washed with 50 ml co-purification buffer and bound protein was eluted using 30 ml glutathione sepharose 4B elution buffer. If required, the sample was incubated with 2 ml TEV protease overnight at 4 °C.

Following affinity purification, and TEV protease cleavage if required, the protein was concentrated in a Vivaspin 20 concentrator with a 50 kDa Mw cut-off (GE Healthcare) via centrifugation at 3,488 x *g* at 4 °C, resuspending every 10 minutes. Once a volume of 1 ml was reached the protein was centrifuged at 20,817 x *g* for 10 minutes and injected onto a superose 6 10/300 GL SEC column in co-purification buffer using a Biorad NCG Chromatography system.

The presence of the target proteins was determined via SDS-PAGE and western blotting against the affinity tags, as described in sections 2.3 and 2.4.

4.2.9 Tandem Affinity Purification: His-tagged Wzz and Strep-tagged Wzy

A membrane pellet, prepared as described in section 2.5, from an *E. coli* pellet obtained from 12 L of *E. coli* cell culture, was solubilised in solubilisation buffer for two hours at 4 °C and ultracentrifuged for one hour at 40,000 rpm at 4 °C using a 50.2 Ti rotor (Beckman Coulter) to pellet any insoluble material. If purifying first via the His-tag of the Wzz and then via the Strep-tag of Wzy, the supernatant was batch-bound for two hours at 4 °C to 3 ml HIS-select cobalt affinity resin (Sigma Aldrich) and washed with 100 ml co-purification nickel wash 1, followed by 100 ml co-purification nickel wash 2. Bound protein was eluted in 30 ml co-purification nickel elution buffer. The sample was dialysed for one hour at 4 °C in co-purification buffer. The dialysis buffer was replaced and dialysis continued overnight at 4 °C. The sample was bound onto 1.5 ml streptactin resin (IBA) for one hour at 4 °C, washed with 50 ml co-purification buffer, and eluted in 30 ml co-purification streptactin elution buffer. If required, the protein was then cleaved with TEV protease at 4 °C for 3 hours.

If purifying first via the Strep-tag of Wzy and then via the His-tag of Wzz, the supernatant was batch-bound for one hour on 1.5 ml streptactin resin (IBA). The resin washed with 100 ml streptactin wash buffer and bound protein was eluted using 30 ml streptactin elution buffer. The sample was dialysed for one hour at 4 °C in co-

purification buffer. The dialysis buffer was replaced and dialysis continued overnight at 4 °C. The sample was batch-bound on to 2 ml HIS-Select cobalt affinity resin (Sigma Aldrich) for two hours at 4 °C. The mixture was applied to an empty column, the resin was washed with 100 ml co-purification buffer and the bound protein was eluted in 30 ml co-purification nickel elution buffer. If required, the protein was cleaved with TEV protease at 4 °C for 3 hours.

Following affinity purification, and TEV protease cleavage if required, the sample was concentrated in a Vivaspin 20 concentrator with a 50 kDa Mw cut-off (GE Healthcare) via centrifugation at 3,488 x *g* at 4 °C, resuspending every 10 minutes. Once a volume of 1 ml was reached the protein was centrifuged at 20,817 x *g* for 10 minutes and injected onto a superose 6 10/300 GL SEC column in co-purification buffer using a Biorad NCG Chromatography system.

The presence of the target proteins was confirmed via SDS-PAGE and western blotting against the affinity tags, as described in sections 2.3 and 2.4, respectively.

4.2.10 Dot blotting for stoichiometry determination

His-tagged Wzz was purified as described in section 4.2.6. Strep-tagged Wzy was purified as described in section 4.2.7. Tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy, purified as described in section 4.2.9. The TEV protease cleavage of all affinity tags was omitted. Three samples (80 ng, 40 ng and 20 ng) of the central fraction of the peak, detected by UV following SEC, obtained for tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy were applied directly to a nitrocellulose membrane (GE Healthcare) in triplicate, along with a serial dilution of either His-tagged Wzz or Strep-tagged Wzy. The serial dilutions of His-tagged Wzz and Strep-tagged Wzy began at 1 μg and each subsequent dilution was two-fold. 1 μg of His-tagged Wzz was included on the dot blot probed with anti-Strep tag antibody and 1 μg of Strep-tagged Wzy was included on the dot blot probed with anti-His tag antibody to ensure

that there was no non-specific interaction with the antibody. The dot blots were continued as described in section 2.4.

The signal for each sample of the standards from the dilution series was quantified using the Biorad ChemiDoc MP imaging software, with a background correction, and used to create a standard curve using Microsoft Excel. As there was non-tagged Wzz present in both the His-tagged Wzz used for the standard curve and the tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy sample, the standard curve, calculated using the signal from the dot blot probed with anti-His tag antibody, was corrected to account for the presence non-tagged Wzz. The correction procedure used was as follows: the signal on an SDS-PAGE gel of the His-tagged Wzz and non-tagged Wzz, present in the His-tagged Wzz sample used for the serial dilution, was quantified using the Biorad ChemiDoc Imaging software, allowing the determination of the ratio of His-tagged Wzz to non-tagged Wzz present in the sample. This ratio was used to correct the standard curve, under the assumption that ratio of His-tagged Wzz to non-tagged Wzz is the same in presence and in absence of Strep-tagged Wzy.

The standard curve was used to determine the mass of both proteins present in the His-tagged Wzz plus Strep-tagged Wzy samples, and was used to calculate the molar ratio of the two proteins present.

4.2.11 SEC-MALS

SEC-MALS was conducted using a superose 6 10/300 SEC column using co-purification buffer, as detailed in section 2.6.

4.2.12 BN-PAGE

Samples were prepared by mixing 6.5 µl of the protein of interest at 1 mg/ml, 2.5 µl of 4 x NativePage sample buffer (Thermo Fisher Scientific) and 1 µl of NativePage 5 % G-250 Sample Additive (Thermo Fisher Scientific). The wells of a NativePAGE Novex

Bis-tris 4-16 % precast gel (Thermo Fisher Scientific) were washed with water prior to loading 3 µl of each sample, along with 5 µl of NativeMark Unstained Protein Standard (Thermo Fisher Scientific). The anode and dark cathode buffers were prepared according to the manufacturer's protocol and cooled at 4 °C prior to use (Thermo Fisher Scientific). The gel was run at 150 V on ice until dye front left the gel.

4.2.13 Western blotting of Blue Native PAGE

BN-PAGE was conducted as described in section 4.2.12, but the dark cathode buffer was exchanged for light cathode buffer (Thermo Fisher Scientific) once the gel had migrated one third of the total distance, as determined by the position of the dye front, and then the BN-PAGE was continued until the dye front reached the bottom of the gel. The gel was washed in 50 ml 2 x NuPAGE transfer buffer plus 0.1 % SDS for 15 minutes at RT. Proteins were transferred from the gel onto a PVDF membrane using 1 x NuPAGE transfer buffer supplemented with 10 % (v/v) methanol and the Biorad semi-dry system, as described in section 2.4. Following transfer, the membrane was air-dried for approximately 45 minutes. The membrane was rinsed in methanol to reactivate the membrane and then rinsed in water to remove the excess methanol. Western blotting, using anti-His tag (Sigma Aldrich) and anti-Strep tag (IBA) antibodies, was carried out as described in section 2.4.

4.2.14 Far western dot blot

Strep-tagged Wzy and His-tagged Wzz were spotted onto nitrocellulose membranes, along with the following controls: 5 µl co-purification buffer, 2 µg of WaaL ligase (from *Helicobacter pylori*), BSA and T4 lysozyme. Table 4.1 shows the amounts of Strep-tagged Wzy and His-tagged Wzz spotted onto the membranes. The membranes were blocked with 5 % (w/v) instant dried skimmed milk powder (Tesco) in PBST for one hour at RT. The membrane bound with Strep-tagged Wzy was probed overnight at 4 °C with 10 µg His-tagged Wzz in PBST supplemented with DDM at 0.024 % (w/v).

The membrane bound with His-tagged Wzz was probed overnight at 4 °C with 10 µg Strep-tagged Wzy in PBST supplemented with DDM at 0.024 % (w/v). The membrane probed with His-tagged Wzz was washed three times, with each wash lasting 10 minutes, in PBST and probed with anti-His tag antibody (Sigma Aldrich) at a 1:10,000 dilution for 5 hours at RT. The membrane probed with Strep-tagged Wzy was washed with PBST 3 times with each wash lasting 10 minutes and probed with anti-Strep tag antibody (IBA) at a 1:20,000 dilution for 5 hours at RT. The membranes were washed again in PBST three times, with each wash lasting 10 minutes and developed using SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific) according to the manufacturer's protocol. Images were visualised using the Biorad ChemiDoc MP system.

Amount of immobilised Wzz (µg)	Amount of immobilised Wzy (µg)
0.20	0.11
0.40	0.22
0.81	0.43
1.21	0.65
1.61	0.87
2.02	1.08
4.03	2.17
8.07	4.33

Table 4.1 The amount of protein immobilised on the membrane during far western dot blotting.

4.2.15 Negative stain EM and cryo-EM

Negative stain EM and cryo-EM were conducted in close collaboration with Dr David Bhella, MRC-University of Glasgow Centre for Virus Research. For initial negative stain EM quality analysis, a JEOL 120 transmission electron microscope (TEM), equipped with a Gatan Orius CCD camera, was used. For cryo-EM, carbon-coated copper grids were used, which were glow discharged, in the presence of amylamine, using an Emitech K100x glow discharge unit. If required, grids were coated with carbon

using a Quorum Technologies Q150T carbon coater, prior to glow discharge. Samples were applied to the grids using a Vitribot (FEI). The grids were blotted to remove excess sample and grids were frozen in liquid ethane using a Vitribot (FEI). Data were collected using a JEOL 2200 200 kV TEM, equipped with a DE20 direct electron detector. Tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy, purified first via the His-tag of Wzz and then via the Strep-tag of Wzy, followed by TEV protease cleavage to remove the affinity tags and SEC, was used to prepare the final cryo-EM samples. The final protein concentration used for cryo-EM data collection was 5 mg/ml, as determined using a Nanodrop using the default 1 Abs = 1 mg/ml setting, as the extinctient coefficient of the complex was not known. Data were processed by Dr David Bhella using Motioncorr (Li *et al*., 2013) and Relion software (Scheres, 2012a, 2012b).

4.2.16 Intact mass spectrometry

Samples of Wzz, Wzy and Wzz plus Wzz, purified as described in sections 4.2.5., 4.2.7., and 4.2.9., respectively, have been submitted for intact mass spectrometry in collaboration with Professor Carol Robinson, University of Oxford.

4.3 Results

It has been proposed that the proteins required for LPS biosynthesis in the Gram-negative bacterial membrane could form a complex (Marolda *et al*., 2006; Nath and Morona, 2015a; Tocilj *et al*., 2008). As the WaaL, Wzx, Wzz and Wzy homologues required for ECA biosynthesis in *P. atrosepticum* could be successfully expressed in our group (Dr Hui Liu and Dr Audrey Le Bas, University of St Andrews, personal communication), this was chosen as a suitable model pathway to explore the protein:protein interactions (PPIs) required for ECA and LPS biosynthesis further. As ECA is also biosynthesised in a Wzy/Wzz-dependent manner, and, akin to the O-antigen, can also be ligated on to lipid A to form ECA_{LPS}, its biosynthesis is expected to follow the same mechanism as LPS biosynthesis (Barr *et al*., 1999; Kuhn *et al*., 1988;

Marolda *et al*., 2006). Furthermore, the components of the ECA repeat unit have been determined, opening up the possibility for substrate production (Kuhn *et al*., 1988).

4.3.1 Purification of GST-tagged Wzz

The method chosen to determine whether Wzz and Wzy interact directly to form a complex was tandem affinity purification (Puig *et al*., 2001; Rigaut *et al*., 1999). This requires the co-expression of the proteins of interest, each fused to a different affinity tag, and in this case coded for by different vectors conveying different antibiotic resistance. Following the successful co-expression, the lysate, or detergent-solubilised membrane in this case, is applied to the first affinity column. The eluate of the first affinity column is then applied to the second affinity column. The presence of both proteins of interest in the eluate of the second column indicates that they form a stable complex (Puig *et al*., 2001; Rigaut *et al*., 1999). An advantage of this technique is that it can be used in combination with other techniques for the investigation of PPIs and the purified protein can be used for structural characterisation. The initial affinity tagging combination chosen was N-terminally GST-tagged Wzz and C-terminally His_{10}-tagged Wzy.

Prior to the tandem affinity purification of Wzz plus Wzy, we needed to ensure that GST-tagged Wzz could be expresse and purified, and that it did not exhibit any non-specific binding to the cobalt affinity resin used. As can be seen in figure 4.8, GST-tagged Wzz could be successfully purified and eluted as a single peak outside of the void volume of a SEC column. The identity of this purified protein was confirmed by mass spectrometry. Furthermore, GST-tagged Wzz did not exhibit non-specific binding to cobalt affinity resin, when tested on a small scale, shown in figure 4.9.

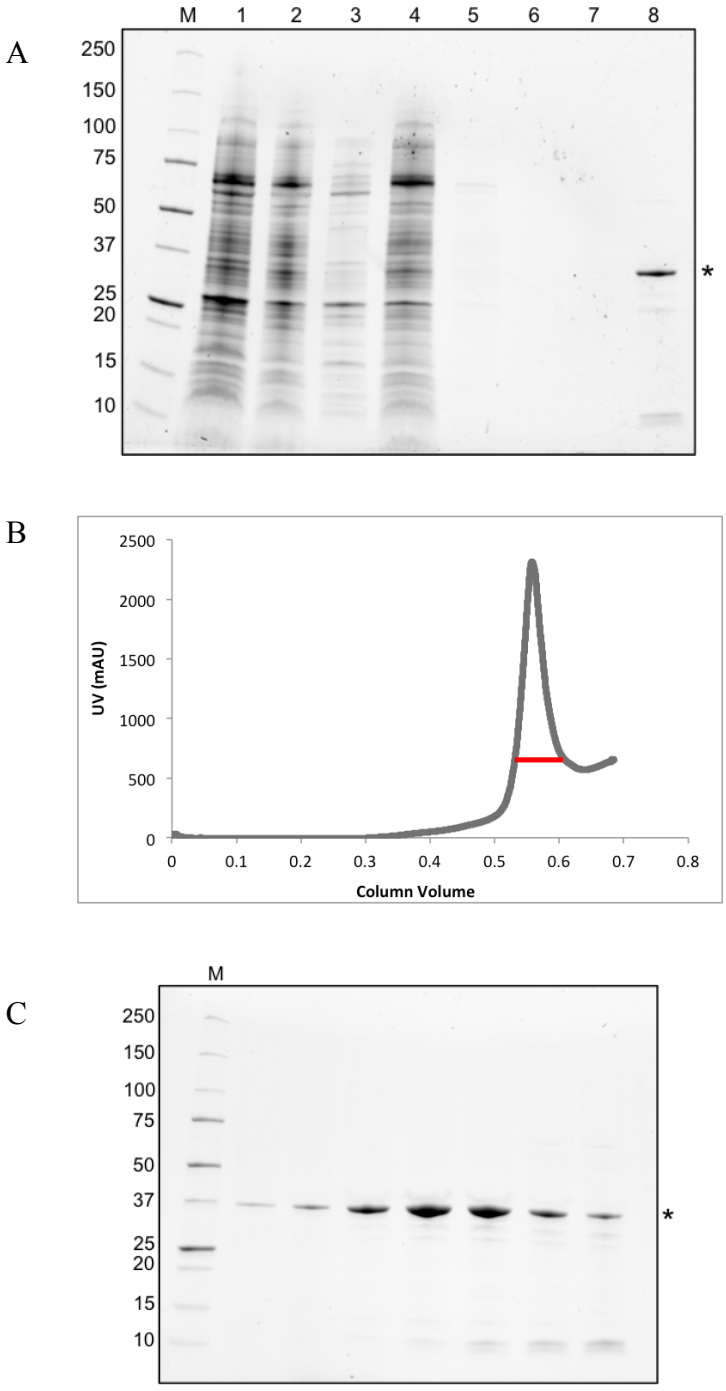


Figure 4.8. Purification of N-terminally GST-tagged Wzz. The SDS-PAGE gel of the glutathione sepharose 4B column is shown in "A" (M: marker; 1: total membrane solubilisation fraction; 2: DDM-soluble membrane fraction; 3: DDM-insoluble membrane fraction; 4: glutathione sepharose 4B FT; 5-7: glutathione sepharose 4B wash fractions 1-3; 8: glutathione sepharose 4B elution via TEV protease-cleavage on column). "B" shows the SEC UV profile and "C" shows the SDS-PAGE gel of the peak fractions. The approximate positions of the fractions in "B" shown in "C" is indicated with a solid red line. The band corresponding to Wzz is indicated with a "*".

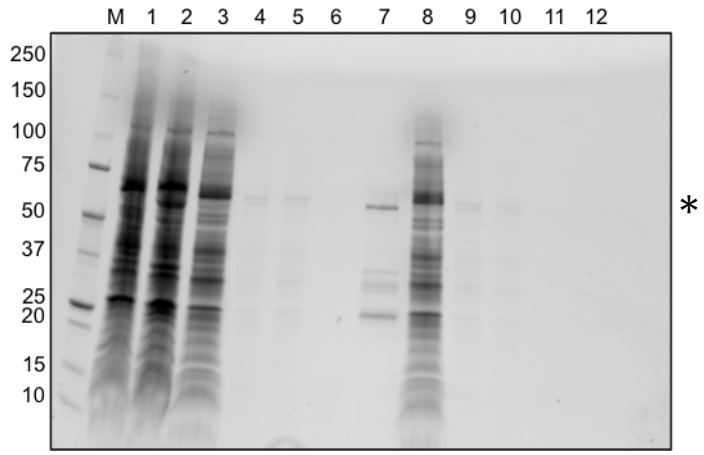


Figure 4.9. SDS-PAGE gel of the small scale purification of GST-tagged Wzz using cobalt affinity resin as a binding control. (M: marker; 1: total membrane fraction; 2: DDM-solubilised membrane fraction; 3: DDM-insoluble membrane fraction; 4: glutathione sepharose 4B FT; 4-6: glutathione sepharose 4B wash fractions 1-3, 7: glutathione sepharose 4B eluate; 8: cobalt affinity column FT, 9-11: cobalt affinity wash fractions 1-3; 12: cobalt affinity eluate). The band at approximately 65 kDa, indicated with a "*", corresponds to GST-tagged Wzz, that at approximately 37 kDa corresponds to Wzz without the N-terminal GST tag and the band at 25 kDa corresponds to the N-terminal GST-tag.

4.3.2 Purification of His-tagged Wzy

As well as confirming that N-terminally GST-tagged Wzz could be expressed and purified, we also needed to determine that C-terminally His-tagged Wzy could be purified. As shown in figure 4.10, His-tagged Wzy could be successfully purified and eluted from the SEC column as a single peak, eluting after the void volume, indicating that this protein is not aggregated. The identity of this protein was confirmed by mass spectrometry. This construct also did not exhibit non-specific binding to the glutathione sepharose 4B column, as can be seen in figure 4.11.

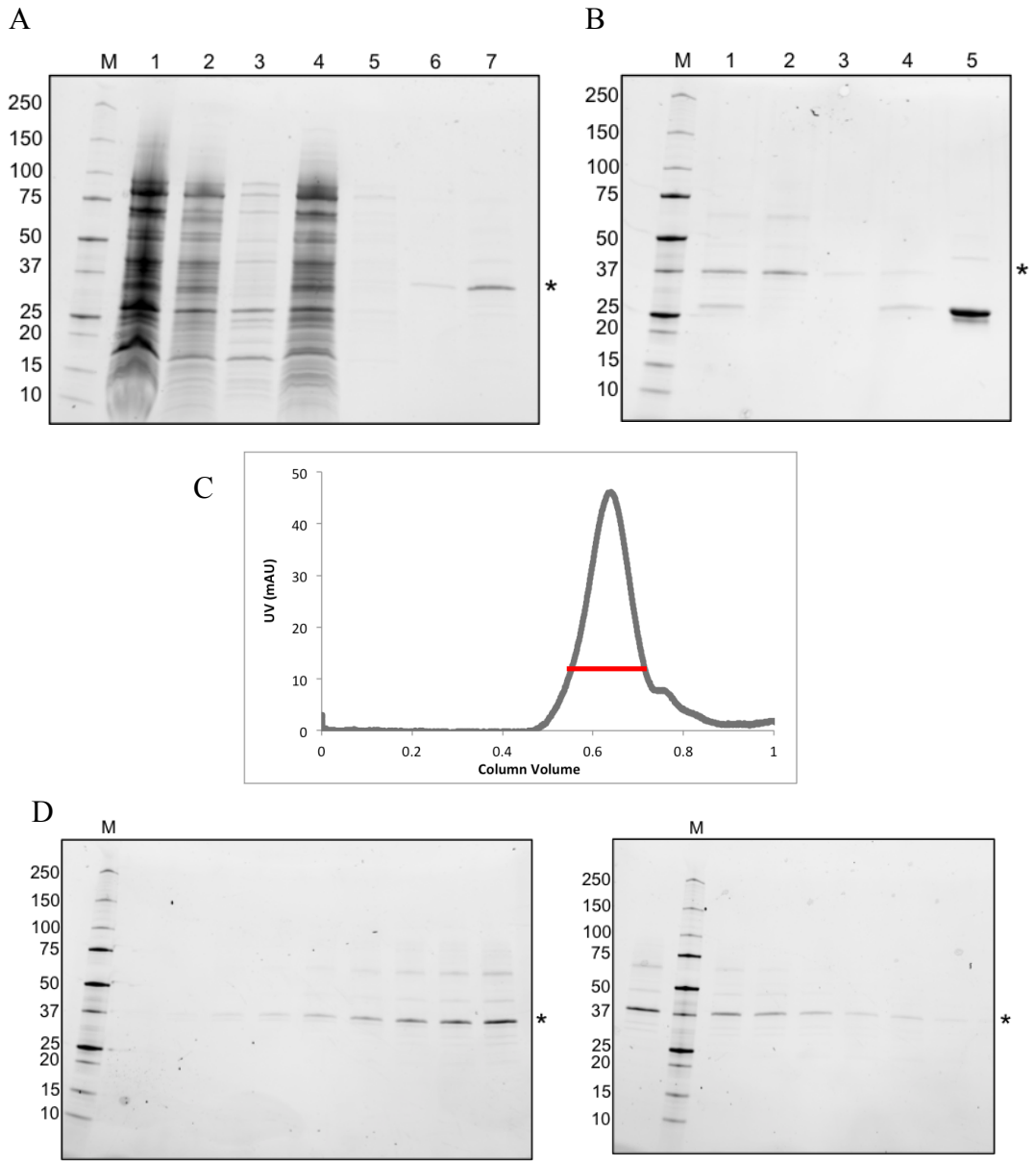


Figure 4.10. Purification of Wzy with a C-terminal His_{10} tag. "A" shows the SDS-PAGE gel of the first cobalt affinity column (M: marker; 1: total membrane fraction; 2: DDM-soluble membrane fraction; 3: DDM-insoluble membrane fraction; 4: cobalt affinity FT; 5: cobalt affinity wash 1; 6: cobalt affinity wash 2; 7: cobalt affinity eluate). "B" shows the SDS-PAGE gel of the second nickel affinity column (1: TEV protease-cleaved sample taken before second nickel affinity column; 2: second nickel affinity column FT; 3: second nickel affinity column wash; 4: second nickel affinity column eluate; 5: TEV protease). "C" shows the SEC UV profile and the SDS-PAGE gel of fractions taken across the peak is shown in "D". The approximate positions of the fractions in "C" shown in "D" is indicated with a solid red line. The band corresponding to Wzy at approximately 37 kDa is indicated with a "*".

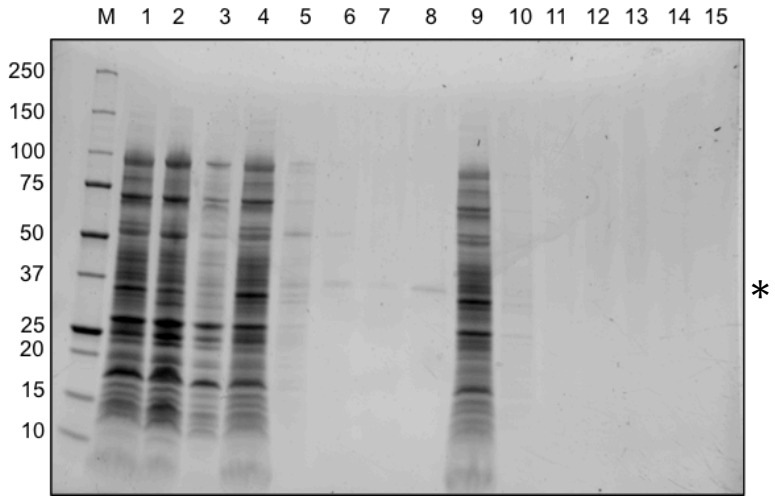


Figure 4.11. SDS-PAGE gel of the small scale purification His-tagged Wzy using glutathione sepharose 4B as a binding control. The band at approximately 37 kDa, indicated with a "*", corresponds to His-tagged Wzy. (M: marker; 1: total membrane fraction; 2: DDM-soluble membrane; 3: DDM-insoluble membrane; 4: cobalt affinity FT; 5: cobalt affinity wash 1; 6: cobalt affinity wash 2 fraction 1; 7: cobalt affinity wash 2 fraction 2; 8: cobalt affinity eluate; 9: glutathione sepharose 4B FT; 10-14: glutathione sepharose 4B wash 1-5; 15: glutathione sepharose 4B eluate.

4.3.3 Tandem affinity purification of His-tagged Wzy and GST-tagged Wzz

GST-tagged Wzz and His-tagged Wzy were co-expressed in *E. coli* C43 (DE3) cells and tandem affinity purified. The SDS-PAGE gel and western blots, probed with anti-His tag and anti-GST tag antibodies, of the tandem affinity purification can be seen in figure 4.12 for the tandem affinity purification via the His-tag of Wzy and then via the GST-tag of Wzz, and in figure 4.13 for the tandem affinity purification via the GST-tag of Wzz and then via the His-tag of Wzy. The SDS-PAGE gel of the tandem affinity purification via the His-tag of Wzy and then via the GST-tag of Wzz was omitted from figure 4.12 due to the low yield, which was below the detection limit of the Stain Free SDS-PAGE gel (Biorad). As can be seen when comparing the western blots, probed with anti-His tag and anti-GST antibodies, shown in figures 4.12 and 4.13, His-tagged Wzy and GST-tagged Wzz were both found in the eluate of the second column,

regardless of which protein was pulled down first, indicative of an interaction between these proteins. Furthermore, the wash of the second column in the tandem affinity purification, shown in figure 4.12, did not contain protein that could be detected using the methods described here, which further indicates a specific interaction between these proteins. However the sample was not concentrated prior to SDS-PAGE and western blotting analysis, which may mean these proteins are present at levels below the detection limit. The wash of the second column of the tandem affinity purification shown in figure 4.13 contained GST-tagged Wzz indicating that, when purifying on a large scale, there was some non-specific binding of GST-tagged Wzz to the cobalt affinity resin. Alternatively, the purified complex is unstable and begins to degrade after a period of time. As can be seen in the SDS-PAGE gel of figure 4.13, the GST tag of Wzz is cleaved either during expression or purification, resulting in free GST, which runs at approximately 25 kDa on the SDS-PAGE gel. This does not pose a problem for the purification of GST-tagged Wzz or the tandem affinity purification of GST-tagged Wzz and His-tagged Wzy, due to the oligomerisation of Wzz.

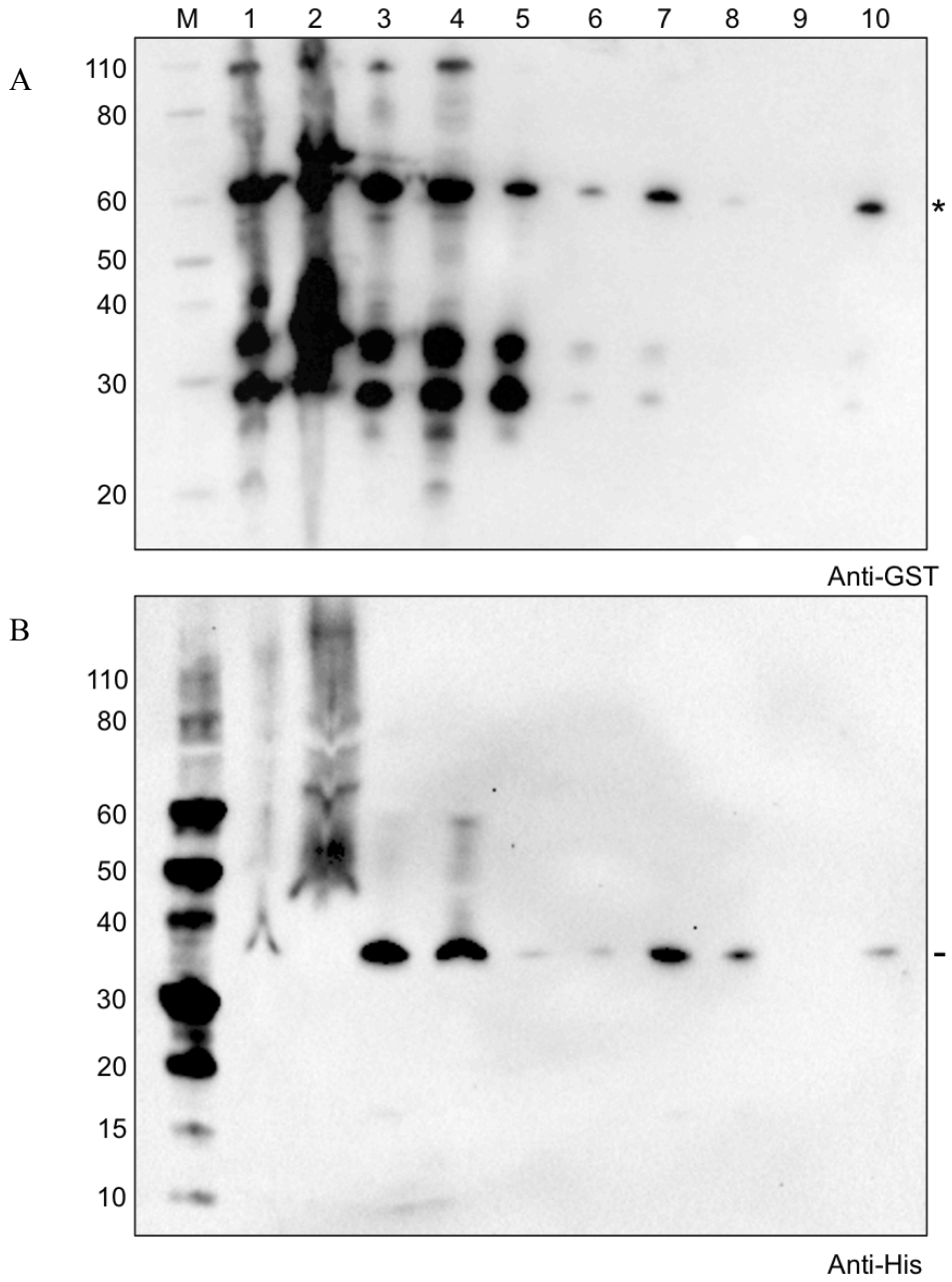


Figure 4.12. Tandem affinity purification of GST-tagged Wzz and His-tagged Wzy using cobalt affinity resin and then glutathione sepharose 4B resin. The western blots, probed with anti-GST tag antibody, shown in "A", and anti-His tag antibody, shown "B", of the tandem affinity purification show that both GST-tagged Wzz and His-tagged Wzy are found in the eluate of the second affinity purification column (lane 10), suggesting an interaction between these proteins. (M: marker; 1: total membrane fraction; 2: DDM-soluble fraction; 3: DDM-insoluble fraction; 4: cobalt affinity FT; 5: cobalt affinity wash 1; 6: cobalt affinity wash 2; 7: cobalt affinity eluate; 8: glutathione sepharose 4B FT; 9: glutathione sepharose 4B wash; 10: glutathione sepharose 4B eluate). The band on the western blot, probed with anti-GST tag antibody, at approximately 60 kDa, indicated with a "*", corresponds to GST-tagged Wzz. The band on the western blot, probed with anti-His tag antibody, at approximately 37 kDa, indicated with a "-", corresponds to His-tagged Wzy.

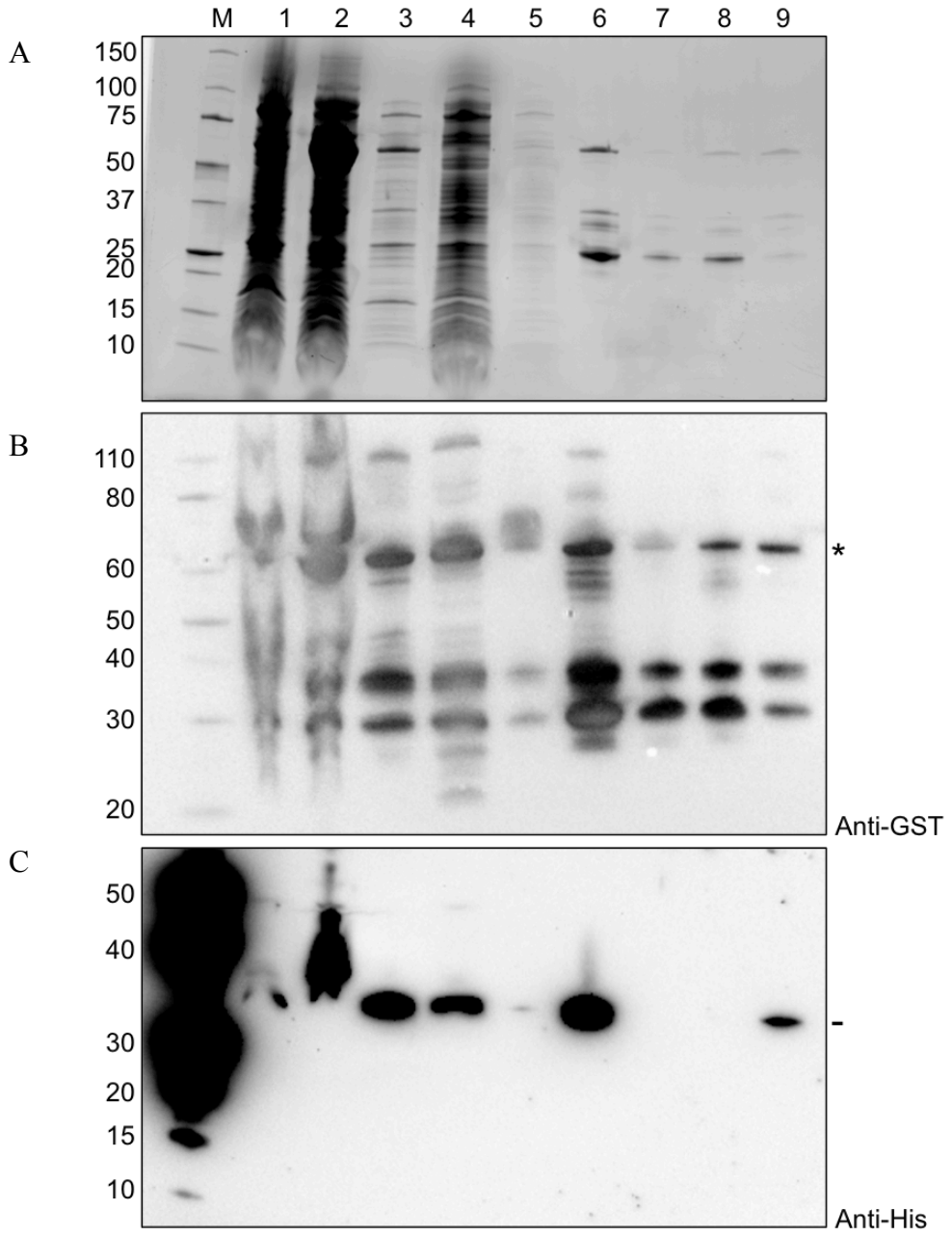


Figure 4.13. Tandem affinity purification of GST-tagged Wzz and His-tagged Wzy using glutathione sepharose 4B resin and then cobalt affinity resin. The SDS-PAGE gel ("A") and western blots, probed with anti-GST tag ("B") and anti-His tag ("C") antibodies, of the tandem affinity purification show that both GST-tagged Wzz and His-tagged Wzy are found in the eluate of the second affinity purification column (lane 9), suggesting an interaction between these proteins. (M: marker; 1: total membrane fraction; 2: DDM-soluble fraction; 3: DDM-insoluble fraction; 4: glutathione sepharose 4B FT; 5: glutathione sepharose 4B wash; 6: glutathione sepharose 4B eluate; 7: cobalt affinity FT; 8: cobalt affinity wash; 9: cobalt affinity eluate). The band on the SDS-PAGE gel and western blot, probed with anti-GST tag antibody, at approximately 60 kDa, indicated with a "*" corresponds to GST-tagged Wzz. The band on the SDS-PAGE gel and western blot, probed with anti-His tag antibody, at approximately 37 kDa, indicated with a "-", corresponds to His-tagged Wzy.

4.3.4 Purification of Strep-tagged Wzy

In order to ensure that the interaction observed via tandem affinity purification is not due to the tagging strategy chosen, a new combination of affinity tags was investigated. This strategy required an N-terminally His-tagged Wzz and C-terminally Strep-tagged Wzy. These proteins initially needed to be purified individually in order to ensure that they could be successfully purified and that they do not exhibit non-specific binding to the affinity column required to pulldown via the other affinity tag. As shown in figure 4.14, Strep-tagged Wzy could be purified and eluted as a single peak, after the void volume, from the SEC column, indicating that the sample was homogeneous with no aggregation, and did not exhibit any non-specific binding to cobalt affinity resin. The identity of this protein was confirmed by mass spectrometry.

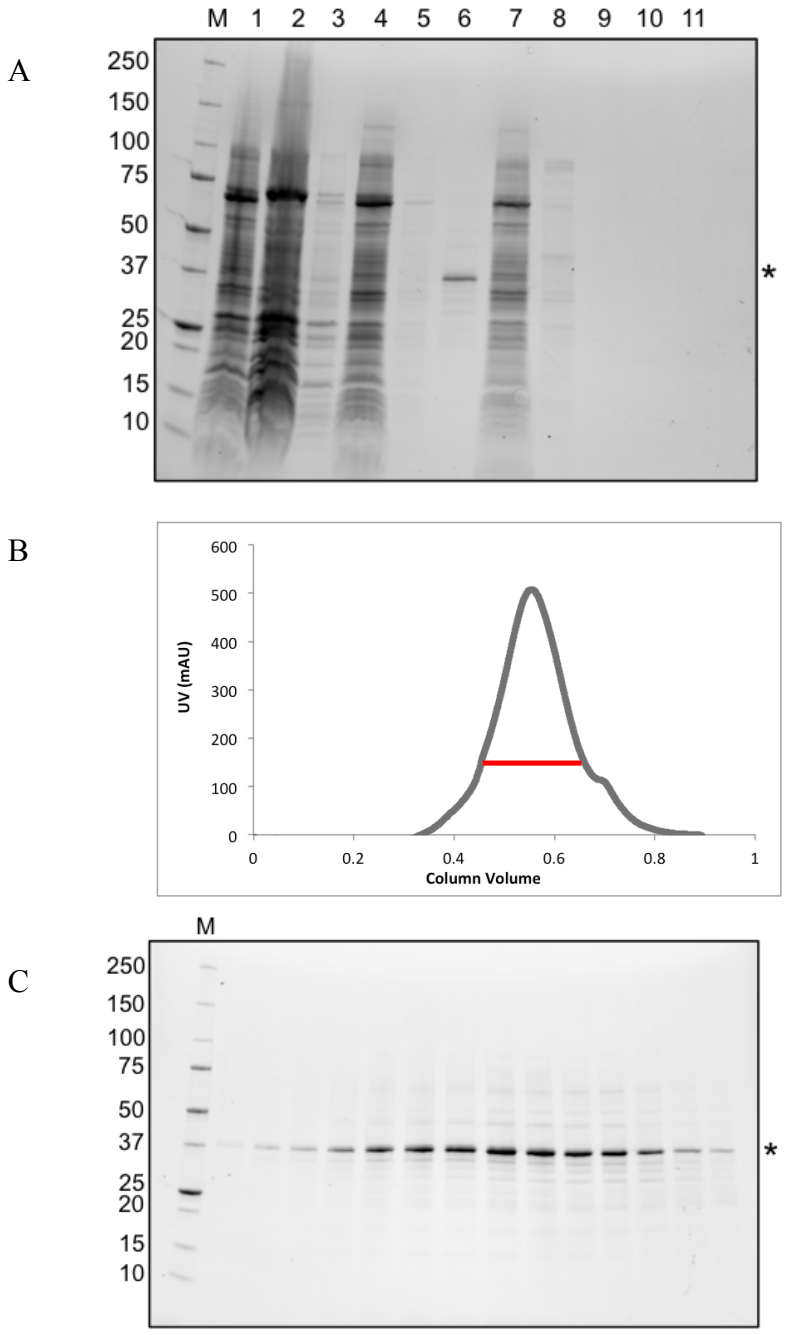


Figure 4.14. Purification of C-terminally Strep-tagged Wzy. The band at approximately 37 kDa, indicated with a "*", corresponds to Strep-tagged Wzy. The SDS-PAGE gel shown in "A" shows that Strep-tagged Wzy could be purified (lane 6) and did not exhibit non-specific binding to cobalt affinity resin (lane 11). (M: marker; 1: total membrane fraction; 2: DDM-soluble membrane fraction; 3: DDM-insoluble membrane fraction; 4: Streptactin FT; 5: Streptactin wash; 6: Streptactin eluate; 7: cobalt affinity FT; 8-10: cobalt affinity wash 1-3; 11: cobalt affinity eluate). "B" shows the SEC UV profile of Strep-tagged Wzy, which elutes as a single peak, and the SDS-PAGE gel of the fractions taken across the central SEC peak is shown in "C". The approximate position of the fractions in "B" shown in "C" is indicated with a solid red line.

4.3.5 Purification of His-tagged Wzz

N-terminally His-tagged Wzz was purified, as shown in figure 4.15, and eluted from the SEC column as a single peak, with no observed aggregation, and did not exhibit any non-specific binding to streptactin resin. The identity of this protein was confirmed by mass spectrometry.

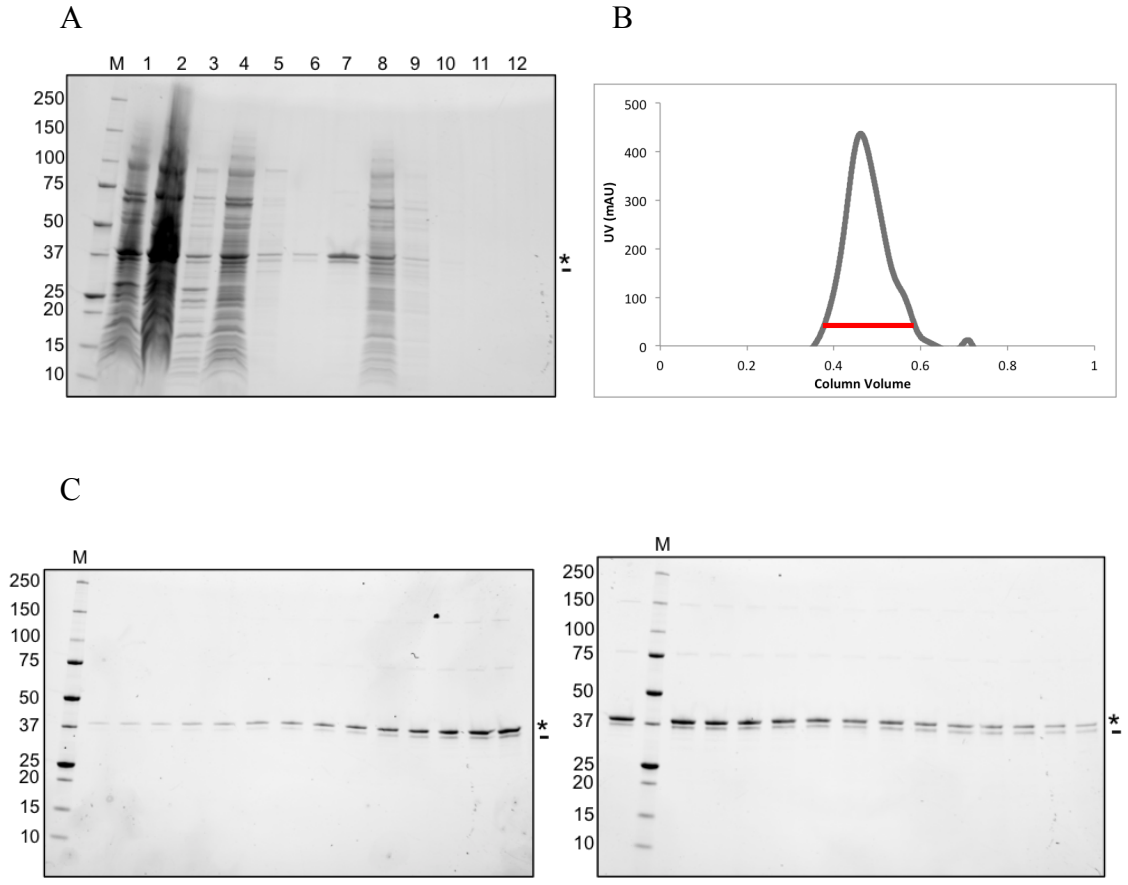


Figure 4.15. Purification of N-terminally His_6-tagged Wzz. The band at approximately 39 kDa, indicated with a "*", corresponds to His-tagged Wzz. The band at approximately 37 kDa, indicated with a "-", corresponds to non-tagged Wzz. The SDS-PAGE gel shown in "A" shows that His-tagged Wzz could be purified (lane 7) and did not exhibit non-specific binding to Streptactin resin (lane 12). (M: marker; 1: total membrane fraction; 2: DDM-soluble membrane fraction; 3: DDM-insoluble membrane fraction; 4: cobalt affinity FT; 5: cobalt affinity wash 1; 6: cobalt affinity wash 2; 7: cobalt affinity eluate; 8: Streptactin FT; 9-11: Streptactin wash 1-3; 12: Streptactin eluate). "B" shows the SEC UV profile of His-tagged Wzz, which elutes as a single peak, and the SDS-PAGE gel of the fractions taken across the central SEC peak is shown in "C". The approximate positions of the fractions in "B" shown in "C" is indicated with a solid red line.

4.3.6 Tandem affinity purification of Strep-tagged Wzy and His-tagged Wzz

In order to determine whether N-terminally His_6-tagged Wzz and C-terminally Strep-tagged Wzy interact, they were co-expressed and tandem affinity purified. The SDS-PAGE gel, western blots, probed with anti-His tag and anti-Strep tag antibodies, of the tandem affinity purification pulling down using a nickel/cobalt column that binds the His-tag of Wzz and then using a column that binds the Strep-tag of Wzz can be seen in figure 4.16. The SDS-PAGE gel and western blots, probed with anti-His tag and anti-Strep tag antibodies, of the tandem affinity purification pulling down via the Strep-tag of Wzy and then via the His-tag of Wzz can be seen in figure 4.17. As can be seen when comparing the results shown in figures 4.16 and 4.17, Strep-tagged Wzy and His-tagged Wzz were both found in the eluate of the second column, regardless of which protein was pulled down first, again indicative of complex formation. This result shows that the observed co-purification is not a feature of a particular affinity tagging strategy as the same results are obtained using both combinations of affinity tags used.

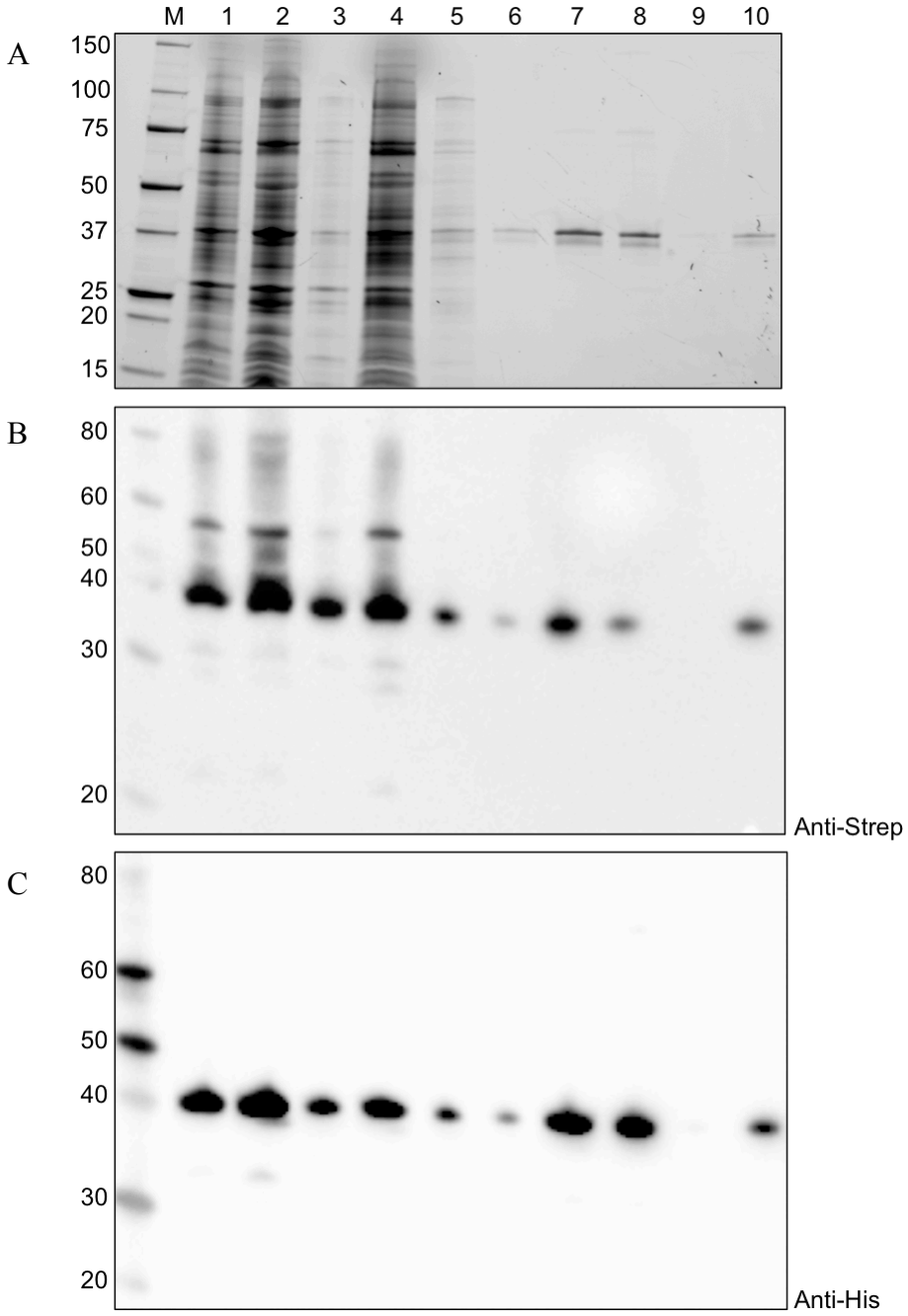


Figure 4.16. Tandem affinity purification of His-tagged Wzz and Strep-tagged Wzy using cobalt affinity resin and then streptactin resin. The SDS-PAGE gel of the tandem affinity purification is shown in "A". The band at approximately 40 kDa on the western blot probed with anti-Strep tag antibody ("B") corresponds to Strep-tagged Wzy. The band at approximately 38 kDa on the western blot probed with anti-His tag antibody ("C") corresponds to His-tagged Wzz. Both His-tagged Wzz and Strep-tagged Wzy are found in the eluate of the second affinity column (lane 10), indicating an interaction between these proteins. (M: marker; 1: total membrane fraction; 2: DDM-soluble membrane fraction; 3: DDM-insoluble membrane fraction; 4: cobalt affinity FT; 5: cobalt affinity wash 1; 6: cobalt affinity wash 2; 7: cobalt affinity eluate; 8: Streptactin FT; 9: Streptactin wash; 10: Streptactin eluate).

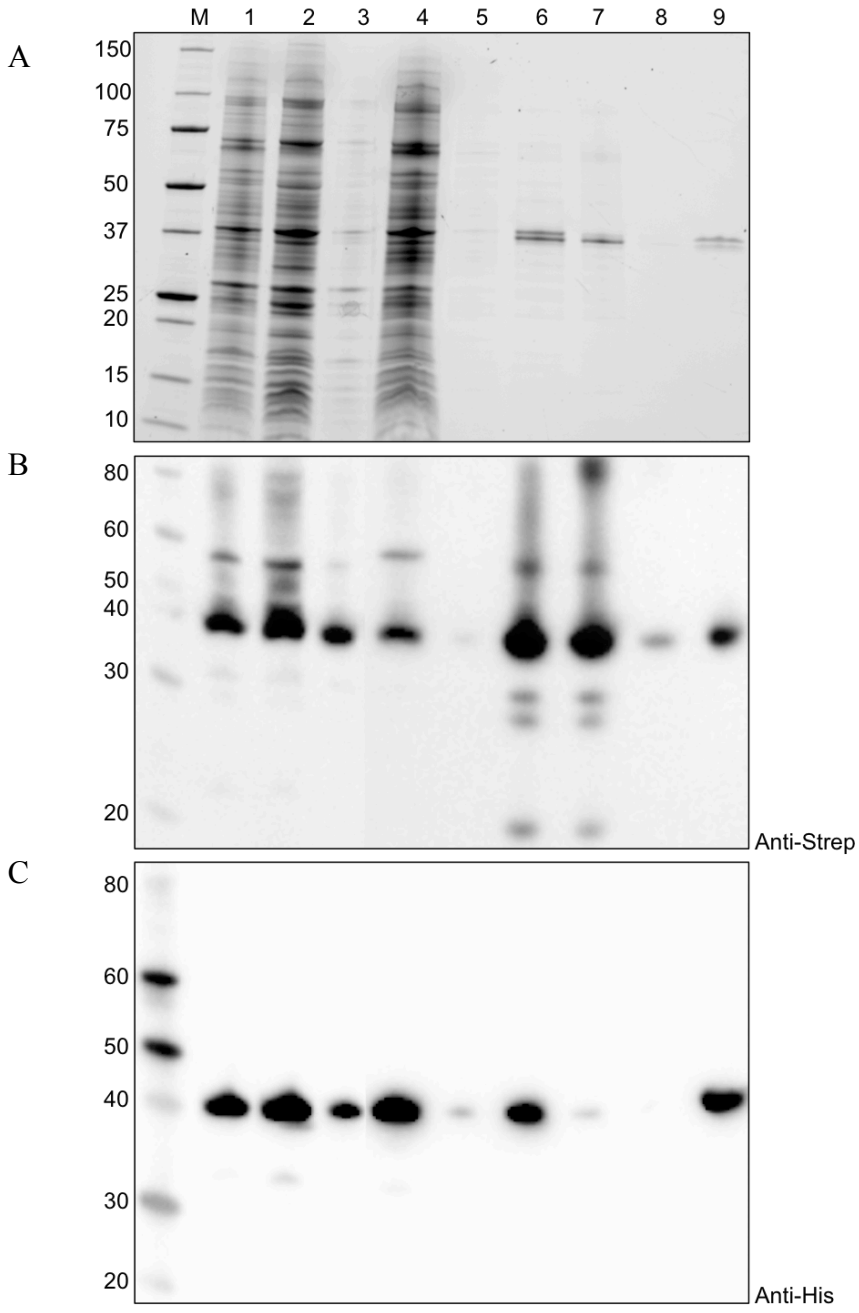


Figure 4.17. Tandem affinity purification of His-tagged Wzz and Strep-tagged Wzy using streptactin resin and then cobalt affinity resin. The SDS-PAGE gel of the tandem affinity purification is shown in "A". The band at approximately 40 kDa on the western blot, probed with anti-Strep tag antibody ("B"), corresponds to Strep-tagged Wzy. The band at approximately 38 kDa on the western blot probed with anti-His tag antibody ("C"), corresponds to His-tagged Wzz. Both His-tagged Wzz and Strep-tagged Wzy are found in the eluate of the second affinity column (lane 9), indicating an interaction between these proteins. (M: marker; 1: total membrane fraction; 2: DDM-soluble membrane fraction; 3: DDM-insoluble membrane fraction; 4: Streptactin FT; 5: Streptactin wash; 6: Streptactin eluate; 7: cobalt affinity FT; 8: cobalt affinity wash; 9: cobalt affinity eluate).

4.3.7 Confirmation of interaction via mass spectrometry

Samples were routinely submitted to the BSRC Mass Spectrometry and Proteomics facility (University of St Andrews) to confirm the presence of Wzz and Wzy in the purified sample. Figure 4.18 shows the SEC UV profile resulting from the tandem affinity purification of His-tagged Wzz and Strep-tagged Wzy, purifying first via the His-tag of the Wzz and then via the Strep-tag of the Wzy, followed by TEV protease cleavage to remove the affinity tags, and the corresponding SDS-PAGE gel of the SEC fractions taken across the peak at 0.56 CV. The peak at 0.78 CV corresponds to the TEV protease. Figures 4.19 and 4.20 show the peptides detected via mass spectrometry for the 37 kDa band and the 39 kDa band, respectively. This confirmed that the 37 kDa band corresponds to Wzz and that the 39 kDa band corresponds to Wzy. Some cross-contamination between the bands is observed, which is to be expected to be due to the close proximity of the bands on the SDS-PAGE gel. Few peptides for Wzy were detected; this is due to the hydrophobic nature of Wzy. The observation of an interaction between Wzz and Wzy after the removal of the affinity tags by TEV protease cleavage confirms that the interaction observed here is not due to an interaction between the affinity tags.

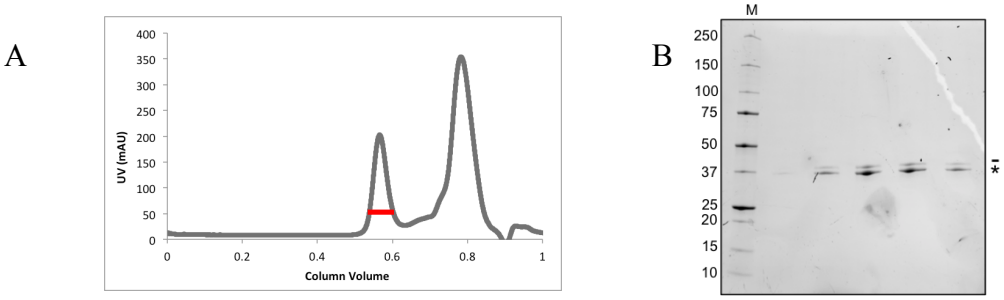


Figure 4.18. Tandem affinity purification of His-tagged Wzz plus Strep-tagged Wzy, followed by TEV protease-cleavage and SEC, for mass spectrometry analysis. The SEC UV profile resulting from the tandem affinity purification of His-tagged Wzz plus Strep-tagged Wzy followed by TEV protease cleavage of the affinity tags prior to the SEC column is shown in "A". The peak at approximately 0.56 CV corresponds to Wzz and Wzy. The peak at 0.78 CV corresponds to TEV protease. The fractions across the peak at 0.56 CV were analysed via SDS-PAGE gel shown in "B". The band at approximately 39 kDa, highlighted with a "-", corresponds to Wzy and that at 37 kDa, highlighted with a "*", corresponds to Wzz. The approximate position of the fractions in "A" shown in "B" is indicated with a red line.

A

MALGQFGGLFVVYLISVIFILSLTWMEFRRVRFNFNVLFSLLYLLTFYFGFPFTCVLVFRFGVEVVPVQYLLQAMLSA
TAFYAVYYVSYKTRLRQKTSVPRAPLLTVNRVEANLTWLLLALIAVATVGIFFLNNGFLLFKLRSYSQIFSSDVSGVAL
KRFFYFFIPAMLVVYFLRQTQRAWLLFLIGTVAFGMLTYVIVGGTRANLIIAFALFLFIGIVRGWITLWMLVAAGIFGIVG
MFWLALKRYGLDVSGDYAFYTFLYLTRDTFSPWENLALLWQNYDKIEFQGLAPIARDFYVFIPSWLWPDRPNLVLN
SANYFTWEVLNNHSGLAISPTLLGSLVVMGGVLFIPLGAIAVGLVIKWFDWVYELGKNDSNRYKAAILQAFCFGAVF
NIIVLTREGVDSFVSRVVFFCLIFGLCLLVAKLLYWLLESAGLIRQRLRRMRATPLAPTPNTVQDPVIKEQILEENLYF
Q

B

GAMVKSENLSTGNALIDNELDIRGLFRLLWRGKVWPIAIGLLFAVVALGYSYLVKQEWGATSITDKPTVNMLG
GYYSQQQFLRNLDARSFSSPQPEQPSISAGAYDEFIMQLAAYDTRRDFWLQTDYYKQRLEGDEKADAALLD
ELVNNIQFTARDDGKKTNDSVKLVAETSSDSNTLLRQYVVFASQRAANHLNEEIKGAWAARTIFMKSQIKRQE
AVAKAIYDREVRSVELALKIAQQQGISRSQTDTPADEIPASEMFLLGRPMLQARLETLQTSGPHYELDYDQNR
AMLATLNVGPTLEASFQTYRYLRTPEEPVKRDSPRRAFLMVMWGAIGVLVGAGVALARRLR

Figure 4.19. Peptides detected via mass spectrometry analysis of the 37 kDa SDS-PAGE gel band shown in figure 4.18b. Wzy peptides detected in the 37 kDa band are highlighted in red in the sequence of Wzy shown in "A". Wzz peptides detected in the 37 kDa band are highlighted in red in the Wzz sequence shown in "B".

A

MALGQFGGLFVVYLISVIFILSLTWMEFRRVRFNFNVLFSLLYLLTFYFGFPFTCVLVFRFGVEVVPVQYLLQAMLSA
TAFYAVYYVSYKTRLRQKTSVPRAPLLTVNRVEANLTWLLLALIAVATVGIFFLNNGFLLFKLRSYSQIFSSDVSGVAL
KRFFYFFIPAMLVVYFLRQTQRAWLLFLIGTVAFGMLTYVIVGGTRANLIIAFALFLFIGIVRGWITLWMLVAAGIFGIVG
MFWLALKRYGLDVSGDYAFYTFLYLTRDTFSPWENLALLWQNYDKIEFQGLAPIARDFYVFIPSWLWPDRPNLVLN
SANYFTWEVLNNHSGLAISPTLLGSLVVMGGVLFIPLGAIAVGLVIKWFDWVYELGKNDSNRYKAAILQAFCFGAVF
NIIVLTREGVDSFVSRVVFFCLIFGLCLLVAKLLYWLLESAGLIRQRLRRMRATPLAPTPNTVQDPVIKEQILEENLYF
Q

B

GAMVKSENLSTGNALIDNELDIRGLFRLLWRGKVWPIAIGLLFAVVALGYSYLVKQEWGATSITDKPTVNMLG
GYYSQQQFLRNLDARSFSSPQPEQPSISAGAYDEFIMQLAAYDTRRDFWLQTDYYKQRLEGDEKADAALL
DELVNNIQFTARDDGKKTNDSVKLVAETSSDSNTLLRQYVVFASQRAANHLNEEIKGAWAARTIFMKSQIKR
QEAVAKAIYDREVRSVELALKIAQQQGISRSQTDTPADEIPASEMFLLGRPMLQARLETLQTSGPHYELDYD
QNRAMLATLNVGPTLEASFQTYRYLRTPEEPVKRDSPRRAFLMVMWGAIGVLVGAGVALARRLR

Figure 4.20. Peptides detected via mass spectrometry analysis of the 39 kDa SDS-PAGE gel band shown in figure 4.18b. Wzy peptides detected in the 39 kDa band are highlighted in red in the sequence of Wzy shown in "A". Wzz peptides detected in the 39 kDa band are highlighted in red in the Wzz sequence shown in "B".

4.3.8 Size of Wzz plus Wzy PDC

SEC-MALS, shown in figure 4.21, determined the molar mass of GST-tagged Wzz plus His-tagged Wzy, with the affinity tags removed by TEV protease-cleavage, in the DDM micelle to be 540 kDa (± 0.3 %). The size of the DDM micelle has been determined to be 72 kDa (Strop and Brunger, 2005). It is possible to use SEC-MALS to determine the protein molar mass, however this requires that the extinction coefficient of the protein is known, which is not the case as we do not know the ratio of the proteins present.

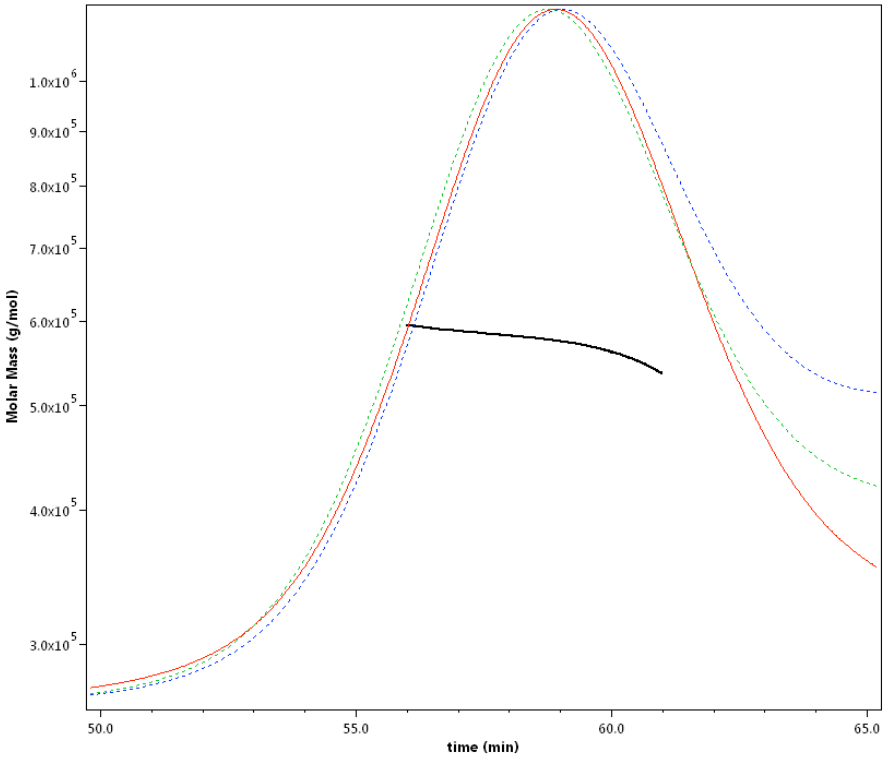


Figure 4.21. SEC-MALS of Wzz plus Wzy. Change in UV over time is shown in green, change in light scattering over time is shown in red and the change in refractive index over time is shown in blue. Change in molar mass (g/mol) over time is shown in black.

4.3.9 Tagging strategy does not affect SEC elution volume of Wzz plus Wzy

An overlay of the SEC elution volumes of TEV protease-cleaved His-Wzz plus TEV protease-cleaved Strep-Wzy (shown in light grey) and TEV protease-cleaved GST-Wzz plus TEV protease-cleaved His-Wzy (shown in dark grey), shown in figure 4.22, shows that the SEC profile is not affected by the tagging strategy chosen. The second peak on the SEC profile resulting from TEV protease-cleaved His-Wzz plus TEV protease-cleaved Strep-Wzy (light grey) at approximately 0.8 CV corresponds to TEV protease.

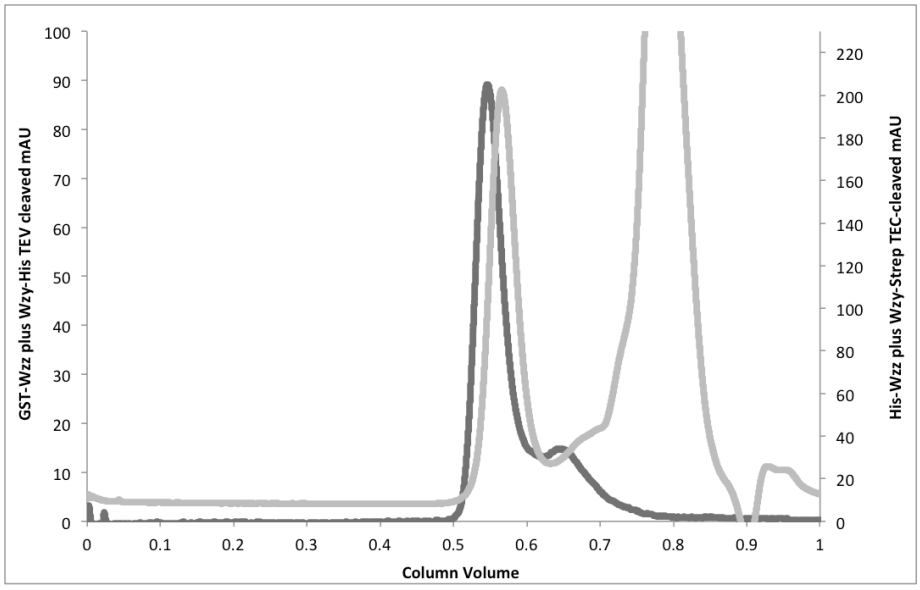


Figure 4.22. Overlay of the SEC elution volume of TEV protease-cleaved GST-Wzz plus His-tagged Wzy and TEV protease-cleaved His-tagged Wzz plus Strep-tagged Wzy. TEV protease-cleaved GST-tagged Wzz plus His-tagged Wzz is shown in dark grey and TEV protease-cleaved His-tagged Wzz plus Strep-tagged Wzy is shown in light grey.

4.3.10 A comparison of the SEC elution volumes of His-tagged Wzz, Strep-tagged Wzy and tandem affinity purified His-tagged Wzz with Strep-tagged Wzy

The SEC elution volumes for His-tagged Wzz, Strep-tagged Wzy and tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy were compared in order to determine whether there was a shift in the elution volumes of the individual proteins upon tandem affinity purification. As shown in figure 4.23, His-tagged Wzz elutes at 0.45 CV and Strep-tagged Wzy elutes at 0.55 CV. Upon tandem affinity purification of His-tagged Wzz plus Strep-tagged Wzy, the elution volume of Strep-tagged Wzy decreases to 0.48 CV, corresponding to an increase in size, whereas the elution volume of His-tagged Wzz slightly increases, corresponding to an apparent reduction in size. This apparent reduction in size could be due the arrangement of the His-tagged Wzz oligomer becoming more compact in the presence of Strep-tagged Wzy.

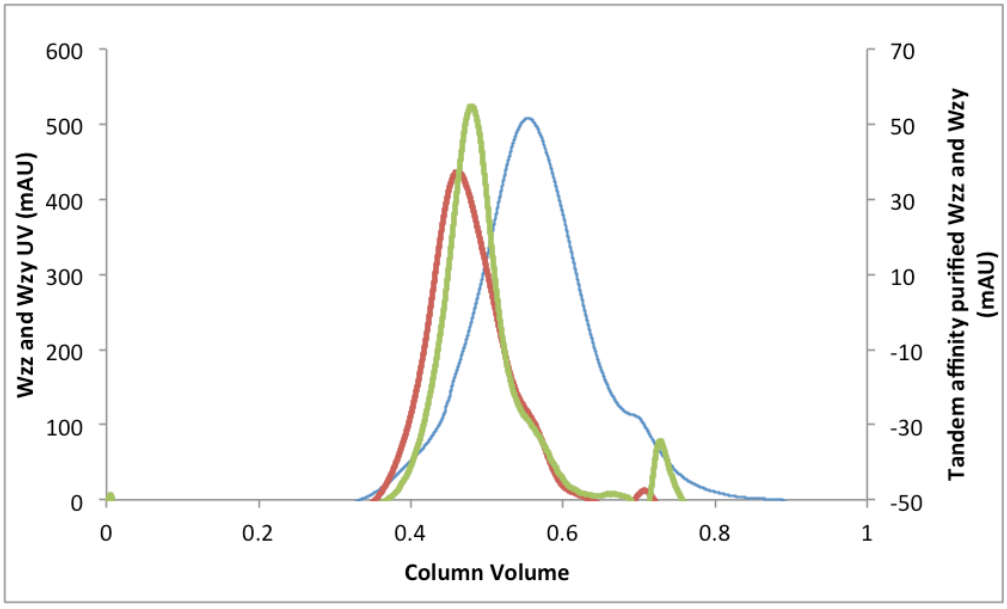


Figure 4.23. Comparison of SEC elution volumes of His-tagged Wzz, Strep-tagged Wzy and tandem affinity purification of His-tagged Wzz plus Strep-tagged Wzy. The SEC elution profile of His-tagged Wzz is shown in red, the SEC elution profile of Strep-tagged Wzy is shown in blue and the SEC elution profile of tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy is shown in green.

4.3.11 BN-PAGE and western blotting of Strep-tagged Wzy and His-tagged Wzz

In order to further compare the sizes of the PDCs of the individual and combined components of this proposed complex and in an attempt to identify a technique that allows a size distinction to be made between Wzz alone and Wzz in complex with Wzy, a combination of BN-PAGE and western blotting was used (figure 4.24). His-tagged Wzz, Strep-tagged Wzy and tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy were run on a BN-PAGE, which was blotted against the His-tag of the Wzz and the Strep-tag of Wzy. Strep-tagged Wzy alone ran as a smear on the BN-PAGE, indicating that Strep-tagged Wzy alone occupies a range of oligomeric states. Upon tandem affinity purification of Strep-tagged Wzy with His-tagged Wzz, both the anti-Strep tag and anti-His tag signal were present in one band. This indicates that these proteins co-purify and provides evidence that they form a complex. Furthermore, His-tagged Wzz alone did not give a signal when probed with the anti-Strep tag antibody, indicating that this anti-Strep tag signal in the His-tagged Wzz plus Strep-tagged Wzy band is not due to non-specific antibody binding to Wzz. However, the PDCs of His-tagged Wzz and His-tagged Wzz plus Strep-tagged Wzy appear to run at approximately the same molecular weight of 720 kDa with no size shift occurring upon co-purification with Strep-tagged Wzy.

4.3.12 Stoichiometry of the interaction between His-tagged Wzz and Strep-tagged Wzy

To ensure that Strep-tagged Wzy did not co-purify with His-tagged Wzz in trace amounts, i.e. there was a defined complex, quantitative dot blotting was used to determine the approximate stoichiometry of the interaction between His-tagged Wzz and Strep-tagged Wzy. Dot blotting was chosen over conventional western blotting due to the ease of quantitation of a dot blot over a western blot as the signal from each dot blot point will not overlap with the neighbouring dot. A serial dilution of His-tagged

Wzz and Strep-tagged Wzy was probed with the appropriate antibody and the signal was used to construct a standard curve of His-tagged Wzz and Strep-tagged Wzy. Three amounts of tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy were loaded in triplicate onto each western dot blot, along with the standard curves, and the signal from the His-tagged plus Strep-tagged Wzy was compared to the standard curves. The dot blots of the serial dilutions of His-tagged Wzz and Strep-tagged Wzy, and the corresponding standard curves, can be seen in figures 4.25 and 4.26, respectively. The signal from His-tagged Wzz was corrected, using the Biorad ChemiDoc ImageLab software, for the minor amount of cleavage of the His-tag, seen in figure 4.15. This was assumed to occur at the same level upon tandem affinity purification. This resulted in an average stoichiometry value of 9.6 (±1.8) His-tagged Wzz to 1 Strep-tagged Wzy.

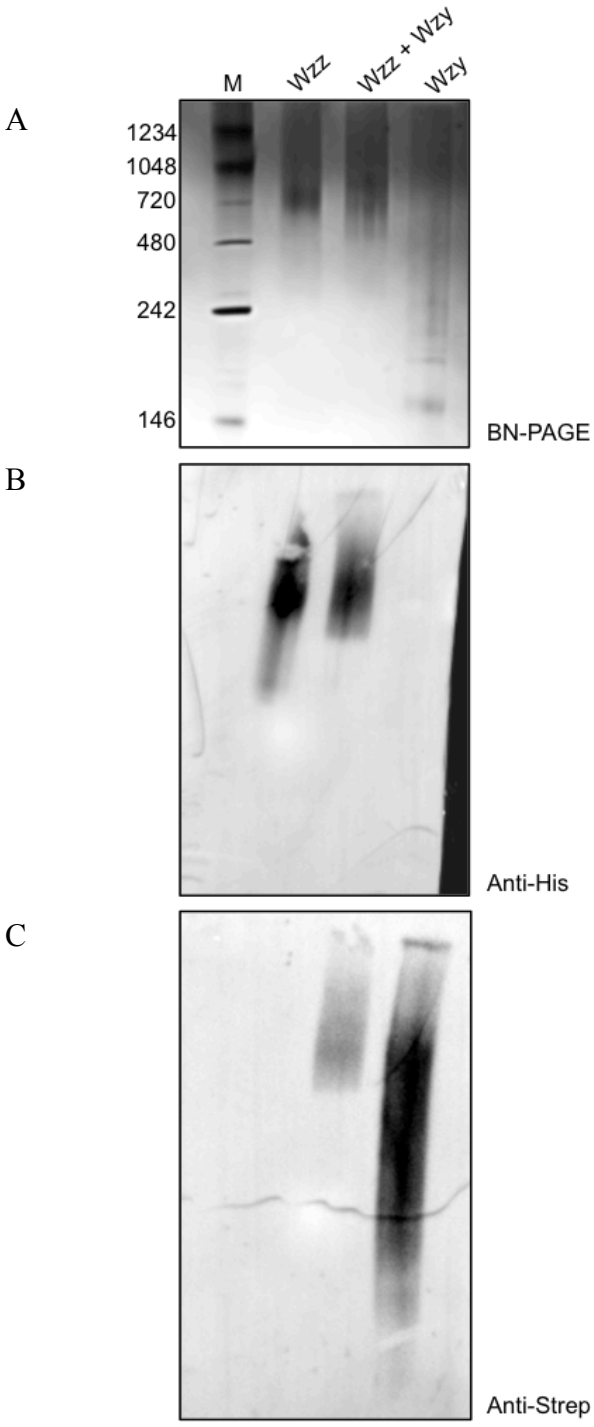


Figure 4.24. BN-PAGE and western blotting of His-tagged Wzz, Strep-tagged Wzy and tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy. Both His-tagged Wzz and tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy run at approximately 720 kDa. In the absence of His-tagged Wzz, Strep-tagged Wzy runs at a range of molecular weights, which restricted to approximately 720 kDa upon tandem affinity purification with His-tagged Wzz. The BN-PAGE gel is shown in "A", the western blot probed with anti-His antibody is shown in "B" and the western blot probed with anti-Strep antibody is shown in "C".

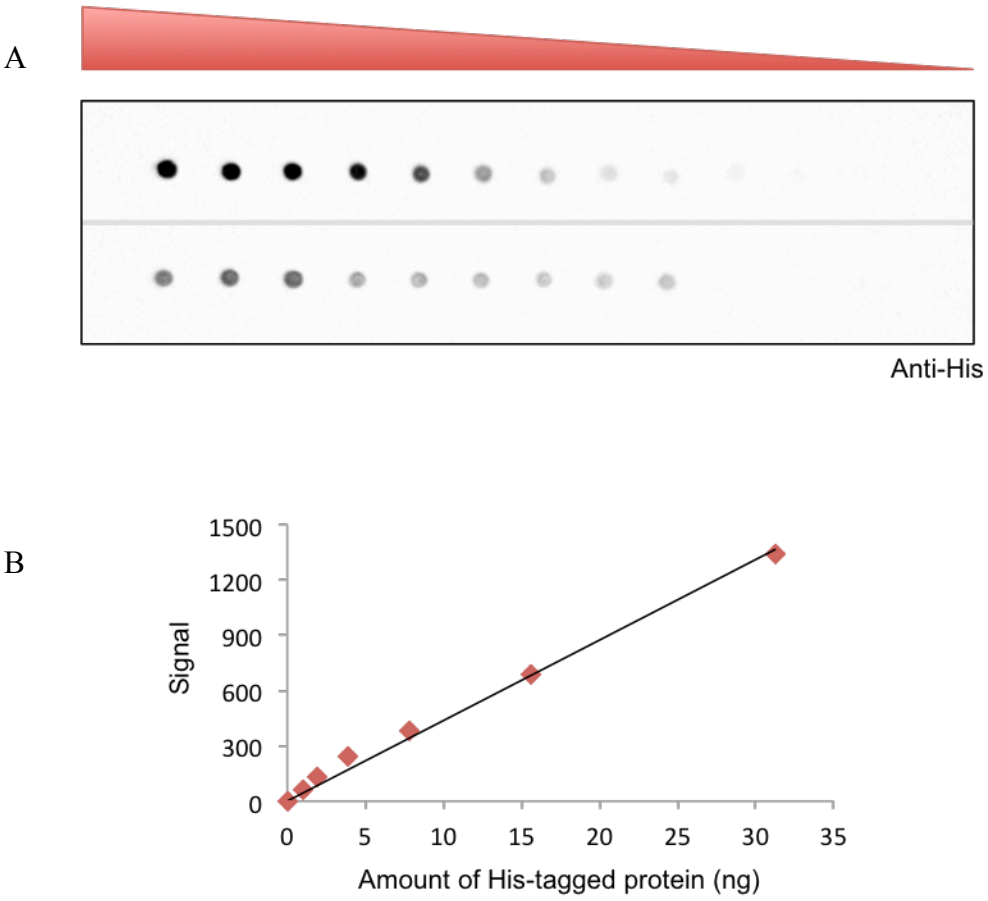


Figure 4.25. Quantitative western blot probed with anti-His tag antibody for stoichiometry determination. The western blot, probed with anti-His tag antibody, of the His-tagged Wzz serial dilution (top line) and the tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy samples (bottom line) is shown in "A". The wedge indicates that the serial dilution of His-tagged Wzz runs from left to right. The standard curve constructed using the signal from the His-tagged Wzz serial dilution, shown in "A", is shown in "B".

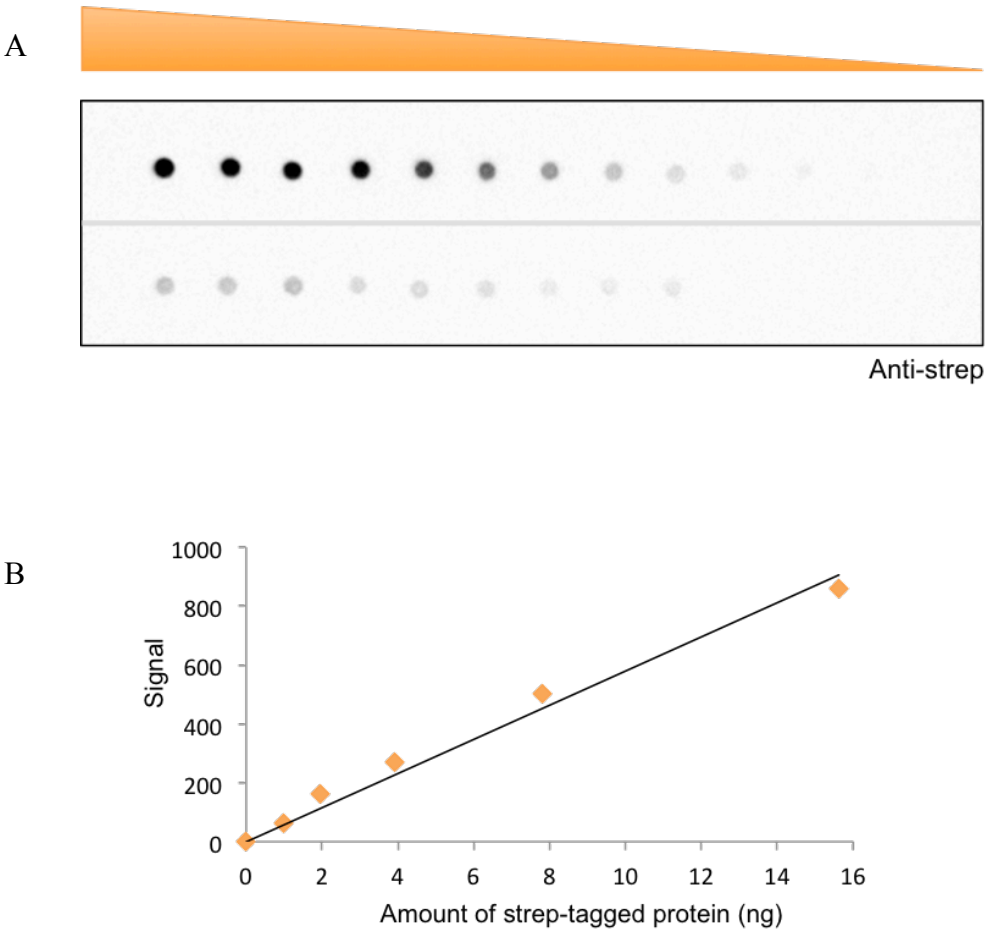


Figure 4.26. Quantitative western blot probed with anti-Strep tag antibody for stoichiometry determination. The western blot, probed with anti-Strep tag antibody, of the Strep-tagged Wzy serial dilution (top line) and the tandem affinity purified His-tagged Wzz plus Strep-tagged Wzy samples (bottom line) is shown in "A". The wedge indicates that the serial dilution of Strep-tagged Wzy runs from left to right. The standard curve constructed using the signal from the Strep-tagged Wzy serial dilution, shown in "A", is shown in "B".

4.3.13 *In vitro* interaction between Strep-tagged Wzy and His-tagged Wzz detected via far western dot blotting

Far western blotting was used to support the model of a Wzz:Wzy complex by determining whether an interaction between Strep-tagged Wzy and His-tagged Wzz, both purified separately as shown in figures 4.14 and 4.15, could be established *in vitro*. Far western dot blotting involves immobilising a "bait" protein onto a membrane, blocking the membrane to avoid non-specific binding and then probing the membrane with an affinity-tagged "prey" protein. The binding of the "prey" protein to the immobilised "bait" protein is detected using an antibody against the affinity tag of the "prey" protein (Edmondson and Dent, 2001). As can be seen in figures 4.27 and 4.28, a signal can be seen on both far western blots, regardless of whether His-tagged Wzz or Strep-tagged Wzy was immobilised on the membrane, with the detected signal increasing with increased amounts of immobilised protein. This indicates an interaction between these proteins. Immobilised BSA, T4 lysozyme and WaaL O-antigen ligase (from *H. pylori*) were used as negative controls, as can be seen in figure 4.28, Strep-tagged Wzy did not bind to the negative controls. However His-tagged Wzz did bind to WaaL O-antigen ligase, shown in figure 4.27, but at a much lower level than that observed for the same amount of immobilised Strep-tagged Wzy.

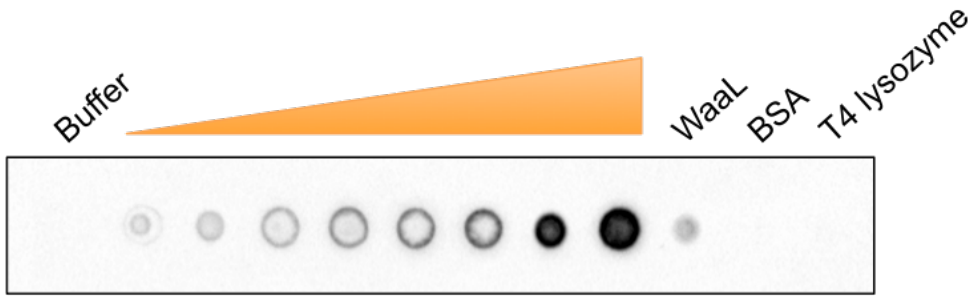


Figure 4.27. Far western dot blot probed with an anti-His tag antibody. A dilution of strep-tagged Wzy was immobilised on the membrane, the membrane was incubated with His-tagged Wzz, followed by the incubation with HRP-conjugated antibody against the His-tag of Wzz. Increasing amounts of immobilised Strep-tagged Wzy (indicated by the wedge) resulted in increasing amounts of signal corresponding to increasing amounts of bound His-tagged Wzz. Co-purification buffer, WaaL (*H. pylori*), BSA and T4 lysozyme were used as negative controls.

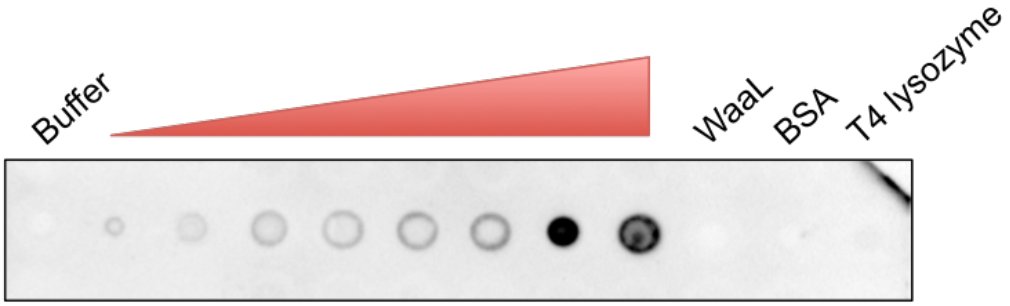


Figure 4.28. Far western dot blot probed with an anti-Strep tag antibody. A dilution of His-tagged Wzz was immobilised on the membrane, the membrane was incubated with Strep-tagged Wzy, followed by the incubation with HRP-conjugated antibody against the Strep-tag of Wzy. Increasing amounts of immobilised His-tagged Wzz (indicated by the wedge) resulted in increasing amounts of signal corresponding to increasing amounts of bound Strep-tagged Wzy. Co-purification buffer, WaaL (from *H. pylori*), BSA and T4 lysozyme were used as negative controls.

4.3.14 Negative stain EM shows a homogeneous sample

As the molar mass of Wzz plus Wzy in DDM was determined by SEC-MALS to be approximately 540 kDa, it was decided that the appropriate method for the structural characterisation of this sample would be cryo-EM. In collaboration with Dr David Bhella (Centre for Virus Research, MRC-University of Glasgow), negative stain EM was used here to determine whether a homogeneous particle results from the purification of GST-tagged Wzz with His-tagged Wzy. As can be seen in figure 4.29, tandem affinity purified Wzz and Wzy form a homogeneous particle, with little aggregation observed. Particles observed via negative stain were approximately 10-20 nm in length.

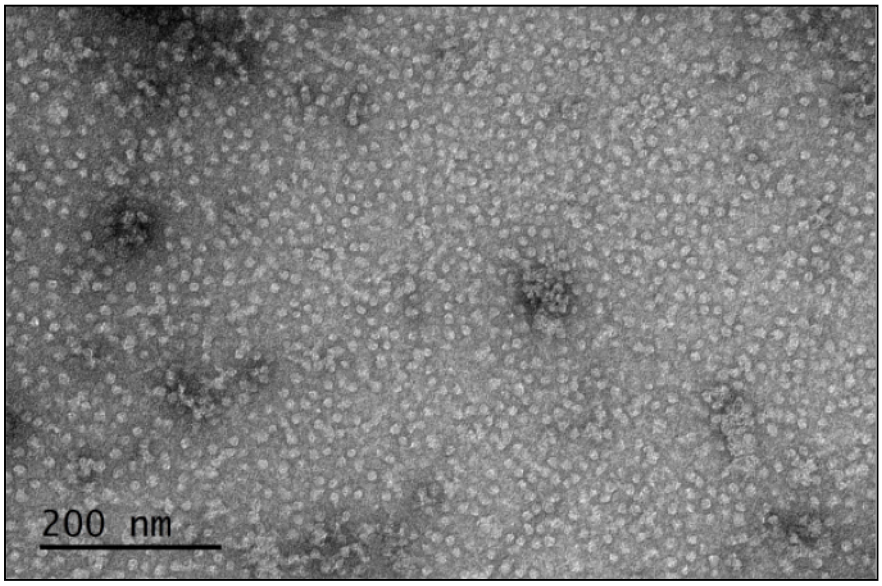


Figure 4.29. Negative stain EM shows that Wzz plus Wzy forms a homogeneous particle. Negative stain EM, performed by Dr David Bhella (Centre for Virus Research, MRC-University of Glasgow).

4.3.15 Cryo-EM

As Wzz plus Wzy were determined by negative stain EM to produce a homogeneous particle with little aggregation observed, we decided to proceed with the structural characterisation of this sample via cryo-EM. Cryo-EM was conducted in collaboration with Dr David Bhella at the University of Glasgow. Dr David Bhella prepared the samples for cryo-EM analysis and I worked closely with him during this process. Cryo-EM data was collected by myself and the data were processed by Dr David Bhella. Initial cryo-EM grids were sub-optimal, as the target protein did not enter the holes in the grids, as shown in figure 4.30. It was therefore decided to coat the grid with a thin carbon film, however, as shown in figure 4.31, this resulted in a preferential orientation of the particle on the grid. The combination of a carbon coating on the grid coupled with glow discharging of the grid in the presence of amylamine, prior to applying the sample on the grid, resulted in a range of orientations of the particle (figure 4.32), but this also resulted in low contrast images. Therefore the original grid preparation conditions were revisited, this time with a higher protein concentration, resulting in higher contrast

images with a range of particle orientations, shown in figure 4.33. However the particles showed a preference for thicker ice, thus reducing the contrast of the images. These optimised grid preparation conditions were subsequently used for data collection. Preliminary 2D classification (prepared by Dr David Bhella using Motioncorr and Relion) of the Wzz plus Wzy sample is shown in figure 4.34. These preliminary 2D classifications show a range of orientations of the particle, with the bell shape of Wzz clearly visible (Tocilj *et al*., 2008). However, no easily identifiable additional density that could correspond to Wzy could be observed in these 2D classifications, when compared to the 2D classifications of Wzz alone, shown in figure 4.35.

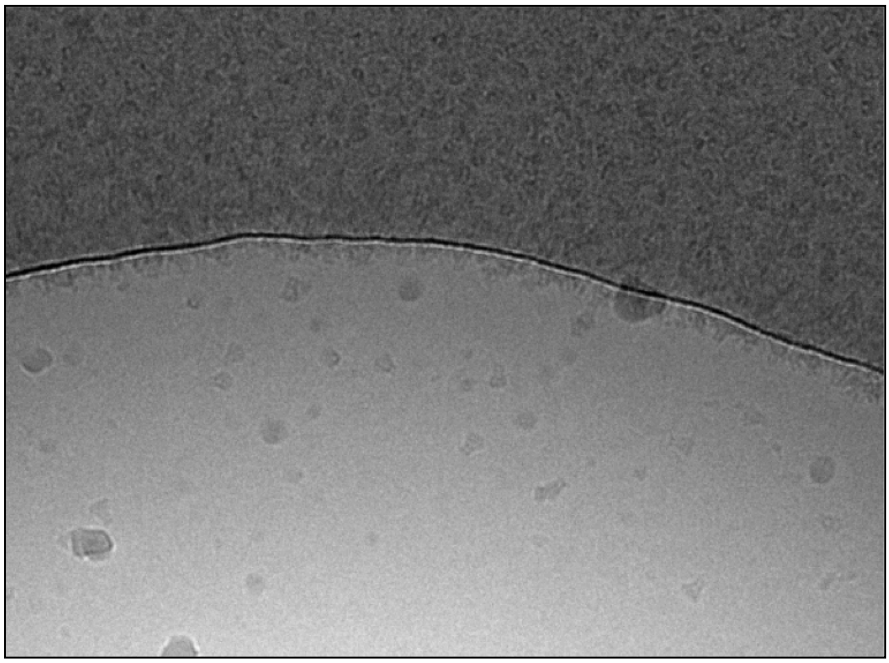


Figure 4.30. Initial cryo-EM images show that Wzz plus Wzy does not enter the holes in the grid. Wzz plus Wzy preferentially interacts with the carbon film of the grid.

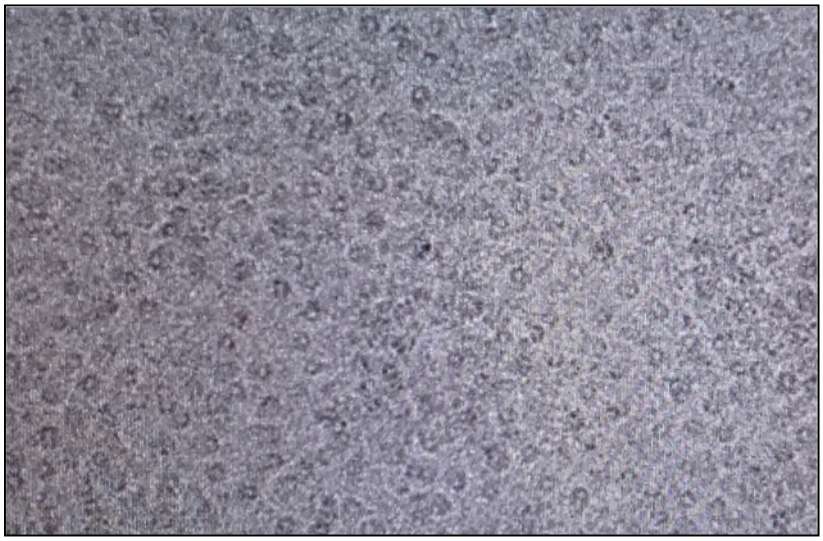


Figure 4.31. Carbon coating of the grid resulted in a preferential orientation of Wzz plus Wzy on the grid.

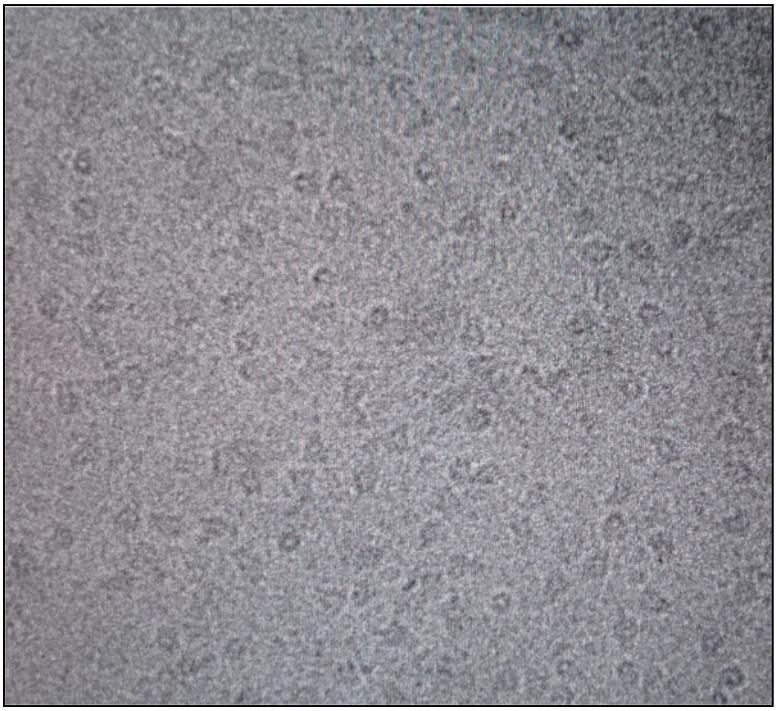


Figure 4.32. The combination of the carbon coating with glow discharging of the grid in the presence of amylamine resulted in a range of orientations of Wzz plus Wzy.

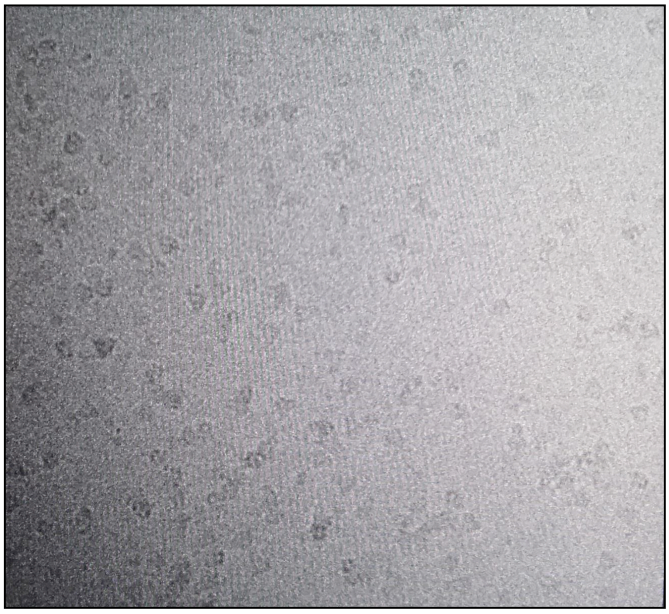


Figure 4.33. Increasing the concentration of Wzz plus Wzy applied to the grid resulted in more Wzz plus Wzy entering the holes in the grid.

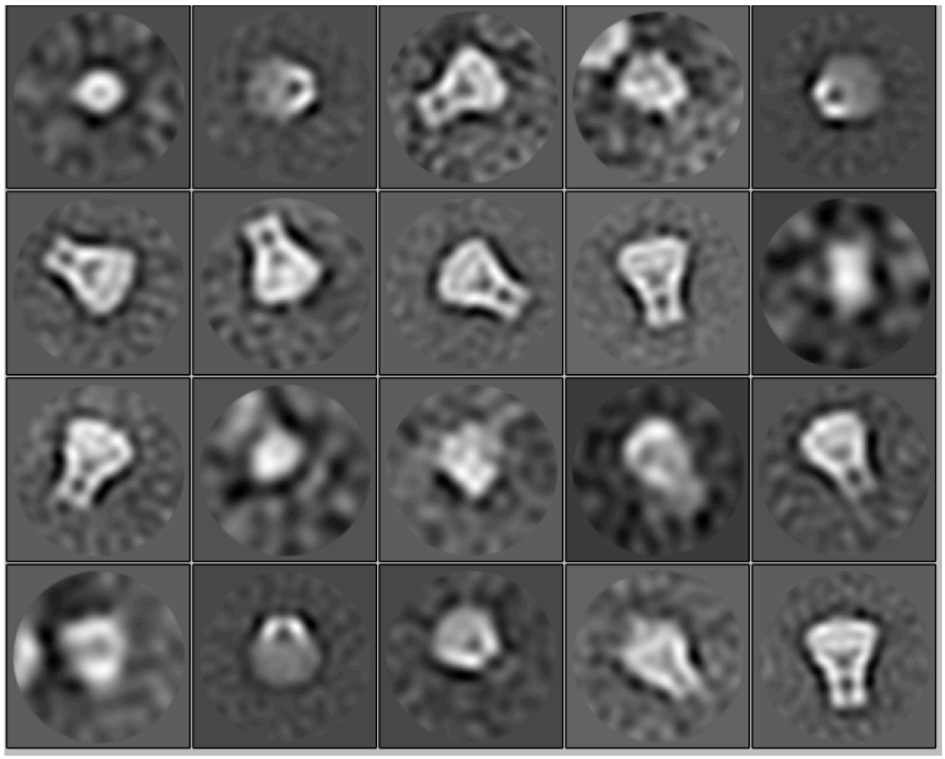


Figure 4.34. Preliminary 2D classifications of Wzz plus Wzy. These 2D classifications, determined by Dr David Bhella, using Relion, show a range of orientations of the particle, with the typical bell shape of Wzz clearly identifiable (Tocilj *et al*., 2008). The wider part of the particle is proposed to correspond to the DDM-coated TMHs (Collins *et al*., 2017).

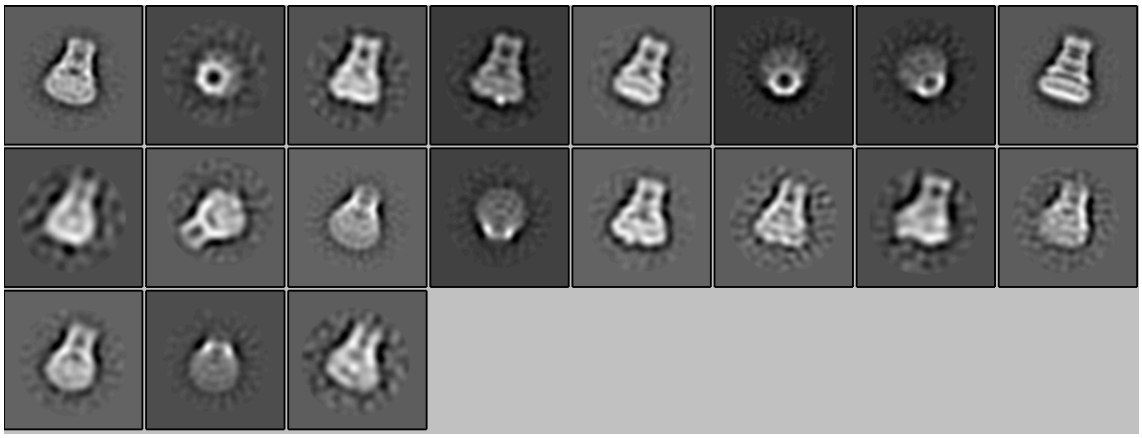


Figure 4.35. 2D classifications of Wzz alone. These 2D classifications were generated by Dr David Bhella, using Relion (Scheres, 2012a, 2012b). The wider part of the particle is proposed to correspond to the DDM-coated TMHs (Collins *et al*., 2017).

4.4 Discussion

Despite its demonstrated importance to bacterial survival, the mechanism of O-antigen chain length regulation by Wzz remains unknown. Numerous mechanisms have been proposed to explain the mode of action of Wzz, in the absence of any obvious catalytic activity, with some of these proposed mechanisms based on a direct interaction between Wzz and Wzy (Bastin *et al*., 1993; Islam and Lam, 2014; Kintz and Goldberg, 2011; Tocilj *et al*., 2008). However, evidence of a direct interaction between Wzz and Wzy was only provided recently by Nath and Morona who showed that untagged Wzz is found in the eluate of a nickel affinity column following *in vivo* crosslinking in the presence of His- and GFP-tagged Wzy (Nath and Morona, 2015a). We therefore aimed to build upon these results.

There are many methods available for the study of PPIs and the method chosen depends both on the mode, be it transient or weak, and stability of the interaction, and also the amount of protein available for study. Tandem affinity purification involves expressing the components of the putative protein complex of interest with different affinity tags. A pulldown against one of the affinity tags is conducted and the eluate from this first

pulldown is applied to a column to pulldown against the other affinity tag (Puig *et al*., 2001; Rigaut *et al*., 1999). This approach was chosen here, using multiple tagging combinations to eliminate any possibility that a PPI could arise from an interaction between the affinity tags. An interaction between Wzz and Wzy could be observed using tandem affinity purification regardless both of tagging strategy and which protein was pulled down first. Furthermore, both affinity tagging combinations implemented resulted in a complex of the same size, as confirmed by comparison of the SEC elution profiles. These results indicate that these proteins form a stable complex. These results also agree with, and build upon, the previous detection of an interaction between Wzz and Wzy (Nath and Morona, 2015a). A direct interaction between Wzz and Wzy is also in agreement with the finding that the polymerase PssT, required for exopolysaccharide biosynthesis, interacts directly with its PCP PssP, as confirmed using a bacterial two-hybrid screen (Marczak *et al*., 2013).

Co-purification of Wzz and Wzy results in a shift in the SEC elution volume of Wzy, corresponding to an increase in size. In contrast, co-purification of Wzz and Wzy results in a slight shift in the SEC elution volume of Wzz, corresponding to a slight reduction in size. This could indicate that the complex formed between Wzz and Wzy is more compact than Wzz alone. As little shift in Wzz SEC elution volume could be observed upon co-purification with Wzy, despite the confirmation of the presence of Wzy using MS, additional methods for studying PPIs were implemented both in an attempt to find a method that would allow the discrimination between Wzz alone, and Wzz plus Wzy, and to further characterise this interaction. As these proteins are both integral membrane proteins, the available methods for studying this PPI are limited due to the possible occlusion of the interaction site(s) by the detergent micelle. Methods implemented here included BN-PAGE and far western dot blotting, both of which confirmed an interaction between Wzz and Wzy.

Given the data supporting an interaction between Wzz and Wzy, as well as the size of the PDC, we decided to proceed with the structural characterisation of this complex using EM, in close collaboration with Dr David Bhella, Centre for Virus Research, MRC-University of Glasgow. Negative stain EM confirmed that the complex formed

between Wzz and Wzy was homogeneous, with little aggregation observed (figure 4.29). Therefore we decided continue the structural characterisation of this complex using cryo-EM. Following the extensive optimisation of grid preparation conditions, data were collected for both Wzz alone and Wzz plus Wzy and 2D classifications were constructed (data processing was conducted by Dr David Bhella). The bell shape of Wzz (Tocilj *et al*., 2008) is clearly recognisable in the 2D classifications of both Wzz alone and Wzz plus Wzy (figures 4.34 and 4.35). The similarity between these 2D classifications and the published structures of Wzz (Collins *et al*., 2017; Kalynych *et al*., 2015; Larue *et al*., 2009; Tocilj *et al*., 2008), and the lack of additional density on the outside of the bell, suggests either that Wzy is lost during cryo-EM sample preparation or that the Wzz oligomer forms around a central Wzy. The 2D class averages obtained by Collins *et al.* for full-length dodecameric Wzz (from *S.* Typhimurium) and the orthogonal projections of the refined EM map (EMDB-3611) (Collins *et al*., 2017) are shown for comparison in figure 4.36.

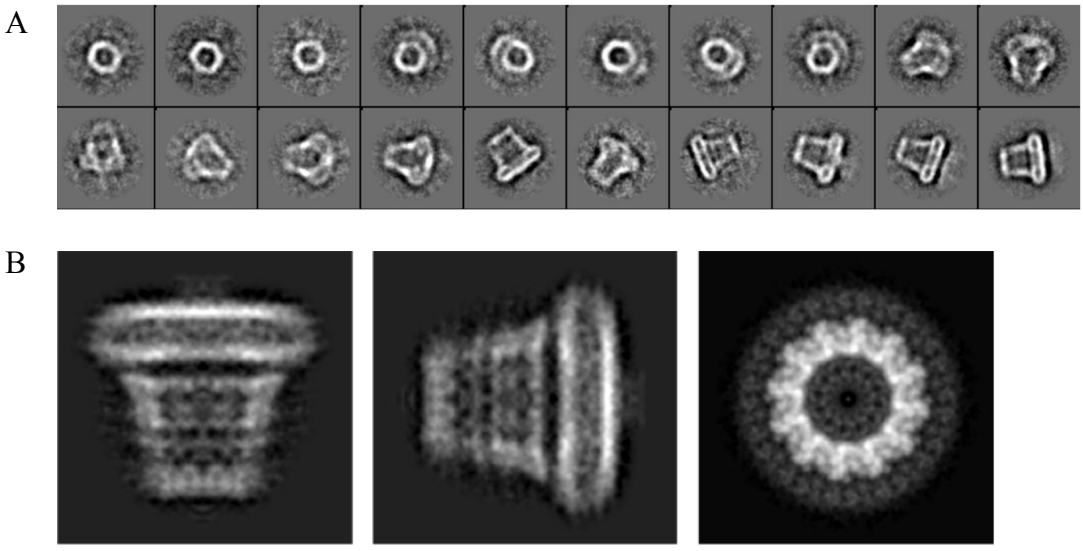


Figure 4.36. 2D classifications of Wzz (from *S.* Typhimurium) and orthogonal projections of the refined EM map. 2D classifications were taken from Collins *et al*. and the EM map projections were obtained from EMDB (EMDB-3611) (Collins *et al*., 2017). The 2D classifications obtained for Wzz alone and Wzz plus Wzy here resemble those from Collins *et al*.

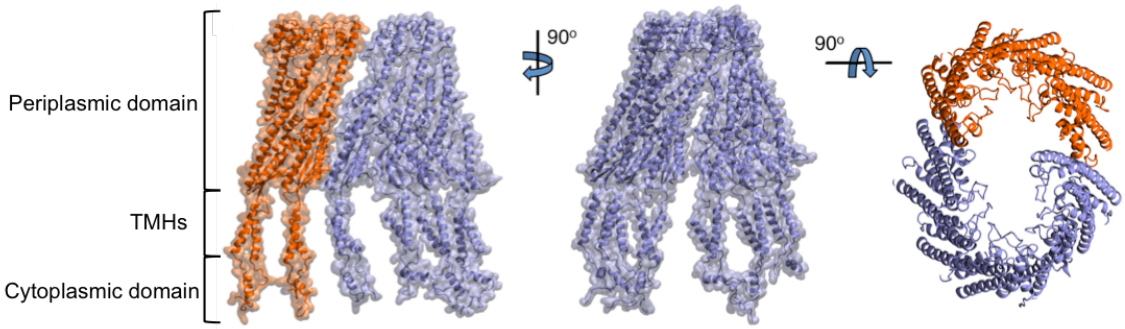


Figure 4.37. Modelled arrangement of Wzz (from *S.* Typhimurium) in lipid bilayer. Modelling by Collins *et al*. suggests that the dodecameric arrangement obtained in their cryo-EM structure is not stable in a lipid bilayer environment, with symmetry quickly lost during simulations (Collins *et al*., 2017). These models show that TMH pairs do not make any contacts with TMH pairs from neighbouring monomers, with a distance of approximately 10 Å between neighbouring TMHs (Collins *et al*., 2017). Image adapted from Collins *et al*. (Collins *et al*., 2017).

The latter proposal, that a Wzz oligomer forms around a central Wzy, is in agreement with the stoichiometry level calculated, as well as the similarity between the SEC elution profiles of Wzz alone and Wzz plus Wzy. This proposal is also in agreement with the low level of native expression of Wzy in comparison to that of Wzz (Carter *et al*., 2009; Daniels *et al*., 1998). These results do not agree with the proposed scaffold model of Wzz O-antigen chain length regulation (Tocilj *et al*., 2008). Futhermore, these results could also contradict some aspects of the model suggested recently by Collins *et al.* (Collins *et al*., 2017). Collins *et al.* suggested a model in which Wzy associates with the outside of a Wzz oligomer, with either the outside or the inside of the Wzz oligomer functioning as a ruler, shown in figure 4.6 (Collins *et al*., 2017). The existence of multiple Wzy binding sites on Wzz (Collins *et al*., 2017), also required for the scaffold model (Tocilj *et al*., 2008), contradicts our observation of only one SEC peak, and one BN-PAGE band, resulting from the tandem affinity purification of Wzz plus Wzy, as the presence of multiple binding sites would imply that multiple Wzy can bind to a Wzz oligomer at any one time, resulting in a range of complex sizes. A model of Wzz (from *S.* Typhimurium), shown in figure 4.37, including both the TMHs and cytoplasmic domain, in a lipid bilayer environment, shows that there are no contacts formed between TMHs of neighbouring monomers, with approximately 10 Å distance between TMHs of

neighbouring monomers (Collins *et al*., 2017). Furthermore, the dodecameric assembly observed in the cryo-EM structure of Wzz (*S.* Typhimurium) was determined, using molecular dynamics simulations, to be unstable in a lipid bilayer environment, with the symmetry quickly disrupted, shown in figure 4.37 (Collins *et al*., 2017). This lack of contact between the TMHs was proposed to provide a possible mode for the interaction between Wzz and Wzy (Collins *et al*., 2017). The insertion of Wzy between the TMHs of Wzz (Collins *et al*., 2017) could agree with the lack of structural differences between Wzz alone and Wzz plus Wzy observed here. The instability of the dodecameric assembly of Wzz in a lipid bilayer environment, as well as the lack of contacts between the TMHs of Wzz support the proposal that the inside of the bell could function as a ruler (Collins *et al*., 2017), which could involve a Wzz oligomer forming around a central Wzy, shown in figure 4.38.

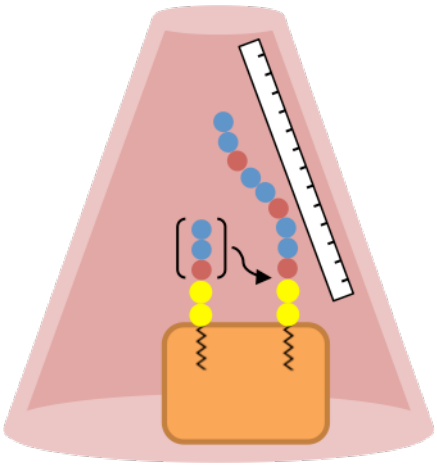


Figure 4.38. A Wzz oligomer could form around a central Wzy.

4.5 Conclusions and future work

Here we describe the tandem affinity purification of the complex formed between Wzz and Wzy (from *P. atrosepticum*) required for ECA biosynthesis, along with the preliminary structural characterisation of this complex. This is the first described example of the heterologous overexpression of this proposed complex. The data obtained here support the proposal that Wzz functions to regulate O-antigen chain

length via a direct interaction with Wzy, however they do not rule out an indirect interaction mediated by the substrate. Our work in this area has just begun and, whilst discussed in the context of the previously proposed mechanisms of Wzz-mediated O-antigen chain length regulation, these results are too preliminary to draw firm conclusions. The continued structural characterisation of this complex via cryo-EM, using the foundations established here, could offer new insights into the mode of O-antigen chain length regulation by Wzz.

4.5.1 Distinguishing between Wzz alone and Wzz plus Wzy

Strong evidence of a direct interaction between Wzz and Wzy is provided in the form of tandem affinity purification using multiple tagging strategies, western blotting and BN-PAGE. Furthermore, an *in vitro* interaction could be detected via far western blotting. SEC, BN-PAGE and cryo-EM were all employed in an attempt to identify a size difference between Wzz alone and tandem affinity purified Wzz with Wzy. However, as neither SEC, BN-PAGE nor cryo-EM could not distinguish between the size of Wzz alone and the size of tandem affinity purified Wzz plus Wzy, future work will initially focus on the confirmation of the presence of Wzy in this complex, as well as methods to distinguish between Wzz alone and Wzz plus Wzy. Intact mass spectrometry, in collaboration with Professor Carol Robinson, University of Oxford, is currently ongoing in an attempt to confirm the presence of Wzy in this complex, as well as the stoichiometry of the interaction between Wzz and Wzy (Laganowsky *et al*., 2014). The employment of a method to distinguish between Wzz alone and Wzz plus Wzy is crucial to the further structural characterisation of this proposed complex, in order to ensure that cryo-EM data collection and subsequent data processing is not complicated by high levels of free Wzz. This would require the tagging of Wzy with a tag that is of sufficient size to be easily identifiable in cryo-EM, but small enough so as to not disrupt the PPI. The additional tag would allow the selection for complexes containing Wzy, over Wzz alone, during the particle picking stage of cryo-EM data collection. For these reasons, *in vitro* tagging strategies are desirable. Fragment antigen binding (FABs) could be used to locate Wzy and has been previously used to tag proteins for cryo-EM

(Wu *et al*., 2012). The use of FABs to locate a specific component of a protein complex offers multiple advantages. FABs are approximately 50 kDa in size and form a distinctive shape that is easily recognisable by cryo-EM (Wu *et al*., 2012). Furthermore, the interaction between FABs and the protein against which they were raised is specific and can be reconstituted *in vitro*, preventing the potential disruption of complex formation that could result from the fusion of a large tag to a target protein. However, FABs exhibit an inherent level of flexibility between domains, in addition to the flexibility of the FAB:target protein complex, that could complicate downstream data processing (Wu *et al*., 2012). Furthermore, if the FAB is raised against Wzy alone, the FAB binding site may no longer be accessible in the complex between Wzz and Wzy. An additional disadvantage of the use of FABs is that they do not offer the possibility of an additional purification step to further select for the complex of interest.

Alternative strategies allow the incorporation of an additional purification step. The tagging of Wzy with an approximately 80 amino acid fragment of the dynein light chain-interacting domain 1 (DID1) followed by the incubation with purified flag-tagged DID2 and yeast dynein light chain (Dyn2), would result in the formation of an approximately 25 nm long tag formed by the interaction between DID1 and DID2 chains bridged by Dyn2 homodimers (Flemming *et al*., 2010). The flag tag of the DID2 chain allows the introduction of an additional purification step (Flemming *et al*., 2010). Furthermore, the fusion of approximately 80 amino acids, which remain unfolded in the absence of DID2 and Dyn2, should not interfere with complex formation (Flemming *et al*., 2010). If the approximately 80 amino acid long tag required for the DID2-Dyn2-DID1 tagging strategy is found to disrupt complex formation, an additional possible *in vitro* tagging strategy involves the incorporation of the unnatural amino acid, *p*-azidophenylalanine, which can then be modified with a tag via click chemistry (Dambacher and Lander, 2015). Dambacher *et al.* describe the tagging of a component of a protein complex, containing the unnatural amino acid *p*-azidophenylalanine, with the 42.5 kDa maltose binding protein (MBP), via click chemistry, which enabled the selection for the MBP-labelled protein complex using amylose resin (Dambacher and Lander, 2015). As well as the additional purification step, a further advantage of both of these *in vitro* tagging strategies over other strategies, for example gold-labelling, is that

the success of both can be confirmed using SDS-PAGE and SEC (Dambacher and Lander, 2015; Flemming *et al*., 2010), whereas successful gold-labelling can only be confirmed via EM. Any future tags should be introduced at the C-terminus of Wzy as C-terminal tags have been demonstrated to be accessible here during tandem affinity purification.

Cryo-EM using non-tagged Wzz plus Wzy, the preliminary results of which are described here, should also be pursued in order to obtain higher resolution data with the aim of obtaining a 3D structure. Particular focus should be on using data processing methods designed for complexes with pseudosymmetry, i.e. complexes with one highly symmetrical element, plus other additional non-symmetrical elements (Scheres, 2016), as is proposed to be the case here, with Wzz forming a symmetrical array with Wzy disrupting this symmetry. This could include particle subtraction, which would require the subtraction of density corresponding to Wzz from the 3D refined particle of Wzz plus Wzy. This would, in theory, leave density corresponding to Wzy, which could then be treated as a "sub-particle" for further refinement. This method would also be suitable if Wzy is present at substoichiometric amounts (Ilca *et al*., 2015; Scheres, 2016). This method, known as "localized reconstruction", has previously been implemented to solve the structure of the non-symmetrical 75 kDa RNA-dependent RNA polymerase, present inside the symmetrical P1 protein shell of the 15 MDa bacteriophage Φ6 polymerisation complex, at a resolution of 7.9 Å (Ilca *et al*., 2015). The RNA-dependent RNA polymerase is not only non-symmetrical, but it is also present at a symmetry axis of the P1 assembly (Ilca *et al*., 2015). Furthermore, it is only present in this location with a partial occupancy (Ilca *et al*., 2015). The 3D refinement of the entire complex, followed by the subtraction of the P1 density allowed the determination of the structure of the RNA-dependent RNA polymerase within the P1 protein assembly, therefore allowing the characterisation of the interaction interface (Ilca *et al*., 2015).

As previously mentioned, these results build upon those obtained by Nath and Morona. (Nath and Morona, 2015a). The detection of an interaction between Wzz and Wzy in *S. flexneri* by Nath and Morona required the use of chemical crosslinking prior to nickel affinity purification via the His tag fused to Wzy (Nath and Morona, 2015a). Although

chemical crosslinking was not required here for the tandem affinity purification of Wzz and Wzy, it should be explored in the future in order to determine whether the stabilisation of this interaction via chemical crosslinking results in a change in the SEC elution profile of Wzz plus Wzy when compared to that of Wzz. This would provide further evidence of complex formation between Wzz and Wzy. An alternative to this approach would be to determine whether a size difference can be seen when using crosslinking in combination with PAGE, for example perfluorooctanoic acid (PFO)-PAGE (Penna, 2008). Chemical crosslinking should also be explored during cryo-EM sample preparation as the lack of obvious density for Wzy could also be due to the loss of Wzy during cryo-EM sample preparation.

4.5.2 Further characterisation of this interaction

Future work will also focus on determining the relevance of this interaction, as it could be an artefact of the overexpression required due to the low levels of native expression of Wzy (Carter *et al*., 2009; Daniels *et al*., 1998). Split-GFP tagging would be able to determine whether these proteins interact *in vivo* (Wilson *et al*., 2004), confirming that the interaction observed is not an artefact of the purification conditions used, for example the extraction of these proteins from the membrane using detergent. Split-GFP tagging involves the fusion of one GFP fragment to one proposed interaction partner and the other fragment of GFP to the other proposed interaction partner. These GFP fragments are unable to fluoresce in the absence of the other fragment. Interaction between the proposed partners *in vivo* would result in the reconstitution of GFP fluorescence (Wilson *et al*., 2004). Furthermore, the relevance of this interaction should also be confirmed by determining whether the addition of the appropriate substrate to tandem affinity purified Wzz and Wzy results in the expected modal distribution of chain lengths, as Woodward *et al.* proposed that O-antigen chain length modality can be restored *in vitro* upon addition of Wzz to Wzy in the presence of substrate, indicating that Wzz and Wzy are the only components of this pathway required for a modal distribution of O-antigen chain lengths (Woodward *et al*., 2010). O-antigen chain length

modality could be detected via a combination of SDS-PAGE coupled with silver staining or western blotting analysis of purified LPS (Davis Jr and Goldberg, 2012).

Far western dot blotting demonstrated the possibility of reconstituting this PPI *in vitro*, despite the presence of the detergent micelle. The reconstitution of this PPI *in vitro* offers the possibility of determining the affinity of this interaction via an enzyme-linked absorbance assay (ELISA) (Biesiadecki and Jin, 2011), as this follows the same principle as far western blotting (Edmondson and Dent, 2001), i.e. the immobilisation of a "bait" protein, followed by the incubation with the "prey" protein, with the readout generated via antibody binding to the "prey" protein (Biesiadecki and Jin, 2011).

The function of Wzy has been studied via an SDM approach (Islam *et al*., 2013; Kim *et al*., 2010; Nath and Morona, 2015b; Nath *et al*., 2015). Multiple studies have identified Wzy mutants that result in an altered O-antigen chain length modality (Islam *et al*., 2013; Nath and Morona, 2015b; Nath *et al*., 2015). However SDM of Wzy could not prevent the *in vivo* crosslinking of Wzy to Wzz (Nath and Morona, 2015a). As the system identified here does not require crosslinking prior to tandem affinity purification, the effect of SDM of Wzy on the interaction observed could be investigated further.

4.5.3 Wzz and Wzy as a component of a larger complex?

The interaction between Wzz and Wzy characterised here could be a component of a larger protein complex. It has been proposed that Wzx, Wzy and Wzz form a multiprotein complex (Marolda *et al*., 2006). Marolda *et al.* found that WzxE, required for the biosynthesis of ECA, could function as the translocase during O16 biosynthesis, but only in the absence of its cognate WzzE and WzyE (Marolda *et al*., 2006). Furthermore, the expression of either WzzE or WzyE resulted in a decrease in the level of O16-linked undecaprenyl pyrophosphate flipped across the inner membrane by WzxE (Marolda *et al*., 2006). The only common factor in the ECA and O16 repeat unit is that they both contain a GlcNAc as their initiating sugar, which is in agreement with

the previous finding that Wzx can flip undecaprenyl-pyrophosphate linked just to the initial sugar of the repeat unit across the inner membrane (Feldman *et al*., 1999; Marolda *et al*., 2006). These results suggest that WzzE and WzyE sequester WzxE in a multiprotein complex (Feldman *et al*., 1999; Marolda *et al*., 2006). The formation of a complex between Wzz, Wzy and Wzx would serve to regulate the somewhat promiscuous activity of Wzx with regards to the components of the sugar repeat unit (Feldman *et al*., 1999; Marolda *et al*., 2006). This would also allow the compartmentalisation of the proteins required for ECA, O-antigen and colanic acid biosynthesis (Marolda *et al*., 2006). Interactions between various components of this pathway could also serve as a feedback loop to further regulate O-antigen biosynthesis, for example the deletion of the periplasmic loop between TMHs 4 and 5 of the initiating enzyme WbaP has been determined to result in a change in O-antigen chain length modality in *S. enterica* strain LT2 (Saldías *et al*., 2008).

Taken together, the results obtained here support a direct interaction between Wzz and Wzy, building upon the work of Nath and Morona. (Nath and Morona, 2015a), but do not rule out an indirect interaction. Our results support the proposal that one Wzy interacts with a Wzz oligomer, with the Wzz oligomer perhaps forming around a central Wzy, and are therefore incompatible with a scaffold model (Tocilj *et al*., 2008). However they do support the hypothesis that Wzz functions as a molecular ruler (Kintz and Goldberg, 2011). These results could also support the "chain feedback" model (Kalynych *et al*., 2012b) and a "chain feedback/ruler" hybrid model proposed by Islam *et al.* (Islam and Lam, 2014), based on the mechanisms proposed by Kintz *et al.* (Kintz and Goldberg, 2011) and Kalynych *et al.* (Kalynych *et al*., 2012b). We, along with Collins *et al.* (Collins *et al*., 2017) propose that the inside of the bell could function as a ruler, see figures 4.6 and 4.38, however further evidence is required to support this model.

5 The structural characterisation of modified heptose biosynthetic enzymes: Parallels with dTDP-L-rhamnose biosynthesis

5.1 Introduction

Modified heptoses, present in the capsules of Gram-negative pathogens including *C. jejuni* and *Y. pseudotuberculosis*, have been shown to play crucial roles in virulence (Ho *et al*., 2008; Kanipes *et al*., 2004; Wong *et al*., 2015). Heptose biosynthesis is not unique to Gram-negative bacteria as examples of heptose biosynthesis are also found in Gram-positive bacteria, for example D-*glycero*-D-*manno*-heptose is found in the S-layer of *Aneurinibacillus thermoaerophilus* DSM 10155 (Kneidinger *et al*., 2001). Like the enzymes required for hexose biosynthesis, the enzymes required for modified heptose biosynthesis represent potential drug targets (Allard *et al*., 2001; Giraud *et al*., 1999, 2000, McCallum *et al*., 2012, 2013), however, in contrast to hexose biosynthetic enzymes, they remain to be structurally characterised.

Modified heptose biosynthesis shares many parallels with the well-characterised dTDP-L-rhamnose biosynthesis pathway and we can use our knowledge of the dTDP-L-rhamnose pathway to further our understanding of and make predictions regarding the mechanisms of modified heptose biosynthesis. The biosynthesis of GDP-6-deoxy-D-*altro*-heptose in *C. jejuni* strain 81-176 proceeds via the DdahA-catalysed C4, C6 dehydration of GDP-D-*manno*-heptose (also known as GDP-D-*glycero*-D-*manno*-heptose) to form GDP-4-keto-6-deoxy-D-*lyxo*-heptose (McCallum *et al*., 2011). This is

similar to the RmlB-catalysed dehydration of dTDP-D-glucose to form dTDP-4-keto-6-deoxy-D-*xylo*-4-hexulose, which occurs as the second step of dTDP-L-rhamnose biosynthesis (Allard *et al*., 2001). Following the mechanism of RmlB-catalysed dehydration proposed by Allard *et al.*, the dehydration of GDP-D-*manno*-heptose is predicted to occur via an oxidation at the C4 position by NAD^+, coupled with the deprotonation of the C4 OH group by an active site base, to form GDP-4-keto-6-deoxy-*manno*-heptose. C5 would be deprotonated by an active site base and an OH group would be released from C6 and subsequently protonated by an active site acid, thereby dehydrating GDP-4-keto-6-deoxy-*manno*-heptose across the C4 and C6 position to form GDP-4-keto-6-deoxy-*manno*-heptose-5,6-ene, which is then reduced at the C5 position to form GDP-4-keto-6-deoxy-D-*lyxo*-heptose (shown in figure 5.1) (Allard *et al*., 2001, 2002). This is epimerised at the C3 position by DdahB to form GDP-4-keto-6-deoxy-D-*arabino*-heptose. This is similar to the RmlC-catalysed epimerisation of dTDP-6-deoxy-D-*xylo*-4-hexulose to form dTDP-6-deoxy-L-*lyxo*-hexulose, which occurs as the second step in dTDP-L-rhamnose biosynthesis (Giraud *et al*., 2000). A predicted mechanism for DdahB-catalysed epimerisation, based on that of RmlC, is shown in figure 5.2 (Dong *et al*., 2007). During DdahB-catalysed epimerisation, the C3 proton is predicted to be abstracted by an active site acid and an active site base would reprotonate the C3 position on the opposite face of the ring to generate the C3 epimer. As is the case with RmlC, this is predicted to occur in a cofactor-independent manner (Dong *et al*., 2007; Giraud *et al*., 2000). Finally, DdahC catalyses the reduction at the C4 position to form GDP-6-deoxy-D-*altro*-heptose (McCallum *et al*., 2012). C4 reduction also occurs as the final step in dTDP-L-rhamnose biosynthesis and is catalysed by RmlD (Graninger *et al*., 1999). Based on the mechanism of the RmlD-catalysed reduction of dTDP-4-keto-6-deoxy-D-*lyxo*-hexulose to form dTDP-L-rhamnose, the conversion of GDP-4-keto-6-deoxy-D-*arabino*-heptose to GDP-6-deoxy-D-*altro*-heptose is predicted to occur via the protonation at the C4 position by NADPH, coupled with the protonation of the 4'-keto group by an active site acid (Blankenfeldt *et al*., 2002). This is shown in figure 5.3.

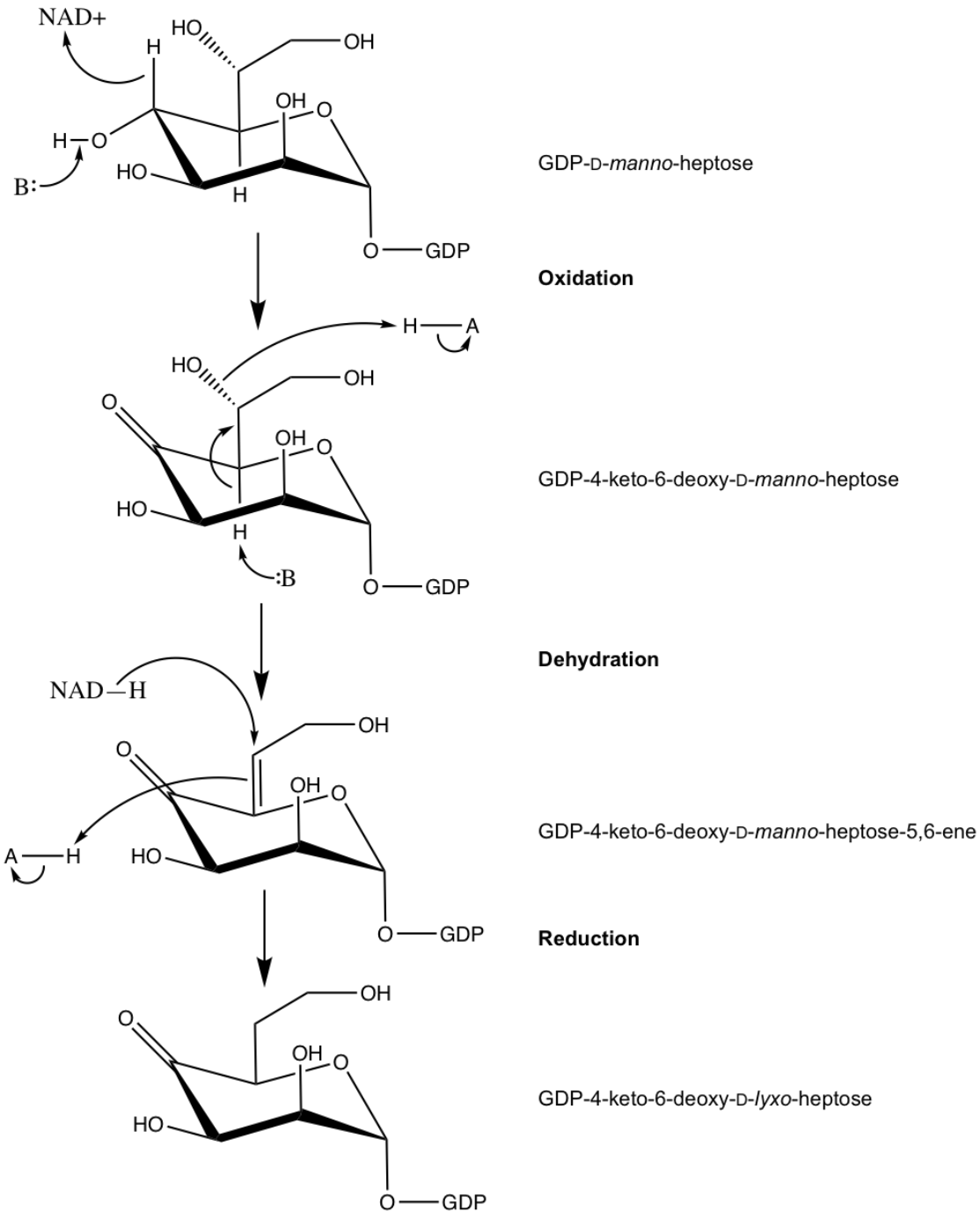


Figure 5.1. Predicted mechanism for the dehydration of GDP-D-*manno*-heptose to form GDP-4-keto-6-deoxy-D-*lyxo*-heptose. This is catalysed by DdahA, during GDP-6-deoxy-D-altro-heptose biosynthesis in *C. jejuni* 81-176 (McCallum *et al*., 2011), and DmhA, during 6-deoxyheptose biosynthesis in *Y. pseudotuberculosis* O:2a (Butty *et al*., 2009; Pacinelli *et al*., 2002). This mechanism is based on that of the dTDP-D-glucose 4,6-dehydratase RmlB (Allard *et al*., 2001, 2002). This image was created using ChemDraw 16.0.

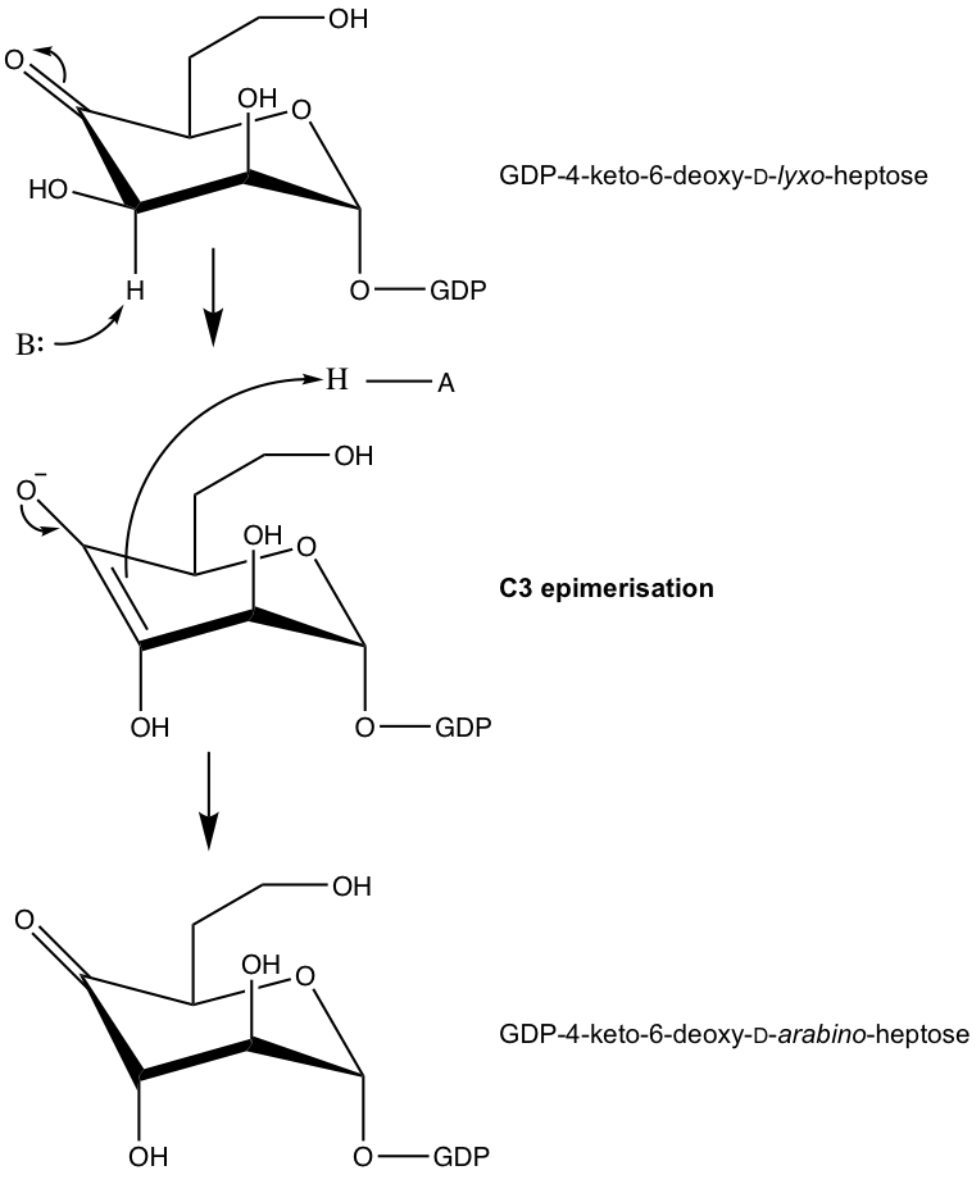


Figure 5.2. Predicted mechanism for the C3 epimerisation of GDP-4-keto-6-deoxy-D-*lyxo*-heptose to form GDP-4-keto-6-deoxy-D-*arabino*-heptose. This is catalysed by DdahB during GDP-6-deoxy-D-*altro*-heptose biosynthesis in *C. jejuni* strain 81-176 (McCallum *et al*., 2012). This predicted mechanism is based on that of RmlC (Dong *et al*., 2003b, 2007). This image was created using ChemDraw 16.0.

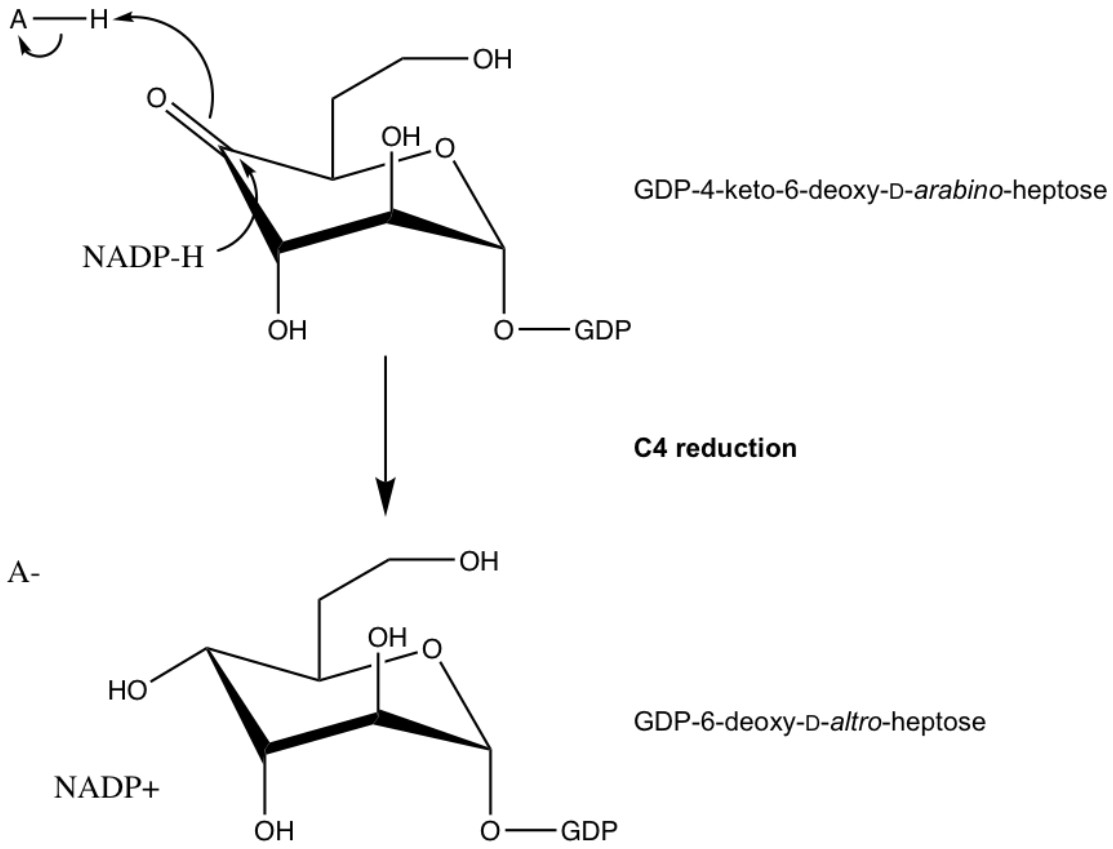


Figure 5.3. Predicted mechanism for the reduction of GDP-4-keto-6-deoxy-D-*arabino*-heptose to form GDP-6-deoxy-D-*altro*-heptose. This is catalysed by DdahC during GDP-6-deoxy-D-*altro*-heptose biosynthesis in *C. jejuni* strain 81-176 (McCallum *et al*., 2012). This mechanism is based on that of RmlD (Blankenfeldt *et al*., 2002)

The biosynthesis of GDP-6-*O*-methyl-L-*gluco*-heptose in *C. jejuni* strain NCTC 11168 follows a similar scheme, but is more complex and less well characterised. As the first step in GDP-6-*O*-methyl-L-*gluco*-heptose biosynthesis, MlghA catalyses the oxidation of GDP-D-*manno*-heptose at the C4 position to form GDP-4-keto-D-*lyxo*-heptose (McCallum *et al*., 2013). The existence of MlghA, and its putative oxidase activity, is inferred from the absence of a dehydrated final product and the finding that MlghB is inactive on GDP-D-*manno*-heptose lacking the 4-keto group, but the gene encoding MlghA remains unidentified (McCallum *et al*., 2013). The methylation of the C6 hydroxyl group, by the methyltransferase MlghD, has been proposed, based on the results of mutagenesis studies (Sternberg *et al*., 2013), to occur as the second step in GDP-*O*-methyl-L-*gluco*-heptose biosynthesis to produce GDP-6-*O*-methyl-4-keto-D-*lyxo*-heptose, however this remains to be biochemically confirmed (McCallum *et al*., 2013). MlghB catalyses the epimerisation at the C3 and C5 positions to form GDP-(6-*O*-methyl)-4-keto-L-*xylo*-heptose, which is reduced at the C4 position by MlghC to form GDP-(6-*O*-methyl)-L-*gluco*-heptose as the final product (McCallum *et al*., 2011, 2012, 2013). The predicted mechanism for the MlghB-catalysed epimerisation of GDP-(6-*O*-methyl)-4-keto-D-*lyxo*-heptose at the C3 and C5 positions, based on that of RmlC (Dong *et al*., 2003b, 2007), is shown in figure 5.4; the order of epimerisation events shown is arbitrary. Like the predicted mechanism of DdahB, a proton would be abstracted from one face of the ring and the same position would be reprotonated on the other face of the ring, generating the corresponding epimer (Dong *et al*., 2003b, 2007). As MlghB epimerises its substrate at two positions, like RmlC, a ring flip is predicted to occur to reorientate the second carbon to enable epimerisation (Dong *et al*., 2007). The mechanism of MlghC is predicted to be the same as that of RmlD, and DdahC, shown in figure 5.3. The identity of the active site residues required for all the mechanisms predicted here remain unconfirmed.

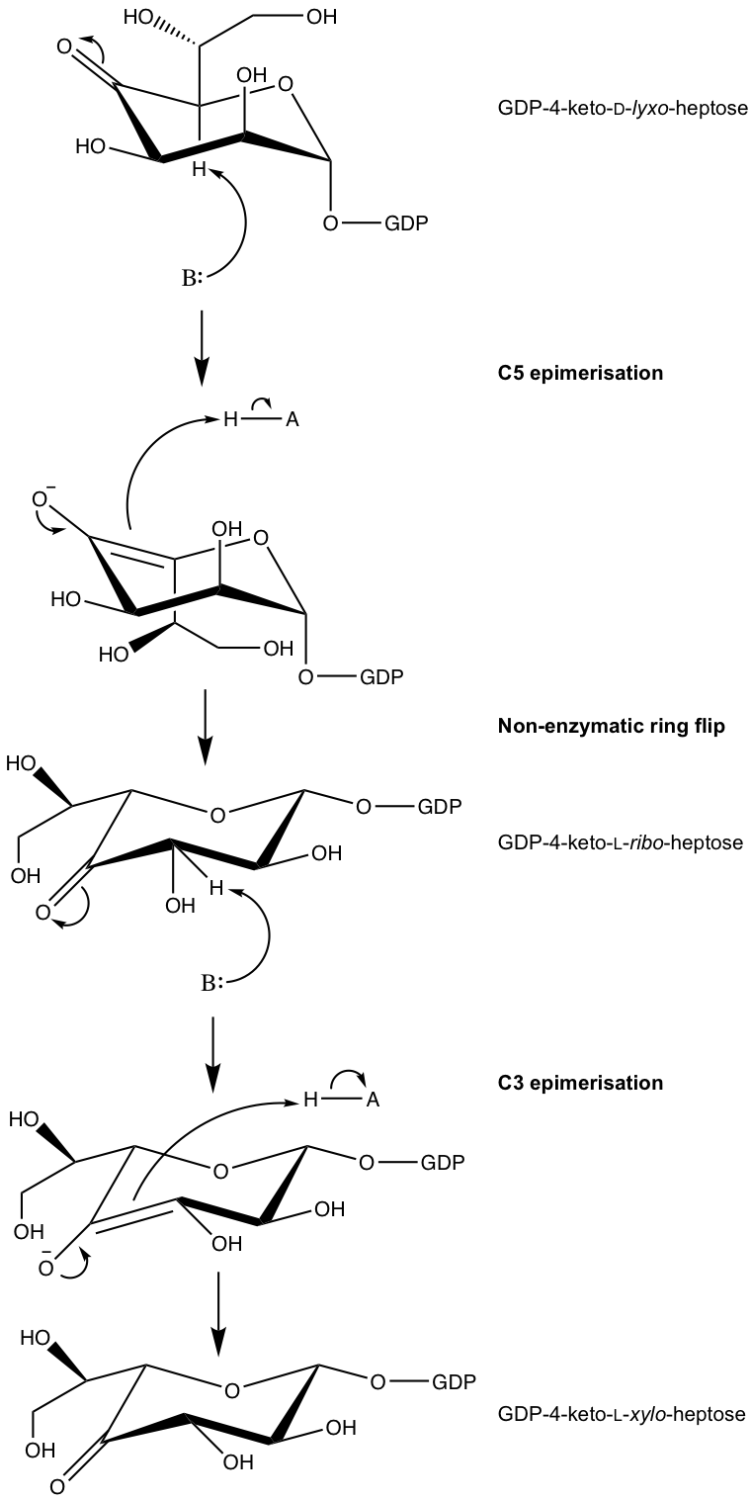


Figure 5.4. Predicted mechanism for the C5/C3 epimerisation of GDP-4-keto-D-*lyxo*-heptose to form GDP-4-keto-L-*xylo*-heptose. This is catalysed by MlghB during GDP-6-*O*-methyl-L-*gluco*-heptose biosynthesis in *C. jejuni* strain NCTC 11168 (McCallum *et al*., 2013). This mechanism is based on that of RmlC and the order of epimerisation events shown here is not confirmed (Dong *et al*., 2007; McCallum *et al*., 2013). This image was created using ChemDraw 16.0.

In addition to the identity of the active site residues, further questions regarding the mechanisms of the steps involved in modified heptose biosynthesis remain unanswered. The dehydration of GDP-D-*manno*-heptose at the C4 and C6 positions to form GDP-4-keto-6-deoxy-D-*lyxo*-heptose, catalysed by DdahA (McCallum *et al*., 2011), also occurs as the first step in the biosynthesis of 6-deoxyheptose in *Y. pseudotuberculosis* strain O:2a and is catalysed by DmhA (Butty *et al*., 2009; Pacinelli *et al*., 2002). DdahA is 78 % identical and 87 % similar to DmhA, shown in figure 5.5 (McCallum *et al*., 2011). DmhA is therefore predicted to follow the same mechanism as DdahA, proposed in figure 5.1. Both DdahA and DmhA bind GDP-mannose and GDP-D-*manno*-heptose with equal affinity, however the k_{cat} of DmhA for GDP-D-*manno*-heptose is approximately 6000-fold higher than that for GDP-mannose (Butty *et al*., 2009) and the k_{cat} of DdahA for GDP-D-*manno*-heptose is 30-fold higher than that for GDP-mannose (McCallum *et al*., 2011). The mechanism that allows the differentiation between GDP-D-*manno*-heptose and GDP-mannose by DdahA and DmhA remains unknown.

```
DmhA       1 MNN-VLITGFTGQVGSQLADYILENTDDHVIGMMRWQESMDNIYHLTDRINKKDRISIQY
DdahA      1 MKKTALITGFTGQVGSQMADFLLENTDYDVIGMMRWQEPMDNIYHLSDRINKKDRISIFY
RmlB       1 MK--ILITGGAGFIGSAVVRHIIKNTQDTVVNIDKLTY-AGNLESLSD-ISESNRYNFEH

DmhA      60 ADLNDLMSLYNLIDTVRPKFIFHLAAQSFPRTSFDIPIETLQTNIIGTANLLECIRKLKQ
DdahA     61 ADLNDYSSLQKLFESQRPDVIFHLAAQSYPKTSFDIPIETLQTNIIGTANILENIRILKA
RmlB      57 ADICDSAEITRIFEQYQPDAVMHLAAESHVDRSITGPAAFIETNIVGTYALLEVARKYWS

DmhA     120 QDGYDP----VVHVCSSSEVYGRAKVGE---------ALNEDTQFHGASPYSISKIGTDY
DdahA    121 KEGYDP----VVHICSSSEVYGKAKVGV---------KLNEETAFHGASPYSISKIGTDY
RmlB     117 ALGEDKKNNFRFHHISTDEVYGDLPHPDEVENSVTLPLFTETTAYAPSSPYSASKASSDH

DmhA     167 LGQFYGEAYGIRTFVTRMGTHTGPRRSDVFFESTVAKQIALIEAGHQEPKLKVGNLASVR
DdahA    168 LGKFYGEAYNIRTFVTRMGTHSGPRRSDVFFESTVAKQIALIEAGYQEPVIKVGNLSSVR
RmlB     177 LVRAWRRTYGLPTIVTNCSNNYGPYH---FPEKLIP--LVILNALEGKPLPIYGKGDQIR

DmhA     227 TFQDARDAVRAYYLLALESGKGNIPNGEVFNIAGDEAFKLPEVI-------ELLLSFSTR
DdahA    228 TFQDCRDAIRAYYLLSLESEKGNIPCGEAFNIAGEEAFKLPEVI-------DILLNFSDM
RmlB     232 DWLYVEDHARALHMVVTEGK-----AGETYNIGGHNEKKNLDVVFTICDLLDEIVPKATS

DmhA     280 --NDIEVVTDTDRLRPIDADYQMFDSTKIKSYINWKPEIKAADMFRDL----L--QHWRN
DdahA    281 G-RGIEVRQVEDRMRPIDADYQMFDNSKIKSFIDWKAEIPVRQMLKDL----L--NHWRN
RmlB     287 YREQIT--YVAD--RPGHDRRYAIDAGKISRELGWKPLETFESGIRKTVEWYLANTQWVN

DmhA     332 EIASGRIP----LNR----
DdahA    334 EIKRGRIP----LNR----
RmlB     343 NVKSGAYQSWIEQNYEGRQ
```

Figure 5.5. Sequence alignment of DmhA (*Y. pseudotuberculosis* strain O:2a), DdahA (*C. jejuni* strain 81-176) and RmlB (*S.* Typhimurium strain LT2). The sequence alignment was created using the T-Coffee server followed by the BoxShade server, provided by the Swiss Institute of Bioinformatics (Notredame *et al*., 2000; Di Tommaso *et al*., 2011).

Both DdahB and MlghB are predicted to be C3/C5 epimerases (Giraud *et al*., 2000; McCallum *et al*., 2012, 2013). MlghB, like RmlC, epimerises the substrate at the C3 and C5 positions (Giraud *et al*., 2000; McCallum *et al*., 2013), whereas DdahB epimerises the substrate only at the C3 position (compare figures 5.2 and 5.4) (McCallum *et al*., 2012). MlghB is thought to epimerise its substrate with no preferential order, with the C3, C5 and C3/C5 epimers all detected by capillary electrophoresis upon addition of MlghB to its substrate (McCallum *et al*., 2013). This difference in observed activities of DdahB and MlghB remains unexplained, given their primary sequences are 81 % identical and 98 % similar (figure 5.6). There are some similarities in DdahB and MlghB activities that suggest a shared mechanism. Neither DdahB nor MlghB epimerises GDP-D-*manno*-heptose, both require a keto group at the C4 position, generated by DdahA and MlghA, respectively, for epimerisation (McCallum *et al*., 2012, 2013). Furthermore, MlghB has been found to epimerise the substrate of DdahB, GDP-4-keto-6-deoxy-D-*lyxo*-heptose, enabling the biochemical characterisation of MlghB in the absence of MlghA (McCallum *et al*., 2013). DdahB and MlghB have been shown to epimerise GDP-mannose, however this occurs at a much slower rate than with the native substrate (Dr Creuzenet, University of Western Ontario, personal communication). The basis of discrimination between heptose and hexose of DdahB and MlghB, as well as the lack of C5 epimerase activity of DdahB, remains unknown.

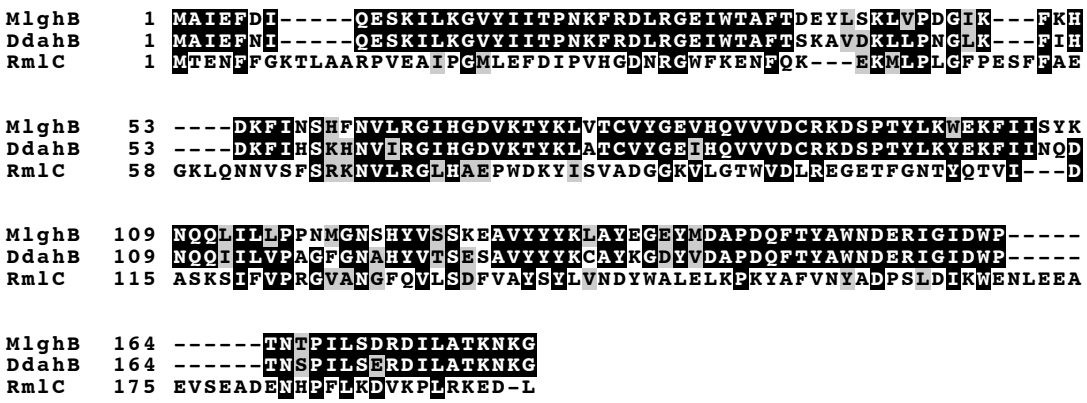


Figure 5.6. A sequence alignment of MlghB (*C. jejuni* strain NCTC 11168), DdahB (*C. jejuni* strain 81-176) and RmlC (*S. suis*). The sequence alignment was created using the T-Coffee server followed by the BoxShade server, provided by the Swiss Institute of Bioinformatics (Notredame *et al*., 2000; Di Tommaso *et al*., 2011).

Both DdahC and MlghC are members of the NAD-dependent epimerase/ dehydratase class (PFAM 01370) and are homologues of GDP-fucose synthetase, which epimerises GDP-4-keto-6-deoxy-mannose at the C3 and C5 positions, and then reduces it at the C4 position to produce GDP-fucose, using only one active site (Finn *et al*., 2016; Somers *et al*., 1998). This is also similar to the GDP-mannose-3,5-epimerase GME (from *Arabidopsis thaliana*), which converts GDP-α-D-mannose to GDP-β-L-galactose during vitamin C biosynthesis (Major *et al*., 2005). DdahC and MlghC are therefore predicted C3/C5 epimerases and C4 reductases (McCallum *et al*., 2012, 2013). However, both DdahC and MlghC appear to only carry out the C4 reduction of their respective substrates (McCallum *et al*., 2012, 2013). This lack of epimerisation activity is also observed in RmlD (Graninger *et al*., 1999). Furthermore, both DdahC and MlghC specifically act on the epimer generated by their cognate epimerase, suggesting a high level of substrate specificity (McCallum *et al*., 2012, 2013). At the level of primary sequence, DdahC and MlghC are 57 % identical and 90 % similar (figure 5.7). The lack of epimerisation activity of DdahC and MlghC and their mode of substrate recognition are open questions.

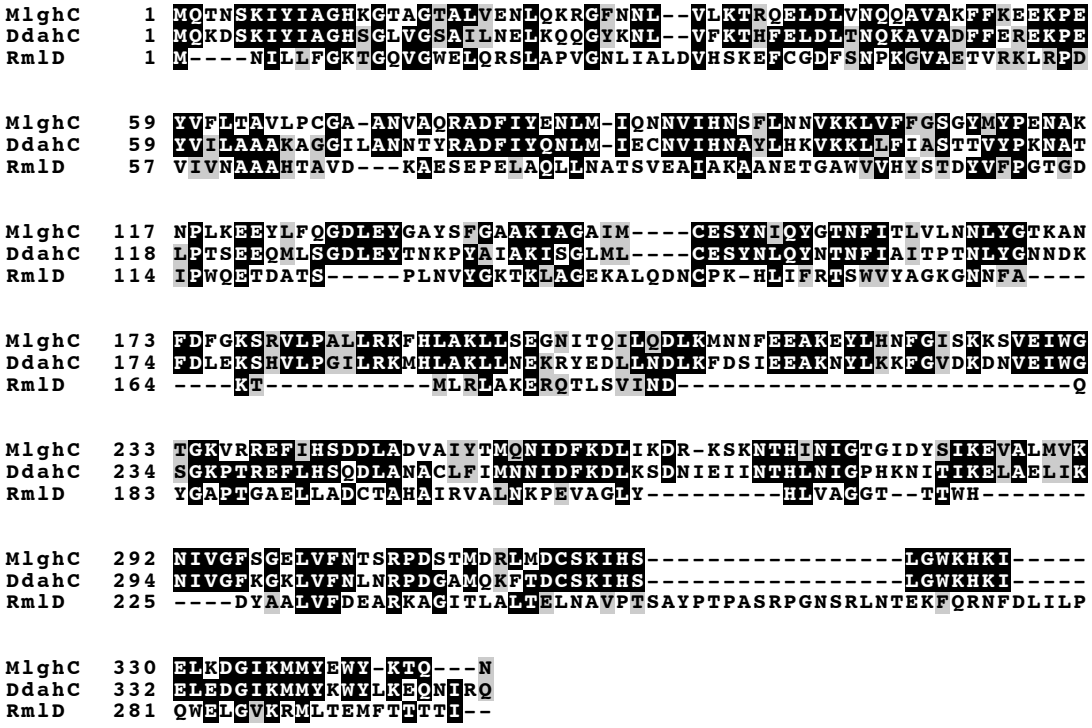


Figure 5.7. A sequence alignment of MlghC (*C. jejuni* strain NCTC 11168), DdahC (*C. jejuni* strain 81-176) and RmlD (*S.* Typhimurium strain LT2). The sequence alignment was created using the T-Coffee server followed by the BoxShade server, provided by the Swiss Institute of Bioinformatics (Notredame *et al*., 2000; Di Tommaso *et al*., 2011).

In order to address these open questions and to allow us to assign the residues required for the mechanisms of modified heptose biosynthesis described here, the enzymes required for the biosynthesis of modified heptoses in *C. jejuni* strain 81-176, namely DdahA, DdahB and DdahC, in *C. jejuni* strain NCTC 11168, namely MlghB and MlghC, as well as DmhA from *Y. pseudotuberculosis* strain O:2a, were expressed, purified and crystallised with the aim of obtaining both their apo and substrate-bound structures via X-ray crystallography. In order to further understand the results obtained, the structures of these enzymes will be compared and contrasted with those of the dTDP-L-rhamnose biosynthetic enzymes.

5.2 Materials and Methods

The recipes for buffers described here can be found in appendix 8.1.3.

5.2.1 Cloning of target genes into expression vectors

ddahB from *C. jejuni* strain 81-176 and *mlghB* from *C. jejuni* strain NCTC-11168 were cloned into the pEHisTEV vector (kanamycin resistant), with a T7 promoter and an N-terminal TEV protease-cleavable His_6 tag (Liu and Naismith, 2009) by Miss Heba Barnawi (University of Western Ontario). *dmhA* from *Y. pseudotuberculosis* strain O:2a had been previously cloned into a pET23 vector with N-terminal non-TEV protease-cleavable His_6 tag (ampicillin resistant) (Butty *et al*., 2009). *ddahA* from *C. jejuni* strain 81-176 had also been previously cloned into the pET23 vector (McCallum *et al*., 2011). *ddahC* from *C. jejuni* strain 81-176 and *mlghC* from *C. jejuni* strain NCTC 11168 had both been previously cloned into the pET23 vector (McCallum *et al*., 2012; McCallum *et al*., 2013). pET-23-*dmhA*, pET-23-*ddahA*, pET-23-*ddahC* and pET-23-*mlghC* were all provided by Dr Carole Creuzenet, University of Western Ontario.

These plasmids were transformed into *E. coli* C43 (DE3) cells, as described in section 2.1, and expression trials were conducted as described in section 2.2. Expression was confirmed via SDS-PAGE and western blotting against the His_6 affinity tag, described in sections 2.3 and 2.4.

5.2.2 SDM of DdahB to remove MS linker

DdahB was cloned by Miss Heba Barnawi (University of Western Ontario) with an extended linker between the TEV protease site and the target sequence, composing of Met-Ser. This MS linker is absent in the MlghB construct. In order to ensure that any observed differences between DdahB and MlghB binding to GDP or GDP-mannose result from differences in their respective target sequences, rather than from a cloning artefact, this MS linker was removed via SDM (Liu and Naismith, 2008), using the primers shown in table 5.1. The following PCR protocol was used: 94 °C for seven

minutes; then 12 cycles of 94 °C for one minute, 50 °C for one minute and 72 °C for 10 minutes; followed by three cycles of 95 °C for one minute, 48 °C for one minute and 72 °C for 10 minutes; then 72 °C for a final 20 minutes (Liu and Naismith, 2008). The PCR product was then incubated with *Dpn*I (Promega) for two hours at 37 °C and the PCR product was purified using a QIAquick gel extraction kit (Qiagen). The purified sample was transformed into *E. coli* DH5 α cells (Bioline), as described in section 2.1. A single colony was then used to inoculate 10 ml LB-medium, supplemented with ampicillin at 100 μg/ml and grown overnight at 37 °C. The cells were then harvested via centrifugation for 10 minutes at 3,488 x *g* and the plasmid was purified using a QIAprep Spin Miniprep kit (Qiagen). The purified plasmid was then sequenced using GATC.

5.2.3 Insertion of a TEV protease site via SDM to allow cleavage of N-terminal His_6-tag from DdahC and MlghC

Primer sequences for the insertion of a TEV protease site, via SDM, between the N-terminal His_6 tag and the DdahC and MlghC protein sequences can be seen in table 5.1. SDM was conducted as described in section 5.2.2.

	Primer Sequence
DdahB MS linker removal Forward	GGG-CGC-CAT-GGC-AAT-AGA-ATT-TAA-TAT-ACA-AGA-ATC-AAA-AA
DdahB MS linker removal Reverse	TAT-TGC-CAT-GGC-GCC-CTG-AAA-ATA-CAG-GTT-TTC-GGT-CG
TEV-DdahC Forward	GAA-AAC-CTG-TAT-TTT-CAG-GGC-ATG-CAA-AAA-GAT-TCT-AAA-ATT-T
TEV-DdahC Reverse	CCT-GAA-AAT-ACA-GGT-TTT-CGC-CCA-TGG-AAC-CGT-GGT-GG
TEV-MlghC Forward	GAA-AAC-CTG-TAT-TTT-CAG-GGC-ATG-CAA-ACA-AAT-TCA-AAA-ATA-T
TEV-MlghC Reverse	CCT-GAA-AAT-ACA-GGT-TTT-CGC-CCA-TGG-AAC-CGT-GGT-GG

Table 5.1. Primers for the removal of the MS linker between the TEV protease cleavage site and DdahB target sequence (from *C. jejuni* strain 81-176), the insertion of a TEV protease cleavage site between the N-terminal His_6 tag and the target sequence of DdahC (from *C. jejuni* strain 81-176) and MlghC (from *C. jejuni* strain NCTC 11168) via SDM. (Liu and Naismith, 2008).

5.2.4 Small-scale purification trials

E. coli cell pellets, obtained from one ml of *E. coli* cell culture, harvested following expression trials, conducted as described in section 2.2, were resuspended 1 ml of PBS and sonicated at 15 microns for five seconds, with one minute off, a total of three times at 4 °C (Soniprep 150). The lysate was centrifuged at 20,817 x *g* at 4 °C for 20 minutes and the supernatant was tested for the presence of soluble His-tagged protein using a Biosprint 15 (Qiagen) as follows: 500 µl supernatant was added to the first well of the Biosprint unit along with 50 µl MagneHis nickel particles (Promega), 500 µl Biosprint wash buffer was added to wells two and three, and 50 µl Biosprint elution buffer was added to wells four and five. A sample was taken from well four (eluate) and analysed via SDS-PAGE, described in section 2.3. The recipes for buffers discussed here can be found in appendix 8.1.3.

5.2.5 Large-scale cell culture

All constructs here were expressed in *E. coli* C43 (DE3) cells using LB growth medium. DdahB and MlghB were expressed at 25 °C overnight (overnight corresponds to approximately 20 hours for all large-scale cell cultures described) following induction at an OD_{600} of 0.6 with 0.5 mM IPTG. The expression of non-TEV protease-cleavable pET-23-MlghC and pET-23-DdahC was induced at an OD_{600} of 0.4 with 1.0 mM IPTG. Growth was continued overnight at 25 °C in the case of non-TEV protease-cleavable pET-23-MlghC and for 3 hours at 37 °C for non-TEV protease-cleavable pET-23-DdahC. Expression of TEV protease-cleavable MlghC and DdahC was induced with 1 mM IPTG at an OD_{600} of 0.4 and the cells were then grown overnight at 25 °C. DmhA and DdahA were both expressed overnight at 25 °C following induction of expression with 0.5 mM IPTG at an OD_{600} 0.4. Following the appropriate growth, cells were harvested at 6,000 rpm for 10 minutes at 4 °C using a JLA-8.1000 rotor (Beckman

Coulter) in a J-26XP centrifuge (Beckman Coulter) and the pellets stored at -80 °C until required.

5.2.6 Protein purification

A cell pellet was resuspended in PBS plus 1 mg DNAse (Sigma Aldrich) and lysed using a cell disrupter at 30 kpsi at 4 °C (the lysate was passed through the cell disrupter twice) (Constant Systems). The cell lysate was centrifuged at 18,000 rpm for 20 minutes at 4 °C using a JA 25.50 rotor (Beckman Coulter) in a J-26XP centrifuge (Beckman Coulter). The supernatant was filtered through a 0.45 µm filter (EMD Millepore) and batch-bound for one hour at 4 °C with 2 ml of nickel sepharose 6 fast flow resin (GE Healthcare), pre-equilibrated with PBS. The mixture was applied to an empty column and the FT was run through the column twice. The column was either washed with 50 CV soluble wash buffer 1, followed by 50 CV soluble wash buffer 2, or washed with just 50 CV soluble wash buffer 2, and bound protein was eluted in 10 CV soluble elution buffer. The presence of target protein was confirmed via SDS-PAGE, described in section 2.3, and this was dialysed for one hour in soluble dialysis buffer at RT. 2 mg of TEV protease was added to the sample and dialysis was continued in TEV protease cleavage buffer overnight at RT. Following TEV protease cleavage, the sample was batch-bound on 1 ml of nickel sepharose 6 fast flow resin (GE Healthcare), pre-equilibrated with TEV protease cleavage buffer, for one hour at RT. The mixture was applied to an empty column and the FT was run through the column twice. The column was washed with 20 ml TEV protease cleavage buffer and non-cleaved protein was eluted using 10 ml soluble elution buffer. The FT and wash were pooled and concentrated using a Vivaspin 20 concentrator (GE Healthcare), with a 10 kDa Mw cut-off for DdahB and MlghB and a 30 kDa Mw cut-off for the remaining proteins studied here, at 3,488 x *g* at 4 °C; the protein was resuspended every 10 minutes during the course of the concentration. Once concentrated to a final volume of 1 ml, the sample was centrifuged for 10 minutes at 20,817 x *g* and then applied to a hiload 16/60 superdex 200 pg column (GE Healthcare), using the Biorad NCG chromatography system, in soluble SEC buffer. The process of the purification of the target protein was

monitored via SDS-PAGE, as described in section 2.3. Fractions containing protein were pooled and concentrated using a Vivaspin 20 concentrator with the appropriate Mw cut-off (GE Healthcare), at 3,488 x *g* at 4 °C. The protein was resuspended every 10 minutes during the concentration process, until the required concentration was reached.

Non-TEV protease-cleavable proteins were purified in essentially the same way, with the following modifications: 50 CV soluble wash buffer 3 was used to wash the first nickel affinity column prior to the elution of the target protein. The TEV protease cleavage step was omitted, with the protein instead being concentrated for SEC immediately after elution from the first nickel affinity column.

SEC-MALS was conducted as detailed in section 2.6.

Protein identity and integrity was determined by mass spectrometry as detailed in section 2.7.

5.2.7 Sitting-drop crystallisation trials

The optimal protein concentration for crystallisation was determined using the PCT (Hampton Research), detailed in section 2.8. Sitting-drop vapour diffusion crystallisation trials were set up at RT as described in section 2.9. DdahB crystallisation trials were set up at 11 mg/ml and MlghB crystallisation trials were set up at 12 mg/ml. Non-TEV protease-cleavable MlghC and DdahC crystallisation trials were set up 6 mg/ml and 8.88 mg/ml, respectively. TEV protease-cleavable MlghC and DdahC crystallisation trials were both set up at 11 mg/ml. DmhA and DdahA crystallisation trials were set up at 10 mg/ml and 12 mg/ml, respectively.

5.2.8 Soaking of DdahB crystals with GDP-mannose

Soaking was conducted in collaboration with Dr Stephen McMahon, University of St Andrews. Crystals of DdahB, grown in JCSG-plus well D1 (24 % (w/v) PEG 1500, 20 % (w/v) glycerol) (Molecular Dimensions), were soaked overnight at 20 °C in mother liquor supplemented with 5 mM GDP-mannose (Sigma Aldrich). Crystals were flash frozen in liquid N_2, without cryoprotection, and screened in-house as described in section 2.10.

5.2.9 Co-crystallisation of MlghB with GDP-mannose

MlghB was concentrated to 12 mg/ml and incubated with 20 mM GDP-mannose (Sigma Aldrich) for 2 hours at RT. The mixture was centrifuged at 20,817 x *g* at 4 °C for 10 minutes prior to pellet any insoluble material prior to crystallisation trials. Crystals grown in in-house St Andrews Screen 1 well B9 (1.18 M sodium tartrate) (Molecular Dimensions) (appendix 8.4.1) were cryo-protected in mother liquor supplemented with 20 % (v/v) glycerol and 20 mM GDP-mannose (Sigma Aldrich). Crystals were flash frozen in liquid N_2 and screened in-house as described in section 2.10.

5.2.10 X-ray diffraction data collection and data processing

DmhA X-ray diffraction data were collected on Beamline I03 at Diamond Light Source and processed using the *autoPROC* pipeline (Vonrhein *et al*., 2011), which incorporates XDS (Kabsch, 2010), Pointless (Evans, 2006), Aimless (Evans and Murshudov, 2013) and STARANISO (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi). The DmhA crystal structure was solved using the Balbes automated molecular replacement pipeline (Long *et al*., 2008) as part of the CCP4 online suite (Winn *et al*., 2011), followed by ARP/wARP for model building (Langer *et al*., 2008). DmhA diffraction data were

analysed using Xtriage as part of the Phenix suite (Adams *et al*., 2010). DdahA X-ray diffraction data were collected on Beamline I024 at Diamond Light Source and processed using the xia2 DIALS pipeline (Gildea *et al*., 2014; Waterman *et al*., 2016).

DdahB and MlghB X-ray diffraction data were both collected on Beamline I04-1 at Diamond Light Source. DdahB X-ray diffraction data were processed using xia2 via the "-3d" pipeline (Winter, 2010), which uses XDS and XSCALE (Kabsch, 2010), and Pointless (Evans, 2006). MlghB X-ray diffraction data was processed using the xia2 DIALS pipeline (Winter, 2010), incorporating Pointless (Evans, 2006), DIALS (Gildea *et al*., 2014; Waterman *et al*., 2016) , and Aimless (Evans and Murshudov, 2013) . The DdahB crystal structure was solved using the Balbes automated molecular replacement pipeline (Long *et al*., 2008), incorporating Molrep (Vagin and Teplyakov, 1997) for molecular replacement, as part of the CCP4 online suite (Winn *et al*., 2011). Buccaneer was used for model building (Cowtan, 2006, 2008). The MlghB crystal structure was solved by molecular replacement using Phaser (Mccoy *et al*., 2007), as part of the CCP4 suite (Winn *et al*., 2011), using DdahB as the search model. The Matthews coefficient of the MlghB crystal was determined using the Matthews coefficient software (Kantardjieff and Rupp, 2003; Matthews, 1968), as part of the CCP4 suite (Winn *et al*., 2011). Automated model building for MlghB was conducted using Buccaneer-autobuild/refine (Cowtan, 2006, 2008) as part of the CCP4 suite (Winn *et al*., 2011).

MlghC X-ray diffraction data were collected on Beamline I04-1 at Diamond Light Source. Data were processed using xia2 pipeline "-3dii" (Winter, 2010), which incorporates XDS and XSCALE (Kabsch, 2010), followed by Scala to merge data (Evans, 2006). The MlghC crystal structure was solved using the Balbes automated molecular replacement pipeline (Long *et al*., 2008) as part of the CCP4 online suite (Winn *et al*., 2011), followed by ARP/wARP for model building (Langer *et al*., 2008).
DdahC X-ray diffraction data were collected in-house, as described in section 2.10. Data were processed using iMosflm (Battye *et al*., 2011), with Pointless used to confirm the space group (Evans, 2006) and Aimless used to scale and merge data (Evans and Murshudov, 2013). The DdahC crystal structure was solved using the Balbes automated

molecular replacement pipeline (Long *et al*., 2008) as part of the CCP4 online suite (Winn *et al*., 2011), followed by ARP/wARP for model building (Langer *et al*., 2008).

X-ray diffraction data for DdahB and MlghB crystals with GDP-mannose were collected in-house, as described in section 2.10. Data were processed using iMosflm (Battye *et al*., 2011), with Pointless used to confirm the space group (Evans, 2006) and Aimless to scale and merge data (Evans and Murshudov, 2013). DdahB and MlghB crystal structures in complex with GDP-mannose were solved by molecular replacement using Phaser (Mccoy *et al*., 2007), as part of the CCP4 suite (Winn *et al*., 2011), using their respective apo structures as search models. Their respective Matthews coefficients were determined using the Matthews coefficient software (Kantardjieff and Rupp, 2003; Matthews, 1968), as part of the CCP4 suite (Winn *et al*., 2011).

All models were completed by manual building in Coot (Emsley *et al*., 2010), followed by refinement conducted using REFMAC5 (Murshudov *et al*., 1997), as part of the CCP4 suite (Winn *et al*., 2011). Appropriate TLS restraints for refinement were determined using the TLSMD server (Painter and Merritt, 2006a, 2006b) and used in refinement using REFMAC5 (Murshudov *et al*., 1997). Models were optimised using the PDBredo server (Joosten *et al*., 2014). The final version of each structure was validated using the MolProbity server (Chen *et al*., 2010; Davis *et al*., 2007) . The native oligomeric state, interfaces and assembly of all structures were determined using PDBePISA ('Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute. (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)) (Krissinel, 2009; Krissinel and Henrick, 2007).

Protein structural alignments were performed using PDBeFold (Protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm)) (Krissinel and Henrick, 2004).

5.2.11 ITC using wild type and active site mutants

Protein: ligand binding was studied using a MicroCal PEAQ-ITC (Malvern). Protein samples were dialysed overnight at 4 °C in soluble SEC buffer prior to the experiment. Sample loading and washing of the cell with soluble SEC buffer was conducted at 20 °C and all experimental runs were conducted at 25 °C with the following injection sequence: 1 x 0.4 µl injection, followed by 18 x 2 µl injections. The following experimental settings were used: reference power (µcal/s) = 5.00; feedback = high; stir speed (rpm) = 750; initial delay (s) = 60; injection spacing (s) = 150; injection duration (s) = 4.0. The concentrations of protein and ligand used were as follows: 16.7 µM DdahB plus 4.13 mM GDP (Sigma Aldrich); 59.5 µM DdahB plus 8.57 mM GDP-mannose (Sigma Aldrich); 59.2 µM MlghB plus 4.36 mM GDP (Sigma Aldrich); 59.2 µM MlghB plus 8.57 mM GDP-mannose (Sigma Aldrich). The protein was in the cell and the ligand in the syringe. The raw data were corrected for dilution effects by removing the background signal generated upon injection of the ligand into soluble SEC buffer using the same experimental parameters. GDP and GDP-mannose (Sigma Aldrich) solutions were prepared in soluble SEC buffer and the concentrations were determined using a Nanodrop spectrophotometer using ε_{GDP} 13800 $M^{-1}cm^{-1}$ at 254 nm.

5.3 Results

5.3.1 Expression and purification of the dehydratases DmhA and DdahA

The first step in GDP-6-deoxy-D-*altro*-heptose biosynthesis in *C. jejuni* strain 81-176 is the dehydration of GDP-D-*manno*-heptose at the C4 and C6 positions to give GDP-4-keto-6-deoxy-D-*lyxo*-heptose, catalysed by DdahA (McCallum *et al*., 2011). This step also occurs as the first step of 6-deoxyheptose biosynthesis in *Y. pseudotuberculosis* strain O:2a and is catalysed by DmhA (Butty *et al*., 2009; Ho *et al*., 2008; Pacinelli *et al*., 2002). In contrast, the first step in GDP-6-*O*-methyl-L-*gluco*-heptose biosynthesis in *C. jejuni* strain NCTC 11168 is predicted to be the MlghA-catalysed oxidation of GDP-D-*manno*-heptose at the C4 position to give GDP-4-keto-D-*lyxo*-heptose (McCallum *et al*., 2013).

N-terminally His_6-tagged DmhA and DdahA are both 40 kDa in Mw and could both be expressed, as confirmed via whole-cell western blotting against the N-terminal His_6 affinity tag, shown in figure 5.8. The presence of the target proteins in the eluate of a small-scale nickel affinity purification, performed using a Biosprint, confirmed that the target proteins were expressed in a soluble form. The SDS-PAGE of the Biosprint eluate samples can be seen in figure 5.9.

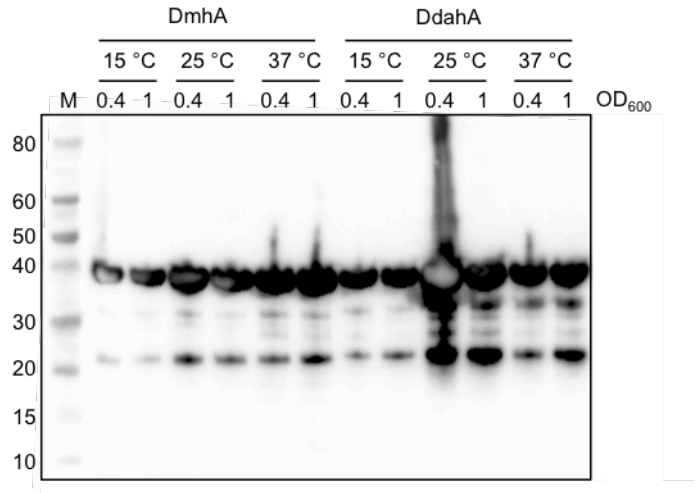


Figure 5.8. Western blot, probed with anti-His tag antibody, of DmhA and DdahA whole cell expression trial samples. Both DmhA and DdahA can be expressed over a range of conditions, with the bands at approximately 40 kDa corresponding to DdahA and DmhA.

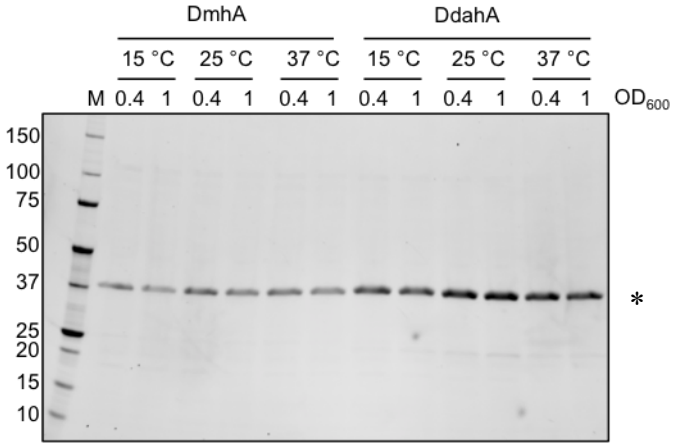


Figure 5.9. SDS-PAGE gel of small-scale purification trials of DmhA and DdahA. Both DdahA and DmhA could be expressed in a soluble form using a range of expression conditions. The bands at approximately 40 kDa, indicated with a "*", correspond to DmhA and DdahA.

Both DmhA and DdahA could be purified in a non-aggregated and homogeneous form (figure 5.10). The identity of these purified proteins was confirmed by mass spectrometry. Both DdahA and DmhA were found to be tetramers in solution, as confirmed via SEC-MALS. SEC-MALS profiles are shown in figure 5.11 and the results in table 5.2. Purification yields of approximately 30 mg/L of *E. coli* cell culture and 20 mg/L of *E. coli* cell culture were obtained for DmhA and DdahA, respectively.

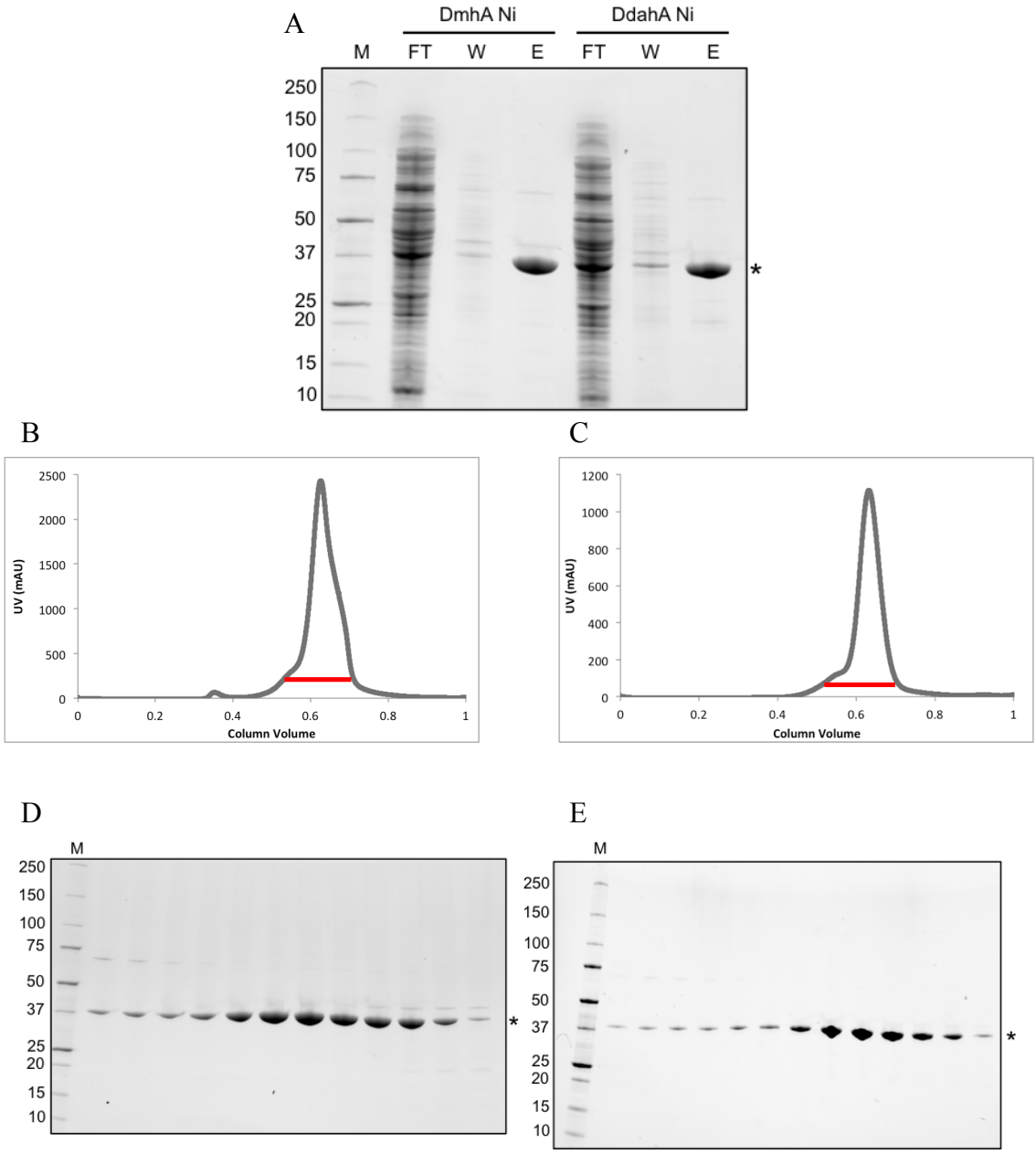


Figure 5.10. The purification of DmhA and DdahA. The SDS-PAGE gel of the nickel affinity purification of DmhA and DdahA can be seen in "A" (M: Marker; FT: FT; W: Wash; E: Elution). The band corresponding to the target proteins is at approximately 40 kDa and is indicated with a "*". "B" and "D" show the DmhA SEC UV chromatogram and SDS-PAGE gel of SEC peak fractions, respectively. "C" and "E" show the DdahA SEC UV chromatogram and SDS-PAGE gel of SEC peak fractions, respectively. The void volume of the SEC column is approximately 0.36 CV. The approximate position of the fractions in "B" and "C" shown in "D" and "E", respectively, is indicated with a solid red line.

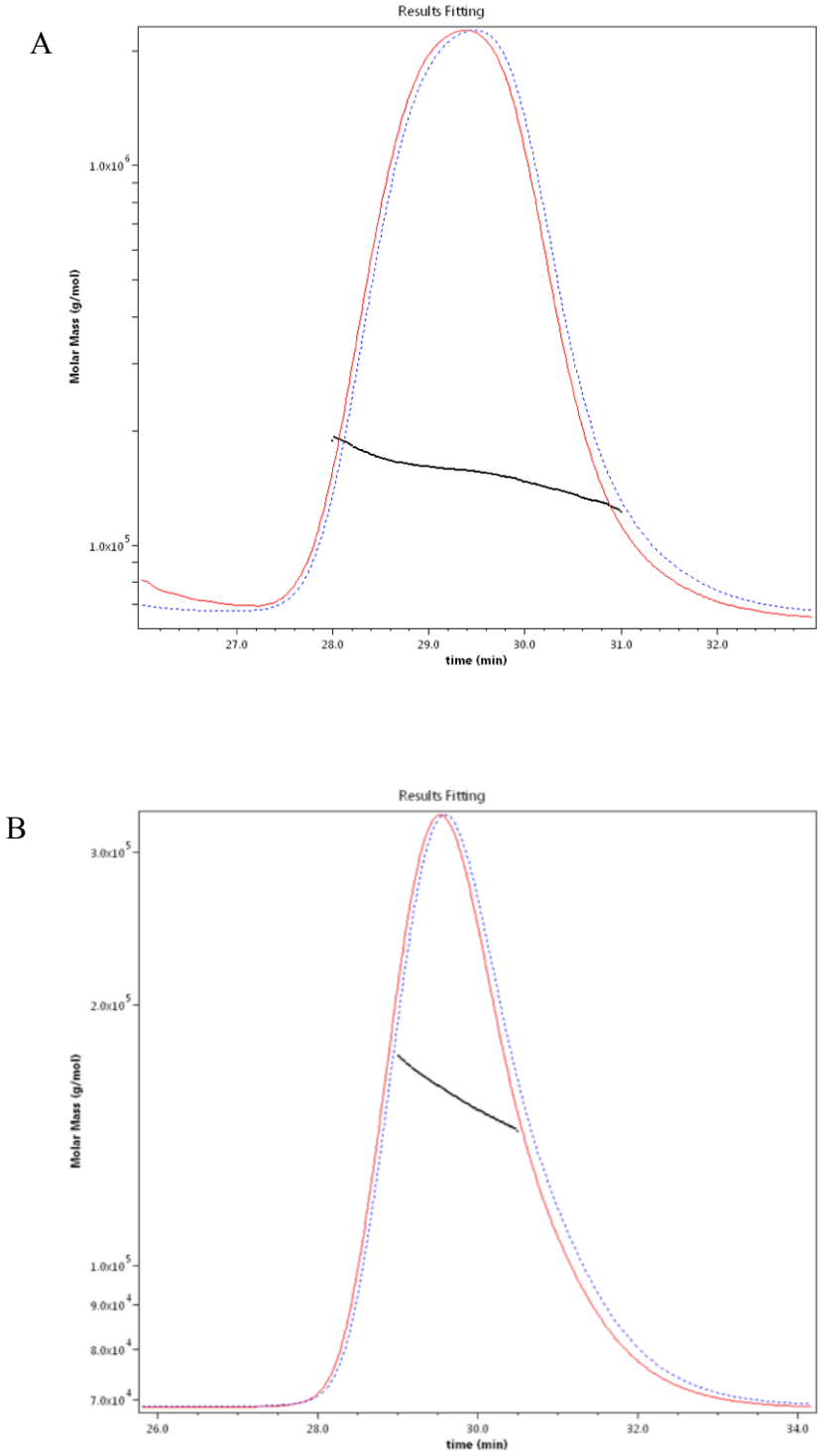


Figure 5.11. SEC-MALS profiles of DmhA and DdahA. The SEC-MALS profile of DmhA is shown in "A" and that of DdahA is shown in "B". The change in light scattering over time is shown in red and the change in refractive index over time is shown in blue. The black line indicates the change in molar mass (g/mol) over time.

	DmhA	DdahA
Molar mass moments (g/mol)		
Mn	1.5 x 10^5 (± 0.5 %)	1.6 x 10^5 (± 0.4 %)
Mw	1.6 x 10^5 (± 0.5 %)	1.7 x 10^5 (± 0.4 %)
Polydispersity		
Mw/Mn	1.0 (± 0.7 %)	1.0 (± 0.5 %)

Table 5.2. DmhA and DdahA SEC-MALS results. SEC-MALS shows that both DmhA and DdahA are both monodisperse tetramers in solution. Mn is defined as: $\Sigma N_i M_i^2 / \Sigma N_i M_i$, where M_i corresponds to the molecular weight of the protein and N_i corresponds to the number of proteins of that molecular weight. Mw is defined as: $\Sigma N_i M_i / \Sigma N_i$. Polydispersity is the range of molecular weights occupied by the sample and is defined as Mw/Mn, with 1.0 corresponding to a monodisperse sample. The higher the polydispersity value, the wider the range of molecular weights occupied by a sample.

5.3.2 Crystallisation of DmhA and DdahA

DmhA crystallised in multiple conditions, with the crystals used for X-ray diffraction data collection appearing after 15 minutes of incubation at 20 °C, in JCSG-plus well G12 (0.1 M Bis-tris pH 5.5, 3 M NaCl) (Molecular Dimensions), when using a drop volume ratio of 1 protein: 2 mother liquor. The crystals reached a full size after one day and can be seen, along with a representative X-ray diffraction pattern, in figure 5.12.

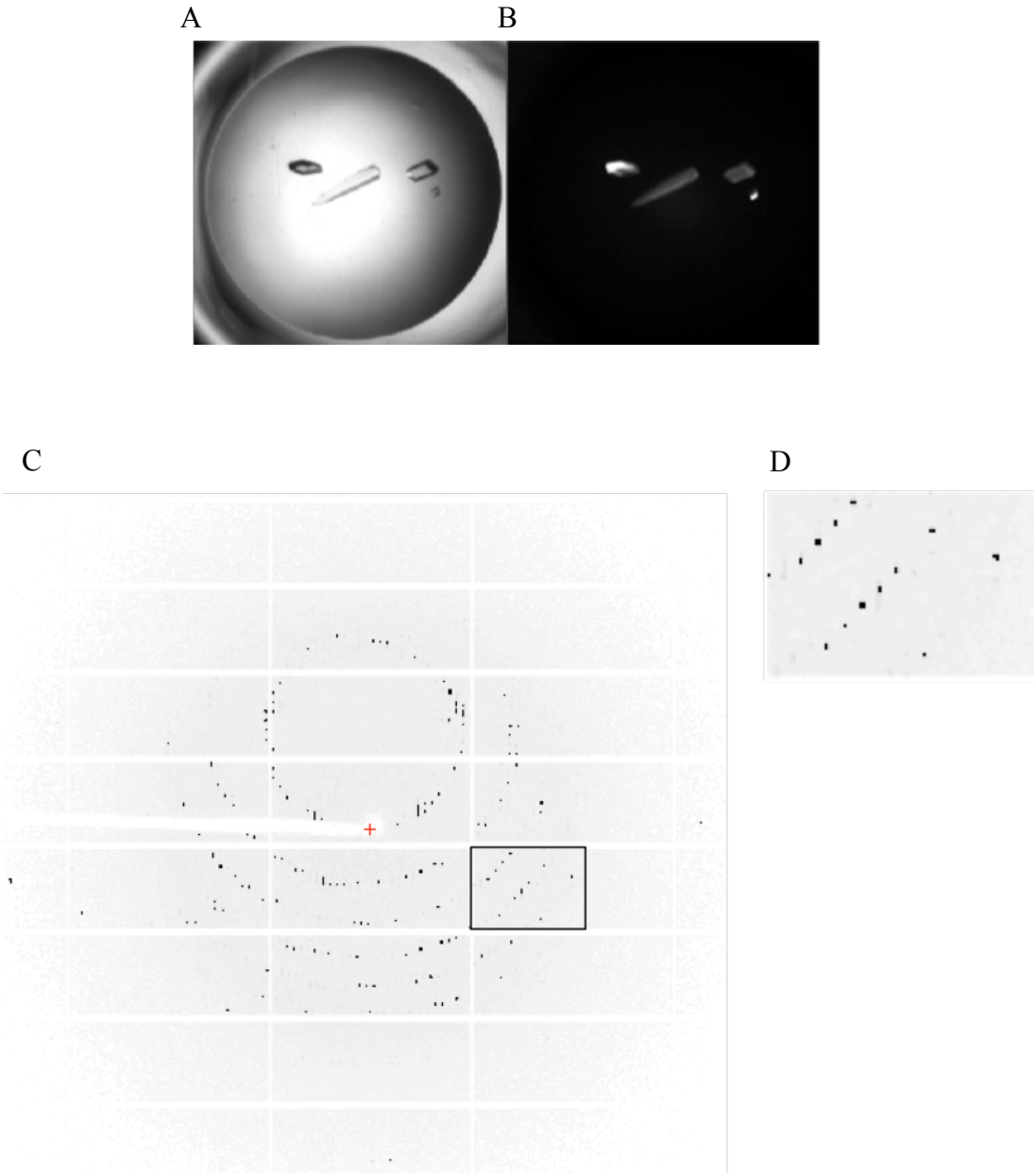


Figure 5.12. DmhA crystals and X-ray diffraction pattern. These crystals were obtained in JCSG-plus well G12 (Molecular Dimensions). The normal image is shown in "A" and the UV image is shown in "B". These images were taken using a Minstrel HT-UV imaging system. A representative X-ray diffraction pattern is shown in "C", with a zoomed view of the area highlighted with a square in "C" shown in "D"

The crystallisation of DdahA proved more problematic. DdahA crystallised in our in-house St Andrews PEG 1 screen well E5 (44.73% PEG 400, 0.1 MOPS pH 6.5, 0.11 M ammonium tartrate) (Molecular Dimensions) (appendix 8.4.5) in a drop volume ratio of 2 protein: 1 mother liquor. These crystals appeared following only one day of incubation at 20 °C, however they took a total of 26 days to reach a full size of approximately 200 x 30 μm, and can be seen in figure 5.13. A similar pattern was observed for well B1 of our in-house St Andrews 2 screen (54.43% PEG 400, 0.1 M sodium cacodylate pH 7.0, 0.09 M lithium sulfate) (Molecular Dimensions). Crystals also appeared in this condition following one day of incubation at 20 °C, but took 12 days to reach a full size of approximately 140 x 60 μm. These crystals diffracted less strongly than those of DmhA, with the best diffraction of 2.93 Å resolution obtained for the crystals in St Andrews PEG 1 well E5, using Beamline I024 at Diamond Light Source, shown in figure 5.13.

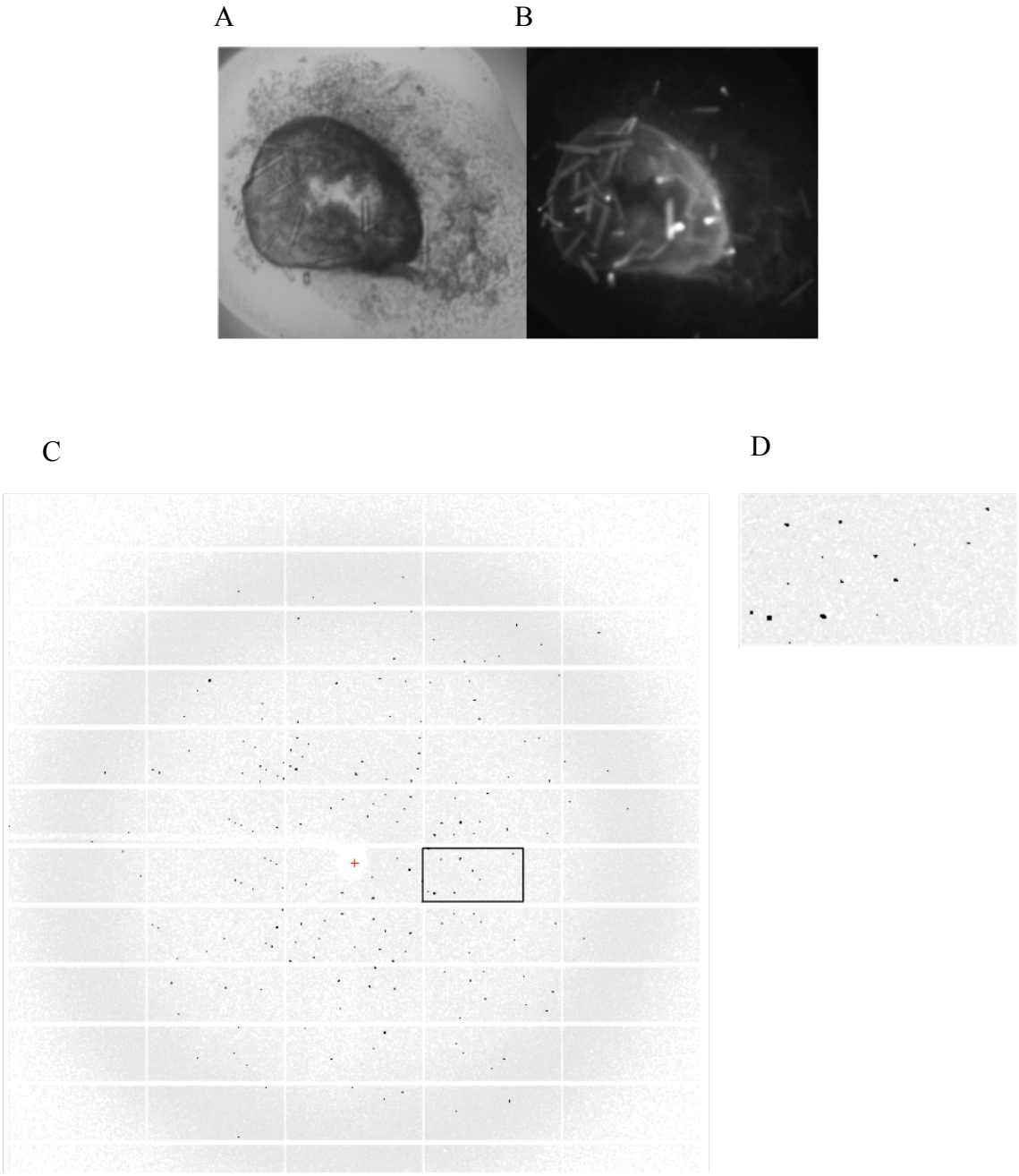


Figure 5.13. DdahA crystals and representative diffraction pattern. These crystals were obtained in our in-house St Andrews PEG 1 screen well E5 (Molecular Dimensions). The normal image is shown in "A" and the UV image is shown in "B". Images were taken using a Minstrel HT-UV imaging system. A representative X-ray diffraction pattern is shown in "C", with a zoomed view of the area highlighted with a square in "C" shown in "D".

5.3.3 Crystal structure of DmhA

The GDP-D-*manno*-heptose 4,6-dehydratase DmhA crystallised in space group C 2 2 2_1, with two chains in the asymmetric unit and diffracted to 2.0 Å resolution. Out of a possible 353 amino acids, the refined model of chain A consists of amino acids 2-136, 145-285 and 303-342 and that of chain B consists of 2-137, 146-288, 292-296 and 300-342. Amino acid 1 corresponds to the N-terminal Met of the target sequence. This and the N-terminal His_6 tag with linker, denoted amino acids -12 to 0, are absent from this structure and therefore presumed disordered. Table 5.3 shows the data collection and refinement statistics. The MolProbity validation report can be seen in figure 5.14 (Chen *et al*., 2010; Davis *et al*., 2007). The relatively high R-factors, shown in table 5.3, are due to the presence of pseudotranslational symmetry, as detected using Xtriage as part of the Phenix suite (Adams *et al*., 2010).

Data Collection	
Wavelength (Å)	0.9763
Space Group	C 2 2 2_1
Unit-cell a, b, c (Å)	61.772, 181.88, 136.27
Unit-cell α, β, γ (°)	90, 90, 90
Resolution range (Å)	136.27 – 2.00 (2.03 – 2.00)
I/σ(I)	10.8 (2.1)
R_{merge} (%)	14.0 (124.6)
$CC_{1/2}$	0.999 (0.873)
Completeness (%)	95.6 (96.6)
Multiplicity	14.7 (14.0)
Refinement	
R-factor (%)	25.8
R_{free} (%)	28.8
R.m.s.d. Bond Lengths (Å)	0.011
R.m.s.d. Bond Angles (°)	1.451
Mean B factor (Å^2)	23.4

Table 5.3. DmhA X-ray diffraction data collection and refinement statistics. Values for the highest resolution shell are shown in brackets. R_{merge} is defined as: $\Sigma_{hkl} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity of reflection hkl. The R-factor is defined as: $\Sigma_{hkl} \| F_{obs} | - | F_{calc} \| / \Sigma_{hkl} | F_{obs} |$, where F_{obs} is the observed structure-factor amplitudes and F_{calc} is the calculated structure-factor amplitudes. R_{free} was calculated, using the R-factor equation, using a 5 % subset of reflections that were not used in refinement.

All-Atom Contacts	Clashscore, all atoms:	3.1		99th percentile* (N=715, 2.00Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	0.36%	Goal: <0.3%
	Favored rotamers	547	97.50%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	614	97.62%	Goal: >98%
	MolProbity score^	1.18		100th percentile* (N=12522, 2.00Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 5237	0.00%	Goal: 0%
	Bad angles:	2 / 7086	0.03%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 23	0.00%	Expected: ≤1 per chain, or ≤5%

Figure 5.14. DmhA MolProbity validation report. The Clashscore and MolProbity score place this structure in the 99th and 100th percentile, respectively, when comparing it to structures of a similar resolution (Chen *et al*., 2010; Davis *et al*., 2007).

DmhA purified as a tetramer, as confirmed by SEC-MALS, shown in figure 5.11 and table 5.2. This oligomeric state was also confirmed using PDBePISA and the physiologically relevant tetramer, constructed using PDBePISA, can be seen in figure 5.15 (Krissinel, 2009; Krissinel and Henrick, 2007).

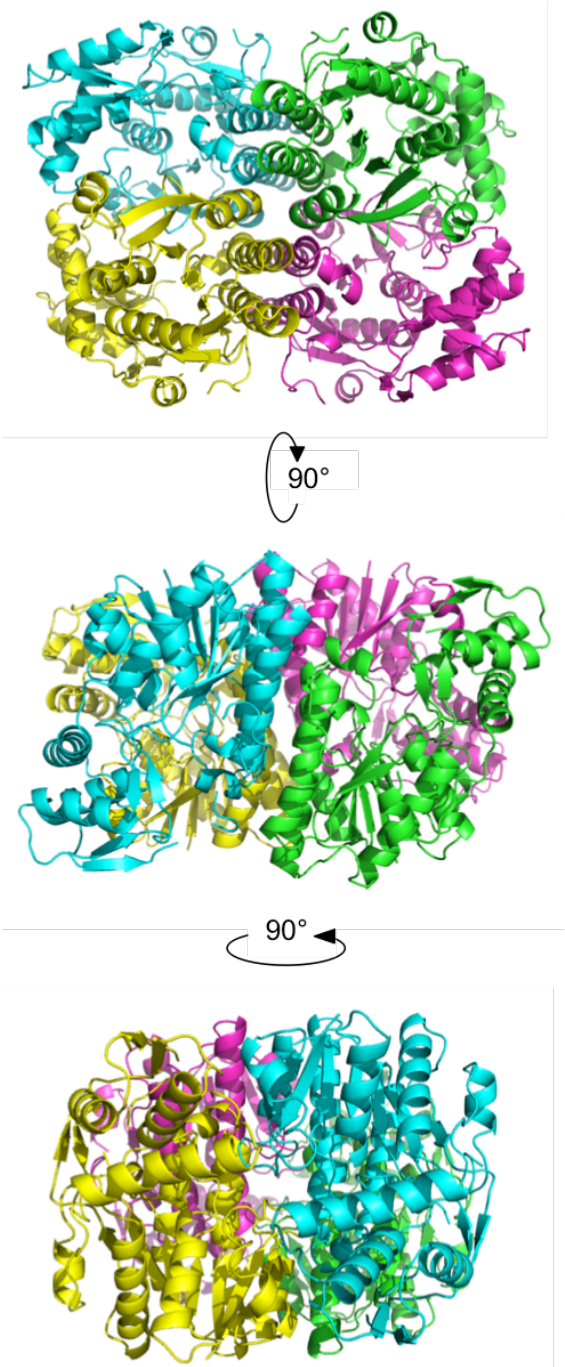


Figure 5.15. DmhA is a tetramer. Three views of the DmhA tetramer, with each monomer shown in a different colour. This biological assembly was determined using the PDBePISA server (Krissinel, 2009; Krissinel and Henrick, 2007). This image was constructed using MacPyMOLEdu.

A DmhA monomer consists of a total of 11 α-helices and 11 β-strands and can be divided into an N- and C-terminal domain. The N-terminal domain forms a Rossmann fold (Hanukoglu, 2015) consisting of a central β-sheet, made up of seven parallel β-strands in the following order: 3-2-1-4-5-6-7, surrounded by eight α-helices. This domain contains a number of conserved motifs, including an YXXXK motif (Kavanagh *et al*., 2008), consisting of Tyr 157 and Lys 161, and a GXXGXXG motif (Kavanagh *et al*., 2008), consisting of Gly 8, Gly 11 and Gly 14, where X is any amino acid. Three α-helices and four β-strands form the globular C-terminal domain. These properties are characteristic of the short chain reductase/dehydrogenase (SDR) family (Kavanagh *et al*., 2008).

5.3.4 Oligomerisation of DmhA

Unusually for SDR family members, DmhA forms a tetramer. Only a few examples of the tetramerisation of SDR family members have been reported in the literature and they include the GDP-D-mannose 4,6-dehydratase MUR1, from *A. thaliana*, required for the conversion of GDP-D-mannose to GDP-4-keto-6-deoxy-mannose during GDP-L-fucose biosynthesis (Mulichak *et al*., 2002), and its homologue GMD, from *P. aeruginosa*, which catalyses the same step during D-rhamnose biosynthesis (Webb *et al*., 2004). DmhA forms a tetramer via forming a "dimer of dimers", with the dimerization interface formed by a four helical bundle, shown in figure 5.16, with each monomer contributing two α-helices, which also form part of the N-terminal Rossmann fold (Hanukoglu, 2015). This helical bundle is formed by helices four and six and the interactions are largely hydrophobic in nature, with hydrogen bonds forming between His 152 of one monomer and Glu 173 of the other, and between Gln 101 of one monomer and Gln 101 of the other. The N-terminal Rossmann fold binds NAD^+ (Hanukoglu, 2015), which can be seen in our crystal structure, and the nature of the NAD^+ binding site will be discussed in more detail in section 5.3.6. As is the case with both MUR1 (Mulichak *et al*., 2002) and GMD (Webb *et al*., 2004), the dimer-dimer interface of DmhA forms via this NAD^+ binding domain, resulting in a distance of 9 Å between the adenine rings of the NAD^+ bound to adjacent monomers. Upon

dimerization of the DmhA dimer, a loop consisting of amino acids 34-41, located between β-strand two and α-helix two, folds over the neighbouring cofactor binding site and not only contributes to the dimerization interface via hydrogen bonds between Thr 46 and Ile 338, both Trp 35 and Tyr 59 with Arg 342 and both Asn 341 and Pro 339 with Asn 50, but also allows the formation of a hydrogen bond between Gln 36 and the pyrophosphate moiety of the neighbouring NAD^+. This tetramerisation interface, highlighting the hydrogen bond between Gln 36 and the NAD^+ of the neighbouring monomer can be seen in figure 5.17.

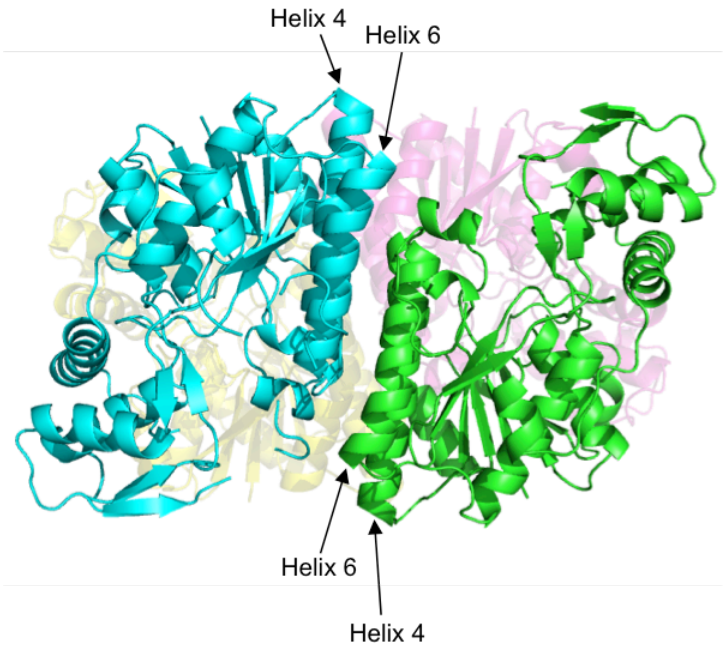


Figure 5.16. DmhA is a "dimer of dimers". DmhA forms a "dimer of dimers", with each dimer formed by the central helical bundle, typical of the SDR family (Kavanagh *et al*., 2008). A dimer is highlighted here, with one monomer in green and the other in blue. This image was created using MacPyMOLEdu.

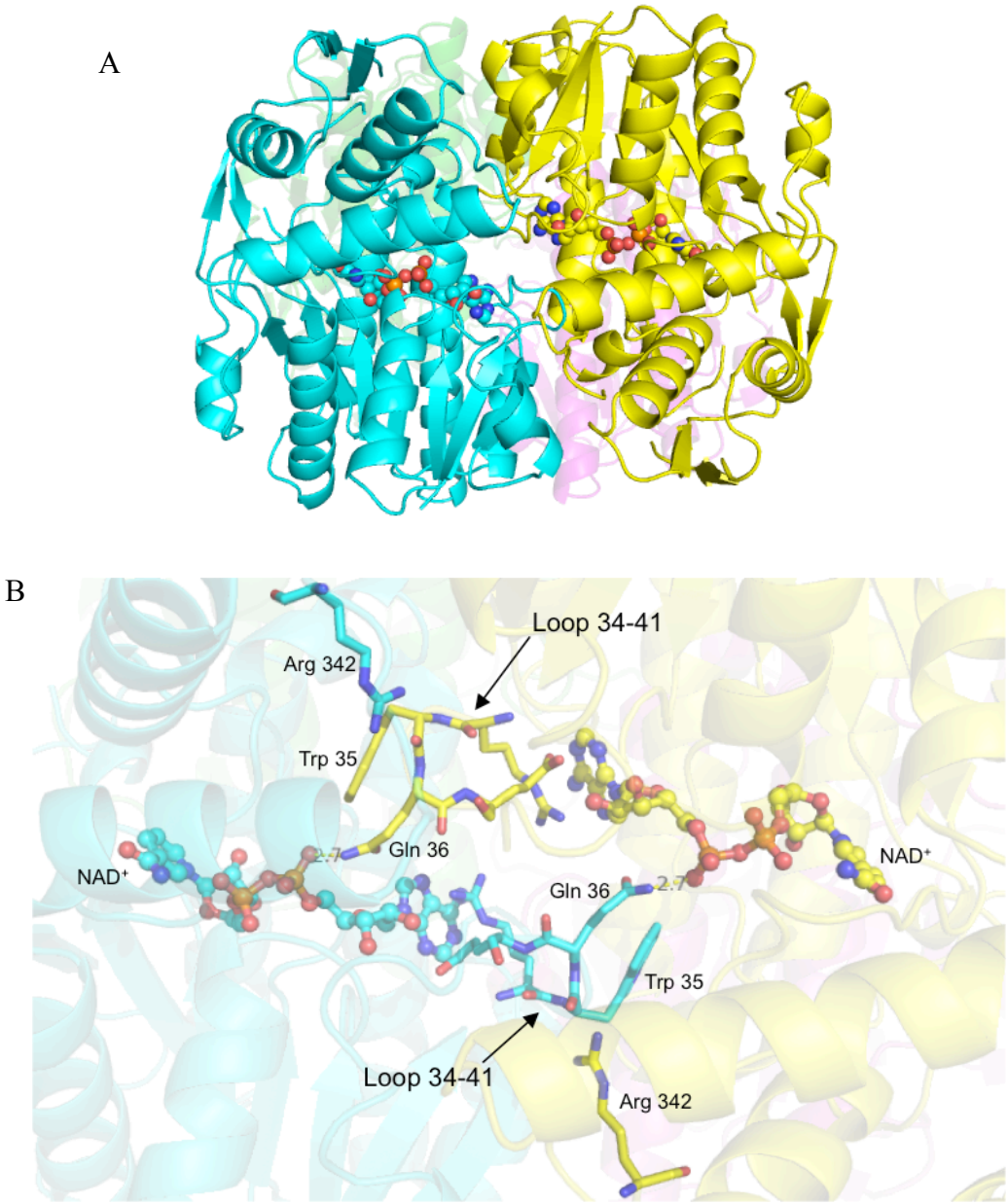


Figure 5.17. DmhA tetramerises via interactions between one monomer and the NAD$^+$ cofactor of the neighbouring monomer. "A" shows the tetramerisation interface between the two dimers and the position of the NAD$^+$ binding site in each monomer. "B" shows the mode of tetramerisation. Upon tetramerisation, loop 34-41 from one monomer folds over the cofactor binding site of the adjacent monomer. This stabilises the tetramer by forming a hydrogen bond between Gln 36, part of loop 34-41, of one monomer and the α-phosphate of the NAD$^+$ of the adjacent monomer. This image was created using MacPyMOLEdu.

5.3.5 A comparison of the structure of DmhA with the structure of the dTDP-D-glucose 4, 6-dehydratase RmlB

Like DmhA, the dTDP-D-glucose 4,6-dehydratase RmlB is a member of the SDR family and their structures can be aligned over the C_α backbone of 297 residues with a root mean square deviation (RMSD) of 1.72 Å for the NAD^+- and dTDP-D-glucose-bound structure of RmlB (from *S.* Typhimurium strain LT2) (PDB: 1KEU) (Allard *et al*., 2002). RmlB and DmhA share a sequence identity of 27 %, shown in figure 5.5 (Krissinel and Henrick, 2004). Both of these enzymes achieve the dehydration of their respective substrates in an NAD^+-dependent manner (Allard *et al*. 2001, 2002). However, as RmlB only forms a homodimer (Allard *et al*., 2001, 2002), they will be compared in the context of the monomer.

5.3.6 Cofactor binding to DmhA and RmlB

Despite not including NAD^+ in the purification buffers, NAD^+ could be modelled into the electron density in both chain. The presence of NAD^+ in the DmhA structure correlates with the previous finding that DmhA copurifies with NAD^+, as confirmed via mass spectrometry, and does not require the addition of this cofactor to catalyse the dehydration of its substrate GDP-D-*manno*-heptose *in vitro* (Butty *et al*., 2009). As expected, the NAD^+ binding site is located in the N-terminal Rossmann fold domain (Hanukoglu, 2015), and could be identified in the F_o-F_c density, contoured at 1.3 σ (Figure 5.18). The N-terminal Rossmann fold of both DmhA and RmlB are formed of seven parallel β-strands and eight α-helices and an overlay of a DmhA monomer and an RmlB monomer is shown in figure 5.19. As also observed upon NAD^+ binding to RmlB, DmhA binds to NAD^+ with the nicotinamide moiety in the *syn* conformation, with a hydrogen-bond between the amide group of the nicotinamide and the α-phosphate (distance: 3.0 Å). In addition to the previously discussed hydrogen bond between Gln 36 and its neighbouring NAD^+, the residues required for the coordination of NAD^+ by DmhA can be seen in figure 5.20. The conserved GXXGXXG and YXXXK motifs are required for NAD^+ binding. Gly 8 and Gly 14 are required for

hydrogen bonding to the pyrophosphate moiety via a bridging water molecule. Tyr 157 and Lys 161 form hydrogen bonds with the 2'OH and 3'OH of the nicotinamide ribose moiety, respectively. RmlB possesses a variation of the GXXGXXG motif, known as the GXGXXG motif, consisting of Gly 8, Gly 10 and Gly 13 (Allard *et al*., 2001). The YXXXK motif in RmlB consists of Tyr 167 and Lys 171 (Allard *et al*., 2001). An overlay of the NAD^+ binding sites of DmhA and RmlB can be seen in figure 5.20.

In addition to forming hydrophobic interactions with many residues, shown in figure 5.20, the nicotinamide moiety of NAD^+ bound to DmhA hydrogen bonds with His 187 and Gln 12. Cys 130 forms a hydrogen bond with the nicotinamide ribose moiety. Gln 12, Val 13, Ala 85, Ser 87 and Ser 193 form hydrogen bonds with the pyrophosphate moiety (Figure 5.20). Little conservation is observed between these residues and those of RmlB, with the exception of Val 13, which corresponds to Ile 12 of RmlB. The ribose of the adenosine moiety forms hydrogen bonds with Thr 10, Gly 11, Met 32, Arg 34 and Leu 62, and again little conservation is observed between these residues and those required for the interaction of RmlB with NAD^+ (figure 5.21) (Allard *et al*., 2001).

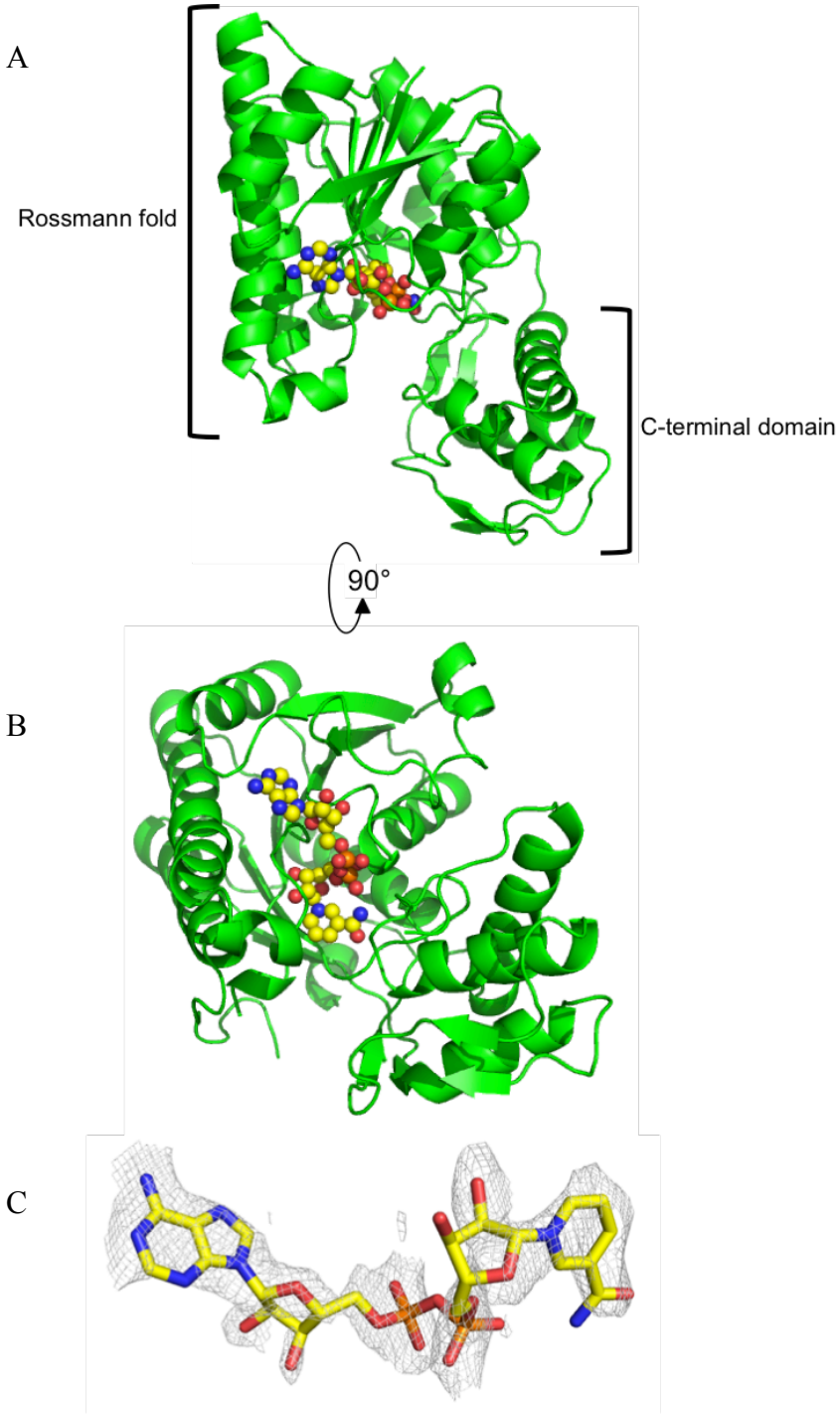


Figure 5.18. NAD$^+$ binds DmhA at the base of the N-terminal Rossmann fold domain. Views of the NAD$^+$ binding site can be seen in "A" and "B", with NAD$^+$ in yellow. "C" shows the F_o-F_c density for one of the NAD$^+$ molecules contoured at 1.3 σ. The F_O-F_C density was calculated by deleting NAD$^+$ from the final refined model, resetting the B factors to 20 Å^2 using the TLSMD server (Painter and Merritt, 2006a, 2006b), refining the model in REFMAC5 (Murshudov *et al*., 1997), with 30 cycles of refinement. The map was created using FFT, as part of the CCP4 suite (Winn *et al*., 2011). This image was created using MacPyMOLEdu.

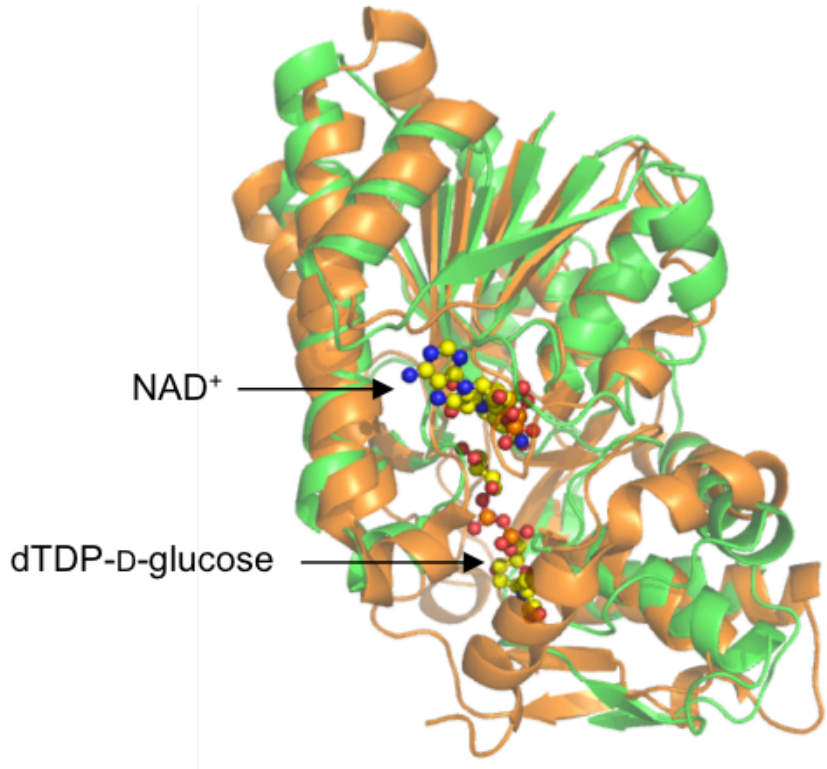


Figure 5.19. An overlay of the DmhA monomer and dTDP-D-glucose-bound RmlB monomer. The substrate binding site is located in the cleft between the two domains. RmlB is shown in orange, DmhA in green and both NAD^+ and dTDP-D-glucose in yellow (PDB: 1KEU) (Allard *et al*., 2002). This image was created using MacPyMOLEdu.

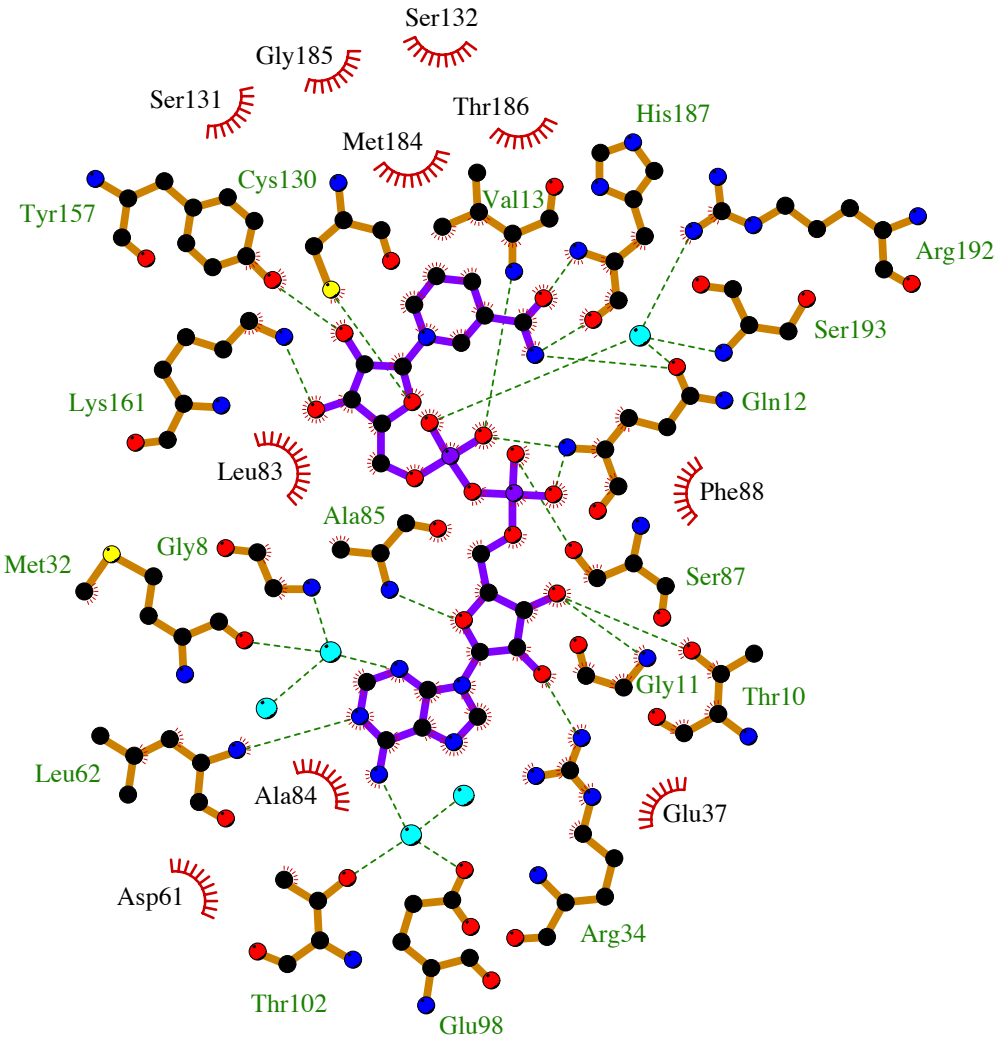


Figure 5.20. Residues required for the coordination of NAD^+ by DmhA. The GXXGXXG and YXXXK motifs are required for the interaction of SDR family members with NAD(P)(H) (Kavanagh et al., 2008). These motifs are conserved in DmhA and correspond to Gly 8, Gly 11 and Gly 14, and Tyr 157 and Lys 161. This figure was created using LigPlot$^+$ (Laskowski and Swindells, 2011). Waters are shown in cyan, carbon is shown in black, nitrogen is shown in blue, oxygen is shown in red, sulfur is shown in yellow and phosphorus is shown in purple. Bonds within the ligand are shown in purple and bonds within the protein are shown in brown. Hydrogen bonds and names of the participating amino acids are shown in green. Hydrophobic interactions are shown in red and the names of the participating amino acids are shown in black.

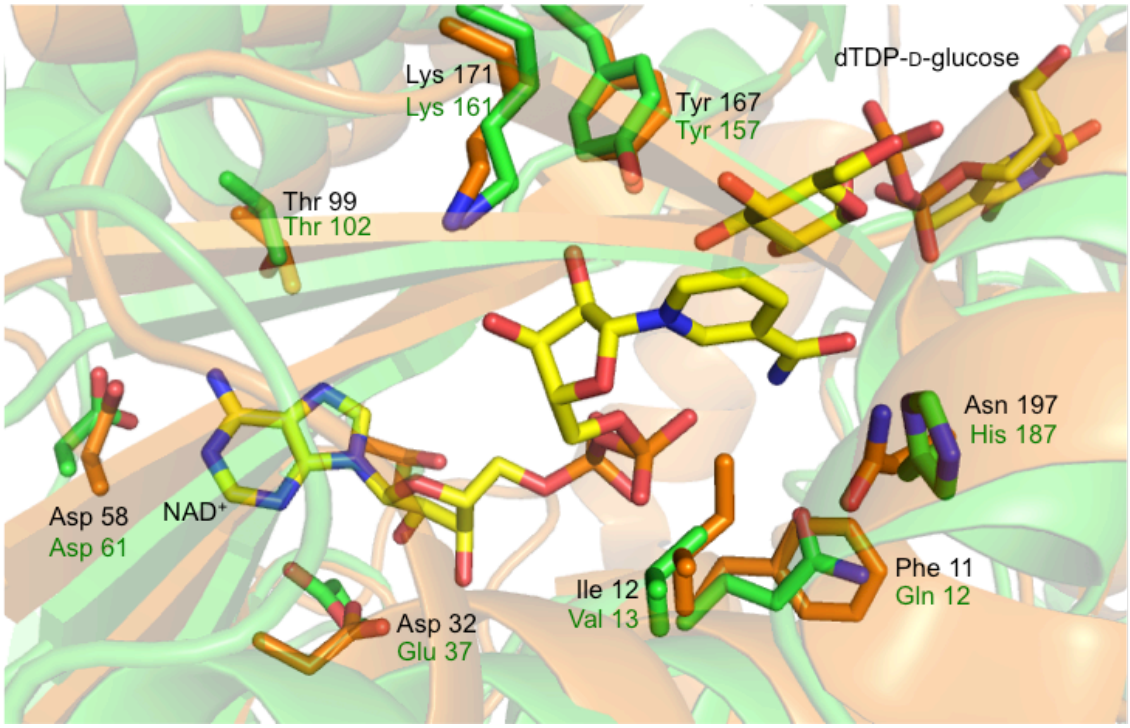


Figure 5.21. A comparison of the NAD(H) binding sites of RmlB and DmhA. DmhA is shown in green and RmlB is shown in orange. Residues required for NAD^+ binding to RmlB that are conserved in DmhA are shown here, with the residue name in black corresponding to RmlB and that in green to DmhA (RmlB PDB: 1KEU) (Allard *et al*., 2002). This image was created using MacPyMOLEdu.

5.3.7 Mode of distinction between NAD(H) and NADP(H)

The specificity of the RmlB-NAD^+ interaction is proposed to result from bidentate hydrogen bonding between the side chain of Asp 32 and the 2'OH and 3'OH groups of the adenosine ribose (Allard *et al*., 2001). This interaction between an acidic amino acid located approximately 20 amino acids C-terminal to the GXXGXXG motif and the hydroxyl groups of the adenosine ribose is conserved throughout the SDR family members that bind NAD(H) (Kavanagh *et al*., 2008). This Asp residue corresponds to Glu 37 in DmhA, and can be seen in figure 5.20 and 5.21.

5.3.8 Locating the active site of DmhA

The alignment of DmhA with the dTDP-D-glucose-bound structure of RmlB (Allard *et al*., 2002) allows the identification of the substrate binding site of DmhA and also possible active site residues. As shown in figure 5.19, the substrate binding site of RmlB is located in the cleft between the N- and C-terminal domains, with the dTDP moiety located in the C-terminal domain of RmlB and the glucose moiety oriented towards the NAD^+ (Allard *et al*., 2002).

Most of the structural variation between DmhA and RmlB is found in this C-terminal substrate-binding domain, as would be expected as they show different substrate specificities (Allard *et al*., 2001; Butty *et al*., 2009). This is demonstrated by the RMSD values for the alignment of the C_α backbone of 1.5 Å and 2.05 Å for the N-terminal domain and C-terminal domain, respectively. This variation is also seen at the primary structure level, as shown in figure 5.22. Arg 231 of RmlB, required for hydrogen bonding to the pyrophosphate moiety of dTDP-D-glucose (Allard *et al*., 2002), is conserved in DmhA and corresponds to Arg 226, which is to be expected as the pyrophosphate moiety is present in both GDP-D-*manno*-heptose and dTDP-D-glucose. Upon dTDP-D-glucose binding to RmlB, residues Glu 290 to His 300 fold over the substrate binding site, enabling hydrogen bond formation between His 300 and the 3'OH of the ribose of the dTDP (Allard *et al*., 2002). Residues 286-302 of DmhA are absent from the NAD^+-bound structure and therefore presumed to be disordered in the absence of substrate.

The active site residues identified in RmlB as Asp134, Glu 135, Tyr 167 and Lys 171 are conserved in DmhA, and correspond to Glu 134, Tyr 157 and Lys 161. The exception is Asp 134, which aligns with Ser 133 in DmhA. Asp 134, Glu 135 and Tyr 167 are all required for hydrogen bonding with the glucose moiety of dTDP-D-glucose (Allard *et al*., 2002). The overlay of the active site residues of RmlB, with those proposed for DmhA, can be seen in figure 5.22.

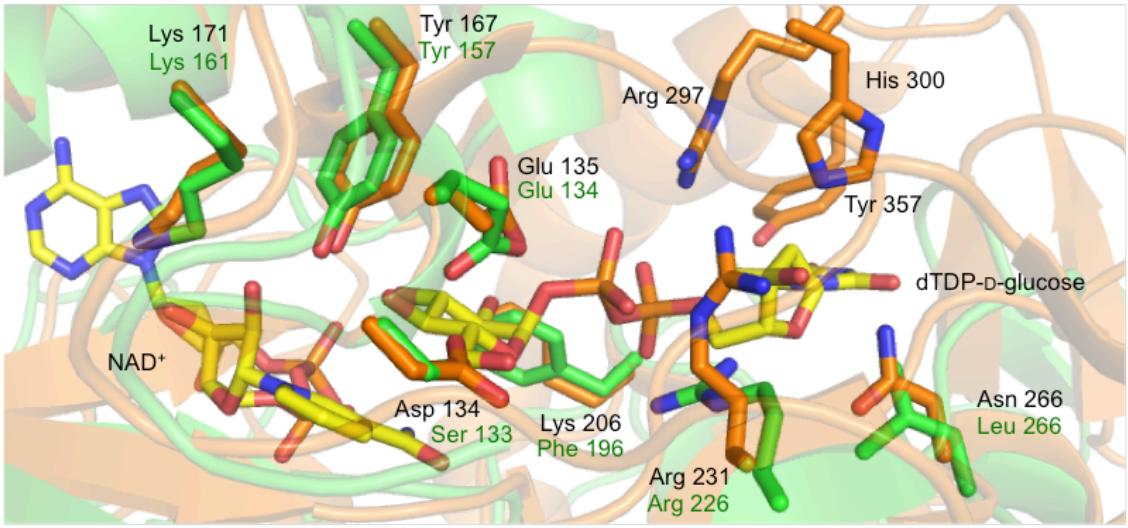


Figure 5.22. A comparison of the substrate binding sites of RmlB and DmhA. DmhA is shown in green and RmlB is shown in orange. Residues required for substrate binding to RmlB that are conserved in DmhA are shown here, with the residue name in black corresponding to RmlB and that in green to DmhA (RmlB PDB: 1KEU) (Allard *et al*., 2002). This image was created using MacPyMOLEdu.

5.3.9 Expression and purification of DdahB and MlghB

The second step in GDP-6-deoxy-D-*altro*-heptose biosynthesis in *C. jejuni* strain 81-176 is the epimerisation of GDP-4-keto-6-deoxy-D-*lyxo*-heptose at the C3 position, catalysed by DdahB, to give GDP-4-keto-6-deoxy-D-*arabino* heptose (McCallum *et al*., 2012). The second step in GDP-6-*O*-methyl-L-*gluco*-heptose biosynthesis in *C. jejuni* strain NCTC 11168 is the epimerisation of GDP-(6-*O*-methyl)-4-keto-D-*lyxo*-heptose at the C3 and C5 position, catalysed by MlghB, to give GDP-(6-*O*-methyl)-4-keto-L-*xylo*-heptose (McCallum *et al*., 2013). MlghB and DdahB could both be expressed, as confirmed via whole-cell western blotting against the N-terminal His_6 affinity tag, shown in figure 5.23. Furthermore, these proteins were expressed in a soluble form, as confirmed by the presence of the target proteins in the eluates of small-scale nickel affinity purifications, as shown in figure 5.24.

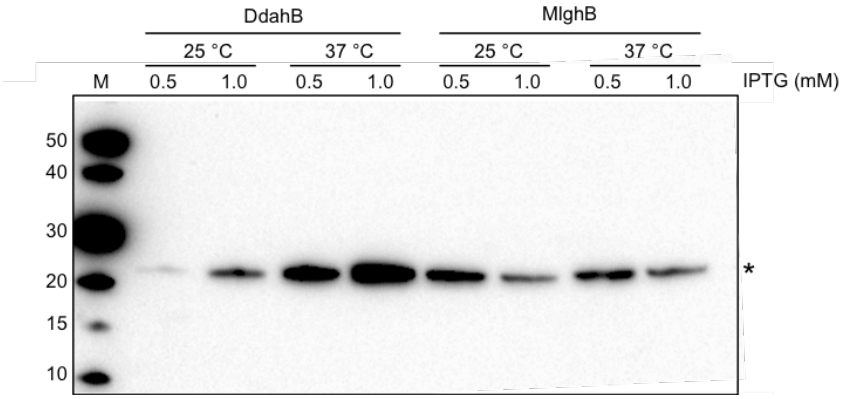


Figure 5.23. Western blot probed with an anti-His tag antibody whole cells following expression trials of DdahB and MlghB. DdahB and MlghB can be expressed over a range of conditions, with the bands at approximately 24 kDa indicated with a "*" corresponding to His-tagged DdahB and His-tagged MlghB.

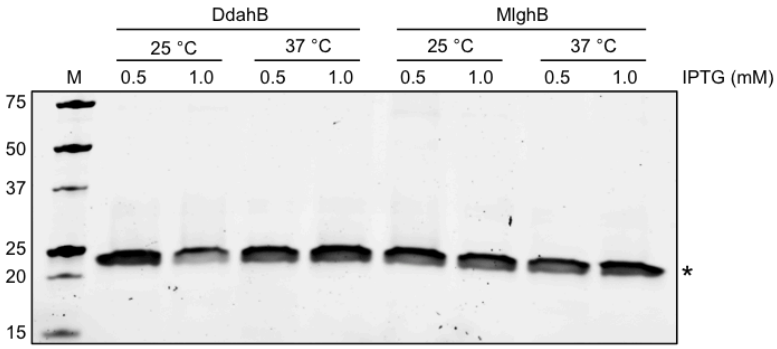


Figure 5.24. SDS-PAGE gel of small-scale purification trials of DdahB and MlghB. Both DdahB and MlghB can be purified using nickel affinity resin from a range of expression conditions, indicating that these proteins are expressed in a soluble form. The bands at approximately 24 kDa, indicated with a "*", correspond to His-tagged DdahB and His-tagged MlghB.

DdahB and MlghB (both 24 kDa in Mw with the N-terminal His_6 tag plus TEV protease cleavage site intact, and 21 kDa in Mw following TEV protease cleavage) were successfully purified. The SDS-PAGE gels of the first and second nickel affinity columns and SEC fractions, with the corresponding UV chromatogram, can be seen in figure 5.25 for DdahB and figure 5.26 for MlghB. The identity of these purified proteins was confirmed by mass spectrometry. Both proteins eluted from the SEC column as monodisperse dimers, as confirmed via SEC-MALS, shown in figure 5.27 and table 5.4. Purification yields of 15-20 mg/L of *E. coli* cell culture were obtained for both DdahB and MlghB.

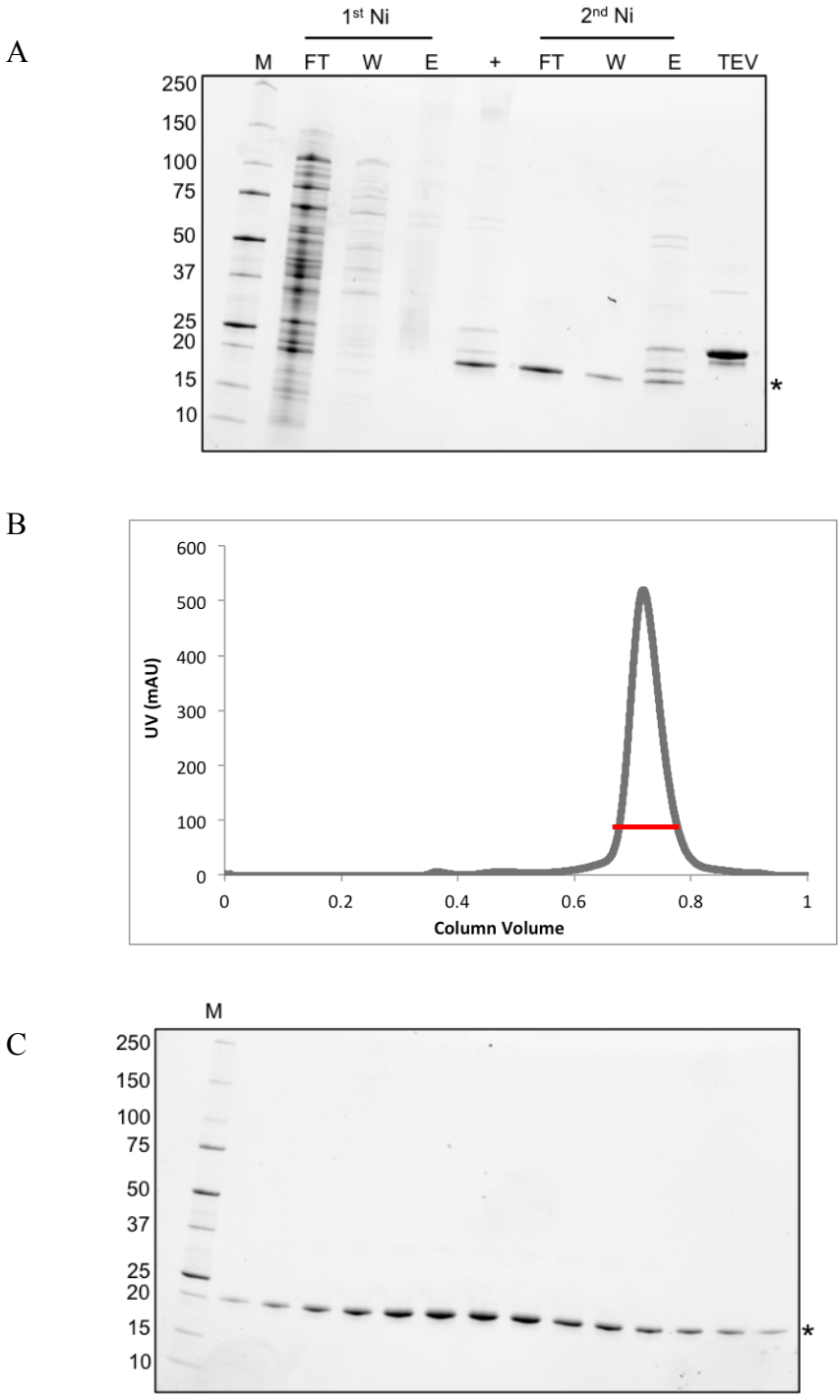


Figure 5.25. Purification of DdahB. The SDS-PAGE gel following the first and second nickel affinity columns can be seen in "A". (M: Marker; FT: FT; W: wash; E: Elution; +: TEV protease-cleaved sample taken prior to second nickel affinity column; TEV: TEV protease). The SEC UV chromatogram can be seen in "B". The void volume equals approximately 0.38 CV. "C" shows an SDS-PAGE gel of SEC fractions. The band at approximately 21 kDa, highlighted with a "*" corresponds to TEV protease-cleaved DdahB. The approximate position of the fractions in "B" shown in "C" is indicated with a red line.

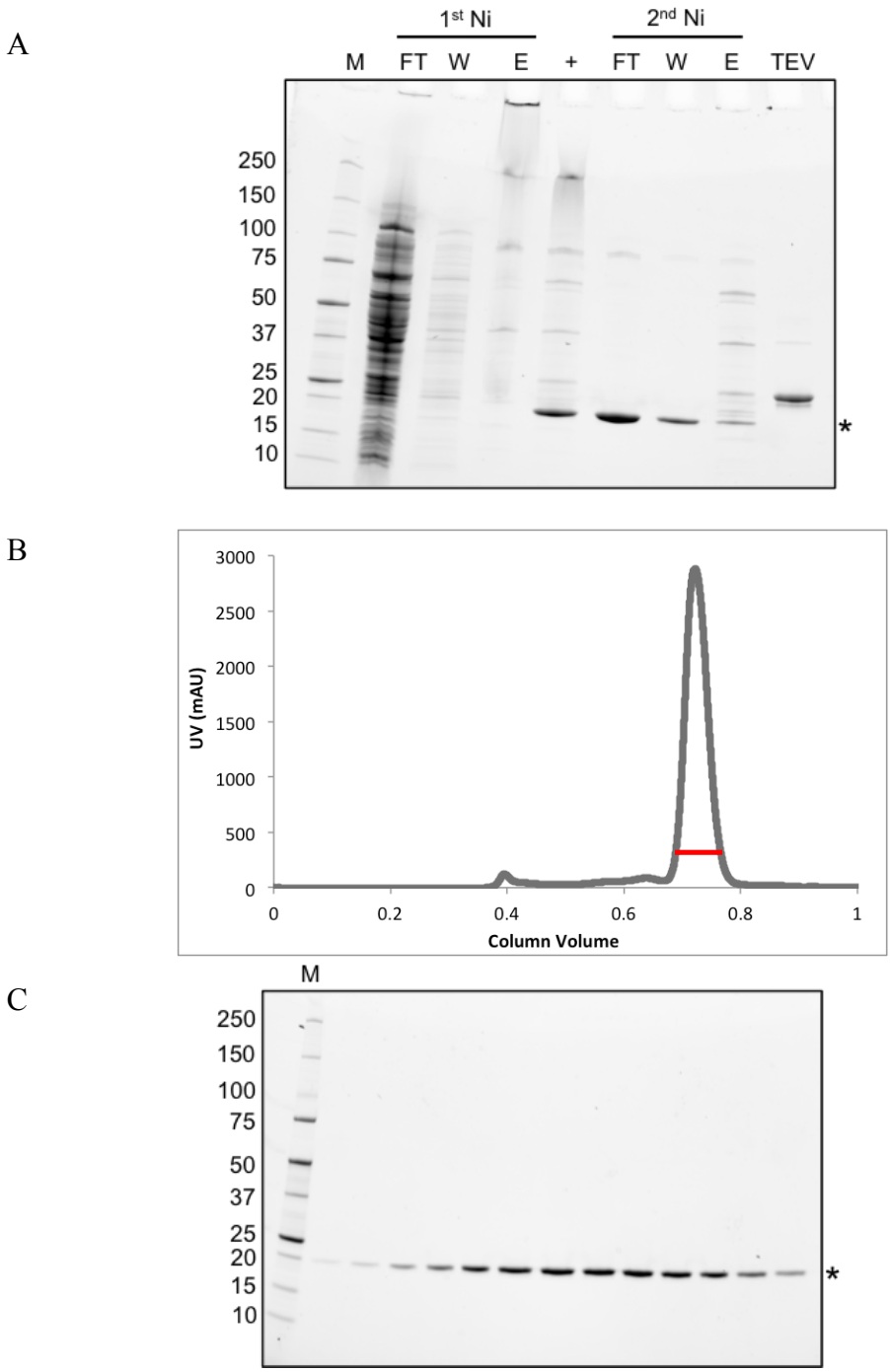
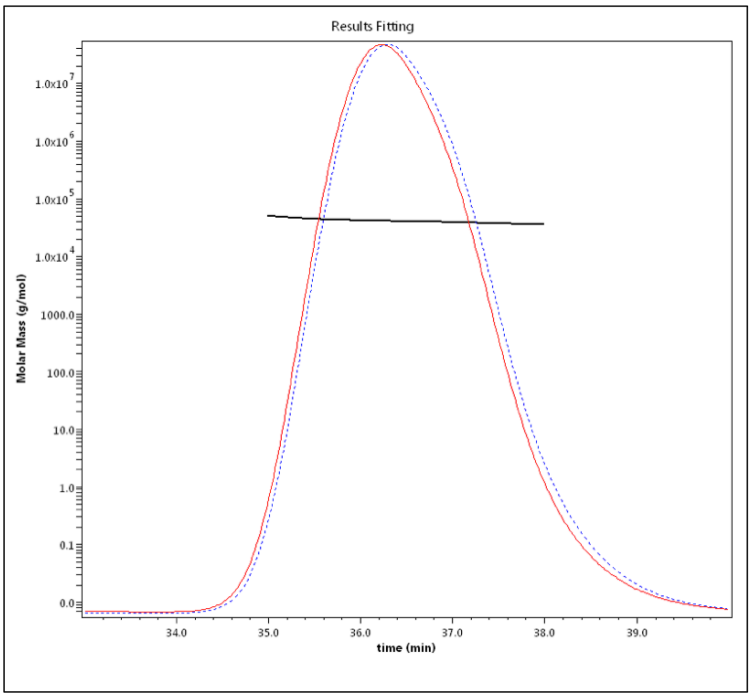


Figure 5.26. Purification of MlghB. The SDS-PAGE gel following the first and second nickel affinity columns can be seen in "A". (M: Marker; FT: FT; W: wash; E: Elution; +: TEV protease-cleaved sample taken prior to second nickel affinity column; TEV: TEV protease). The SEC UV chromatogram can be seen in "B". The void volume equals approximately 0.4 CV. "C" shows an SDS-PAGE gel of SEC fractions. The band at approximately 21 kDa, highlighted with a "*", corresponds to TEV protease-cleaved MlghB. The approximate position of the fractions in "B" shown in "C" is indicated with a red line.

A



B

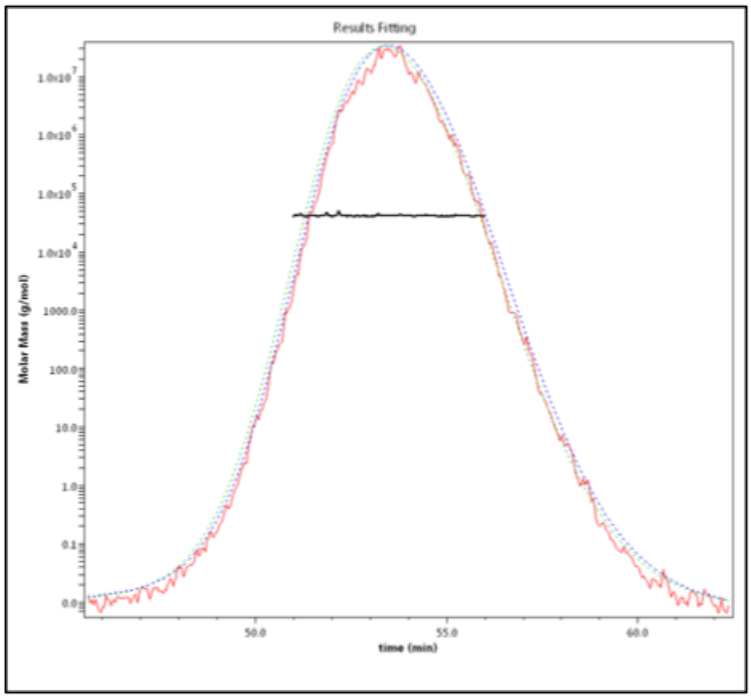


Figure 5.27. SEC-MALS profiles of DdahB and MlghB. Both DdahB (A) and MlghB (B) were confirmed by SEC-MALS to be monodisperse dimers in solution. The SEC-MALS profile in (A) corresponds to DdahB without the Met-Ser linker between the N-terminal TEV protease-cleavable His_6 tag and the target sequence. The change in light scattering over time is shown in red, the change in refractive index over time is shown in blue, the change in UV over time is shown in green (shown in B only) and the change in molar mass (g/mol) over time is shown in black.

	DdahB	MlghB
Molar mass moments (g/mol)		
Mn	4.1×10^4 (± 0.4 %)	4.2×10^4 (± 3.4 %)
Mw	4.1×10^4 (± 0.4 %)	4.2×10^4 (± 3.5 %)
Polydispersity		
Mw/Mn	1.0 (± 0.6 %)	1.0 (± 4.9 %)

Table 5.4. DdahB and MlghB SEC-MALS results. The monomeric molecular weight of both DdahB and MlghB is 21 kDa, therefore values of 41 kDa and 42 kDa, for DdahB and MlghB, respectively, for both the number average molecular weight (Mn) and the weight average molecular weight (Mw), show that DdahB and MlghB are both dimers in solution. Mn is defined as: $\Sigma N_i M_i^2 / \Sigma N_i M_i$, where M_i corresponds to the molecular weight of the protein and N_i corresponds to the number of proteins of that molecular weight. Mw is defined as: $\Sigma N_i M_i / \Sigma N_i$. Polydispersity is the range of molecular weights occupied by the sample and is defined as Mw/Mn, with 1.0 corresponding to a monodisperse sample. The higher the polydispersity value, the wider the range of molecular weights occupied by a sample.

5.3.10 Crystallisation of DdahB and MlghB

DdahB crystallised in JCSG-plus (Molecular Dimensions) well D1 (24 % (w/v) PEG 1500, 20 % (v/v) glycerol) and MlghB crystallised in St. Andrews PEG 1 (appendix 8.4.5.) well E3 (31.05 % (w/v) PEG 1500, 0.25 M sodium-potassium phosphate, 3.83 % (v/v) 1,4-dioxane). A representative X-ray diffraction image of DdahB can be seen in figure 5.28. A crystal of MlghB and a X-ray representative diffraction image can be seen in figure 5.29.

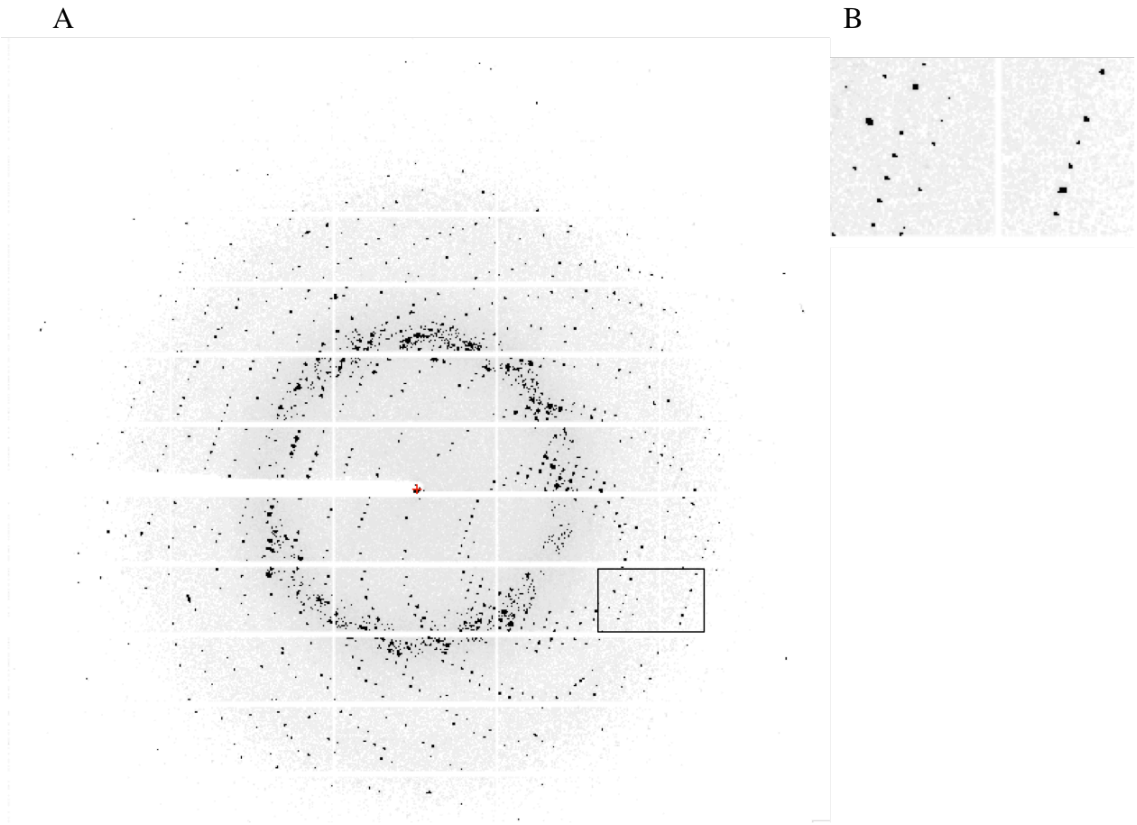


Figure 5.28. DdahB X-ray diffraction pattern. A representative X-ray diffraction pattern of DdahB can be seen in "A" and a zoomed view of the area highlighted in "A" with a square is shown in "B".

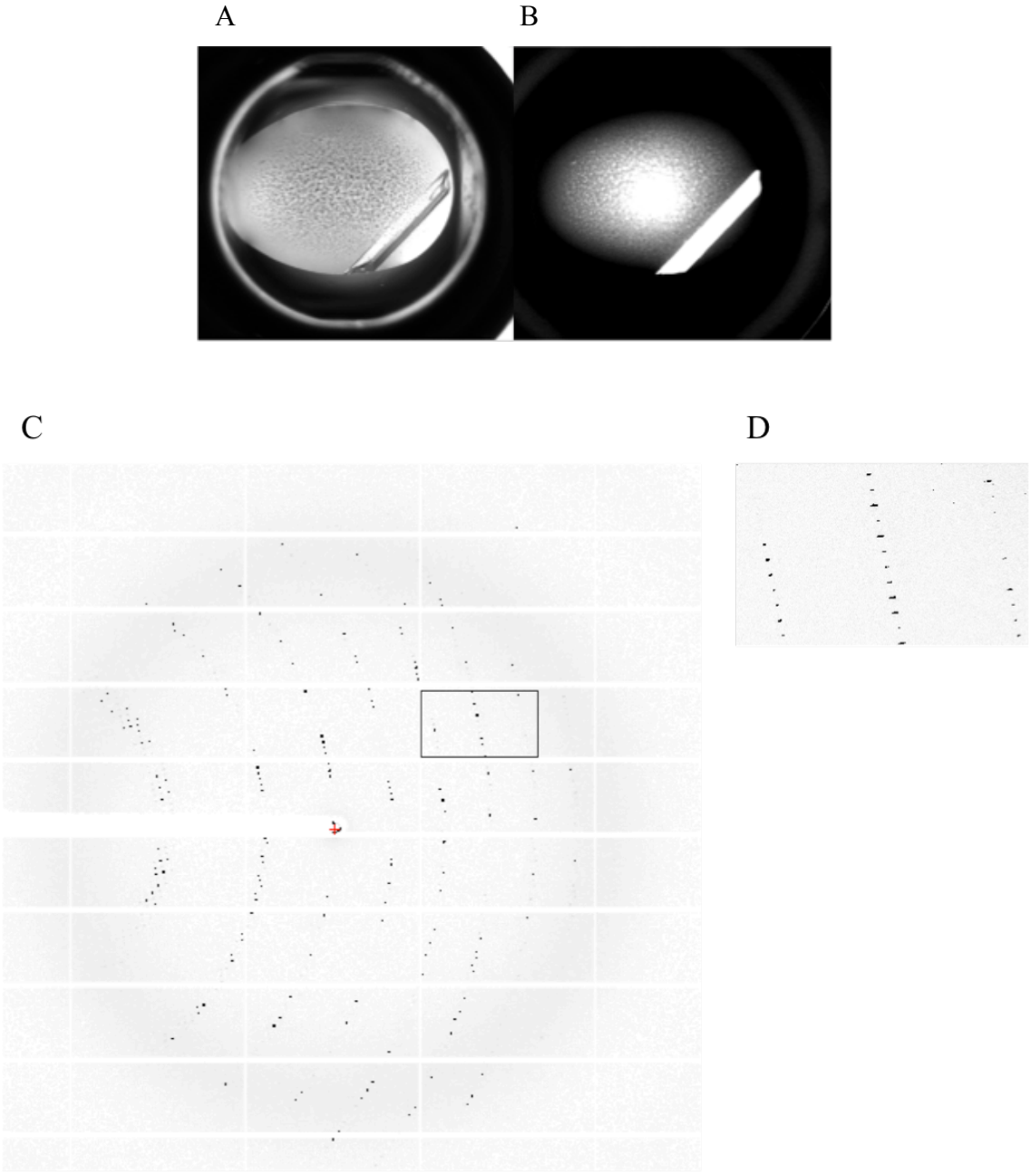


Figure 5.29. MlghB crystal and X-ray diffraction pattern. This crystal was obtained in St Andrews PEG 1 (Molecular Dimensions). The normal image is shown in "A" and the UV image is shown in "B". These images were taken using a Minstrel HT-UV imaging system. A representative X-ray diffraction pattern is shown in "C", with a zoomed view of the area highlighted with a square in "C" shown in "D".

5.3.11 The crystal structures of DdahB and MlghB

The crystal structures of DdahB and MlghB were both solved via molecular replacement to 1.30 Å and 2.14 Å resolution, respectively. DdahB was solved using the structure of the dTDP-6-deoxy-D-*xylo*-4-hexulose 3,5-epimerase RmlC, from *P. aeruginosa*, in complex with dTDP-xylose (PDB: 2IXI) as a search model (Dong *et al*., 207). MlghB was solved using DdahB as a search model. DdahB crystallised in space group P 1 2_1 1, with two chains in the asymmetric unit and MlghB crystallised in space group P 2_1 2_1 2_1, with four chains in the asymmetric unit. Tables 5.5 and 5.6 show the data collection and refinement statistics for DdahB and MlghB, respectively. The MolProbity validation report (Chen *et al*., 2010; Davis *et al*., 2007) can be seen in figures 5.30 and 5.31 for DdahB and MlghB, respectively.

Data Collection	
Wavelength (Å)	0.91587
Space Group	P 1 2_1 1
Unit-cell a, b, c (Å)	47.75, 68.19, 53.67
Unit-cell α, β, γ (°)	90, 91.36, 90
Resolution range (Å)	34.1 – 1.30 (1.32 – 1.30)
I/σ(I)	13.9 (1.2)
R_{merge} (%)	4.2 (57.0)
$CC_{1/2}$	0.999 (0.667)
Completeness (%)	95.4 (72.5)
Multiplicity	2.7 (1.4)
Refinement	
R-factor (%)	14.8
R_{free} (%)	17.5
R.m.s.d. Bond Lengths (Å)	0.013
R.m.s.d. Bond Angles (°)	1.603
Mean B factor (Å^2)	22.3

Table 5.5. X-ray diffraction data collection and refinement statistics for DdahB. Values for the highest resolution shell are shown in brackets. R_{merge} is defined as: $\Sigma_{hkl} \Sigma_i | I_i(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity of reflection hkl. The R-factor is defined as: $\Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where F_{obs} is the observed structure-factor amplitudes and F_{calc} is the calculated structure-factor amplitudes. R_{free} was calculated, using the R-factor equation, using a 5 % subset of reflections that were not used in refinement.

All-Atom Contacts	Clashscore, all atoms:	2.42		98th percentile* (N=386, 1.30Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	309	98.41%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	336	98.25%	Goal: >98%
	MolProbity score^	1.02		99th percentile* (N=2276, 1.30Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.30%	Goal: 0
	Bad bonds:	0 / 2956	0.00%	Goal: 0%
	Bad angles:	2 / 4014	0.05%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 14	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	2 / 331	0.60%	Goal: <0.05%

Figure 5.30. MolProbity validation report for the refined structure of DdahB. The Clashscore and MolProbity score place this structure in the 98th and 99th percentile, respectively when comparing this structure to those of similar resolution (Chen *et al*., 2010; Davis *et al*., 2007).

Data Collection	
Wavelength (Å)	0.91587
Space Group	P 2_1 2_1 2_1
Unit-cell a, b, c (Å)	42.26, 121.67, 153.80
Unit-cell α, β, γ (°)	90, 90, 90
Resolution range (Å)	95.43 – 2.14 (2.18 – 2.14)
I/σ(I)	12.7 (1.3)
R_{merge} (%)	6.0 (76.2)
$CC_{1/2}$	0.999 (0.463)
Completeness (%)	100.0 (99.7)
Multiplicity	5.9 (4.2)
Refinement	
R-factor (%)	22.1
R_{free} (%)	24.6
R.m.s.d. Bond Lengths (Å)	0.011
R.m.s.d. Bond Angles (°)	1.540
Mean B factor (Å^2)	36.3

Table 5.6. X-ray diffraction data collection and refinement statistics for MlghB. Values for the highest resolution shell are shown in brackets. R_{merge} is defined as: $\Sigma_{hkl} \Sigma_i | I_i(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity of reflection hkl. The R-factor is defined as: $\Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where F_{obs} is the observed structure-factor amplitudes and F_{calc} is the calculated structure-factor amplitudes. R_{free} was calculated, using the R-factor equation, using a 5 % subset of reflections that were not used in refinement.

All-Atom Contacts	Clashscore, all atoms:	2.05		100th percentile* (N=555, 2.14Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	12	1.93%	Goal: <0.3%
	Favored rotamers	575	92.44%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	655	98.64%	Goal: >98%
	MolProbity score^	1.19		100th percentile* (N=11227, 2.14Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 5805	0.00%	Goal: 0%
	Bad angles:	6 / 7872	0.08%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 29	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	4 / 643	0.62%	Goal: <0.05%

Figure 5.31. MolProbity validation report for the refined structure of MlghB. Both the Clashscore and MolProbity score place this structure in the 100th percentile when comparing to structures of similar resolution (Chen *et al*., 2010; Davis *et al*., 2007)**.**

DdahB is a homodimer, as confirmed by SEC-MALS, shown in figure 5.27 and table 5.4. Analysis of the refined model using the PDBePISA server also suggests that a dimer is the most stable oligomeric state in solution (Complexation significance score (CSS) = 1) (Krissinel, 2009; Krissinel and Henrick, 2007). Out of a possible 183 amino acids, with the first two amino acids corresponding to the linker between the TEV protease-cleavable N-terminal His_6 tag and the target sequence, denoted -1 to 0, the refined model of chain A consists of amino acids 1-139 and 145-174. The refined model of chain B consists of amino acids 3-181. Each monomer consists of 14 β-strands and two short α-helices, with nine β-strands, arranged in two β-sheets, forming a small anti-parallel β-barrel, known as a jellyroll, shown in figure 5.28. The dimerization interface is stabilised by hydrogen bond interactions between the amine and carbonyl groups of the backbone of β-strand four of one monomer and β-strand five of the other, as highlighted in figure 5.32. These contribute to the 3,960 $Å^2$ of solvent-exposed surface area buried upon dimerization out a total solvent-exposed surface area of 16,820 $Å^2$ (Krissinel, 2009; Krissinel and Henrick, 2007).

Like DdahB, MlghB is also a homodimer, as confirmed by SEC-MALS, shown in figure 5.27 and table 5.4. This was also confirmed by the analysis of the refined model using the PDBePISA server (Krissinel, 2009; Krissinel and Henrick, 2007). Out of a possible 183 amino acids, the first two of which correspond to the linker between the TEV protease-cleavable N-terminal His_6 tag and the target sequence, with the N-terminal Met denoted as residue 1, the refined model of chain A consists of amino acids 2-178, chain B of 2-140 and 149-176, chain C of 3-138 and 147-173, and chain D of 3-142 and 148-178. PDBePISA constructed a physiologically relevant dimer, with a CSS score of 1, using chains A and D, therefore these chains will be discussed in more detail (Krissinel, 2009; Krissinel and Henrick, 2007). Each monomer consists of 13 β-strands and 2 α-helices and, as in DdahB, β-strands 4-13 form a jellyroll, consisting of two 4-stranded antiparallel β-sheets, shown in figure 5.32. MlghB dimerizes in the same way as DdahB, with hydrogen bonds between the amine and carbonyl groups of the backbone of β-strand three of one monomer and β-strand four of the other, highlighted in figure 5.33. Similar to DdahB, 3,190 $Å^2$ out of a total surface area of 15,930 $Å^2$ is buried upon dimerization (Krissinel, 2009; Krissinel and Henrick, 2007).

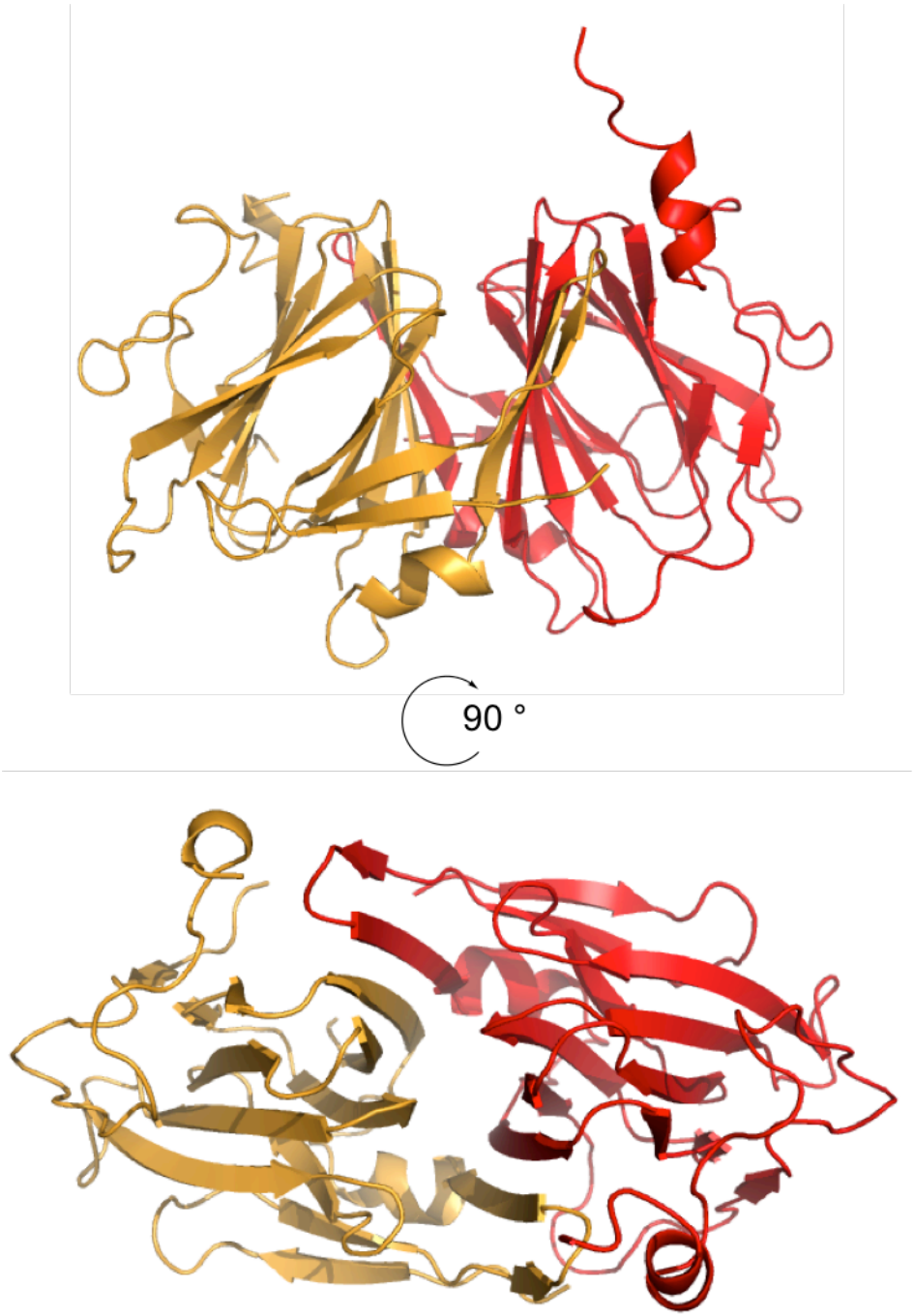


Figure 5.32. DdahB is a homodimer. One monomer is shown in orange and the other in red, with both monomers adopting a RmlC-like cupin fold. DdahB dimerizes via β-strand-strand interactions. These images were created using MacPyMOLEdu.

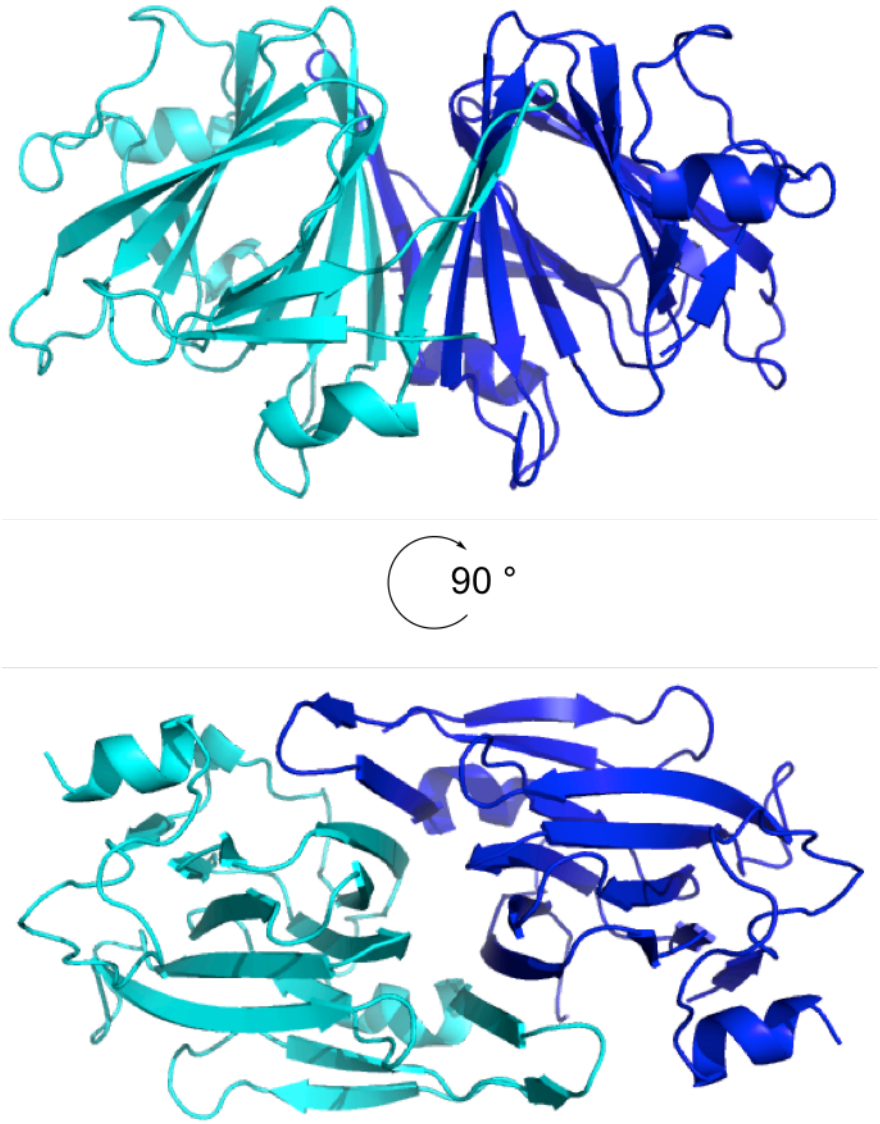


Figure 5.33. MlghB is a homodimer. One monomer is shown in cyan and the other in dark blue. Each monomer adopts a RmlC-like cupin fold and the monomers dimerize via β-strand-strand interactions. These images were created using MacPyMOLEdu.

The described structural, and sequence, similarity between DdahB and MlghB, despite the difference in epimerisation activity, is confirmed by a RMSD of 0.63 Å for the alignment of the C_α backbone over 160 residues of DdahB (chain B) and MlghB (chain B) (Krissinel and Henrick, 2004). An overlay of the DdahB and MlghB homodimers is shown in figure 5.34.

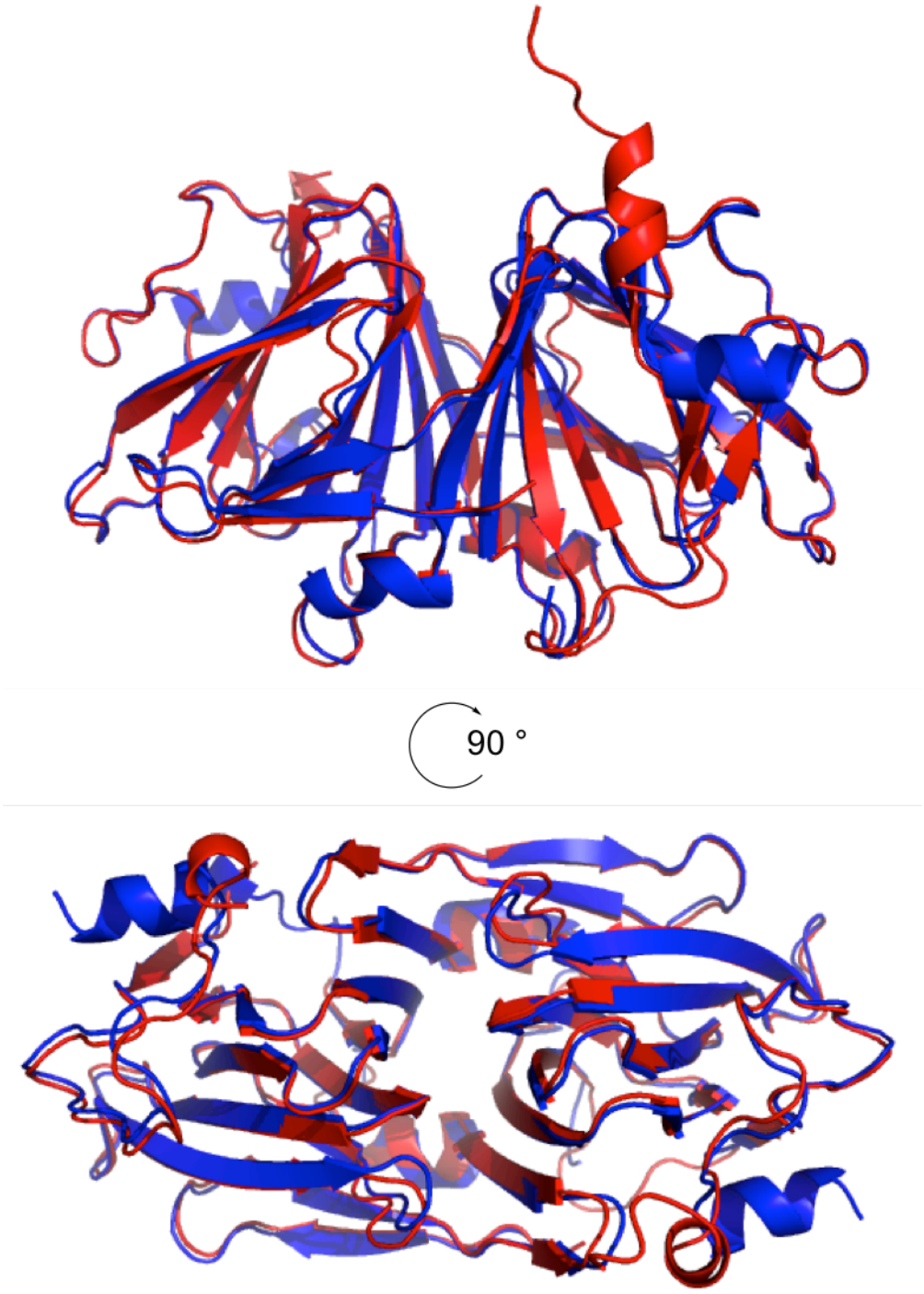


Figure 5.34. An overlay of the structures of DdahB and MlghB. DdahB is shown in red and MlghB in blue. This image was created using MacPyMOLEdu.

Neither DdahB nor MlghB cocrystallised with a cofactor, which correlates with the previous findings that a cofactor is required neither for RmlC-catalysed epimerisation (Giraud *et al*., 2000) nor DdahB/MlghB-catalysed epimerisation (McCallum *et al*., 2012, 2013).

5.3.12 A comparison of the structures of DdahB and MlghB with the dTDP-6-deoxy-D-*xylo*-4-hexulose 3,5-epimerase RmlC

The jellyroll fold seen in both DdahB and MlghB homodimers is also known as an RmlC-like jellyroll or an RmlC-like cupin fold (Giraud *et al*., 2000). RmlC catalyses the epimerisation of dTDP-6-deoxy-D-*xylo*-4-hexulose at the C3 and C5 positions to form dTDP-6-deoxy-L-*lyxo*-hexulose as the second step of dTDP-L-rhamnose biosynthesis (Giraud *et al*., 2000). Despite a shared sequence identity between DdahB and RmlC (from *Streptococcus suis*) (PDB: 2IXL, chain D) of just 20.8 % (figure 5.6), the C_α backbone of 148 residues out of 169 for DdahB (chain A) and 196 for RmlC (chain D) could be aligned with an RMSD of 1.52 Å (figure 5.35) (Krissinel and Henrick, 2004). MlghB and RmlC share a similarly low level of sequence identity of just 21.5 % (figure 5.6), however these proteins could also be aligned with an RMSD of 1.51 Å over the C_α backbone of 155 residues out of 167 for MlghB (chain B) and 196 for RmlC (chain D) (figure 5.36) (Krissinel and Henrick, 2004).

The structure of RmlC (from *S. suis*) (PDB: 2IXL) has been solved in complex with the product analogue dTDP-rhamnose (Dong *et al*., 2007). This has allowed the identification of the substrate binding site and the proposal of the active site residues of MlghB and DdahB. The active site of DdahB and MlghB, as well as RmlC (Dong *et al*., 2007), is located at the entrance of the jellyroll, shown in figures 5.35 and 5.36, respectively. RmlC epimerises its substrate using a His-Tyr dyad, corresponding to His 73 and Tyr 140 in RmlC, with a Lys residue, corresponding to Lys 82, proposed to stabilise the transition state (Dong *et al*., 2007). This is also believed to be the case for both DdahB and MlghB (Dr Creuzenet, University of Western Ontario, unpublished

data), with the active site His corresponding to His 67 in both DdahB and MlghB and the active site Tyr corresponding to Tyr 134 in both DdahB and MlghB. Lys 82, required for transition state stabilisation, corresponds to Lys 74 in both DdahB and MlghB. Ile 66, located in the proposed active site of both DdahB and MlghB forms a cis peptide (see figures 5.30 and 5.31). There is little observed difference in the positions of the proposed active site residues in the DdahB and MlghB apo structures and the positions of the confirmed active site residues in the structure of RmlC in complex with dTDP-rhamnose, as shown in figure 5.35 for DdahB and 5.36 for MlghB. Therefore, the apo structures of DdahB and MlghB were not sufficient to explain the observed differences in DdahB (McCallum *et al*., 2012) and MlghB epimerisation activity (McCallum *et al*., 2013).

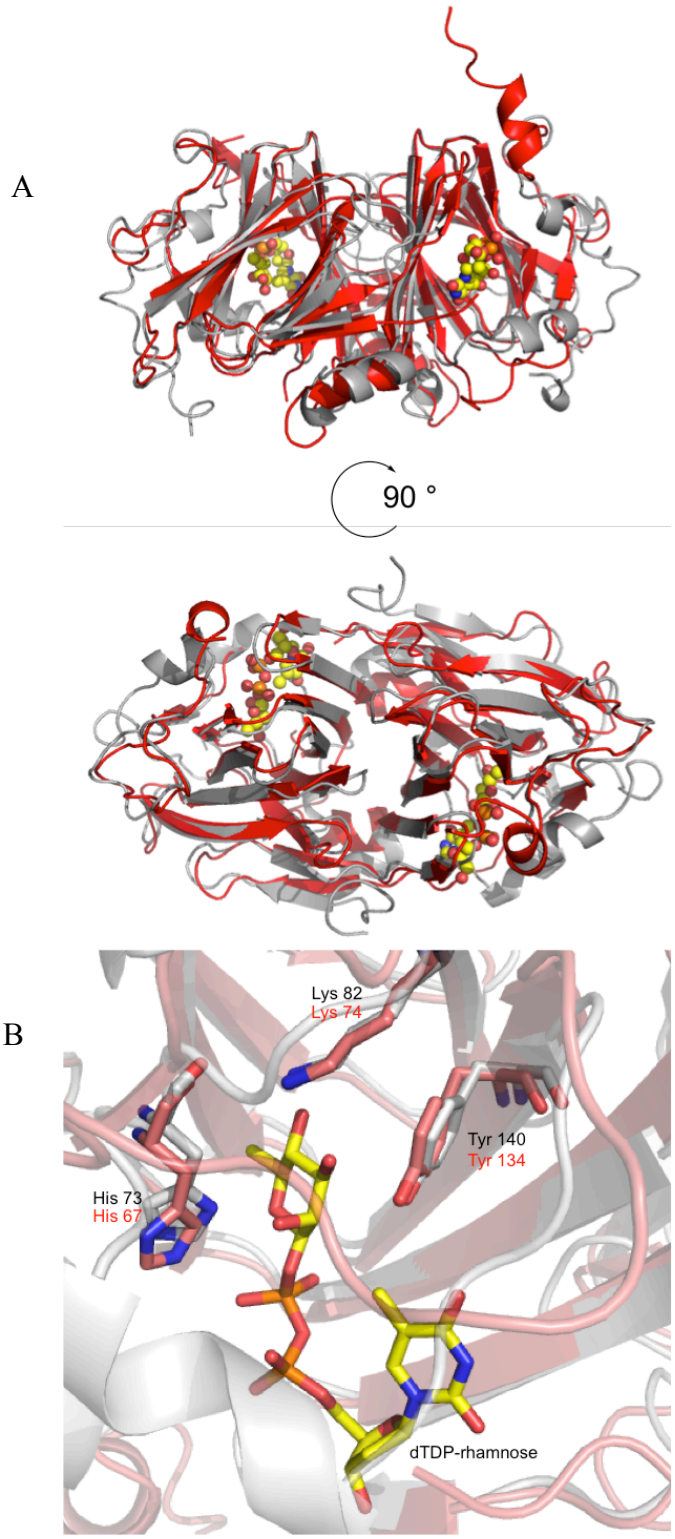


Figure 5.35. A comparison of the DdahB and RmlC active sites. An overlay of dTDP-rhamnose-bound RmlC (PDB: 2IXL) (Dong *et al*., 2007) and DdahB apo structure suggests that DdahB binds its substrate at the entrance of the jellyroll (A). The active site residues of RmlC are conserved in DdahB (B). DdahB is shown in red, RmlC in grey and dTDP-rhamnose in yellow. The labels in red correspond to DdahB and those in black to RmlC. This image was made using MacPyMOLEdu.

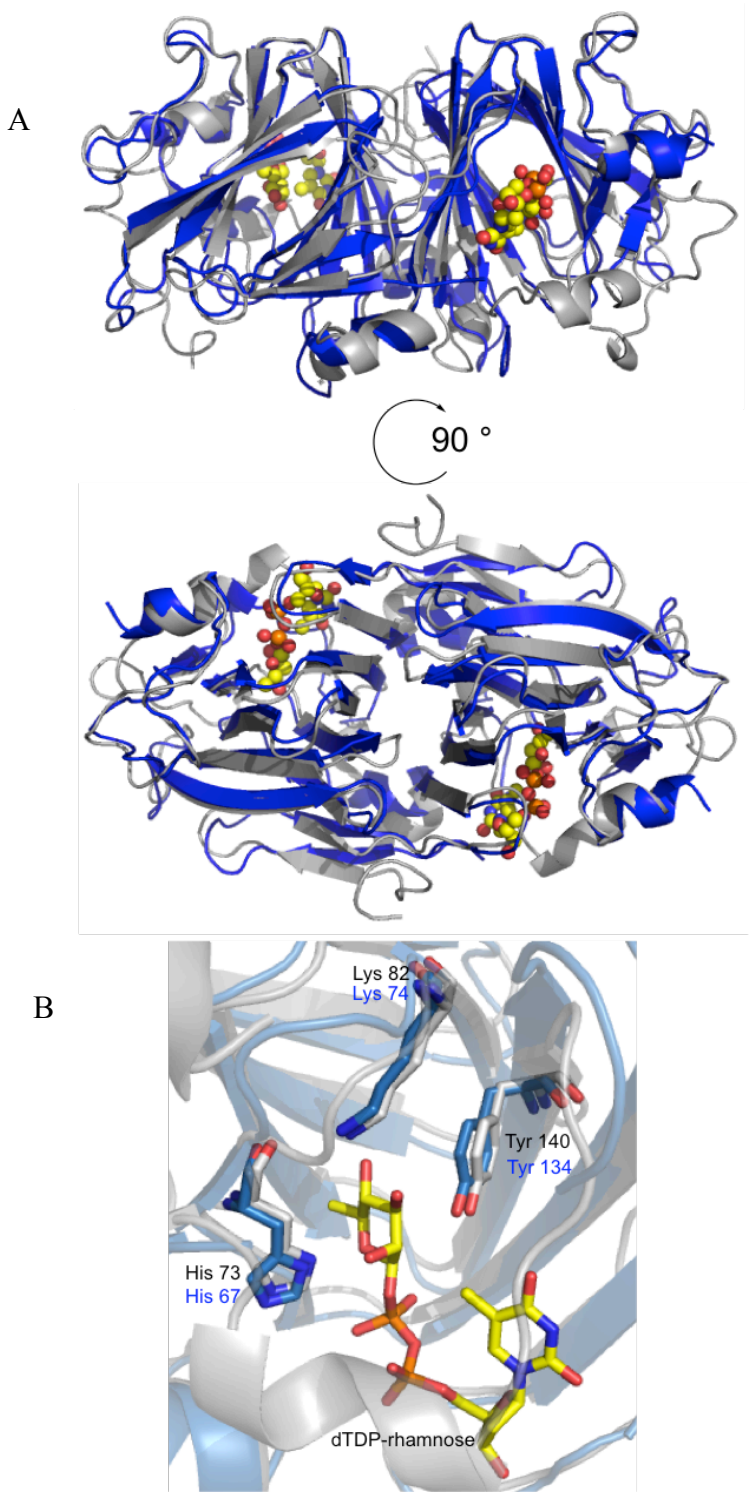


Figure 5.36. A comparison of the MlghB and RmlC active sites. An overlay of dTDP-rhamnose-bound RmlC (PDB: 2IXL) (Dong *et al*., 2007) and MlghB apo structure suggests that MlghB binds its substrate at the entrance of the jellyroll (A). The active site residues of RmlC are conserved in MlghB (B). MlghB is shown in blue, RmlC in grey and dTDP-rhamnose in yellow. The labels in blue correspond to MlghB and those in black to RmlC. This image was made using MacPyMOLEdu.

5.3.13 The structure of DdahB with GDP and GDP-mannose

Although both DdahB and MlghB are modified heptose epimerases, they have also been shown to epimerise GDP-mannose (Dr Creuzenet, University of Western Ontario, personal communication). Therefore in order gain a better understanding of substrate binding by DdahB and MlghB than that offered by the alignment of the apo structures of DdahB and MlghB with dTDP-rhamnose-bound RmlC (PDB: 2IXL) (Dong *et al*., 2007), we attempted to solve the structures of DdahB and MlghB in the presence of GDP-mannose via both soaking of GDP-mannose into apo crystals and co-crystallisation.

The structure of DdahB with GDP-mannose was solved to a resolution of 2.35 Å, in the space group P 1 2_1 1, which the same space group as the apo structure. A representative X-ray diffraction pattern can be seen in figure 5.37. Two chains are present in the asymmetric unit, with the refined model containing the following residues: 2-143, 148-177 in chain A and 2-138, 148-174 in chain B. The data collection and refinement statistics can be seen in table 5.7, with the final MolProbity validation report (Chen *et al*., 2010; Davis *et al*., 2007) shown in figure 5.38. Although the crystal was soaked in a solution of GDP-mannose prepared in mother liquor, GDP-mannose could only be confidently modelled in one chain, with GDP modelled in the other. The F_o-F_c density, contoured at 1.5 σ, for the GDP and GDP-mannose can be seen in figure 5.39.

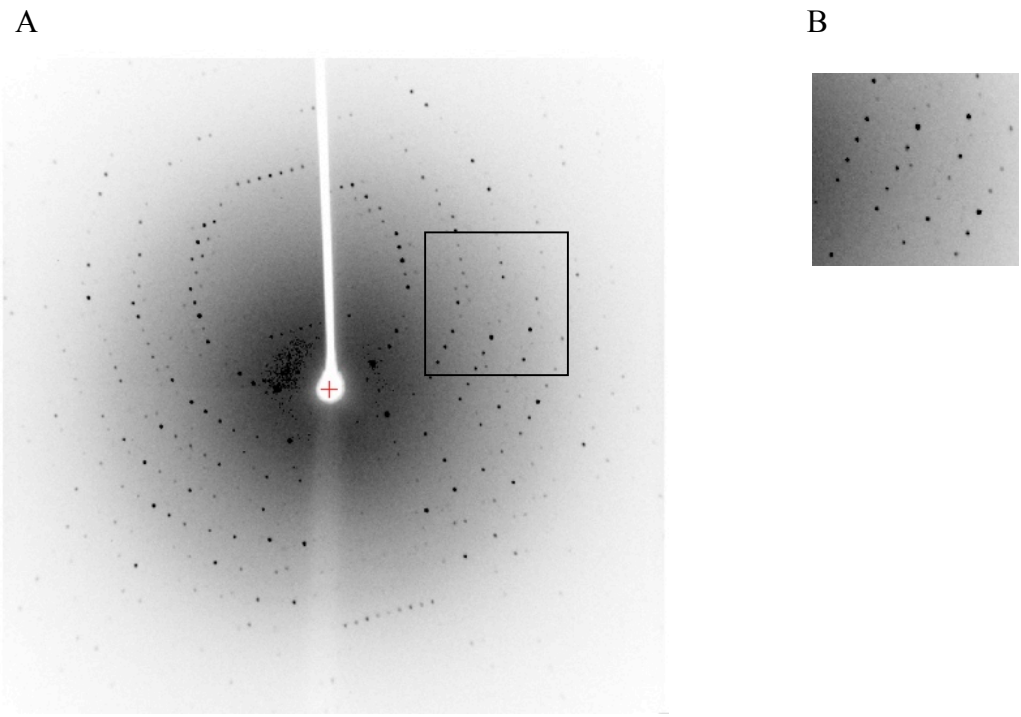


Figure 5.37. Representative X-ray diffraction pattern obtained following soaking of DdahB crystal with GDP-mannose. Entire X-ray diffraction pattern is shown in A, with a zoomed view of the area highlighted with a square in "A" shown in B.

All-Atom Contacts	Clashscore, all atoms:	0.54		100th percentile* (N=335, 2.35Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	0.34%	Goal: <0.3%
	Favored rotamers	292	98.32%	Goal: >98%
	Ramachandran outliers	1	0.31%	Goal: <0.05%
	Ramachandran favored	319	98.15%	Goal: >98%
	MolProbity score^	0.68		100th percentile* (N=9377, 2.35Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 2812	0.00%	Goal: 0%
	Bad angles:	0 / 3809	0.00%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	2 / 318	0.63%	Goal: <0.05%

Figure 5.38. MolProbity validation report for the refined model of DdahB bound to GDP-mannose. Both the clashscore and MolProbity score place this structure in the 100th percentile (Chen *et al*., 2010; Davis *et al*., 2007).

Data Collection	
Wavelength (Å)	1.542
Space Group	P 1 2_1 1
Unit-cell a, b, c (Å)	48.001, 67.81, 53.14
Unit-cell α, β, γ (°)	90, 91.79, 90
Resolution range (Å)	39.17 – 2.35 (2.43 – 2.35)
I/σ(I)	16.3 (6.1)
R_{merge} (%)	6.1 (16.9)
$CC_{1/2}$	0.997 (0.945)
Completeness (%)	98.4 (89.1)
Multiplicity	3.6 (2.7)
Refinement	
R-factor (%)	18.5
R_{free} (%)	24.3
R.m.s.d. Bond Lengths (Å)	0.007
R.m.s.d. Bond Angles (°)	1.302
Mean B factor (Å^2)	21.3

Table 5.7. X-ray diffraction data collection and refinement statistics for DdahB crystal soaked with GDP-mannose. Values for the highest resolution shell are shown in brackets. R_{merge} is defined as: $\Sigma_{hkl} \Sigma_i | I_i(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity of reflection hkl. The R-factor is defined as: $\Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where F_{obs} is the observed structure-factor amplitudes and F_{calc} is the calculated structure-factor amplitudes. R_{free} was calculated, using the R-factor equation, using a 5 % subset of reflections that were not used in refinement.

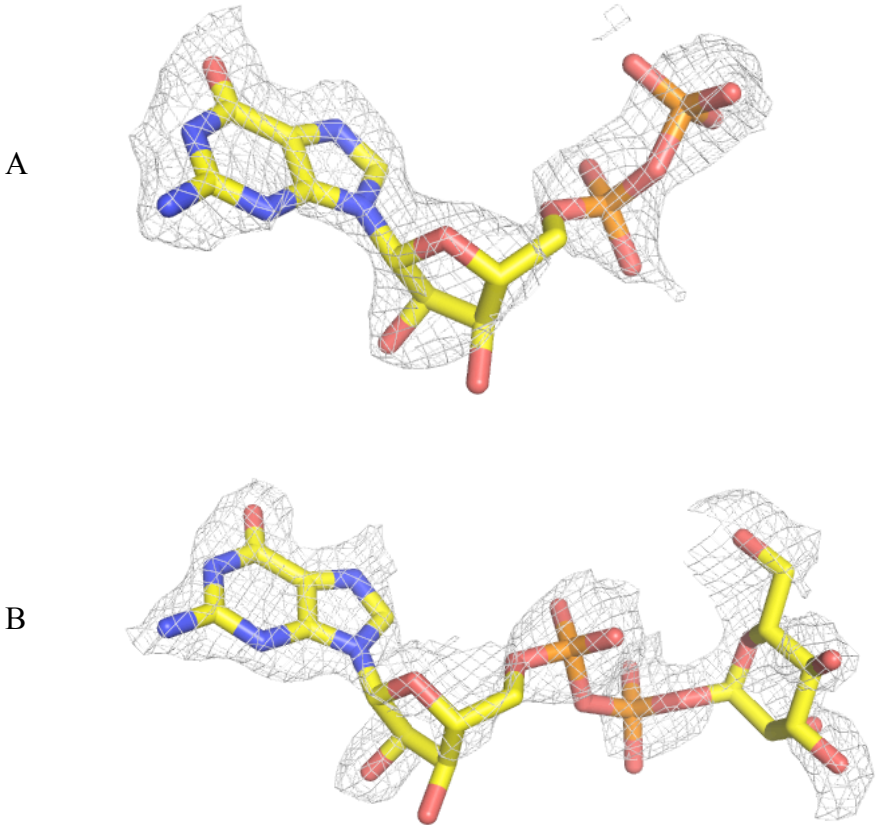


Figure 5.39. F_O-F_C density for GDP and GDP-mannose bound to DdahB. The F_O-F_C density, contoured at 1.5 σ, for GDP and GDP-mannose is shown in "A" and "B", respectively. The F_O-F_C density was calculated by deleting GDP and GDP-mannose from the final refined model, resetting the B factors to 20 $Å^2$ using the TLSMD server (Painter and Merritt, 2006a, 2006b), refining in REFMAC5 (Murshudov *et al*., 1997),, with 50 cycles of refinement. The map was created using FFT, as part of the CCP4 suite (Winn *et al*., 2011). This image was created using MacPyMOLEdu.

The C_α backbone of DdahB in the absence (chain A) and presence (chain B) of GDP-mannose can be aligned over 171 residues with an RMSD of 0.24 Å, indicating that little change in secondary structure occurs upon substrate binding (Krissinel and Henrick, 2004). This is illustrated by the overlay of the apo structureand substrate-bound structure, shown in figure 5.40. As predicted by the overlay of the DdahB apo structure, with the dTDP-rhamnose-bound RmlC structure, the GDP-mannose binding site can be seen at the entry of the central jellyroll (Dong *et al*. 2003; Dong *et al*. 2007). However, in the case of dTDP-rhamnose binding to RmlC, the sugar moiety is oriented

into the jellyroll (Dong *et al*., 2007), whereas here the mannose moiety remains on the surface of the protein, pointing away from the predicted active site, shown in figure 5.41. The binding mode of GDP to DdahB more closely resembles that of dTDP-rhamnose binding to RmlC (Dong *et al*., 2007), with the β-phosphate of GDP oriented towards the centre of the jellyroll shown in figure 5.41.

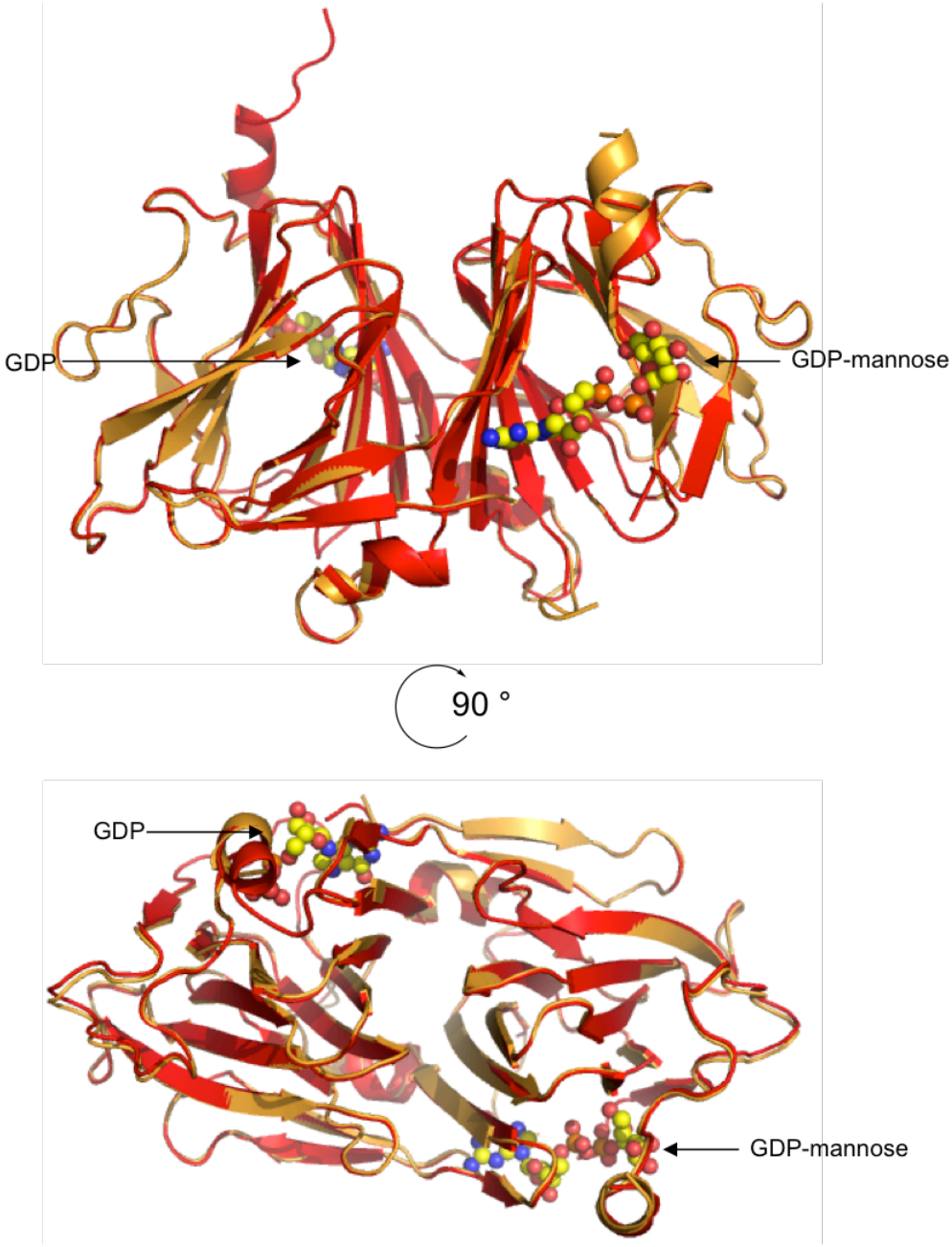


Figure 5.40. DdahB binding to GDP/GDP-mannose. An overlay of the DdahB apo structure, in red, with the GDP/GDP-mannose-bound structure, shown in yellow, shows that little conformational change occurs upon substrate binding. The substrate binding site is at the entry of the jellyroll. This image was constructed using MacPyMOLEdu.

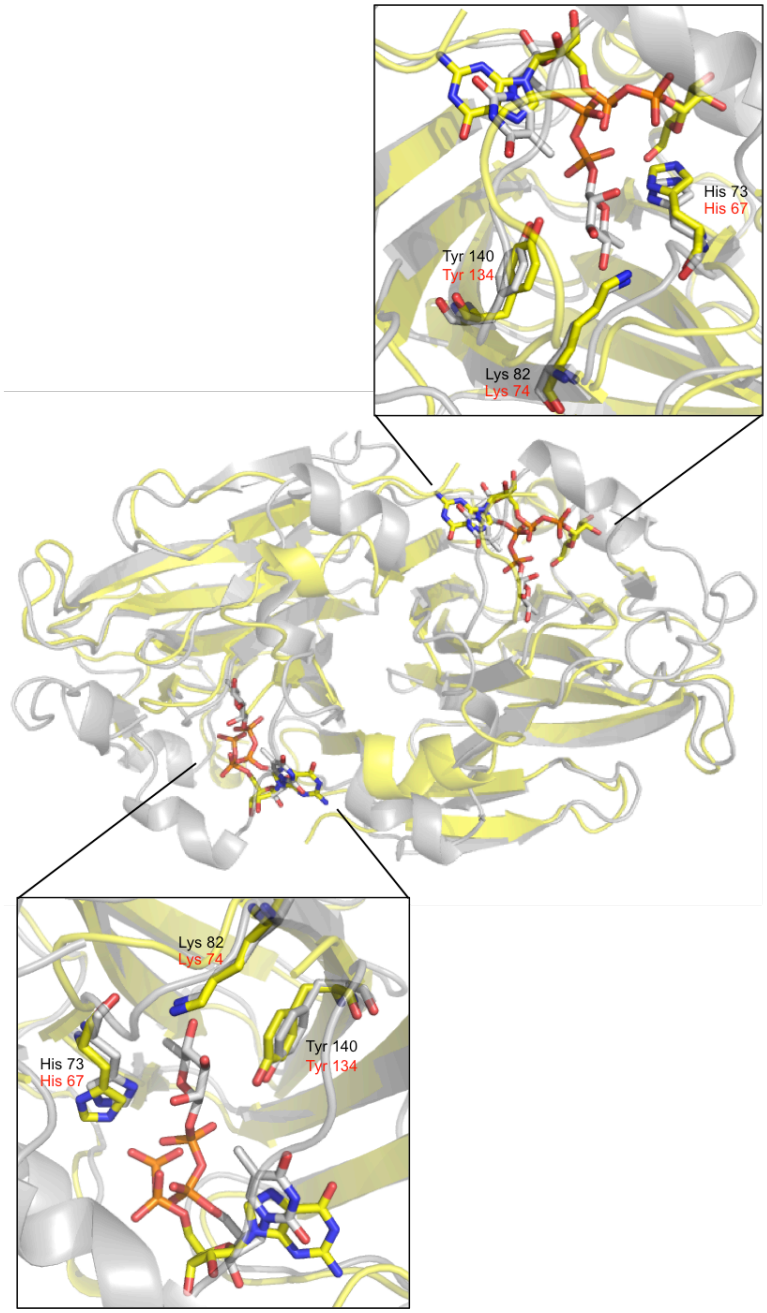


Figure 5.41. A comparison of GDP-mannose/GDP binding to DdahB with dTDP-rhamnose binding to RmlC. Like dTDP-rhamnose binding to RmlC, both GDP and GDP-mannose bind to DdahB at the entrance of the jellyroll, but the mannose moiety of GDP-mannose is oriented away from the active site when bound to DdahB (shown top right), whereas the rhamnose moiety of dTDP-rhamnose is oriented into the jellyroll, towards the active site His-Tyr dyad (Dong *et al*., 2007). The binding mode of GDP to DdahB more closely resembles that of dTDP-rhamnose binding to RmlC (shown bottom left). RmlC and dTDP-rhamnose are shown in grey and DdahB, GDP and GDP-mannose are shown in yellow. The residue names in black correspond to RmlC and those in red to DdahB (RmlC-dTDP-rhamnose PDB: 2IXL) (Dong *et al*., 2007). This image was created using MacPyMOLEdu.

The majority of contacts between DdahB and GDP-mannose are made via the guanosine moiety. Hydrogen bonds are formed between the guanine and Lys 54, and Thr 33 and Asn 22, bridged by two water molecules, as can be seen in figure 5.42. Hydrophobic interactions are also formed between the guanine and Ile 3, Lys 23, Ile 31 and Trp 32. The side chain of Phe 24 stacks with the ribose moiety and electron density is only visible for this side-chain when bound to substrate. These hydrogen bonds and hydrophobic interactions are also formed between DdahB and GDP alone, shown in figure 5.43, indicating that the guanosine binding site is the same in both cases (compare figures 5.42 and 5.43). Due to the orientation of the GDP into the jellyroll, additional hydrogen bonds are formed between the β-phosphate and Arg 28, Arg 64 and the predicted catalytic residue His 67 upon GDP binding to DdahB, shown in figure 5.43. In contrast, only the predicted catalytic His 67 and Gly 65 form hydrogen bonds, via the side chain and main chain, respectively, with the mannose moiety upon DdahB binding to GDP-mannose. Hydrophobic interactions are also formed between the mannose moiety and Arg 64 and Leu 169, seen in figure 5.42.

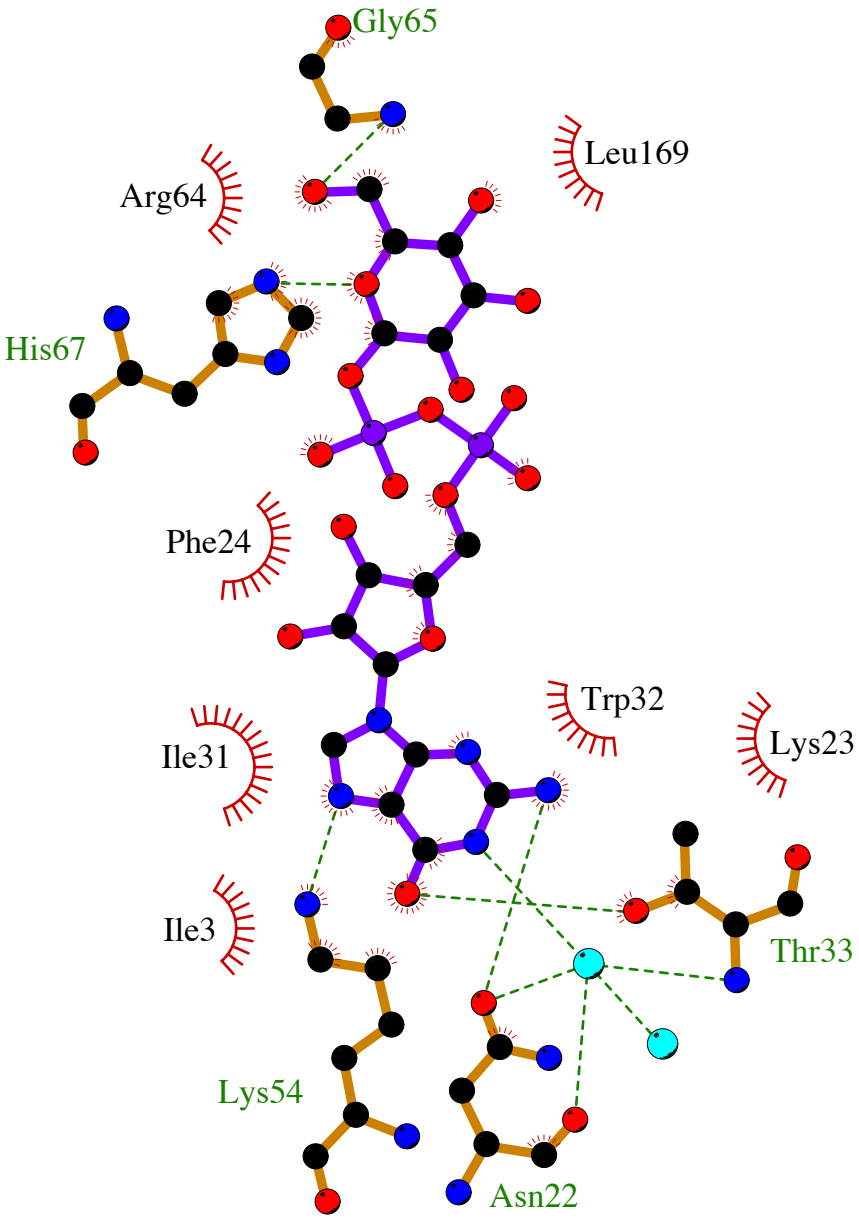


Figure 5.42. The amino acids required for DdahB binding to GDP-mannose. This figure was created using LigPlot+ (Laskowski and Swindells, 2011). Waters are shown in cyan, carbon is shown in black, nitrogen is shown in blue, oxygen is shown in red, and phosphorus is shown in purple. Bonds within the ligand are shown in purple and bonds within the protein shown in brown. Hydrogen bonds and names of the participating amino acids are shown in green. Hydrophobic interactions are shown in red and the names of the participating amino acids are shown in black.

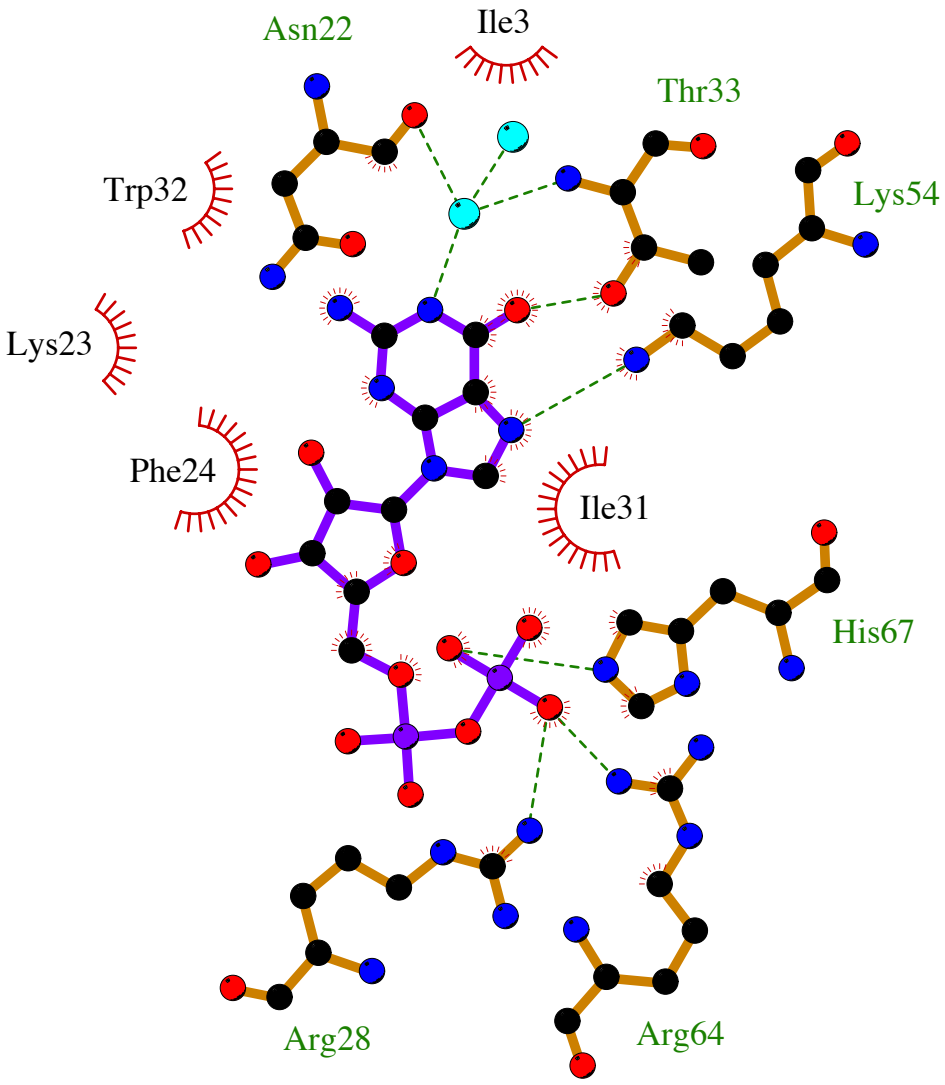


Figure 5.43. The amino acids required for DdahB binding to GDP. This figure was created using LigPlot+ (Laskowski and Swindells, 2011). Waters are shown in cyan, carbon is shown in black, nitrogen is shown in blue, oxygen is shown in red, and phosphorus is shown in purple. Bonds within the ligand are shown in purple and bonds within the protein are shown in brown. Hydrogen bonds and names of the participating amino acids are shown in green. Hydrophobic interactions are shown in red and the names of the participating amino acids are shown in black.

Residues required for the DdahB:GDP-mannose binding mode observed are contributed by both chains, with residues from one chain largely required for binding to the guanosine moiety and residues from the other chain largely required for binding to the mannose moiety. This rationalises the requirement of dimerization to the function of this class of epimerase (Dong *et al*. 2007; Giraud *et al*. 2000; Dong *et al*. 2003). This can be seen in figure 5.44.

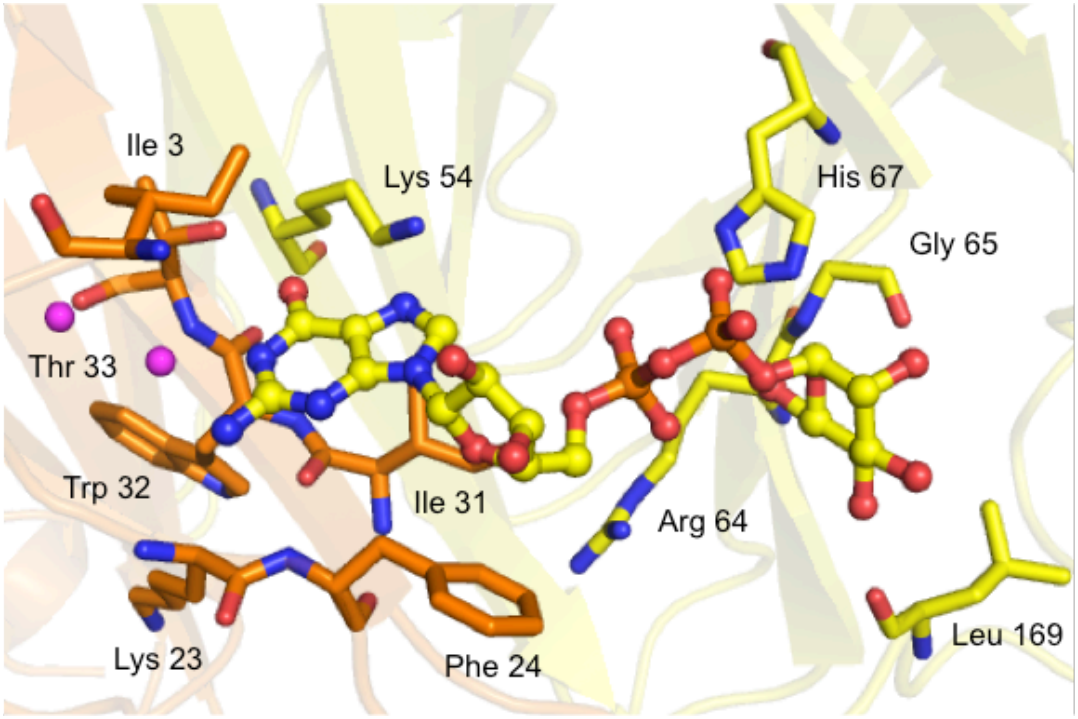


Figure 5.44. The importance of DdahB dimerization to GDP-mannose binding. Residues from both chains of the dimer are required for GDP-mannose binding to DdahB; residues from one chain are shown in orange and the other in yellow. GDP-mannose is shown in yellow, using the ball and stick representation, and water is shown in pink. This image was created using MacPyMOLEdu.

As previously mentioned, the mannose moiety of GDP-mannose is oriented away from the putative active site and therefore little conformational change is observed in the active site residues of DdahB upon GDP-mannose, or GDP, binding, shown in figure 5.45.

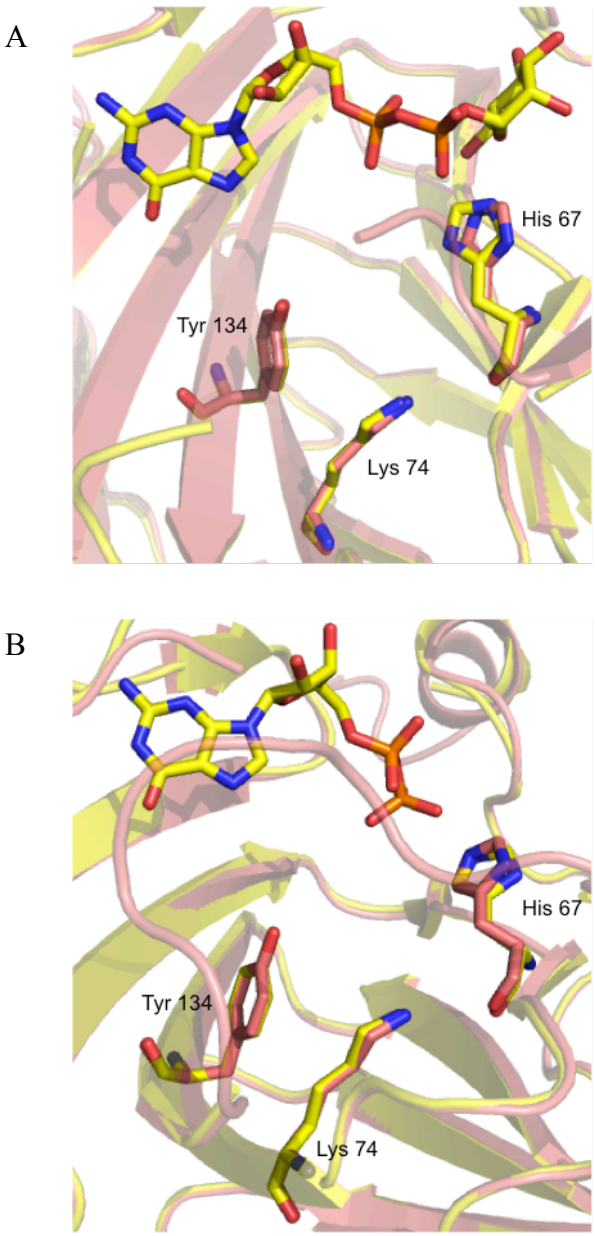


Figure 5.45. DdahB active site conformational changes upon substrate binding. The conformational changes that occur in the active site residues upon DdahB binding to GDP-mannose are shown in "A" and those that occur upon DdahB binding to GDP are shown in "B". The DdahB apo structure is shown in light red and the GDP/GDP-mannose-bound DdahB structure is shown in yellow. This image was created using MacPyMOLEdu.

This lack of specific contacts between DdahB and the mannose moiety of GDP-mannose, and the resulting flexibility meaning that it could only be confidently built into one chain, correlates with ITC data confirming that DdahB interacts with GDP with more than 10-fold higher affinity than GDP-mannose, with observed K_D values of 49.4 μM and 649 μM, respectively. This indicates that the mannose moiety hinders binding. The raw ITC data and fitting can be seen in table 5.8 and figure 5.46.

	DdahB		MlghB	
	GDP	GDP-mannose	GDP	GDP-mannose
N (sites)	1.00	1.00	1.00	1.00
K_D (M)	32.6e-6 ± 2.59e-6	419e-6 ± 41.5e-6	48.2e-6 ± 1.43e-6	410e-6 ± 30.1e-6
ΔH (kcal/mol)	-7.12 ± 0.204	-8.61 ± 0.662	15.2 ± 0.526	-18.52 ± 9.62
ΔG (kcal/mol)	-6.12	-4.61	-5.88	-4.62
-TΔS (kcal/mol)	0.998	4.00	9.31	13.9

Table 5.8. ITC output data for the interactions of DdahB and MlghB with GDP and GDP-mannose.

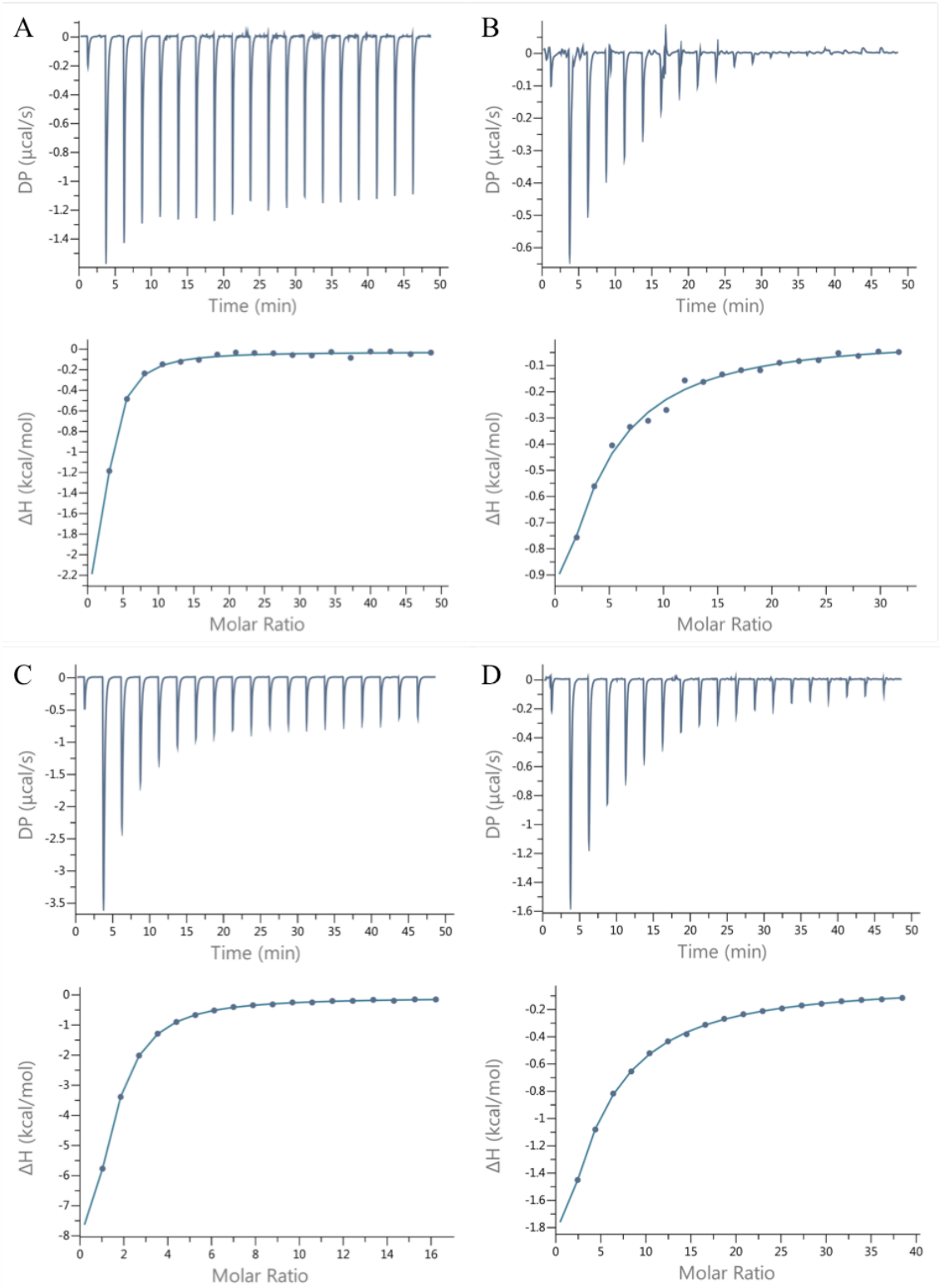


Figure 5.46. ITC data for GDP and GDP-mannose binding to DdahB and MlghB. The ITC curve for GDP binding to DdahB is shown in "A" and the ITC curve for GDP-mannose binding to DdahB is shown in "B". The ITC curve for GDP binding to MlghB is shown in "C" and the ITC curve for MlghB binding to GDP-mannose is shown in "D". These curves have been background corrected.

5.3.14 The structure of MlghB with GDP

MlghB crystallised in the presence of GDP-mannose and the resulting complex structure was solved to a resolution of 2.60 Å in space group P $2_1\,2_1\,2_1$, with four chains in the asymmetric unit. The refined model contains the following residues: 0-177 in chain A (residue 0 denotes the final residue of the linker prior to the target sequence), 2-178 in chain B, 1-177 in chain C and 2-177 in chain D. A representative X-ray diffraction pattern can be seen in figure 5.47. The data collection and refinement statistics are reported in table 5.9, with the final MolProbity validation report (Chen *et al*., 2010; Davis *et al*., 2007) shown in figure 5.48. Despite incubating MlghB with GDP-mannose, only GDP could be confidently built into the four chains. The F_o-F_c density, contoured at 2.5 σ, for a representative GDP molecule can be seen in figure 5.49. Future HPLC analysis of soaked crystals would confirm whether GDP-mannose is in fact bound to the crystals, but is not visible due to flexibility resulting from a lack of contacts between the mannose moiety and MlghB. The lack of observed density for the mannose moiety correlates with ITC data showing that MlghB binds GDP much tighter than GDP-mannose, with K_D values of 49.2 μM and 410 μM, respectively, shown in figure 5.46 and table 5.8. This lower affinity of MlghB for GDP-mannose than GDP mirrors that seen with DdahB and also indicates that the mannose moiety is a hindrance to binding. The K_D values for DdahB and MlghB binding to GDP are almost identical, indicating a shared mode of GDP binding. As is the case with DdahB binding to GDP-mannose, little conformation change occurs upon MlghB binding to GDP (figure 5.50). The C_α backbone of MlghB in absence (chain D) and in presence (chain B) of GDP can be aligned over 177 residues with an RMSD of 0.37 Å (Krissinel and Henrick, 2004). As is the case with RmlC and DdahB, the GDP moiety is bound at the entrance to the central jellyroll, however the β-phosphate is oriented away from the active site, when comparing the orientation of the GDP bound to MlghB to that of the dTDP-rhamnose bound to RmlC (Dong *et al*., 2007), shown in figure 5.51. This orientation of the GDP away from the active site is similar to the observed binding mode of DdahB to GDP-mannose (compare figures 5.41 and 5.51). The residues required for GDP binding to

MlghB are shown in figure 5.52. GDP binds in the same location on MlghB as DdahB, as seen when comparing figures 5.52 and 5.43, however additional hydrogen bonds are found between Lys 108 and the guanine moiety, Arg 25 and the 3'OH of the ribose and Arg 172 and the α-phosphate. No hydrogen bonds are formed between MlghB and the β-phosphate. As the β-phosphate of the GDP is pointing away from the active site, no conformational change occurs in the active site residues upon binding of MlghB to GDP, as shown in figure 5.53.

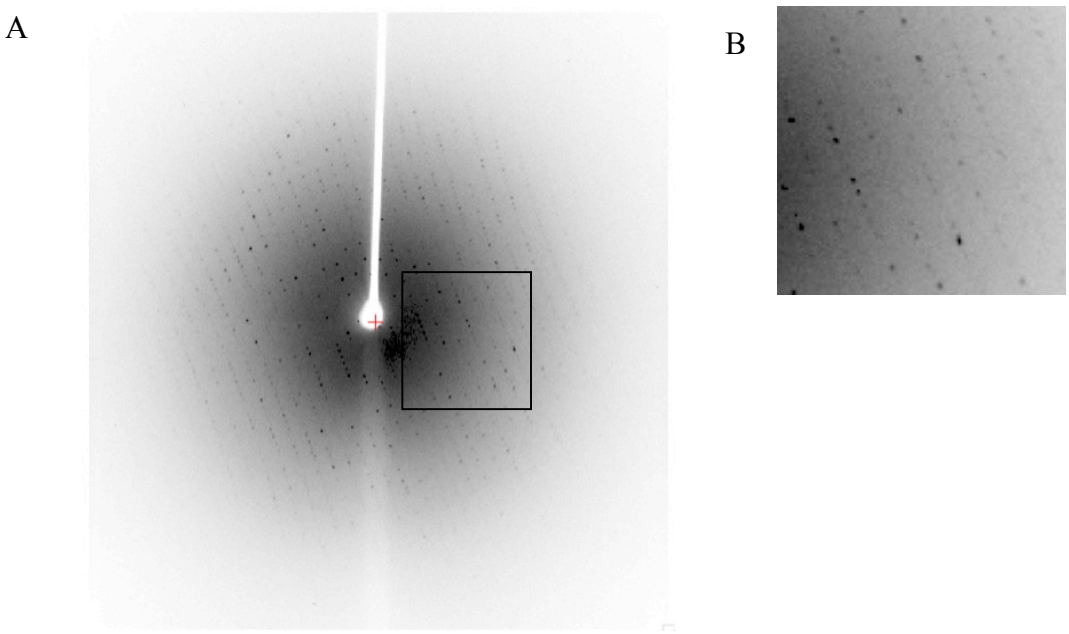


Figure 5.47. Representative X-ray diffraction image obtained from a crystal of MlghB grown in the presence of GDP-mannose. Entire X-ray diffraction pattern is shown in "A", with a zoomed view of the area highlighted in "A" with a square shown in "B".

Data Collection	
Wavelength (Å)	1.542
Space Group	P 2_1 2_1 2_1
Unit-cell a, b, c (Å)	42.35, 121.88. 154.05
Unit-cell α, β, γ (°)	90, 90, 90
Resolution range (Å)	56.67-2.60 (2.67-2.60)
I/σ(I)	12.5 (3.4)
R_{merge} (%)	16.2 (66.9)
$CC_{1/2}$	0.992 (0.812)
Completeness (%)	99.3 (91.8)
Multiplicity	7.0 (6.8)
Refinement	
R-factor (%)	21.0
R_{free} (%)	23.5
R.m.s.d. Bond Lengths (Å)	0.010
R.m.s.d. Bond Angles (°)	1.570
Mean B factor (Å^2)	26.5

Table 5.9. X-ray diffraction data collection and refinement statistics for MlghB crystallised in the presence of GDP-mannose. Values for the highest resolution shell are shown in brackets. R_{merge} is defined as: $\Sigma_{hkl} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection *hkl* and $\langle I(hkl) \rangle$ is the average intensity of reflection *hkl*. The R-factor is defined as: $\Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where F_{obs} is the observed structure-factor amplitudes and F_{calc} is the calculated structure-factor amplitudes. R_{free} was calculated, using the R-factor equation, using a 5 % subset of reflections that were not used in refinement.

All-Atom Contacts	Clashscore, all atoms:	2.04		100th percentile* (N=227, 2.60Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	617	95.96%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	696	99.43%	Goal: >98%
	MolProbity score^	0.97		100th percentile* (N=6237, 2.60Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 6022	0.00%	Goal: 0%
	Bad angles:	4 / 8174	0.05%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 32	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	4 / 672	0.60%	Goal: <0.05%
Low-resolution Criteria	CaBLAM outliers	8	1.15%	Goal: <1.0%
	CA Geometry outliers	3	0.43%	Goal: <0.5%

Figure 5.48. MolProbity validation report for the refined model of MlghB with GDP-mannose. Both the Clashscore and MolProbity score place this structure in the 100th percentile (Chen *et al*., 2010; Davis *et al*., 2007).

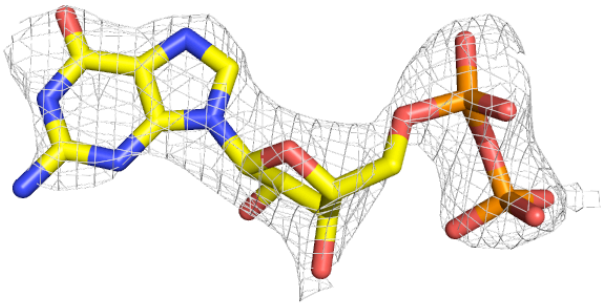


Figure 5.49. F_O-F_C density for GDP bound to MlghB. The F_O-F_C electron density for GDP bound to MlghB is contoured at 2.5 σ. The F_O-F_C density was calculated by deleting GDP from the final refined model, resetting the B factors to 20 $Å^2$ using the TLSMD server (Painter and Merritt, 2006a, 2006b), and refining in REFMAC5 (Murshudov *et al*., 1997), with 50 cycles of refinement. The map was created using FFT, as part of the CCP4 suite (Winn *et al*., 2011). This image was created using MacPyMOLEdu.

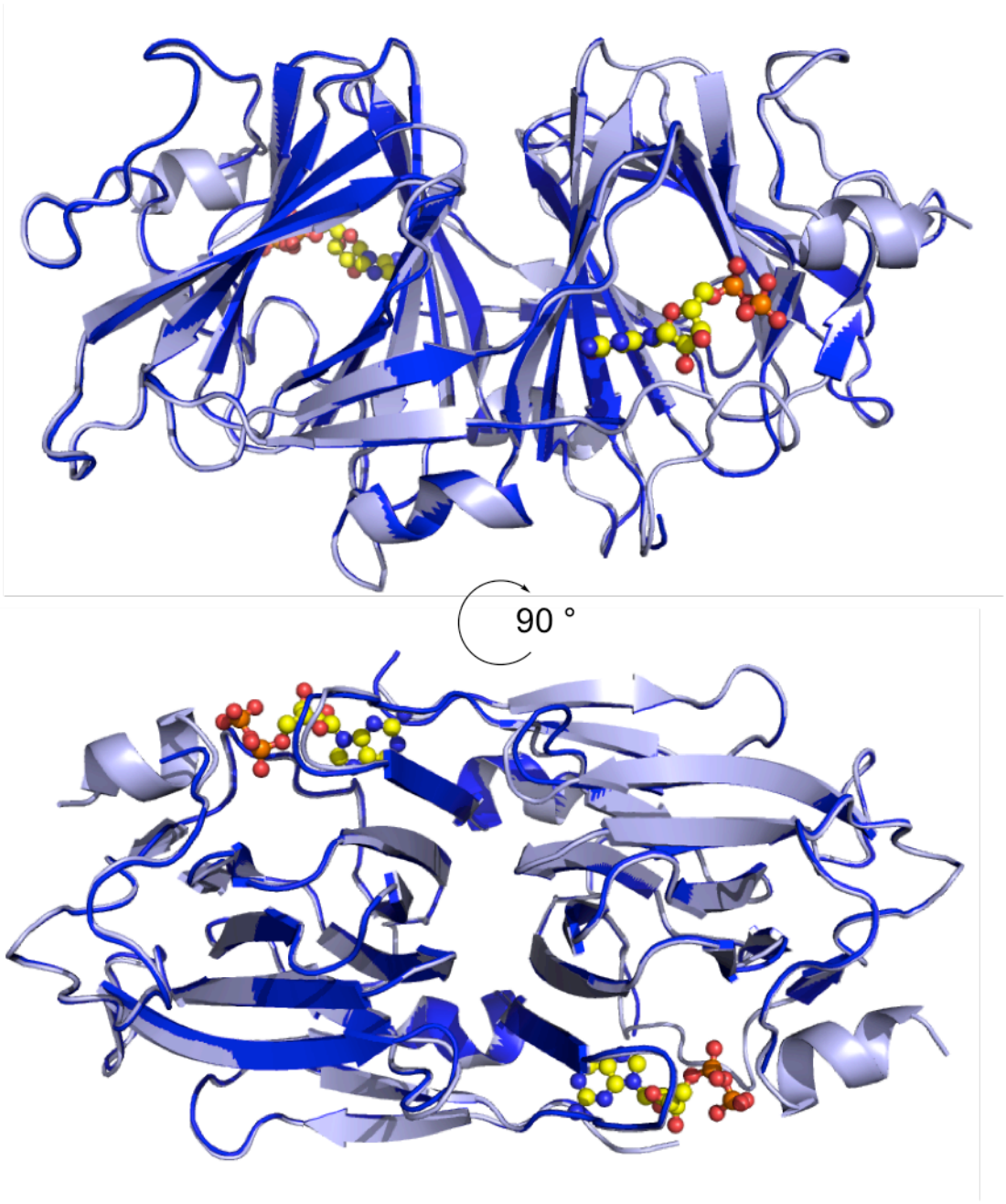


Figure 5.50. Overlay of MlghB apo structure and GDP-bound MlghB structure. MlghB apo structure is shown in dark blue and the GDP-bound structure in lilac. GDP is shown in yellow and is bound at the entrance of the jellyroll. This image was created using MacPyMOLEdu.

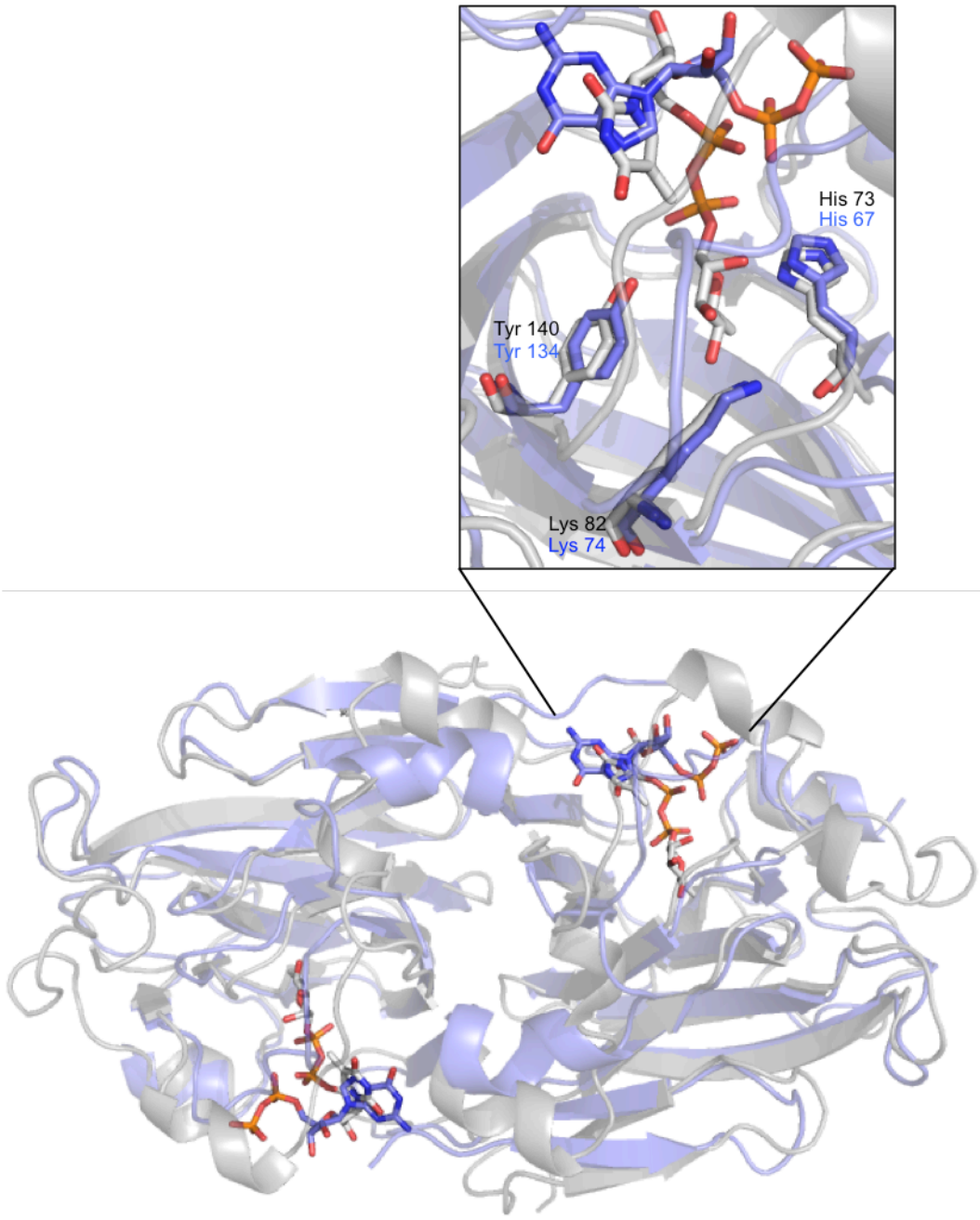


Figure 5.51. A comparison of the RmlC dTDP-rhamnose binding site with the MlghB GDP binding site. The rhamnose moiety of dTDP-rhamnose is oriented into the jellyroll of RmlC (Dong *et al*., 2007), whereas the GDP is oriented away from the jellyroll of MlghB. RmlC and dTDP-rhamnose are both shown in grey. MlghB and GDP are both shown in blue. (RmlC PDB: 2IXL (Dong *et al*., 2007)). The residue names in black correspond to RmlC and those in blue to MlghB. This image was created using MacPyMOLEdu.

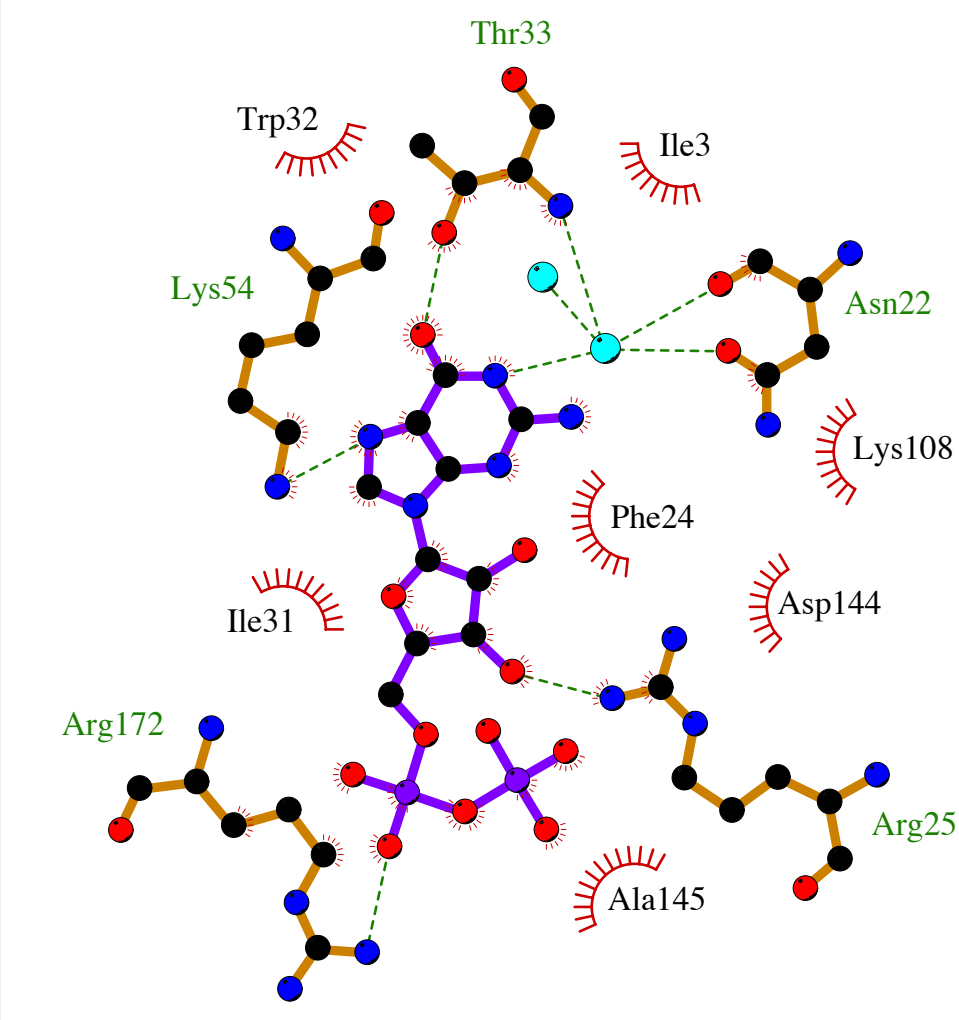


Figure 5.52. The amino acids required for MlghB binding to GDP. This figure was created using LigPlot+ (Laskowski and Swindells, 2011). Waters are shown in cyan, carbon is shown in black, nitrogen is shown in blue, oxygen is shown in red, and phosphorus is shown in purple. Bonds within the protein are shown in brown and bonds within the ligand are shown in purple. Hydrogen bonds and names of the participating amino acids are shown in green. Hydrophobic interactions are shown in red and the names of the participating amino acids are shown in black.

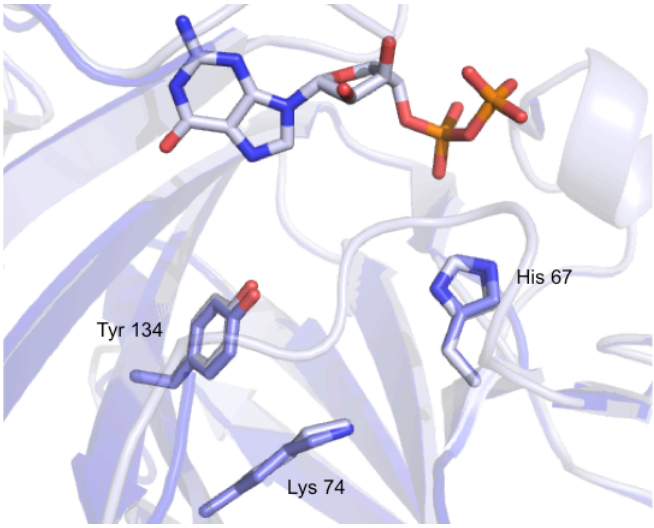


Figure 5.53. GDP binding to MlghB does not induce an active site conformational change. The positions of the active site residues in the apo structure are shown in dark blue and those upon GDP binding are shown in light blue. This image was constructed using MacPyMOLEdu.

5.3.15 Expression, purification and crystallisation of non-TEV protease-cleavable His_6-tagged DdahC and MlghC

Following the DdahB-catalysed epimerisation of GDP-4-keto-6-deoxy-D-*lyxo*-heptose at the C3 position to give GDP-4-keto-6-deoxy-D-*arabino*-heptose in GDP-6-deoxy-D-*altro*-heptose biosynthesis and the MlghB-catalysed epimerisation of GDP-(6-*O*-methyl)-4-keto-D-*lyxo*-heptose at the C3 and C5 positions to give GDP-(6-*O*-methyl)-4-keto-L-*xylo*-heptose during GDP-*O*-methyl-L-*gluco*-heptose biosynthesis, both products of epimerisation are reduced at the C4 positions by DdahC and MlghC to give GDP-6-deoxy-D-*altro*-heptose and GDP-(6-*O*-methyl)-L-*gluco*-heptose, respectively. Expression of DdahC (42 kDa) and MlghC (41 kDa), both with a non-TEV protease-cleavable N-terminal His_6-tag, was confirmed via western blotting against the N-terminal His_6 affinity tag and can be seen in figures 5.54 and 5.55, respectively. The successful purification of DdahC and MlghC was confirmed via SDS-PAGE gel, shown in figure 5.56. The identity of these purified proteins was confirmed by mass spectrometry. Both DdahC and MlghC purify as monodisperse dimers, as confirmed via SEC-MALS, shown in figure 5.57 and table 5.10. Purification yields of approximately 8

mg/L of *E. coli* culture and 15 mg/L *E. coli* culture were obtained for DdahC and MlghC, respectively.

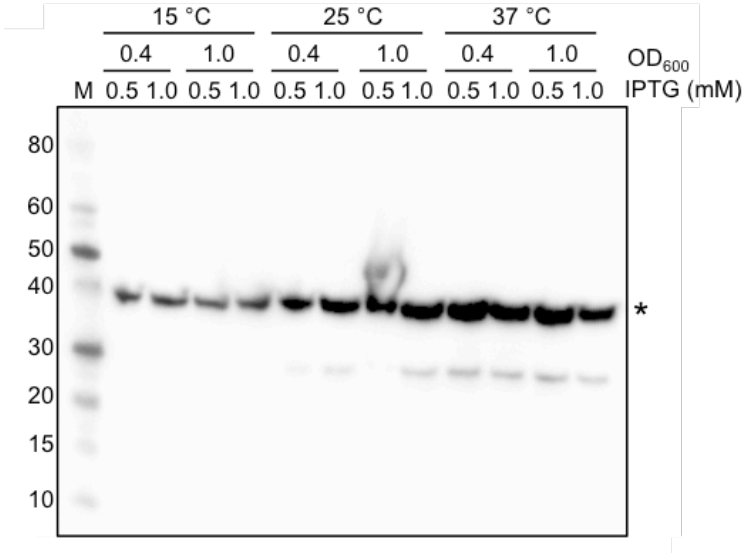


Figure 5.54. Western blot probed with anti-His tag antibody of whole cells expressing non-TEV protease-cleavable His_6-tagged DdahC under different conditions. The band at approximately 40 kDa, indicated with a "*", corresponds to His_6-tagged DdahC.

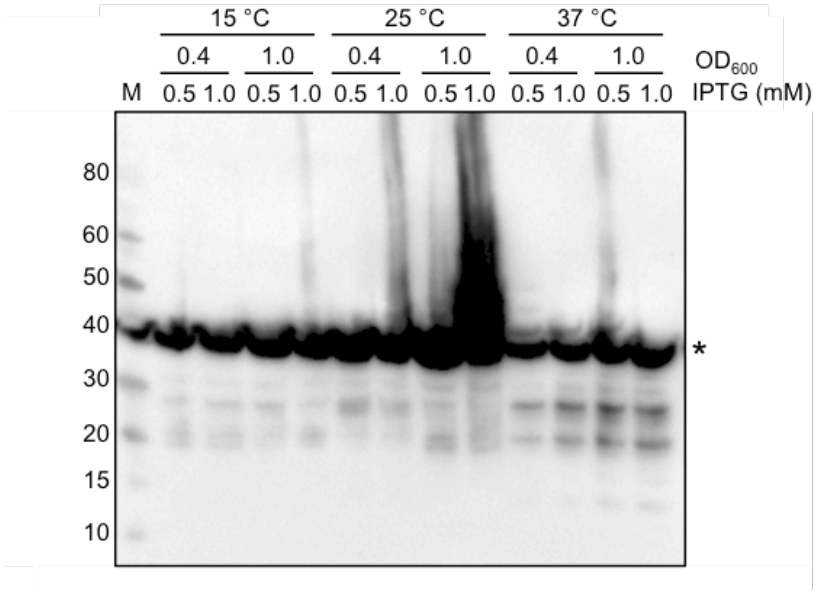


Figure 5.55. Western blot probed with anti-His tag antibody of whole cells expressing non-TEV protease-cleavable His_6-tagged MlghC under different conditions. The band at approximately 40 kDa, indicated with a "*", corresponds to His_6-tagged MlghC..

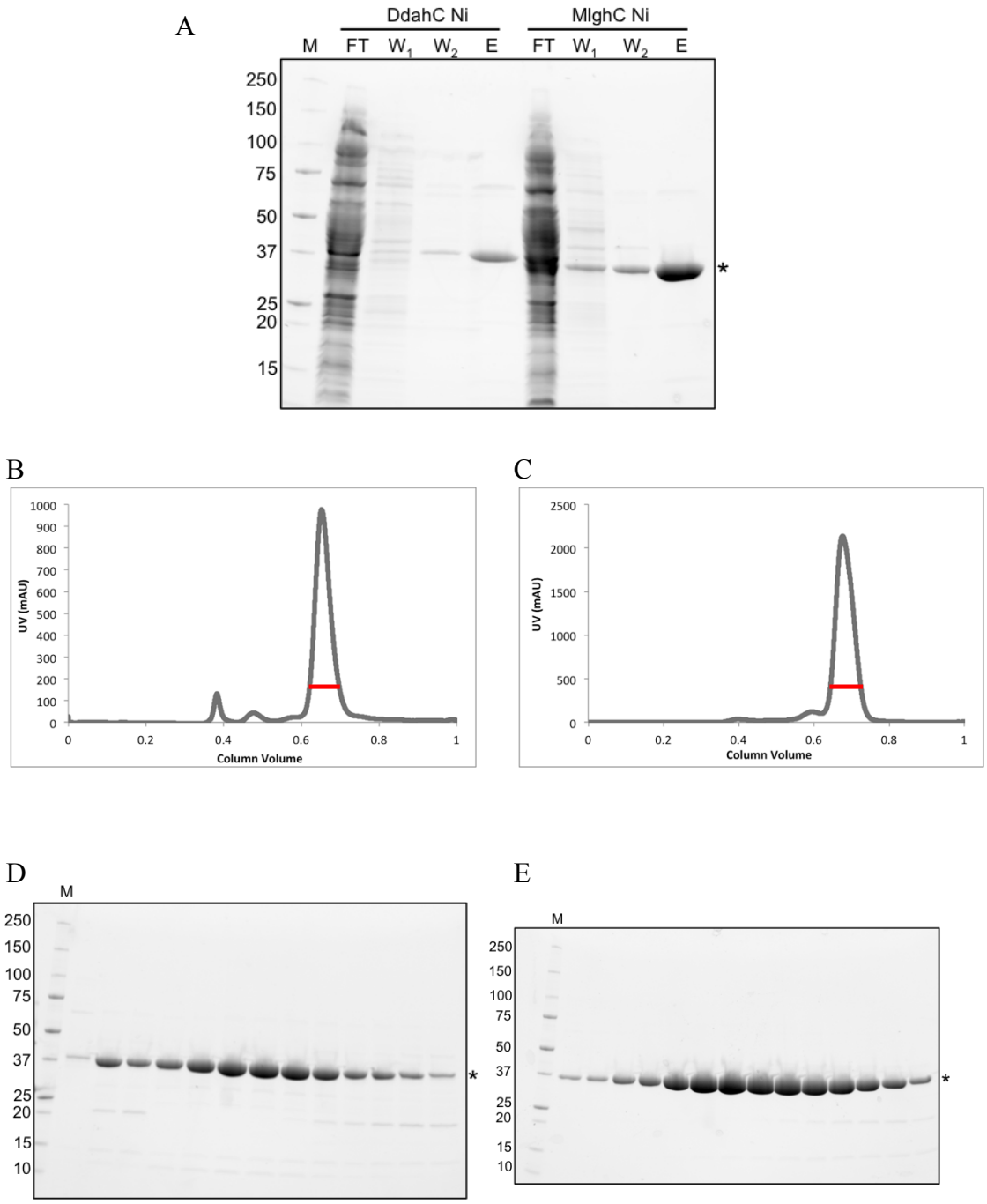


Figure 5.56. Purification of non-TEV protease-cleavable His_6-tagged DdahC and His_6-tagged MlghC. The SDS-PAGE gel of the nickel affinity purification of DdahC and MlghC is shown in "A". (M: marker; FT: FT; W_1: wash 1; W_2: wash 2; E: eluate). "B" and "C" show the SEC UV elution profiles of DdahC and MlghC, respectively. "D" and "E" show the SDS-PAGE gel of the SEC fractions across the peaks shown in "B" and "C", with "D" corresponding to DdahC and "E" to MlghC. The void volume of the SEC column equals approximately 0.38 CV. The bands at approximately 37 kDa, indicated with a "*", correspond to DdahC and MlghC. The approximate position of the fractions on "B" and "C" shown in "D" and "E", respectively, is indicated with a red line.

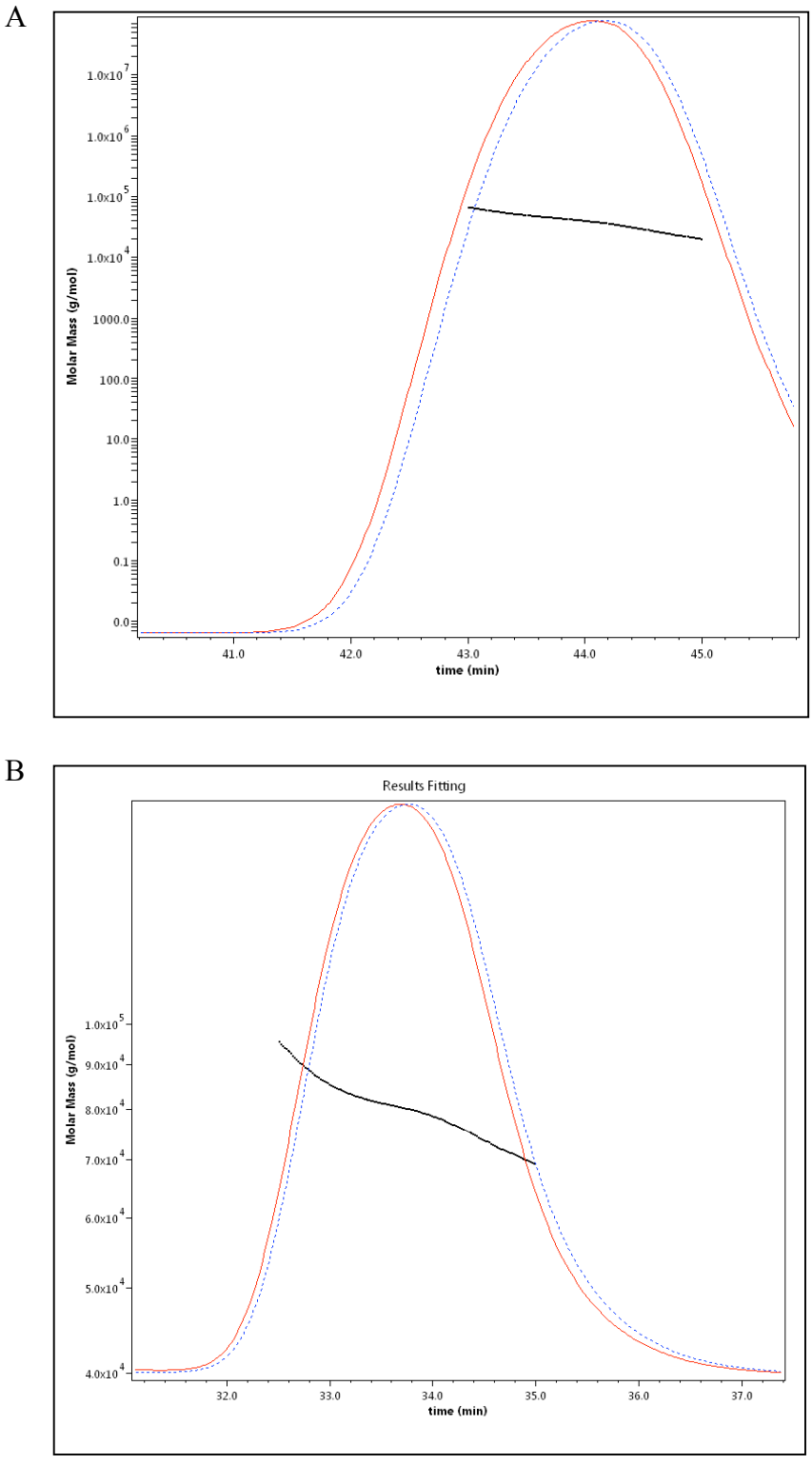


Figure 5.57. SEC-MALS profiles of non-TEV protease-cleavable His_6-tagged DdahC and MlghC. Both DdahC (shown in "A") and MlghC (shown in "B") were found to form monodisperse dimers in solution. The change in light scattering over time is shown in red and the change in refractive index over time is shown in blue. The change in molar mass (g/mol) over time is shown in black.

	DdahC	MlghC
Molar mass moments (g/mol)		
Mn	9.7×10^4 (± 0.7 %)	7.9×10^4 (± 0.4 %)
Mw	9.8×10^5 (± 0.7 %)	8.0×10^4 (± 0.4 %)
Polydispersity		
Mw/Mn	1.0 (± 0.998 %)	1.0 (± 0.6 %)

Table 5.10. SEC-MALS results for non-TEV protease-cleavable His_6-tagged DdahC and MlghC. The monomeric molecular weight of DdahC is 42 kDa and the monomeric molecular weight of MlghC is 41 kDa. Therefore values of 97 kDa and 79 kDa for the number average molecular weight (Mn), respectively, and values of 98 kDa and 80 kDa for the weight average molecular weight (Mw), respectively, show that DdahC and MlghC are both dimers in solution. Mn is defined as: $\Sigma N_i M_i^2 / \Sigma N_i M_i$, where M_i corresponds to the molecular weight of the protein and N_i corresponds to the number of proteins of that molecular weight. Mw is defined as: $\Sigma N_i M_i / \Sigma N_i$. Polydispersity is the range of molecular weights occupied by the sample and is defined as Mw/Mn, with 1.0 corresponding to a monodisperse sample. The higher the polydispersity value, the wider the range of molecular weights occupied by a sample.

As both target proteins were monodisperse, they were set up in sparse matrix sitting-drop vapour diffusion crystallisation trials. Crystals were obtained for MlghC in our in-house St Andrews 3 screen well B1 (26.2 % (w/v) PEG 1500, 0.1 M MES pH 6.0, 0.14 M sodium-potassium phosphate) (Molecular Dimensions) (appendix 8.4.3) in a drop volume ratio of 1 protein: 1 mother liquor. These crystals appeared full size (330 µm x 30 µm) following 1 day of incubation at 20 °C and can be seen in figure 5.58. Crystals could be obtained for DdahC in Wizard 1 & 2 Classic well B9 (20 % (w/v) PEG 8000, 0.1 M HEPES pH 7.5) (Molecular Dimensions). Crystals appeared in this condition, in a drop containing protein and mother liquor in a 1:1 volume ratio, after 1 day, reaching a full size of 180 µm in length and 20 µm in width after 3 days of incubation at 20 °C, and can be seen in figure 5.59. Following optimisation of crystallisation conditions, these crystals diffracted poorly when screened using our in-house source.

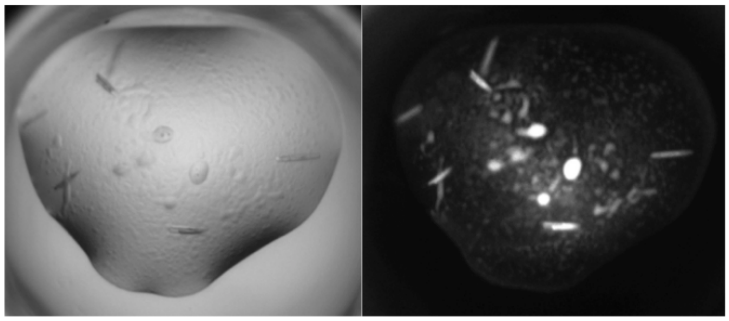


Figure 5.58. Crystals of non-TEV protease-cleavable His_6-tagged MlghC. These crystals were obtained in our in-house St Andrews 3 screen well B1 (Molecular Dimensions). The normal image is shown on the left and the UV image on the right. Images were taken using a Minstrel HT-UV imaging system.

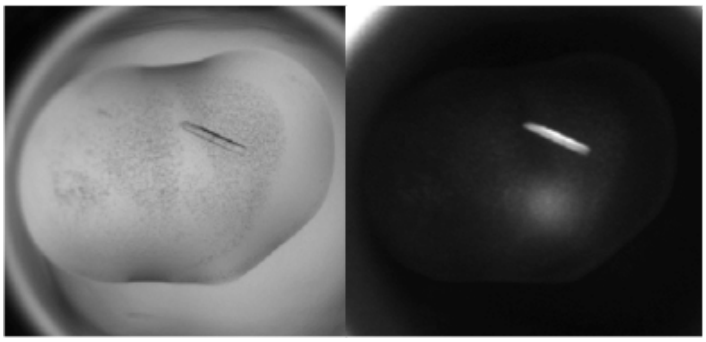


Figure 5.59. Crystals of non-TEV protease-cleavable His_6-tagged DdahC. These crystals were obtained in Wizard 1 & 2 Classic well B9 (Molecular Dimensions). The normal image is shown on the left and the UV image on the right. Images were taken using a Minstrel HT-UV imaging system.

5.3.16 Expression and purification of TEV protease-cleavable DdahC and MlghC

As only poorly diffracting crystals could be obtained with the non-TEV protease cleavable constructs of DdahC and MlghC, the inherent flexibility of the N-terminal His_6 tag was proposed to disrupt crystal contacts, resulting in poor crystal packing. A TEV protease protease site was therefore introduced, via SDM, between the N-terminal His_6 tag and the target protein sequence (Liu and Naismith, 2008). Expression of TEV protease-cleavable DdahC (Mw: 40 kDa) and MlghC (Mw: 39 kDa) was confirmed via western blotting against the N-terminal His_6 tag, shown in figures 5.60 and 5.61, respectively. These constructs could be purified and the results of the purification of TEV protease-cleavable DdahC and TEV protease-cleavable MlghC can be seen in figure 5.62. As well as removing the flexible His_6 tag to promote crystallisation, the introduction of the TEV protease site allowed the addition of a second nickel affinity column, allowing further purification of the target protein, as can be seen when comparing figures 5.56 and 5.62. However, this also resulted in a decreased purification yield of approximately 2-5 mg/L of *E. coli* cell culture for both TEV protease-cleaved DdahC and TEV protease-cleaved MlghC, as a large proportion of the protein was not cleaved by the TEV protease and was therefore present in the eluate of the second nickel affinity column.

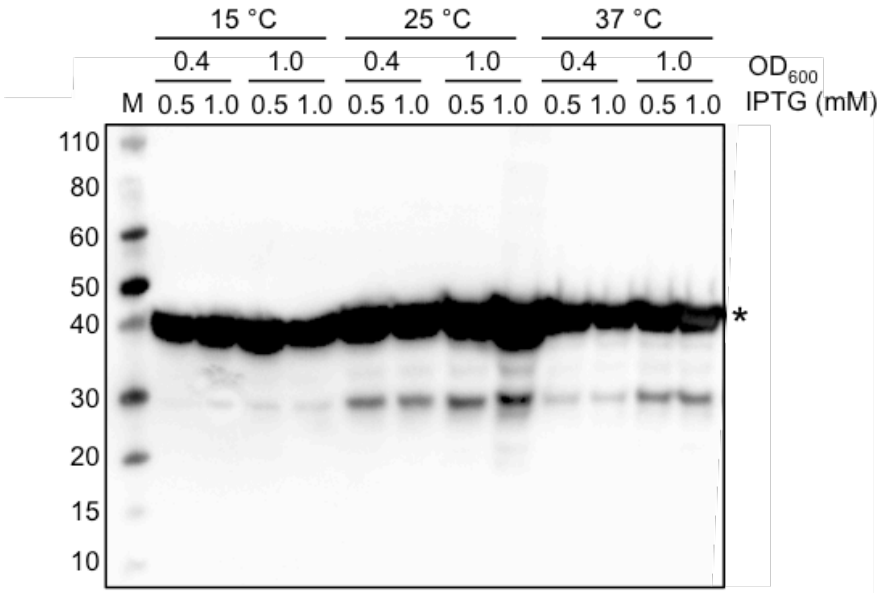


Figure 5.60. Western blot probed with an anti-His tag antibody of whole cells expressing TEV protease-cleavable DdahC under different conditions. The band at approximately 40 kDa, indicated with a "*", corresponds to DdahC.

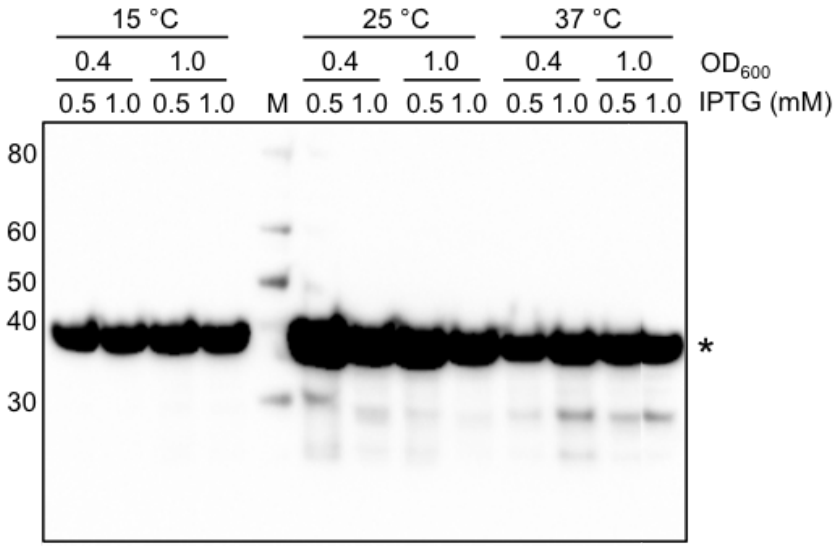


Figure 5.61. Western blot probed with an anti-His tag antibody of whole cells expressing TEV protease-cleavable MlghhC under different conditions. The band at approximately 38 kDa, indicated with a "*", corresponds to MlghC.

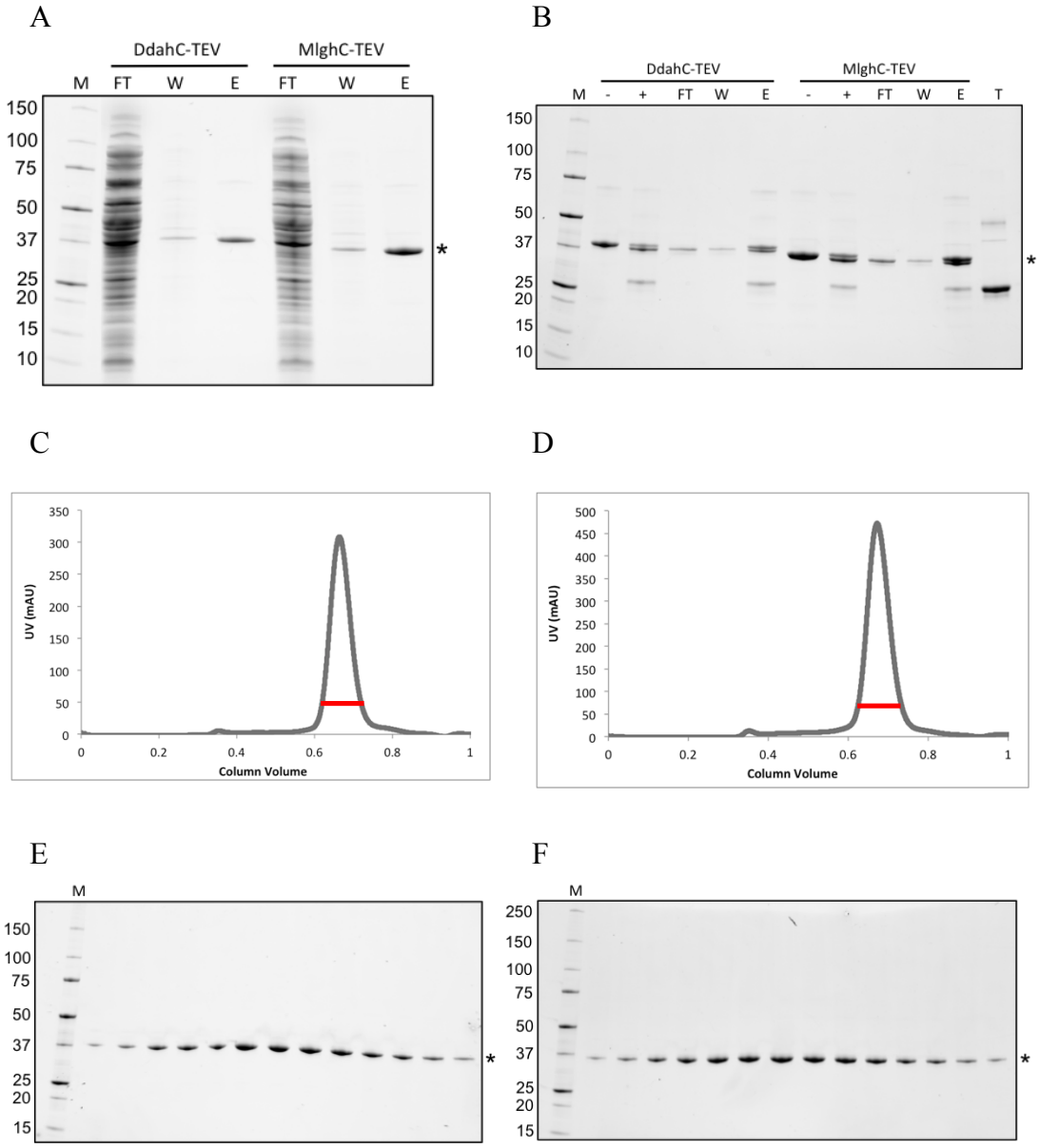


Figure 5.62. Purification of TEV protease-cleavable DdahC and MlghC. The SDS-PAGE gel of the first nickel affinity column is shown in "A". The SDS-PAGE gel of the second nickel affinity column after TEV protease-cleavage is shown in "B" (M: marker; -: first nickel affinity eluate; +: TEV protease-cleaved sample taken before the second nickel affinity column; FT: FT; W: wash; E: eluate). The SEC UV chromatograms for DdahC and MlghC are shown in "C" and "D", respectively. The SDS-PAGE gel of SEC fractions taken across the peak is shown in "E" for DdahC and "F" for MlghC. The approximate positions of the fractions on "C" and "D" shown in "E" and "F", respectively, is indicated with a red line.

5.3.17 Crystallisation of TEV protease-cleaved DdahC and MlghC

The removal of the N-terminal His_6-tag from DdahC and MlghC not only resulted in purer protein. TEV protease-cleaved DdahC crystallised in our in-house St Andrews 16 screen well F5 (53.38 % (w/v) PEG 400, 0.1 M HEPES pH 7.5, 0.08 M ammonium citrate) (Molecular Dimensions) (appendix 8.4.4). Crystals appeared after six days and reached a full size of approximately 450 μm x 70 μm after a total of eight days incubation at 20 °C, in a drop volume ratio of 1 protein: 1 mother liquor. A TEV protease-cleaved DdahC crystal, along with a representative X-ray diffraction image, can be seen in figure 5.63.

TEV protease-cleaved MlghC crystallised in our in-house St Andrews PEG 1 screen wells B10 (24.63 % (w/v) PEG 8000, 0.1 M Bicine pH 8.5, 0.12 M sodium citrate 0.05 % (w/v) LDAO) and H4 (24.43 % (w/v) PEG 6000, 0.1 M sodium citrate pH 5.5, 0.06 M ammonium phosphate) (Molecular Dimensions) (appendix 8.4.5). Crystals obtained in St Andrews PEG 1 B10 (Molecular Dimensions) appeared after one day and reached a full size of 200 μm in length after a total of two days incubation at 20 °C, in a drop volume ratio of 1 protein: 1 mother liquor. Crystals of TEV protease-cleaved MlghC, along with a representative diffraction image, can be seen in figure 5.64.

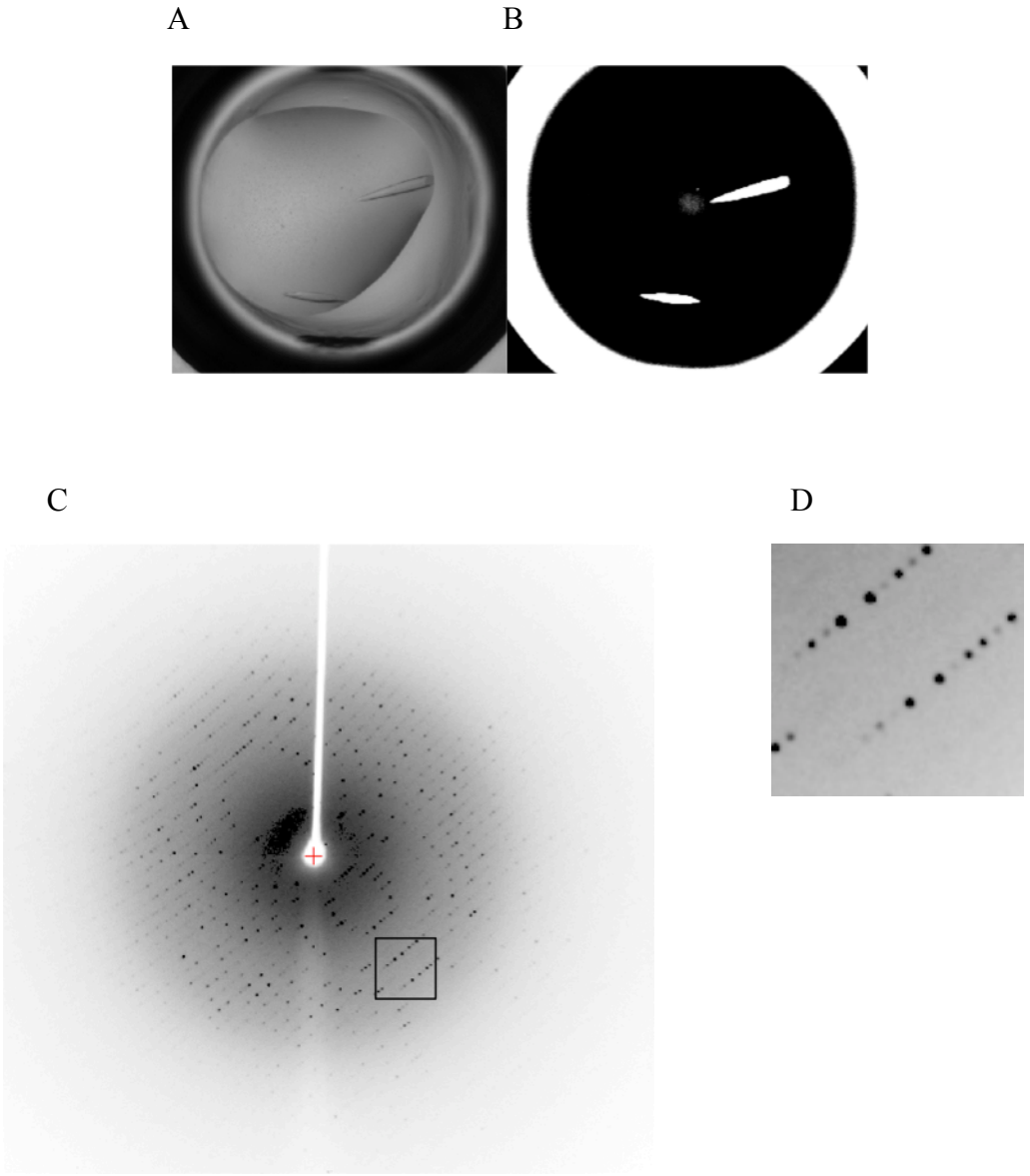


Figure 5.63. Crystals and representative X-ray diffraction image of TEV protease-cleaved DdahC. These crystals were obtained in our in-house St Andrews 16 screen well F5 (Molecular Dimensions). The normal image is shown in "A" and the UV image in "B". The images in "A" and "B" were taken using a Minstrel HT-UV imaging system. A representative X-ray diffraction image is shown in "C", with a zoomed view of the area highlighted in "C" with a square shown in "D".

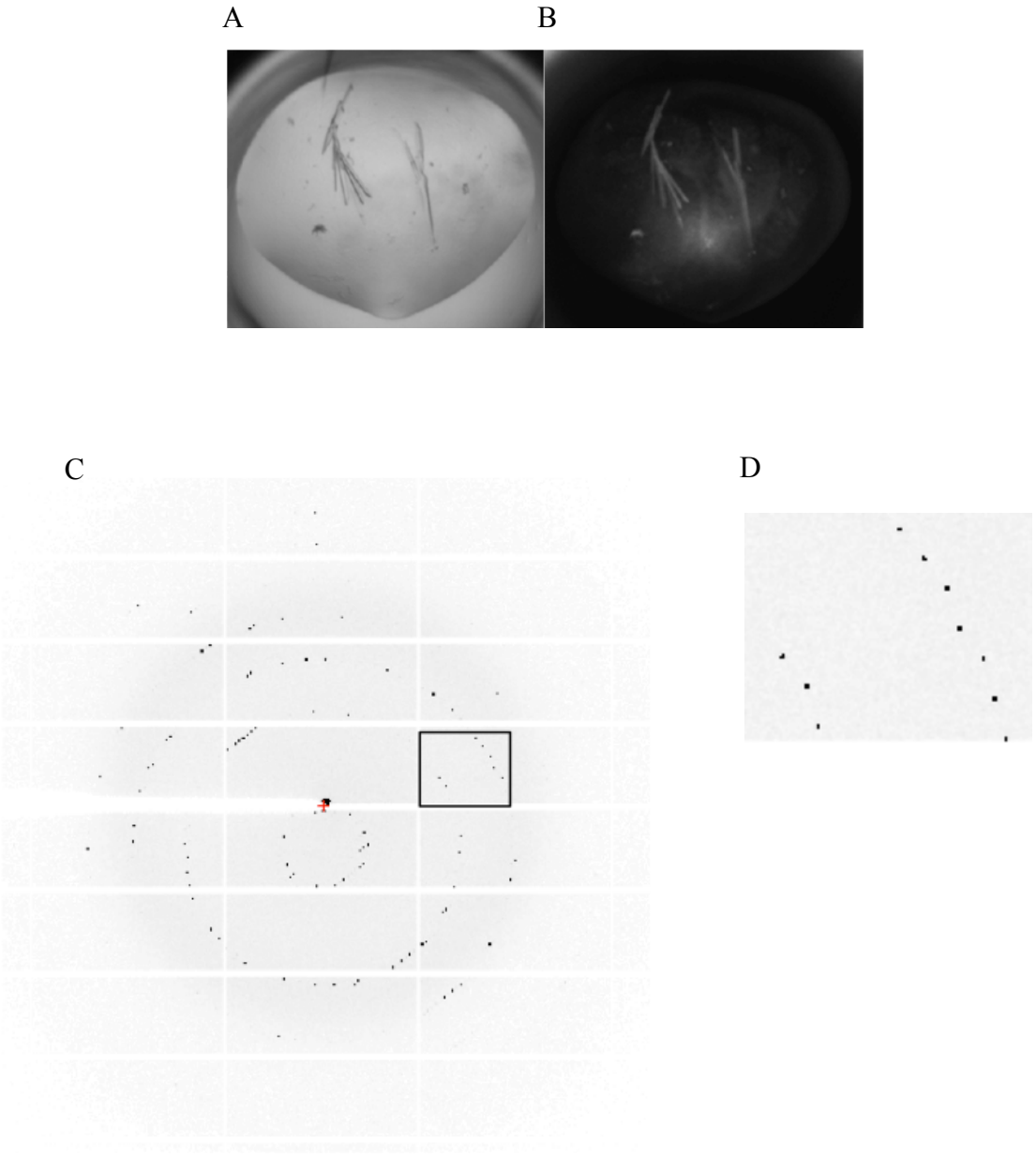


Figure 5.64. Crystals and representative X-ray diffraction image of TEV protease-cleaved MlghC. These crystals were obtained in our in-house St Andrews PEG 1 screen well B10 (Molecular Dimensions). The normal image is shown in "A" and the UV image in "B". The images in "A" and "B" were taken using a Minstrel HT-UV imaging system. A representative X-ray diffraction image is shown in "C", with a zoomed view of the area highlighted with a ssquare in "C" shown in "D".

5.3.18 The crystal structures of TEV protease-cleaved DdahC and MlghC

The structures of TEV protease-cleaved DdahC and MlghC were both solved via molecular replacement. The crystal of TEV protease-cleaved MlghC diffracted to 1.66 Å resolution and belonged to space group P 1 2_1 1, with two chains in the asymmetric unit. TEV protease-cleaved DdahC diffracted to 2.08 Å resolution, in space group 4_2 2_1 2, with only one chain in the asymmetric unit. Tables 5.11 and 5.12 show the data collection and refinement statistics for MlghC and DdahC, respectively. MolProbity validation reports (Chen *et al*., 2010; Davis *et al*., 2007) can be seen in figures 5.65 and 5.66 for MlghC and DdahC, respectively.

Data Collection	
Wavelength (Å)	0.982
Space Group	P 1 2_1 1
Unit-cell a, b, c (Å)	57.80, 131.96, 59.33
Unit-cell α, β, γ (°)	90, 105.84, 90
Resolution range (Å)	65.98 – 1.66 (1.70 – 1.66)
I/σ(I)	11.8 (1.2)
R_{merge} (%)	5.9 (76.0)
$CC_{1/2}$	0.999 (0.519)
Completeness (%)	99.2 (99.1)
Multiplicity	3.7 (3.1)
Refinement	
R-factor (%)	18.0
R_{free} (%)	20.2
R.m.s.d. Bond Lengths (Å)	0.012
R.m.s.d. Bond Angles (°)	1.487
Mean B factor (Å^2)	26.0

Table 5.11. X-ray diffraction data collection and refinement statistics for TEV protease-cleaved MlghC. Values for the highest resolution shell are shown in brackets. R_{merge} is defined as: $\Sigma_{hkl} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity of reflection hkl. The R-factor is defined as: $\Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where F_{obs} is the observed structure-factor amplitudes and F_{calc} is the calculated structure-factor amplitudes. R_{free} was calculated, using the R-factor equation, using a 5 % subset of reflections that were not used in refinement.

All-Atom Contacts	Clashscore, all atoms:	1.04		99th percentile* (N=791, 1.66Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	543	97.66%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	624	98.42%	Goal: >98%
	MolProbity score^	0.80		100th percentile* (N=8776, 1.66Å ± 0.25Å)
	Cβ deviations >0.25Å	3	0.50%	Goal: 0
	Bad bonds:	0 / 5286	0.00%	Goal: 0%
	Bad angles:	3 / 7111	0.04%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	2 / 10	20.00%	Expected: ≤1 per chain, or ≤5%

Figure 5.65. MolProbity validation report for the refined model of TEV protease-cleaved MlghC. The Clashscore and MolProbity score place this structure in the 99th and 100th percentile, respectively (Chen *et al*., 2010; Davis *et al*., 2007).

Data Collection	
Wavelength (Å)	1.542
Space Group	P $4_2 2_1 2$
Unit-cell a, b, c (Å)	131.801, 131.801, 50.17
Unit-cell α, β, γ (°)	90, 90, 90
Resolution range (Å)	33.05 – 2.08 (2.14 – 2.08)
I/σ(I)	11.5 (3.3)
R_{merge} (%)	16.3 (61.0)
$CC_{1/2}$	0.946 (0.798)
Completeness (%)	99.9 (99.0)
Multiplicity	12.5 (7.9)
Refinement	
R-factor (%)	19.0
R_{free} (%)	23.2
R.m.s.d. Bond Lengths (Å)	0.019
R.m.s.d. Bond Angles (°)	1.89
Mean B factor (Å^2)	26.2

Table 5.12. X-ray diffraction data collection and refinement statistics for TEV protease-cleaved DdahC. Values for the highest resolution shell are shown in brackets. R_{merge} is defined as: $\Sigma_{hkl} \Sigma_i \left| I_i(hkl) - \langle I(hkl) \rangle \right| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity of reflection hkl. The R-factor is defined as: $\Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where F_{obs} is the observed structure-factor amplitudes and F_{calc} is the calculated structure-factor amplitudes. R_{free} was calculated, using the R-factor equation, using a 5 % subset of reflections that were not used in refinement.

All-Atom Contacts	Clashscore, all atoms:	2.12		99th percentile* (N=577, 2.08Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	0.65%	Goal: <0.3%
	Favored rotamers	296	96.42%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	335	98.24%	Goal: >98%
	MolProbity score^	0.98		100th percentile* (N=11640, 2.08Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 2864	0.00%	Goal: 0%
	Bad angles:	2 / 3852	0.05%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	1 / 8	12.50%	Expected: ≤1 per chain, or ≤5%

Figure 5.66. MolProbity validation report for the refined model of TEV protease-cleaved DdahC. The Clashscore and the MolProbity score place this structure in the 99th and 100th percentile, respectively (Chen *et al*., 2010; Davis *et al*., 2007).

Out of a possible 347 amino acids, the refined MlghC model consists of amino acids 0-263, 268-304 and 310-344 in chain A and amino acids 0-172, 179-263, 268-305 and 310-344 in chain B (0 denotes the final amino acid of the TEV protease-site between the N-terminal His_6-tag and the target sequence). An MlghC monomer consists of eight β-strands and 15 α-helices. Out of a possible 353 amino acids, the refined DdahC model consists of amino acids 0-308 and 313-352 (0 again denotes the final amino acid of the TEV protease-site between the N-terminal His_6-tag and the target sequence). The DdahC monomer consists of seven β-strands and 16 α-helices. Both the DdahC and MlghC monomers can be divided into an N- and C-terminal domain. The N-terminal domain is a Rossmann fold (Hanukoglu, 2015). The central β-sheet of the Rossmann fold of DdahC is formed by five parallel β-strands, with a strand topology of 1-2-3-4-6 and is surrounded by seven α-helices. The central β-sheet of the Rossmann fold of MlghC consists of four parallel β-strands, with a strand topology of 1-2-3-4 and is surrounded by six α-helices (Hanukoglu, 2015). The globular C-terminal domain of both DdahC and MlghC consists of five α-helices and two β-strands and is kinked relative to the N-terminal domain.

SEC-MALS confirms that DdahC and MlghC are both dimers in solution (figure 5.57 and table 5.9). Analysis of the assembly using the PDBePISA server also confirms this. The CSS score of the DdahC assembly generated by PDBePISA is 1 and that of the MlghC assembly is 0.593 for the interaction between each chain and 0.407 for the interaction between each chain and its cofactor (Krissinel, 2009; Krissinel and Henrick, 2007). DdahC and MlghC homodimers are shown in figures 5.67 and 5.68, respectively. DdahC and MlghC dimerize around a two-fold axis and the dimerization interface consists of moderate and weak electrostatic interactions, as well as hydrophobic interactions, between α-helices 3, 4 and 7, which form a helical bundle. These helices also form part of the N-terminal Rossmann fold (Hanukoglu, 2015). MlghC also dimerises via additional moderate to weak electrostatic interactions between the loops connecting α-helices 6 and 7, and 12 and 13. The presence of the N-terminal Rossmann fold, containing conserved YXXXK and GXXGXXG motifs, as well as the described dimerization mode, allow DdahC and MlghC to be assigned, along with DmhA and

RmlB previously described (Allard *et al*., 2001, 2002), as members of the SDR family (Kavanagh *et al*., 2008).

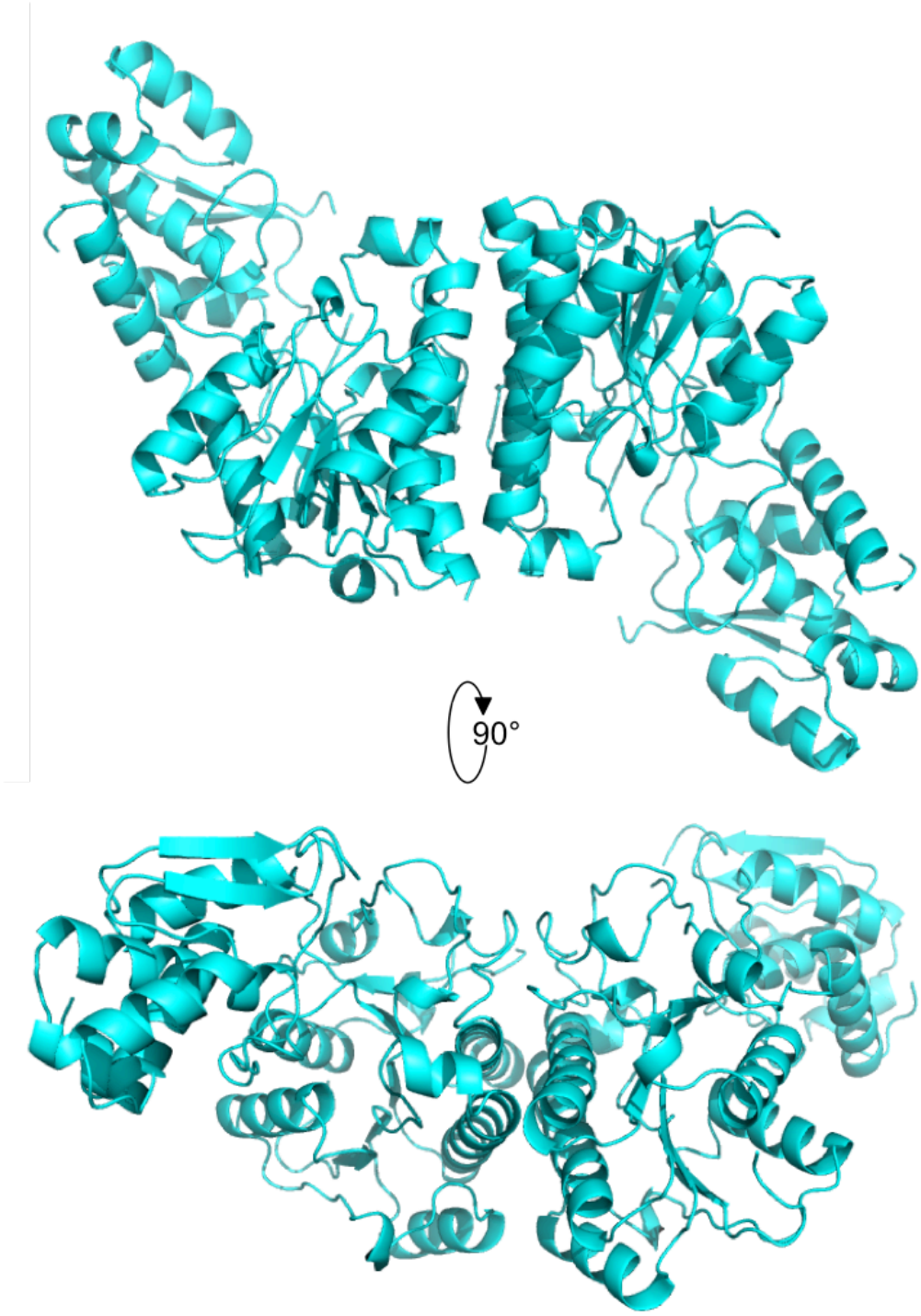


Figure 5.67. DdahC is a homodimer. DdahC dimerizes via an α-helical bundle formed between α-helices 3, 4 and 7. This image was created using MacPyMOLEdu.

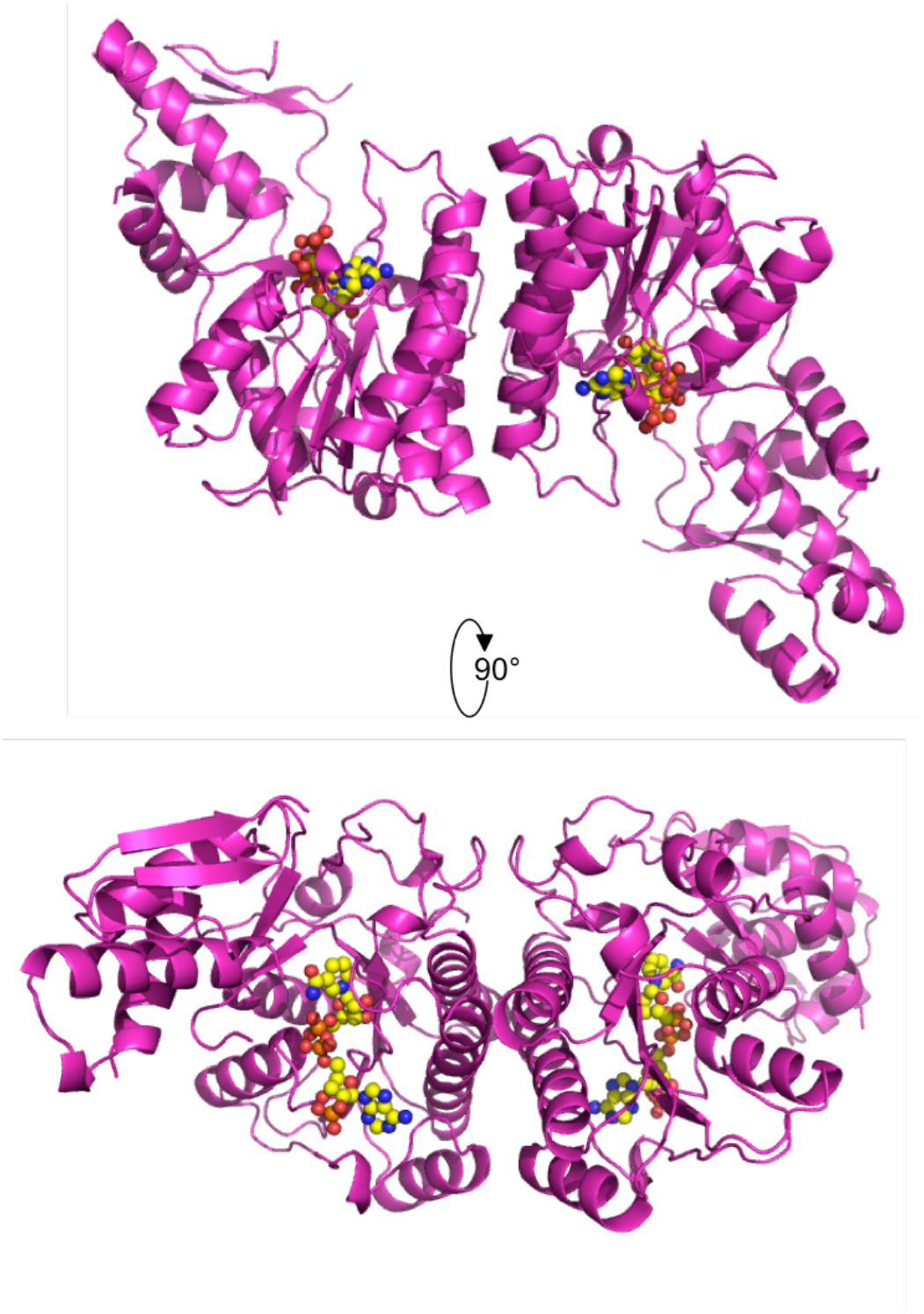


Figure 5.68. MlghC is a homodimer. MlghC dimerizes via an α-helical bundle formed between α-helices 3, 4 and 7. These α-helices also form part of the Rossmann fold (Hanukoglu, 2015). $NADP^+$ is shown in yellow ball and stick representation and binds at the base of the Rossmann fold. This image was constructed using MacPyMOLEdu.

5.3.19 A comparison of the structures of DdahC and MlghC with the dTDP-6-deoxy-L-*lyxo*-hexulose reductase RmlD

The DdahC- and MlghC-catalysed reduction of their respective substrates at the C4 position, as the final step in modified heptose biosynthesis, shares parallels with the RmlD-catalysed reduction of dTDP-6-deoxy-L-*lyxo*-hexulose at the C4 position to form dTDP-L-rhamnose, which occurs as the final step of dTDP-L-rhamnose biosynthesis (Graninger *et al*., 1999). These enzymes all reduce their respective substrates in an NAD(P)H-dependent manner (Graninger *et al*., 1999; McCallum *et al*., 2012, 2013). RmlD is also a member of the SDR family and can therefore also be divided into an N-terminal Rossmann fold and a C-terminal globular domain (Blankenfeldt *et al*., 2002; Kavanagh *et al*., 2008). The central parallel β-sheet of RmlD contains two more β-strands than that of MlghC and one more β-strand than that of DdahC. The structure of RmlD has been solved in the presence of both the cofactor NADPH and its product dTDP-L-rhamnose (PDB: 1KC3) (Blankenfeldt *et al*., 2002). This structure of RmlD can be aligned with DdahC over the C_α backbone of 266 residues with an RMSD of 2.84 Å (Krissinel and Henrick, 2004) and they share a 15 % sequence identity (figure 5.7). This structure of RmlD can also be aligned with MlghC over the C_α backbone of 258 residues with an RMSD of 2.40 Å (Krissinel and Henrick, 2004). MlghC and RmlD share a sequence identity of 16 % (figure 5.7). The previously described dimerization mode of MlghC and DdahC is typical of the SDR family (Kavanagh *et al*., 2008) and is seen in other members, for example GDP-fucose synthetase (Somers *et al*., 1998) and the dTDP-D-glucose 4,6-dehydratase RmlB (Allard *et al*., 2001). In contrast, and also atypical of the SDR family, RmlD dimerises via a Mg^{2+} ion, coordinated by Glu residues located in the C-terminal domain (Blankenfeldt *et al*., 2002), rather than via the helical bundle formed by α-helices of the N-terminal domain, as observed in both DdahC and MlghC. This structure of RmlD in complex with cofactor and product (Blankenfeldt *et al*., 2002) allows the prediction of the DdahC and MlghC substrate binding sites and the proposal of their active site residues. As the observed dimerization

modes differ between DdahC/MlghC and RmlD, the cofactor and substrate binding sites will be discussed in the context of a monomer (Blankenfeldt *et al*., 2002).

5.3.20 NADP$^+$ binding to MlghC

The DdahC- and MlghC-catalysed reduction of their respective substrates occurs in an NADPH-dependent manner (McCallum *et al*., 2012, 2013). MlghC copurified and cocrystallised with its cofactor NADP$^+$ and the F_o-F_c density, contoured at 3σ, for one of the NADP$^+$ molecules is shown in figure 5.69. NADP$^+$ binds in the previously described N-terminal Rossmann fold (Hanukoglu, 2015) of MlghC and is shown in figure 5.68. The residues required for NADP$^+$ binding by MlghC can be seen in figure 5.70. This NADP$^+$ binding site location is also shared with RmlD (Blankenfeldt *et al*., 2002), shown in figure 5.71. An overlay of the cofactor binding sites of RmlD and MlghC is shown in figure 5.72. DdahC, on the other hand, did not cocrystallise with its cofactor NADP$^+$. As is the case with RmlD binding to NADPH (Blankenfeldt *et al*., 2002), MlghC binds NADP$^+$ in its *syn* conformation with a hydrogen bond formed between the amide of the nicotinamide moiety and the α-phosphate, with a distance of 2.7 Å between the aforementioned amide group and the α-phosphate of NADP$^+$ when bound to both RmlD (Blankenfeldt *et al*., 2002) and MlghC. Both MlghC and RmlD interact with the nicotinamide moiety largely via hydrophobic interactions contributed to by Val 154 in RmlD (Blankenfeldt *et al*., 2002) and Leu 163-Asn 165, Phe 105 and Gly 106 in MlghC.

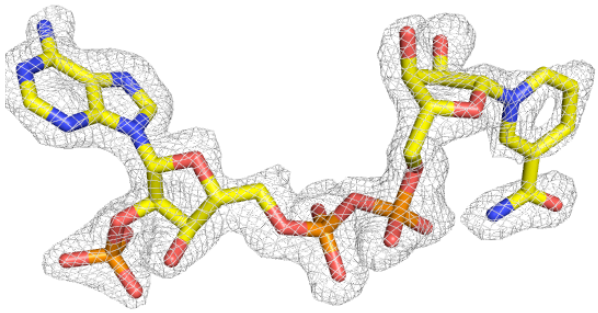


Figure 5.69. F_O-F_C density for NADP$^+$ bound to MlghC. The F_O-F_C density is contoured at 3 σ. The F_O-F_C density was calculated by deleting NADP$^+$ from the final refined model, resetting the B factors to 20 Å^2 using the TLSMD server (Painter and Merritt, 2006a, 2006b), and refining in REFMAC5 (Murshudov *et al*., 1997), with 50 cycles of refinement. The map was created using FFT, as part of the CCP4 suite (Winn *et al*., 2011). This image was created using MacPyMOLEdu.

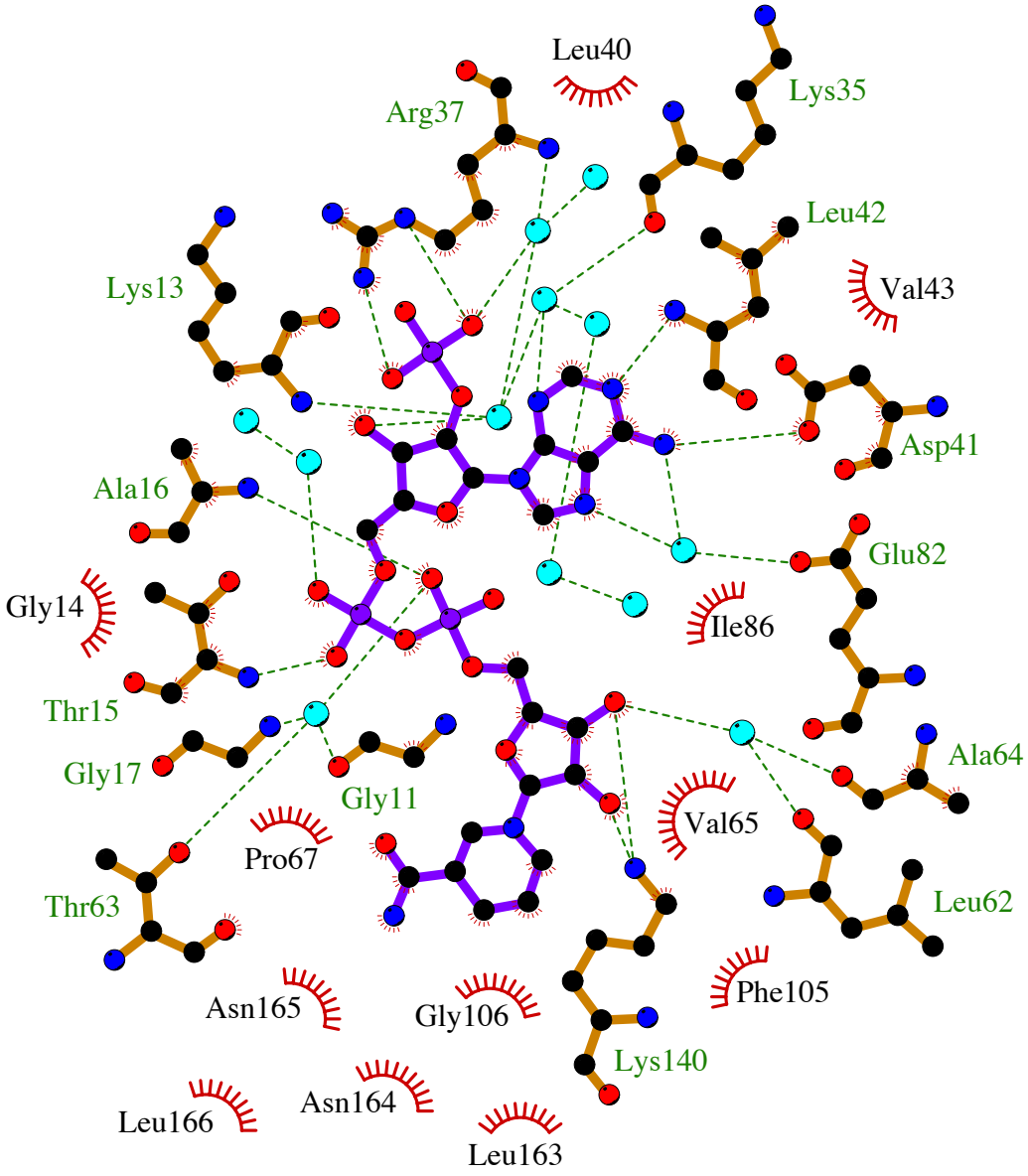


Figure 5.70. Amino acids required for MlghC binding to NADP$^+$. This image was created using LigPlot$^+$ (Laskowski and Swindells, 2011). Waters are shown in cyan, carbon is shown in black, nitrogen is shown in blue, oxygen is shown in red and phosphorus is shown in purple. Bonds within the ligand are shown in purple and bonds within the protein are shown in brown. Hydrogen bonds and names of the participating amino acids are shown in green. Hydrophobic interactions are shown in red and the names of the participating amino acids are shown in black.

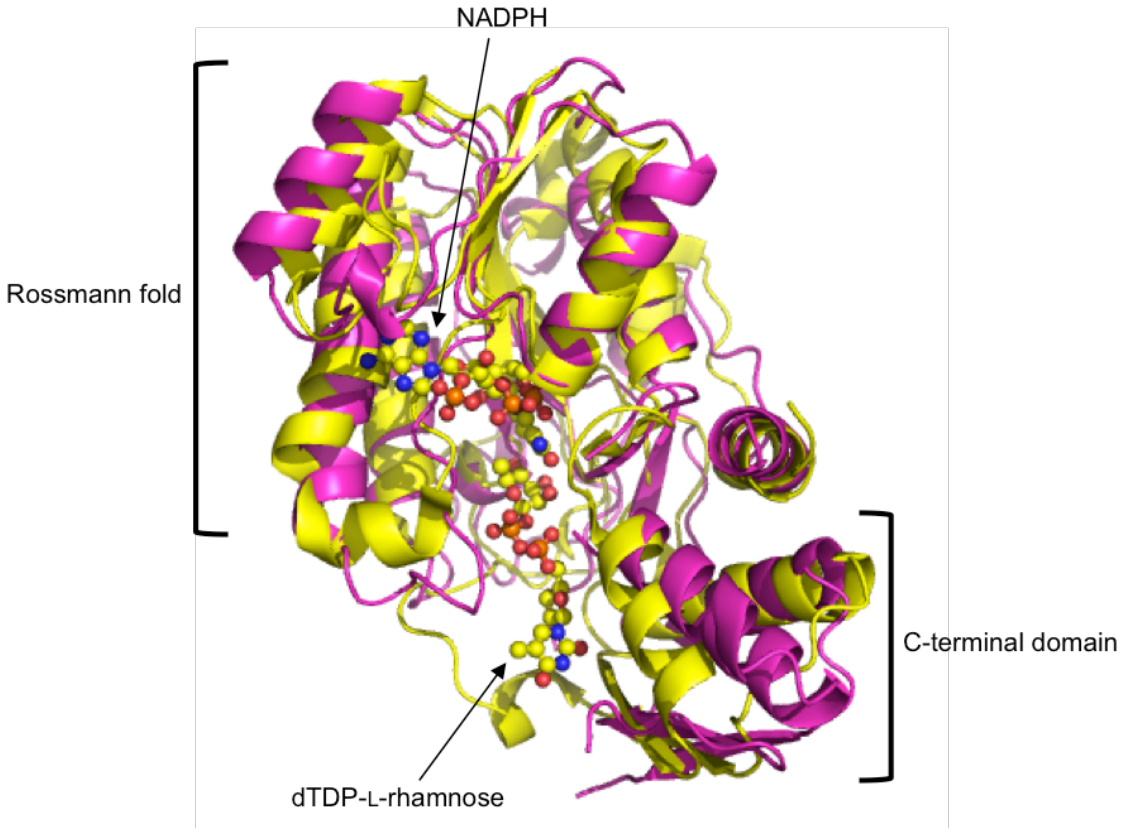


Figure 5.71. Overlay of MlghC and RmlD monomers. The NADPH cofactor binds in the N-terminal Rossmann fold and the substrate of RmlD, dTDP-L-rhamnose, binds in the C-terminal domain, with the glucose moiety oriented towards the nicotinamide moiety of the NADPH (RmlD PDB: 1KC3) (Blankenfeldt *et al*., 2002). MlghC is shown in magenta and RmlD in yellow. This image was created using MacPyMOLEdu.

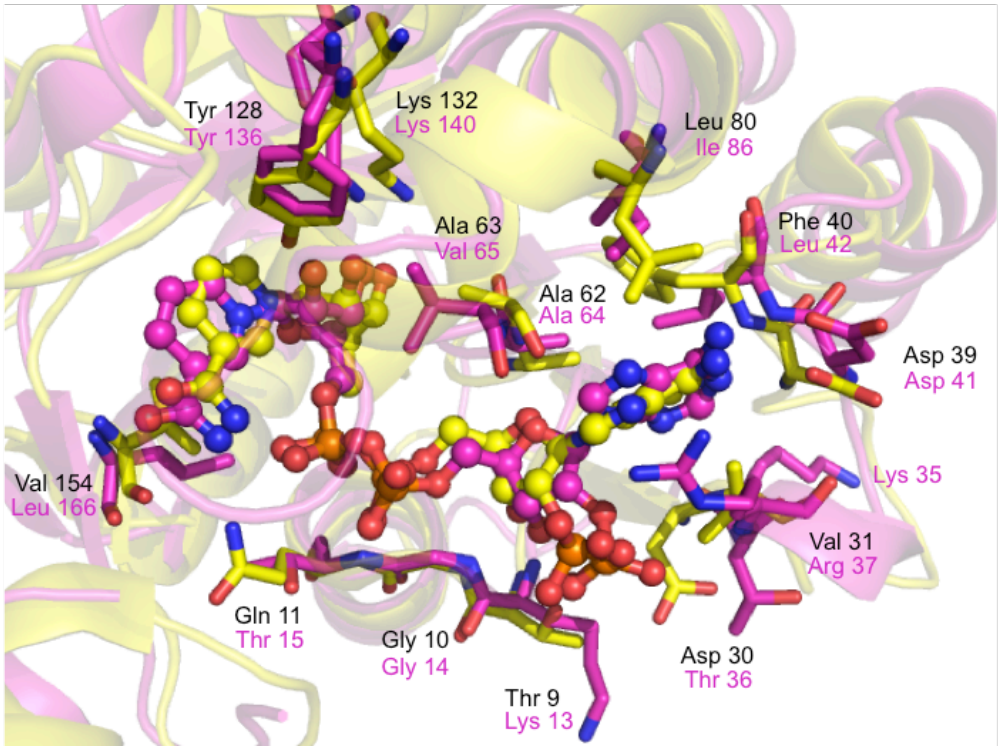


Figure 5.72. Overlay of MlghC and RmlD NADP$^+$ binding sites. MlghC, and the NADP$^+$ bound to MlghC, is shown in magenta. RmlD, and the NADPH bound to RmlD, is shown in yellow. The labels in magenta correspond to MlghC and the labels in black to RmlD (RmlD PDB: 1KC3) (Blankenfeldt *et al*., 2002). This image was created using MacPyMOLEdu.

Lys 140 of the YXXXK motif in MlghC hydrogen bonds with the 2'OH and 3'OH of the nicotinamide ribose moiety, shown in figure 5.70 (Kavanagh *et al*., 2008). This interaction is conserved amongst SDR family members and is also seen in RmlD (Lys 132), with an additional hydrogen bond present between Tyr 128 of the YXXXK motif of RmlD and the 2' OH of the nicotinamide ribose moiety (Blankenfeldt *et al*., 2002), which is absent from MlghC, shown in figure 5.72. Hydrogen bonds are also formed between the 3' OH of the nicotinamide ribose moiety and the carbonyl groups of Leu 62 and Ala 64 in MlghC, seen in figure 5.70.

The conserved GXXGXXG motifs of MlghC and DdahC consist of Gly 11, Gly 14 and Gly 17, with Gly 11 and Gly 17 of MlghC being required for hydrogen bonding with the pyrophosphate moiety of $NADP^+$, via bridging water molecules, shown in figure 5.70. This motif is also present in RmlD and is formed by Gly 7, Gly 10 and Gly 13 (Blankenfeldt *et al*., 2002). Hydrogen bonds are also formed between Thr 15, Ala 16 and Thr 63, via a bridging water molecule in the case of Thr 63, and the pyrophosphate moiety, shown in figure 5.70. Gly 10 and Gln 11 fulfil this role in RmlD (Blankenfeldt *et al*., 2002), shown in figure 5.72.

The adenine ring of $NADP^+$ of MlghC forms hydrogen bonds with the carbonyl group of the side chain of Asp 41 and the amine group of Leu 42 in MlghC, which corresponds to Asp 39 and Phe 40 in RmlD, shown in figures 5.70 and 5.72. Additional hydrogen bonds are formed between the adenine ring and Lys 35 and Glu 82 via bridging water molecules, shown in figure 5.70. Hydrophobic interactions also form between the adenine ring and Ile 36, Leu 40 and Val 43, shown in figure 5.70.

5.3.21 DdahC crystallised in absence of $NADP^+$

HPLC analysis of purified DdahC did not detect the presence of $NADP^+$ (data not shown), indicating that DdahC did not copurify with $NADP^+$. Despite the lack of $NADP^+$ cofactor in the DdahC structure, the structures of MlghC and DdahC can be aligned with an RMSD of 1.31 Å over the C_α backbone of 333 amino acids out of 336 amino acids for MlghC and 347 amino acids for DdahC (Krissinel and Henrick, 2004). This indicates that there is little structural change that occurs upon $NADP^+$ binding, see figure 5.73 The residues required for $NADP^+$ binding in MlghC are largely conserved in DdahC, with substitutions of Arg 37 (present in MlghC) for His 37 (present in DdahC) and of Lys 13 (present in MlghC) for Ser 13 (present in DdahC). An overlay of the cofactor binding sites of DdahC and MlghC can be seen in figure 5.74. The lack of $NADP^+$ copurification with DdahC indicates that this interaction is weak. This is supported by the finding that purified RmlD had to be incubated with NADPH at 20

mM, prior to cocrystallisation, in order to generate sufficient electron density to allow the cofactor to be modelled (Blankenfeldt *et al*., 2002).

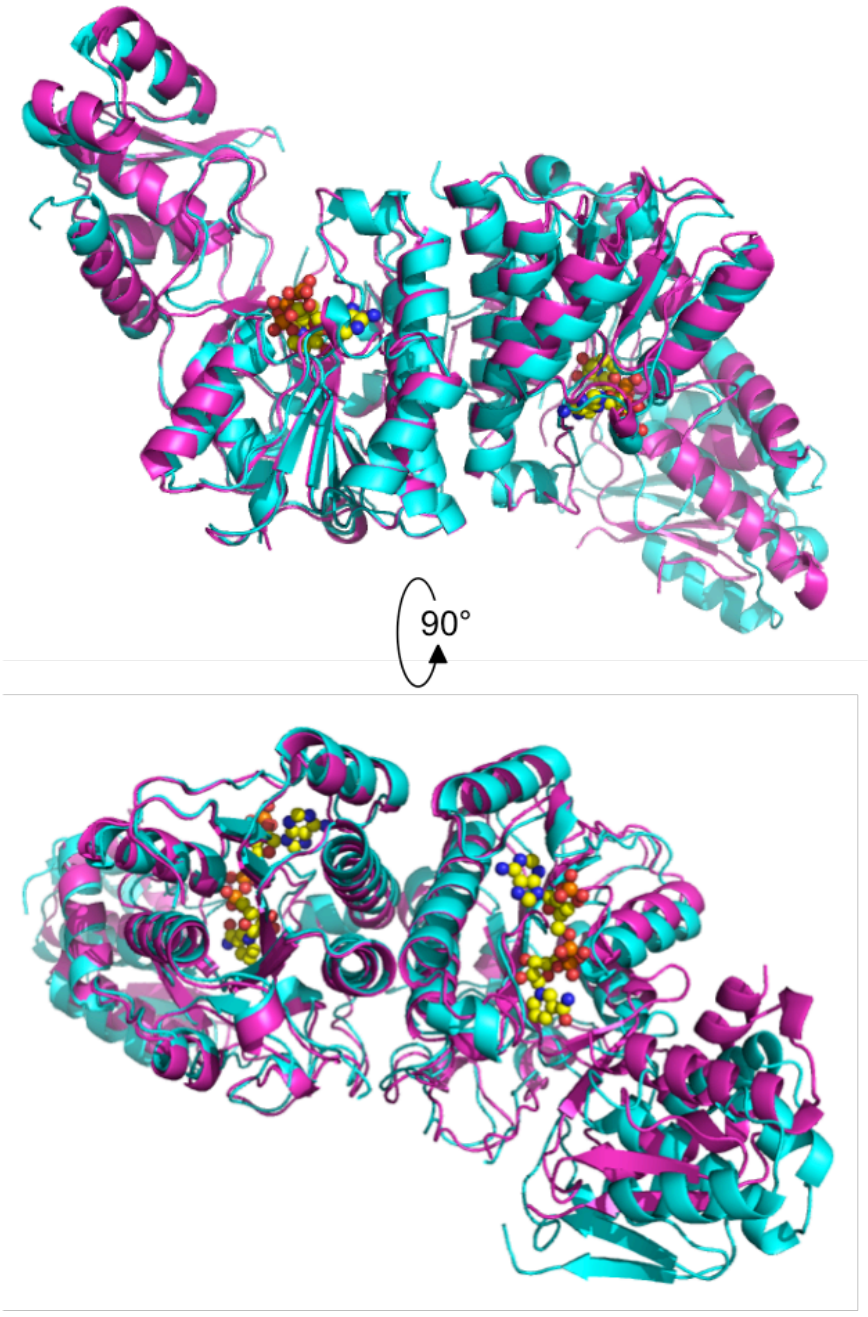


Figure 5.73. Overlay of $NADP^+$-bound MlghC and apo DdahC. Little structural changes occur upon cofactor binding. DdahC is shown in cyan and MlghC in magenta. $NADP^+$ is shown in yellow in ball and stick representation. This image was created using MacPyMOLEdu.

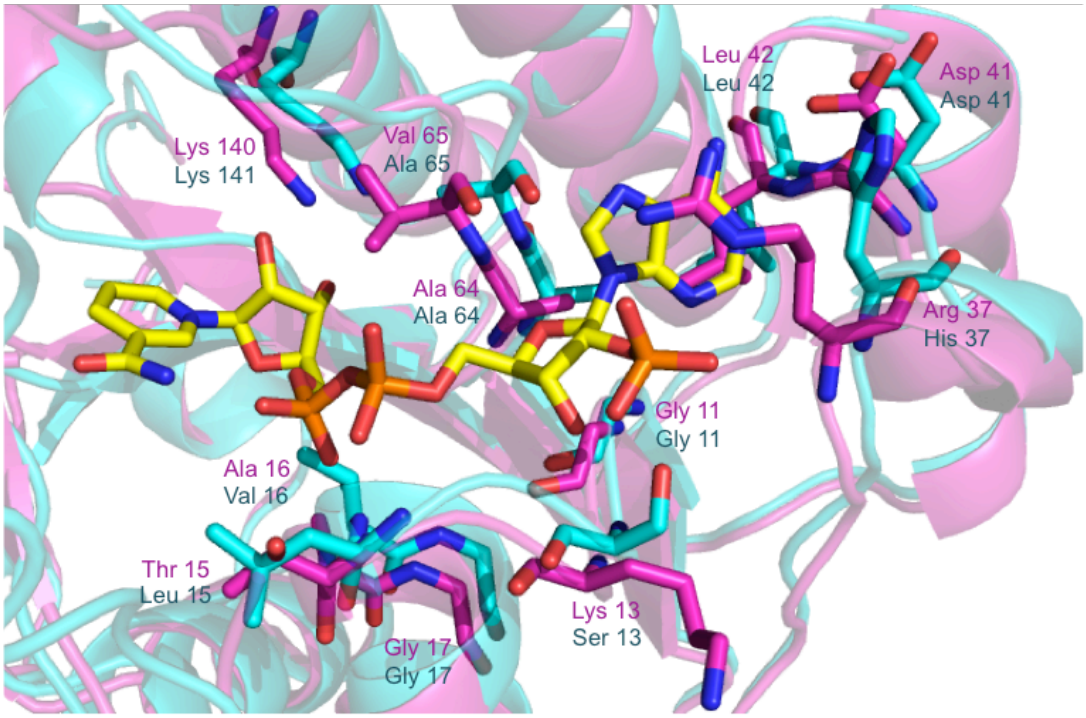


Figure 5.74. Overlay of the MlghC NADP$^+$ binding site and the proposed DdahC NADP$^+$ binding site. Residues required by MlghC to bind NADP$^+$ are largely conserved in DdahC. MlghC is shown in magenta, DdahC is shown in cyan and NADP$^+$ is shown in yellow. The residue names shown in green correspond to those of DdahC and those in magenta correspond to MlghC. This image was created using MacPyMOLEdu.

5.3.22 Mode of distinction between NADP(H) and NAD(H)

Two main factors are involved in the distinction between NAD(H) and NADP(H) binding by SDR family members. The presence of a conserved Asp residue, approximately 20 residues C-terminal to the GXXGXXG motif, usually located at the C-terminus of the second β-strand of the central β-sheet of the Rossmann fold indicates a preferential interaction with NAD(H) (Kavanagh *et al*., 2008; Lesk, 1995). This corresponds to Asp 30 in RmlD and is required for hydrogen-bonding to the 2' OH and 3' OH of the ribose of NAD(H) (Blankenfeldt *et al*., 2002). As MlghC and DdahC preferentially bind NADP(H) (McCallum *et al*., 2012, 2013), this conserved Asp is

absent. SDR family members that preferentially interact with NADP(H) have been found to possess a basic residue, within the conserved GXXGXXG motif, required for hydrogen bonding with the 2' phosphate group of the adenosine ribose of NADP(H) (Kavanagh *et al*., 2008; Lesk, 1995). Lys 13 in MlghC, which hydrogen bonds via two bridging water molecules with the 2' phosphate group of the adenosine ribose, fulfils this requirement, shown in figures 5.70 and 5.72. This Lys is not conserved in RmlD and aligns with Thr 9 in the NADPH- and dTDP-L-rhamnose-bound structure of RmlD (Blankenfeldt *et al*., 2002), shown in 5.72. This 2'-phosphate group also hydrogen bonds with Arg 37 in MlghC. This residue is also absent in RmlD and aligns with Val 31, shown in figure 5.72. We propose these positively charged residues underpin MlghC's preference for NADP(H) over NAD(H) (McCallum *et al*., 2013). In the absence of these positively charged residues in RmlD, there are no interactions between the additional phosphate group of NADPH and RmlD, enabling the interaction of RmlD with both NADP(H) and NAD(H) (Blankenfeldt *et al*., 2002). Instead, the reorientation of the side chain of Asp 30 is proposed to allow the discrimination between NADP(H) and NAD(H) (Blankenfeldt *et al*., 2002).

5.3.23 Locating the active site of DdahC and MlghC

The overlay of MlghC with RmlD, shown in figure 5.71, predicts the substrate binding site to be in a solvent-exposed cleft between the N- and C-terminal domains, with the substrate adopting an elongated conformation and the rhamnose moiety oriented towards the nicotinamide ring of the NADPH, bound in the N-terminal Rossman fold domain (Blankenfeldt *et al*., 2002). Due to the differences in substrate specificity, the majority of the structural differences between these proteins are found in this C-terminal domain. This is highlighted by the difference between the RMSD values for the alignment of the C_α backbone of the N-terminal domains and that of the C-terminal domains. The RMSD for the C_α backbone alignment of the N-terminal cofactor binding domains of MlghC and RmlD is 2.13 Å, which is lower than that of the alignment across the entire chain of 2.40 Å. The RMSD is 2.56 Å for the alignment of the C-terminal substrate binding domains (Krissinel and Henrick, 2004). A similar pattern is

observed when aligning the C_α backbone of DdahC and RmlD. The RMSD for the C_α backbone alignment of the N-terminal cofactor binding domains is 2.20 Å, which is again lower than that of the alignment across the whole chain of 2.84 Å (Krissinel and Henrick, 2004). The RMSD for the C_α backbone alignment of the C-terminal domains of DdahC and RmlD is 2.44 Å, however only 44 of 105 residues could be aligned, indicating the presence of a greater structural divergence than represented by this RMSD value (Krissinel and Henrick, 2004). Due to this higher level of structural divergence between DdahC and RmlD, and the absence of the cofactor in the DdahC structure, the substrate binding site will only be discussed and compared in detail for MlghC and RmlD.

In RmlD, the dTDP binding pocket, shown in figure 5.75, is located in the C-terminal domain and is formed by Ser 177, Val 178 and Ile 179 (Blankenfeldt *et al*., 2002). This corresponds to the loop between β-strands six and seven of MlghC, consisting of amino acids 232-237, which would be expected to fold over the GDP binding pocket upon substrate binding. Arg 260 of RmlD is required for hydrogen bonding to the pyrophosphate moiety of dTDP-L-rhamnose (Blankenfeldt *et al*., 2002). A number of candidate Arg residues in MlghC exist for this interaction, including Arg 306 and Arg 309, both of which are found in the loop between β-strand eight and α-helix 14, which is flexible and partially disordered in the absence of substrate, with electron density absent for amino acids 306-309. Arg 179, found at the C-terminal end of the loop linking β-strand five and α-helix eight, could also fulfil this role.

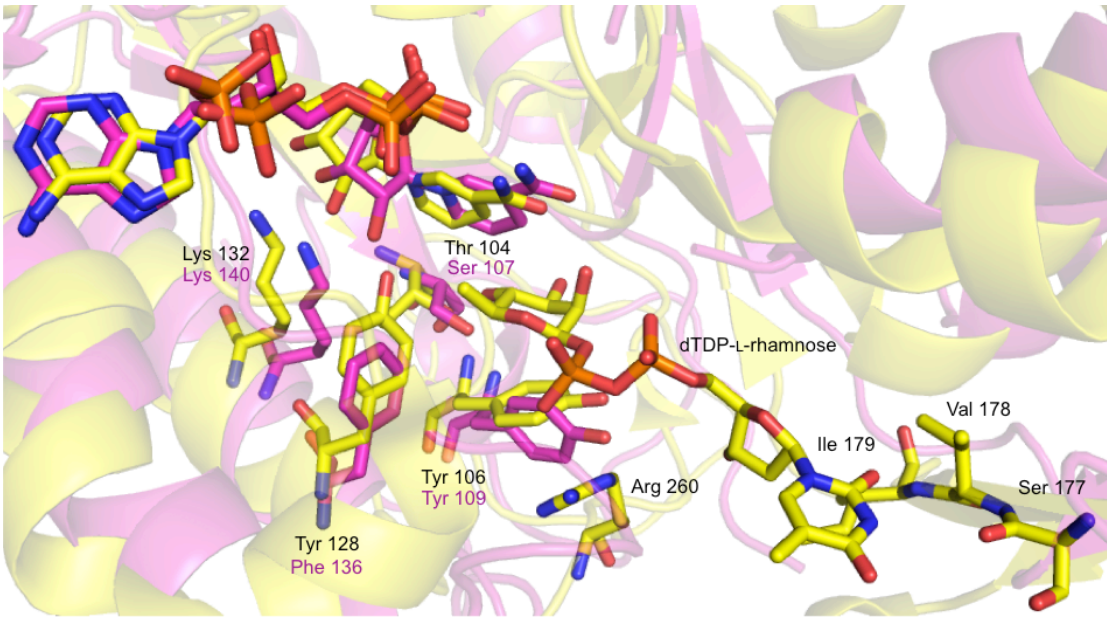


Figure 5.75. A comparison of the substrate binding sites of RmlD and MlghC. MlghC, and its cofactor $NADP^+$, are shown in magenta. RmlD, and its cofactor NADPH and product dTDP-L-rhamnose, are shown in yellow. The residues required for dTDP-L-rhamnose binding in RmlD, and the corresponding residues in MlghC are shown here, with the black labels corresponding to RmlD and the magenta labels to MlghC (RmlD PDB: 1KC3) (Blankenfeldt *et al*., 2002). This figure was created using MacPyMOLEdu.

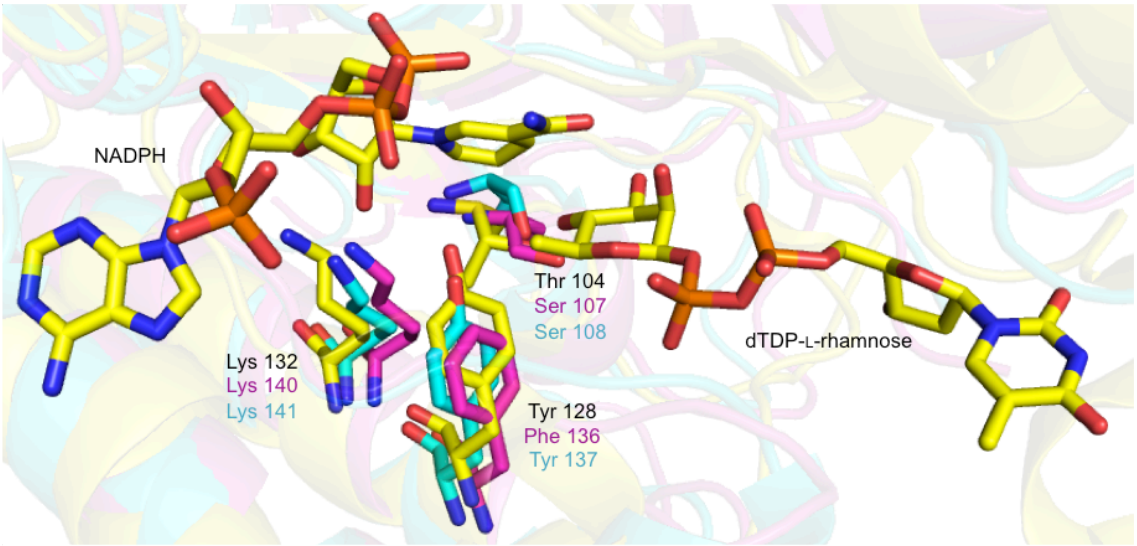


Figure 5.76. A comparison of the active site residues of RmlD with the putative active site residues of MlghC and DdahC. RmlD is shown in yellow, DdahC in cyan and MlghC in magenta. The labels in black correspond to RmlD, those in cyan to DdahC and those in magenta to MlghC. NADPH and dTDP-L-rhamnose are both bound to RmlD (RmlD PDB: 1KC3) (Blankenfeldt *et al*., 2002). This figure was created using MacPyMOLEdu.

The rhamnose moiety is located at the base of the Rossmann fold of RmlD, stacked between the nicotinamide moiety of the NADPH and the side chain of Tyr 106. This Tyr is also conserved in MlghC and corresponds to Tyr 109, but in the absence of sugar it is present as a different rotamer to that observed in the dTDP-L-rhamnose-bound RmlD structure, shown in figure 5.75. The active site residues of RmlD, namely Thr 104, Tyr 128 and Lys 132 (Blankenfeldt *et al*., 2002), align with the putative active site residues of DdahC and MlghC, shown in figure 5.76. The structural alignment of RmlD and DdahC identifies the putative active site residues as Ser 108, Tyr 137 and Lys 141 and the structural alignment of RmlD and MlghC identifies the putative active site residues of MlghC as Ser 107, Phe 136 and Lys 140 (Dr Creuzenet, University of Western Ontario, personal communication). Thr 104 and Tyr 128 also form hydrogen bonds with the rhamnose moiety in the dTDP-L-rhamnose-bound structure of RmlD (Blankenfeldt *et al*., 2002), highlighting a possible role in substrate binding for the corresponding residues in DdahC and MlghC.

5.4 Discussion

Despite the demonstrated role of modified heptoses in the virulence of *C. jejuni* and *Y. pseudotuberculosis*, the enzymes required for their biosynthesis have remained uncharacterised (Ho *et al*., 2008; Wong *et al*., 2015). Here we describe the structures of the biosynthetic enzymes that catalyse the main steps of modified heptose biosynthesis in *C. jejuni* strain 81-176, namely DdahB and DdahC (McCallum *et al*., 2012), and *C. jejuni* strain NCTC 11186, namely MlghB and MlghC (McCallum *et al*., 2013), and that of DmhA, which catalyses the first step of 6-deoxyheptose biosynthesis in *Y. pseudotuberculosis* strain O:2a (Pacinelli *et al*., 2002). Whilst hexose biosynthetic enzymes have been well characterised, (Allard *et al*. 2001; Giraud *et al*. 2000; Giraud *et al*. 1999; Allard *et al*. 2002; Blankenfeldt *et al*. 2000; Dong *et al*. 2007; Dong *et al*. 2003), this study offers new insights into the biosynthesis of modified heptoses.

The structure of the dehydratase DmhA, in the presence of its cofactor NAD^+, has been solved to a resolution of 2.0 Å. The structure of the epimerase DdahB has been solved

in the absence and presence of the substrate mimic GDP/GDP-mannose to a resolution of 1.84 Å and 2.35 Å, respectively. The structure of the epimerase MlghB has been solved in absence and presence of GDP to a resolution of 2.31 Å and 2.60 Å, respectively. The structures of the reductases MlghC, in presence of $NADP^+$, and the apo form of DdahC have been solved to 1.66 Å and 2.08 Å resolution, respectively. These structures, along with their comparison to the enzymes of the dTDP-L-rhamnose biosynthetic pathway, have allowed us to answer some questions about the mechanisms of action for each of the enzymes.

5.4.1 Mechanism of DmhA-catalysed dehydration of GDP-D-*manno*-heptose

A dehydration step signals the beginning of modified heptose biosynthesis in *C. jejuni* strain 81-176 (McCallum *et al*., 2012), 6-deoxyheptose biosynthesis in *Y. pseudotuberculosis* O:2a (Butty *et al*., 2009; Pacinelli *et al*., 2002) and dTDP-L-rhamnose biosynthesis (Allard *et al*., 2001). The comparison of the structure of the GDP-D-*manno*-heptose 4,6-dehydratase DmhA in complex with NAD^+ with that of the dTDP-D-glucose 4,6-dehydratase RmlB in complex with NAD^+ and dTDP-D-glucose (Allard *et al*., 2002) has allowed the proposal of a mechanism for the DmhA-catalysed dehydration of GDP-D-*manno*-heptose. This mechanism, as well as the structure of DmhA, is shown in figure 5.77. Dehydration is known to take place via three concerted steps (Allard *et al*., 2001, 2002). The first step is oxidation, during which a hydride is transferred from the C4 position to the NAD^+ cofactor, coupled with the deprotonation of the C4 OH group by the active site Tyr, Tyr 167 in the case of RmlB (Allard *et al*., 2001, 2002) and Tyr 157 in the case of DmhA. The positively charged environment surrounding the active site Tyr, created by the Lys 171 of RmlB and Lys 161 of DmhA, and the positively charged nicotinamide moiety of the NAD^+, results in a decrease in the pK_a of the hydroxyl group of the Tyr side chain, allowing it to perform the role of catalytic base (Allard *et al*., 2001, 2002). This results in the conversion of dTDP-D-glucose to dTDP-4-keto-glucose during dTDP-L-rhamnose biosynthesis (Allard *et al*., 2001, 2002) and GDP-D-*manno*-heptose to GDP-4-keto-6-deoxy-*manno*-heptose during

6-deoxyheptose biosynthesis. The second step is the dehydration step, which can now occur due to the decrease in the pK_a value of the C5 proton, resulting from the oxidation at the C4 position (Allard *et al*., 2001, 2002). Dehydration occurs via the deprotonation of the C5 catalysed by Glu 135 of RmlB, followed by the protonation of the hydroxyl group released from the C6 position, catalysed by Asp 134 of RmlB, resulting in the elimination of water across the C5 and C6 positions (Allard *et al*., 2001, 2002). Glu 135 of RmlB is conserved in DmhA (Glu 134 in DmhA). However, Asp 134 of RmlB aligns with Ser 133 of DmhA, therefore the identity of this catalytic acid remains unconfirmed, as the high pK_a value of the Ser side chain hydroxyl group makes it unlikely to act as a catalytic acid. This dehydration step converts dTDP-4-keto-glucose to dTDP-4-keto-glucose-5,6-ene during dTDP-L-rhamnose biosynthesis (Allard *et al*., 2001, 2002) and GDP-4-keto-*manno*-heptose to GDP-4-keto-6-deoxy-*manno*-heptose-5,6-ene during 6-deoxyheptose biosynthesis. The final step is a reduction step, which involves the reduction of the C5 position by the catalytic Tyr (Tyr 167 in the case of RmlB and Tyr 157 in the case of DmhA), regenerating the tyrosinate group (Allard *et al*., 2001, 2002). This is followed by the reduction of the C6 position by the NADH, regenerating the NAD^+ cofactor (Allard *et al*., 2001, 2002). This yields dTDP-4-keto-6-deoxy-D-glucose as the final product of the RmlB-catalysed dehydration of dTDP-D-glucose (Allard *et al*., 2001, 2002) and GDP-4-keto-6-deoxy-D-*lyxo*-heptose as the final product of the DmhA-catalysed dehydration of GDP-D-*manno*-heptose. The DdahA-catalysed dehydration of GDP-D-*manno*-heptose to form GDP-4-keto-6-deoxy-D-*lyxo*-heptose would also be expected to follow the same mechanism as DmhA (McCallum *et al*., 2011).

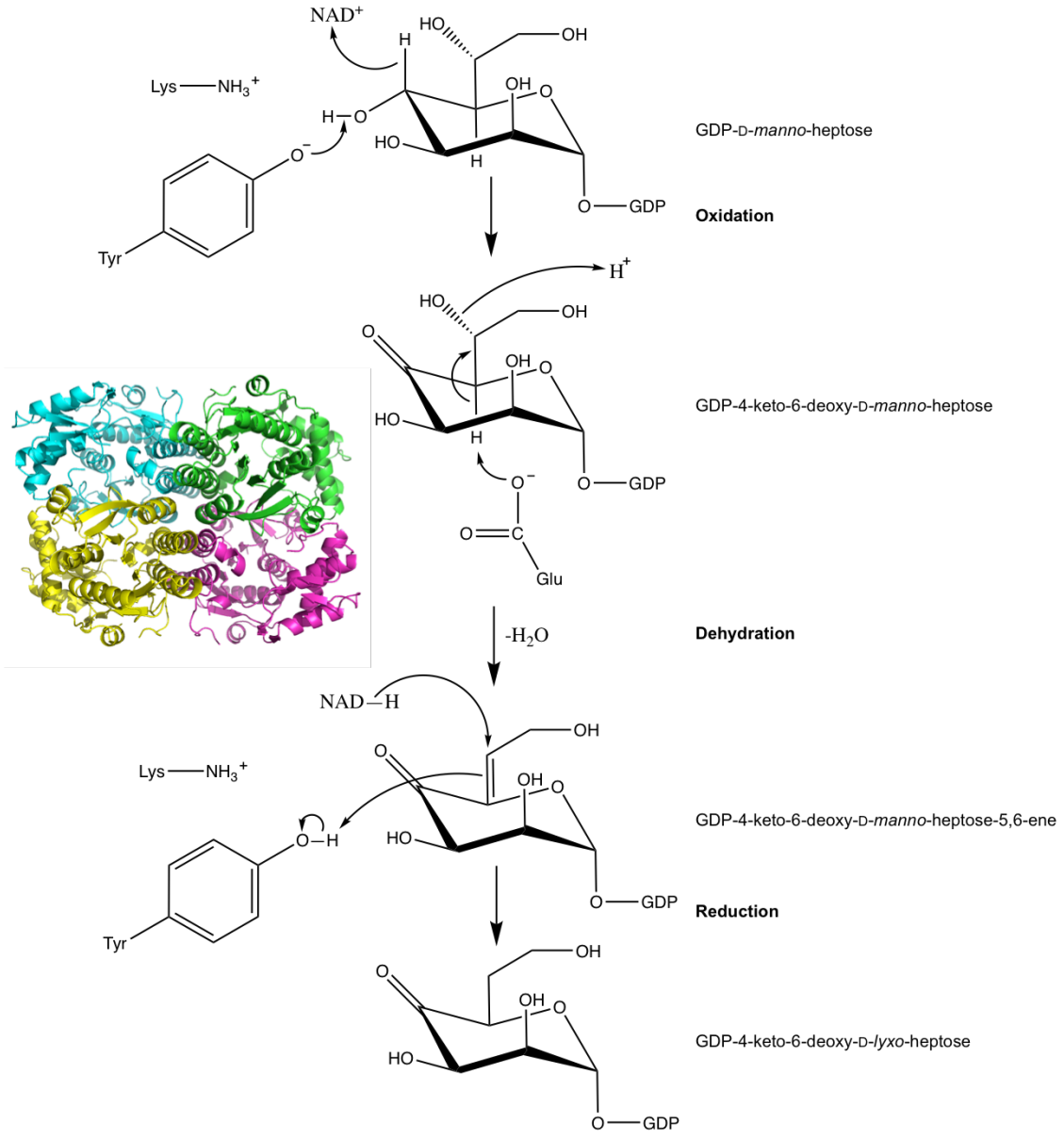


Figure 5.77. Proposed mechanism of DmhA-catalysed dehydration of GDP-D-*manno*-heptose. This mechanism is based on that of RmlB (Allard *et al*., 2001, 2002). The structure of DmhA is seen inset. This image was created using ChemDraw 16.0 and MacPyMOLEdu.

5.4.2 Mechanism of DdahB- and MlghB-catalysed epimerisation

The second step in both modified heptose biosynthesis in *C. jejuni* and dTDP-L-rhamnose biosynthesis is epimerisation (Giraud *et al*., 2000; McCallum *et al*., 2012, 2013). Both GDP-(6-*O*-methyl)-L-*gluco*-heptose biosynthesis in *C. jejuni* strain NCTC 11168 and dTDP-L-rhamnose biosynthesis occur via a C3,C5 epimerisation step, catalysed by MlghB and RmlC, respectively (Giraud *et al*., 2000; McCallum *et al*., 2013), whereas the biosynthesis of GDP-6-deoxy-D-*altro*-heptose requires the epimerisation at only the C3 position, catalysed by DdahB (McCallum *et al*., 2012). The proposed mechanism of DdahB-catalysed C3 epimerisation of GDP-4-keto-6-deoxy-D-*lyxo*-heptose to form GDP-4-keto-6-deoxy-D-*arabino*-heptose, along with the structure of DdahB, is shown in figure 5.78. The proposed mechanism of MlghB-catalysed C5, C3 epimerisation of GDP-(6-*O*-methyl)-4-keto-D-*lyxo*-heptose to form GDP-(6-*O*-methyl)-4-keto-L-*xylo*-heptose, along with the structure of MlghB, is shown in figure 5.79. The order of epimerisation events shown in figure 5.79 is arbitrary as the order of the epimerisation events catalysed by MlghB remains uncertain (McCallum *et al*., 2013). Higher rates of deuterium incorporation in RmlC (from *S.* Typhimurium)-bound dTDP-6-deoxy-D-*xylo*-4-hexulose observed for the C5 position compared to the C3 position suggest that C5 is epimerised first (Dong *et al*., 2007). However, upon addition of MlghB to the substrate mimic GDP-4-keto-6-deoxy-D-*lyxo*-heptose, three products are observed using capillary electrophoresis, corresponding to the C3 epimer, C5 epimer and C3, C5 epimers, suggesting there is no preferential order of MlghB-catalysed epimerisation (McCallum *et al*., 2013).

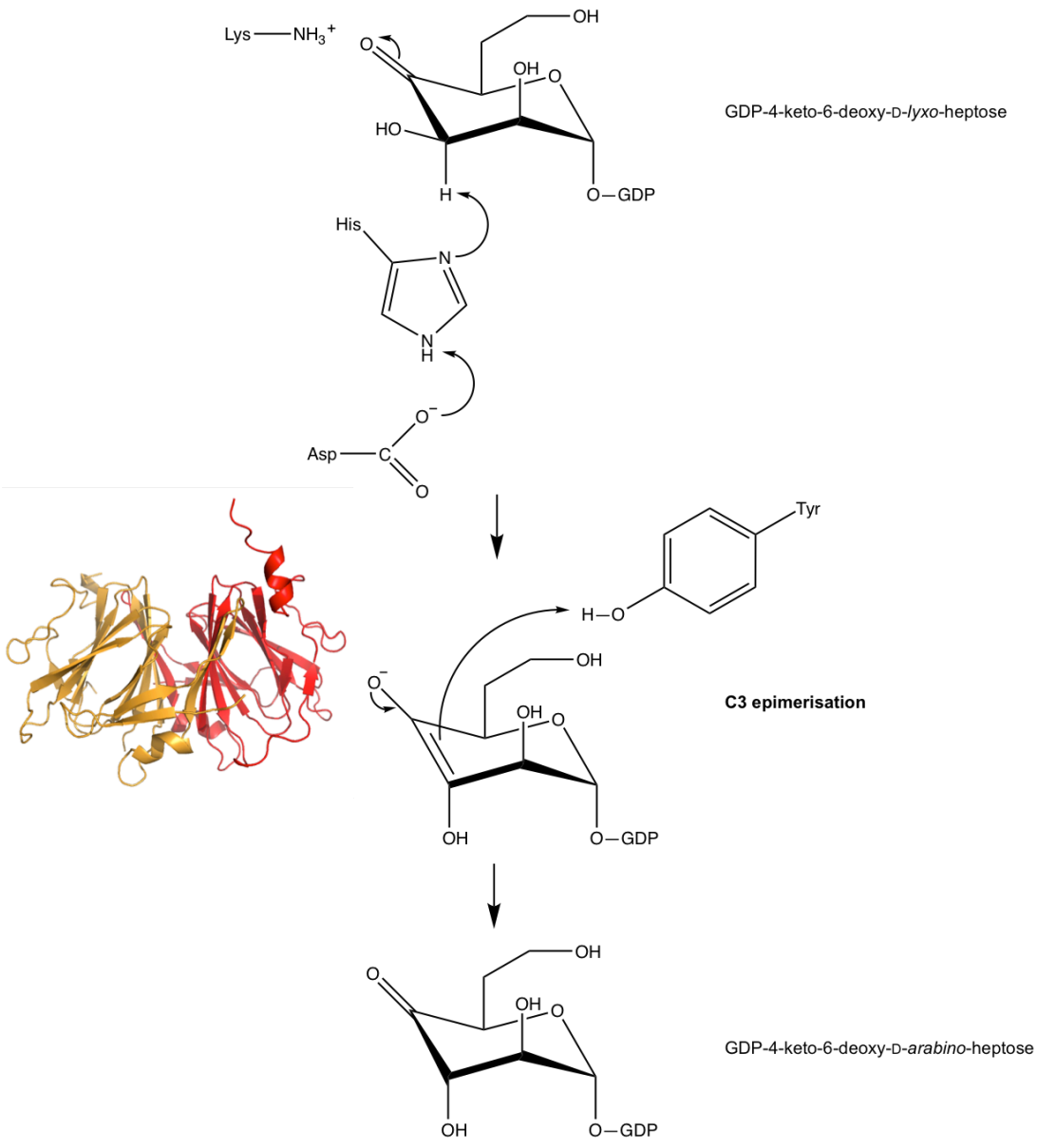


Figure 5.78. Proposed mechanism of DdahB-catalysed C3 epimerisation. This proposed mechanism is based on that of the C3, C5 epimerase RmlC (Dong *et al*., 2003b, 2007). The structure of DdahB is shown inset. This image was created using ChemDraw 16.0 and MacPyMOLEdu.

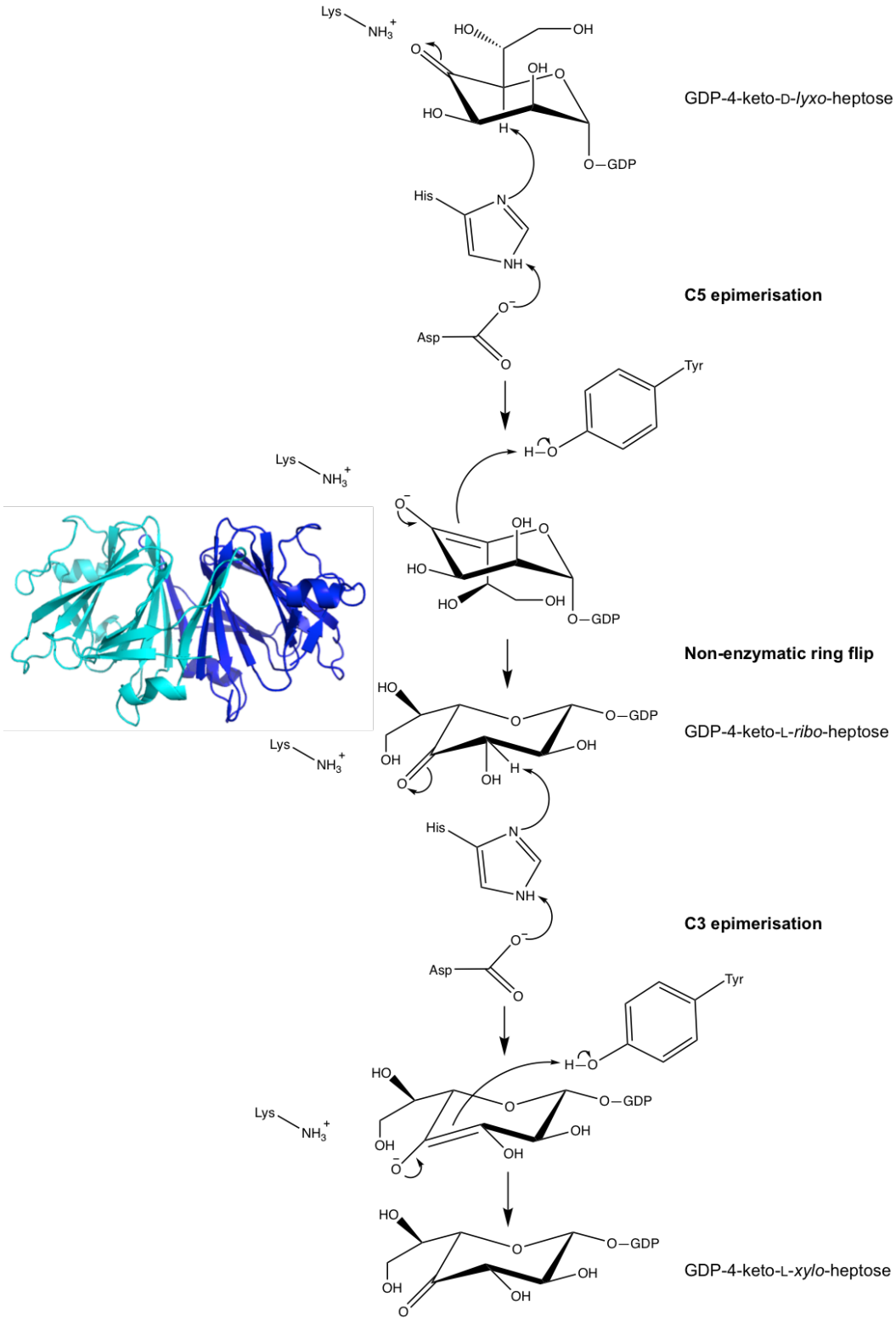


Figure 5.79. Proposed mechanism of MlghB-catalysed C5/C3 epimerisation. This proposed mechanism is based on that of RmlC and the order of epimerisation shown here is arbitrary (Dong *et al*., 2003b, 2007; McCallum *et al*., 2013). The structure of MlghB is shown inset. This image was created using ChemDraw 16.0 and MacPyMOLEdu.

His 76 in RmlC, corresponding to His 67 in both DdahB and MlghB, acts as the active site base, extracting a proton from C3 in the case of DdahB and either C5 or C3 in the case of both RmlC and MlghB. This is facilitated by the abstraction of a proton from the imidazole side chain of the His residue, catalysed by Asp 180 in the case of RmlC (Dong *et al*., 2003b, 2007). This Asp residue is conserved in MlghB and aligns with Asp 171. The identity of the corresponding residue in DdahB cannot be confirmed, as the C-terminal α-helix in which Asp 180 in RmlC is located (Dong *et al*., 2003b, 2007), is oriented away from the active site in the case of DdahB, but Asp 173 in this C-terminal α-helix could fulfil this role if a conformational change, resulting in the reorientation of this C-terminal α-helix, occurs upon substrate binding. The transition state has an enolate at the C4 position, which is stabilised by Lys 82 in RmlC (Dong *et al*., 2003b, 2007), corresponding to Lys 74 in both DdahB and MlghB. Tyr 140 of RmlC, corresponding to Tyr 134 in both DdahB and MlghB, then acts as the catalytic base and reprotonates the transition state on the opposite face of the ring to generate the C3 or C5 epimer (Dong *et al*., 2003b, 2007). Prior to the second epimerisation reaction, proposed to be at the C3 position in the case of RmlC and catalysed by MlghB and RmlC only, the enzyme must be regenerated by transferring protons to and from the surrounding solvent (Dong *et al*., 2003b, 2007). A ring flip is also required to orient the C3 proton in *trans* with respect to the C4 keto group (Dong *et al*., 2007). The His residue of the His-Asp dyad acts as the active site base, abstracting a proton from the C3 position (Dong *et al*., 2007). This is thought to occur via a twist boat conformation in the case of RmlC (Dong *et al*., 2007). The enolate intermediate is stabilised by the conserved Lys residue and the ring is reprotonated on the opposite face of the ring by the conserved Tyr, resulting in the C3 epimer (Dong *et al*., 2003b, 2007).

5.4.3 DdahB as a monoepimerase

The absence of the C5 epimerisation activity of DdahB remains unexplained. Other examples of RmlC-like monoepimerases have been structurally-characterised, including the dTDP-4-keto-6-deoxy-D-glucose-5-epimerase EvaD, required for the biosynthesis of dTDP-epivancosamine in *Amycolatopsis orientalis* (Merkel *et al*., 2004) and the

dTDP-4-keto-6-deoxy-glucose-3-epimerase ChmJ, required for dTDP-6-deoxy-D-allose biosynthesis in *Streptomyces bikiniensis* (Kubiak *et al*., 2012). Modelling of a sugar ring into the active site of the dTMP-bound EvaD structure, based on the active site of dTDP-D-xylose- and dTDP-D-glucose-bound structures of RmlC (from *S. suis*) (Dong *et al*., 2003b), allowed the proposal that the monoepimerase activity of EvaD arises from a difference in the orientation of the catalytic Tyr with respect to that of dTDP-sugar-bound RmlC (Merkel *et al*., 2004). This reorientation was proposed to be facilitated by the substitution of a Tyr (Tyr 138), two residues towards the N-terminus from the catalytic Tyr in RmlC, for a Met in EvaD (Merkel *et al*., 2004). However this Tyr, proposed to block the reorientation of the catalytic Tyr, is conserved in both DdahB and MlghB (Tyr 132) and is found in a similar conformation when comparing the apo and substrate-bound structures of DdahB and MlghB with dTDP-rhamnose-bound RmlC (from *S. suis*) (Dong *et al*., 2007). Furthermore, the orientation of the catalytic Tyr in dTDP-quinovose-bound ChmJ is similar to that of dTDP-rhamnose-bound RmlC (from *Mycobacterium tuberculosis*) (Kubiak *et al*., 2012). This therefore does not explain the difference in epimerisation activities seen here.

5.4.4 Mechanism of DdahC- and MlghC-catalysed reduction

The final step in both modified heptose biosynthesis in *C. jejuni* and dTDP-L-rhamnose biosynthesis is the reduction of the substrate sugar at the C4 position to yield the final product. This is catalysed by MlghC and DdahC during modified heptose biosynthesis in *C. jejuni* strains NCTC 11168 and 81-176, respectively (McCallum *et al*., 2012, 2013), and RmlD during dTDP-L-rhamnose biosynthesis (Graninger *et al*., 1999).

During the RmlD-catalysed reduction of dTDP-6-deoxy-L-*lyxo*-4-hexulose at the C4 position to form dTDP-L-rhamnose, NAD(P)H protonates dTDP-6-deoxy-L-*lyxo*-4-hexulose at the C4 position and the keto group at the C4 position is protonated by Tyr 128, resulting in dTDP-L-rhamnose, $NAD(P)^+$ and a tyrosinate (Blankenfeldt *et al*., 2002). The $NAD(P)^+$ must be released to allow new NAD(P)H binding during the regeneration of the enzyme (Blankenfeldt *et al*., 2002). Tyr 128 is able to act as the

active site acid due to the lowering of the pK_a of the Tyr side chain hydroxyl group resulting from the surrounding positive environment, created by both Lys 132 and the nicotinamide moiety of the NAD(P)H cofactor (Blankenfeldt *et al*., 2002). As the active site residues of RmlD are conserved in DdahC and correspond to Tyr 137 and Lys 141, the DdahC-reduction of GDP-4-keto-6-deoxy-D-*arabino*-heptose during GDP-6-deoxy-D-*altro*-heptose biosynthesis is expected to occur via the same mechanism as that of RmlD (Blankenfeldt *et al*., 2002). However, the identity of the active site acid in MlghC remains unidentified, as this Tyr is substituted for a Phe in MlghC. The mechanism of the DdahC-catalysed reduction of GDP-4-keto-6-deoxy-D-*arabino*-heptose at the C4 position to form GDP-6-deoxy-D-*altro*-heptose is shown in figure 5.80.

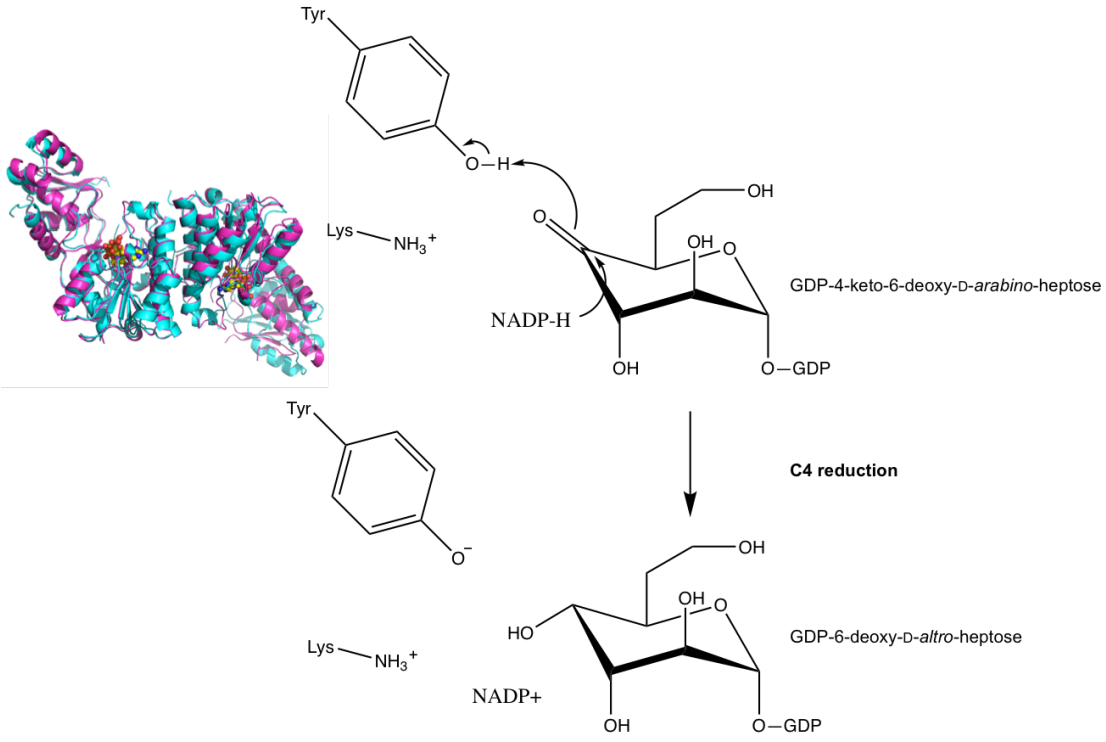


Figure 5.80. Proposed mechanism of DdahC-catalysed C4 reduction. This mechanism is based on that of RmlD (Blankenfeldt *et al*., 2002). An overlay of apo DdahC, in cyan, with $NADP^+$-bound MlghC, in magenta, is shown inset, with $NADP^+$ shown in yellow. This image was created using ChemDraw 16.0 and MacPyMOLEdu.

The absence of $NADP^+$ copurification with DdahC here is in agreement with the proposal that the interaction between this class of enzymes and NAD(P)H is weak. As previously described, this weak interaction allows NAD^+, generated following RmlD-catalysed reduction of dTDP-6-deoxy-L-*lyxo*-4-hexulose, to be released to allow new cofactor binding, in order to restore functionality to the enzyme following the reduction of its substrate (Blankenfeldt *et al*., 2002).

Despite its conservation throughout the SDR family, including DdahC and MlghC, the role of the Ser, or Thr in the case of RmlD, remains unknown. It has been proposed to be required for the stabilisation of a reaction intermediate (Blankenfeldt *et al*. 2002; Giraud *et al*. 1999; Graninger *et al*. 1999).

5.4.5 MlghC and DdahC as inactive epimerases

Other members of the SDR family catalyse the epimerisation of their substrates prior to the reduction at the C4 position. An example of this is GDP-fucose synthetase, which catalyses the epimerisation of GDP-4-keto-6-deoxy-D-mannose at the C3 and C5 positions to form GDP-4-keto-5-deoxy-L-glucose, which it then reduces at the C4 position to form GDP-L-fucose, during the biosynthesis of GDP-L-fucose from GDP-D-mannose (Lau and Tanner, 2008; Somers *et al*., 1998). UDP-galactose 4-epimerase (GalE), which forms UDP-glucose from UDP-galactose during galactose metabolism in humans and biofilm formation in bacteria, is also a member of the SDR family (Beerens *et al*., 2015). During the GalE-catalysed conversion of UDP-galactose to UDP-glucose, UDP-galactose is oxidised at the C4 position, in an NAD^+-dependent manner, catalysed by the Tyr of the conserved YXXXK motif, generating NADH and UDP-4-keto-glucose. The 4-keto-glucose moiety then rotates in the active site, allowing the reduction of the keto group on the opposite face of the ring, thereby resulting in epimerisation of UDP-galactose at the C4 position, to give UDP-glucose, and NAD^+ (Beerens *et al*., 2015; Holden *et al*., 2003; Thoden *et al*., 1996). Despite the conservation of the YXXXK motif, with the Tyr substituted for Phe in the case of MlghC, this epimerisation is not observed in DdahC- or MlghC-catalysed reduction

during modified heptose biosynthesis (McCallum *et al*., 2012, 2013) or in RmlD-catalysed reduction during dTDP-L-rhamnose biosynthesis (Blankenfeldt *et al*., 2002; Graninger *et al*., 1999). The lack of epimerisation activity in RmlD was proposed to be due to the presence of a Tyr 106, required for the correct orientation of the substrate sugar moiety in the active site (Blankenfeldt *et al*., 2002). This bulky Tyr is believed to prevent the required rotation of the sugar moiety (Blankenfeldt *et al*., 2002). This Tyr is not conserved in GalE (Thoden *et al*., 1996), however it is conserved in MlghC (Tyr 109), offering a possible explanation as to why this enzyme also does not epimerise its substrate. An alignment of the RmlD and MlghC active sites, showing this additional conserved Tyr, and its position relative to dTDP-L-rhamnose is shown in figure 5.75. This lack of epimerisation activity could also be an evolutionary aspect of this pathway as these pathways also encode separate epimerases (Giraud *et al*., 2000; McCallum *et al*., 2012, 2013).

5.5 Conclusion and future work

The presence of modified heptoses in the *C. jejuni* and *Y. pseudotuberculosis* capsule has been demonstrated to be crucial to the invasion of host cells during infection and the subsequent persistence of the infection (Ho *et al*., 2008; Wong *et al*., 2015). However the enzymes required for their biosynthesis remained uncharacterised. The first structural characterisation of enzymes required for modified heptose biosynthesis is described here, and the comparison of these structures to those of dTDP-L-rhamnose biosynthetic enzymes, has facilitated new insights into the mechanisms of modified heptose biosynthesis in two pathogenic strains of *C. jejuni*, as well as *Y. pseudotuberculosis*. A mechanism was predicted here, based on those of the dTDP-L-rhamnose biosynthesis pathway, for each step required for modified heptose biosynthesis. These predicted mechanisms were built upon via the structural characterisation of modified heptose biosynthetic enzymes and the structural alignments of these enzymes and those required for dTDP-L-rhamnose biosynthesis. This approach has allowed the identification of the active site residues of the modified heptose biosynthetic enzymes described here. The determination of the structures of these

enzymes in complex with their native substrates or products, in combination with mutagenesis studies, is required to confirm these proposed mechanisms.

Despite the successful structural characterisation of these proteins, a number of questions remain open. The previously described difference in epimerisation activity between DdahB and MlghB, despite the high level of sequence identity and structural similarity, remains unexplained. Furthermore, the discussed substrate specificity of all of these enzymes for heptoses rather than hexoses also remains unexplained. The determination of the structures of these enzymes in complex with their native substrates or products would hopefully answer these remaining questions. A deeper understanding of the modes of action of these enzymes, resulting from the protein:substrate/product complex structure determination coupled with mutagenesis studies, based on the foundations laid here, will ultimately lead to inhibitor development and the expansion of the molecular toolkit available for the chemienzymatic synthesis of novel sugars.

6 Concluding Statement

The discovery of penicillin by Alexander Fleming in 1928 (Fleming, 1929) is hailed as the beginning of the "antibiotic era" (Zaman et al., 2017), however Fleming himself first warned about the potential for the arisal of antibiotic resistance, due to incorrect dosage and ease of access, during his Nobel lecture in 1945 (Fleming, 1945).

We are now on course for a "post-antibiotic era", during which medical procedures that, thanks to the "antibiotic era", have become routine, are predicted to become high risk (WHO, 2017a) [Accessed 28/02/2018]. Currently, approximately 700,000 deaths annually are attributed to antibiotic resistance worldwide (OECD, 2016) [Accessed 28/02/2018]. Levels of antibiotic resistance are also on the increase, with the probability of encountering an antibiotic resistant infection in 23 of 26 member countries of the Organisation for Economic Co-operation and Development (OECD) increasing from an average of 10 % in 2005 to 15 % in 2014 (OECD, 2016) [Accessed 28/02/2018]. This increase in the level of antibiotic-resistant infections is predicted to cost the OECD $2.9 trillion by 2050 (OECD, 2016) [Accessed 28/02/2018].

There are two modes by which antibiotics can enter their target bacteria: porin-mediated uptake, used primarily by small hydrophilic antibiotics, or diffusion across the membrane, used primarily by hydrophobic antibiotics (Delcour, 2009). In contrast to Gram-positive bacteria, Gram-negative bacteria possess an additional asymmetric membrane, termed the outer membrane (Silhavy et al., 2010). The presence of LPS in the outer leaflet of the Gram-negative outer membrane renders this membrane impermeable to many antibiotics (Delcour, 2009), and is responsible for the higher levels of innate antibiotic resistance of Gram-negative bacteria than Gram-positive bacteria (Silhavy et al., 2010). Nine of the twelve antibiotic-resistant bacteria identified by the World Health Organisation (WHO) as representing the highest risk to human health are Gram-negative bacteria (WHO, 2017b) [Accessed: 28/02/2018]. Despite the

demonstrated increased risk associated with a Gram-negative bacterial infection, none of the five new classes of antibiotic introduced since the year 2000 are able to treat a Gram-negative bacterial infection (OECD, 2016) [Accessed 28/02/2018].

Polymyxins are often referred to as the antibiotics of "last resort" in the treatment of an antibiotic-resistant Gram-negative bacterial infection (Olaitan, 2014). This class of antibiotics functions to permeabilise the LPS leaflet of the outer membrane by displacing the divalent cations required for the close packing of negatively-charged LPS (Delcour, 2009). However, cases of resistance to this "last resort" class of antibiotics have been detected (Delcour, 2009) and have been determined to result from alterations in LPS biosynthesis (Delcour, 2009; Olaitan, 2014). This demonstrates our requirement for a greater knowledge of the LPS biosynthetic pathway in the fight against antiobiotic resistance. The research described here goes some way to improving our understanding of the biosynthesis of LPS and also provides a foundation to build upon.

7 Bibliography

Abeyrathne, P.D., and Lam, J.S. (2007). WaaL of *Pseudomonas aeruginosa* utilizes ATP in *in vitro* ligation of O antigen onto lipid A-core. Mol. Microbiol. *65*, 1345–1359.

Abeyrathne, P.D., Daniels, C., Poon, K.K.H., Matewish, M.J., and Lam, J.S. (2005). Functional characterization of WaaL, a ligase associated with linking O-antigen polysaccharide to the core of *Pseudomonas aeruginosa* lipopolysaccharide. J. Bacteriol. *187*, 3002–3012.

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 213–221.

Allard, S.T.M., Giraud, M., Whitfield, C., Graninger, M., Messner, P., and Naismith, J.H. (2001). The Crystal Structure of dTDP-D-Glucose 4, 6-Dehydratase (RmlB) from *Salmonella enterica* Serovar Typhimurium, the Second Enzyme in the dTDP-L-Rhamnose Pathway. J. Mol. Biol. *307*, 283–295.

Allard, S.T.M., Beis, K., Giraud, M., Hegeman, A.D., Gross, J.W., Wilmouth, R.C., Whitfield, C., Graninger, M., Messner, P., Allen, A.G., et al. (2002). Toward a Structural Understanding of the Dehydratase Mechanism. Structure *10*, 81–92.

Allos, B.M. (2001). *Campylobacter jejuni* Infections: Update on Emerging Issues and Trends. Clin. Infect. Dis. *32*, 1201–1206.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Bacon, D.J., Szymanski, C.M., Burr, D.H., Silver, R.P., Alm, R.A., and Guerry, P. (2001). A phase-variable capsule is involved in virulence of *Campylobacter jejuni* 81-176. Mol. Microbiol. *40*, 769–777.

Bai, X., McMullan, G., and Scheres, S.H.W. (2015). How cryo-EM is revolutionizing structural biology. Trends Biochem. Sci. *40*, 49–57.

Barr, K., Klena, J., and Rick, P.D. (1999). The modality of enterobacterial common antigen polysaccharide chain lengths is regulated by *O349* of the *wec* gene cluster of *Escherichia coli* K-12. J. Bacteriol. *181*, 6564–6568.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L.S., and Subramaniam, S. (2015). 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. Science *348*, 1147–1151.

Bastin, D.A., Stevenson, G., Brown, P.K., Haase, A., and Reeves, P.R. (1993). Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol. Microbiol. *7*, 725–734.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM : a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. Sect. D Biol. Crystallogr. *D67*, 271–281.

Beerens, K., Soetaert, W., and Desmet, T. (2015). UDP-hexose 4-epimerases: a view on structure, mechanism and substrate specificity. Carbohydr. Res. *414*, 8–14.

Bergman, M.P., Engering, A., Smits, H.H., van Vliet, S.J., van Bodegraven, A.A.,

Wirth, H.-P., Kapsenberg, M.L., Vandenbroucke-Grauls, C.M.J.E., van Kooyk, Y., and Appelmelk, B.J. (2004). *Helicobacter pylori* Modulates the T Helper Cell 1/T Helper Cell 2 Balance through Phase-variable Interaction between Lipopolysaccharide and DC-SIGN. J. Exp. Med. *200*, 979–990.

Biesiadecki, B.J., and Jin, J.P. (2011). A high-throughput solid-phase microplate protein-binding assay to investigate interactions between myofilament proteins. J. Biomed. Biotechnol. *2011*, 1–8.

Blankenfeldt, W., Asuncion, M., Lam, J.S., and Naismith, J.H. (2000). The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA). EMBO *19*, 6652–6663.

Blankenfeldt, W., Kerr, I.D., Giraud, M., McMiken, H.J., Leonard, G., Whitfield, C., Messner, P., Graninger, M., and Naismith, J.H. (2002). Variation on a Theme of SDR : dTDP-6-Deoxy-L-lyxo-4-Hexulose Reductase (RmlD) Shows a New Mg^{2+} -Dependent Dimerization Mode. Structure *10*, 773–786.

Boivin, S., Kozak, S., and Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. Protein Expr. Purif. *91*, 192–206.

Botos, I., Majdalani, N., Mayclin, S.J., McCarthy, J.G., Lundquist, K., Wojtowicz, D., Barnard, T.J., Gumbart, J.C., and Buchanan, S.K. (2016). Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. Structure *24*, 965–976.

Branigan, E., Pliotas, C., Hagelueken, G., and Naismith, J.H. (2013). Quantification of free cysteines in membrane and soluble proteins using a fluorescent dye and thermal unfolding. Nat. Protoc. *8*, 2090–2097.

Bushell, S.R., Mainprize, I.L., Wear, M.A., Lou, H., Whitfield, C., and Naismith, J.H. (2013). Wzi Is an Outer Membrane Lectin that Underpins Group 1 Capsule Assembly in

Escherichia coli. Structure *21*, 844–853.

Butty, F.D., Aucoin, M., Morrison, L., Ho, N., Shaw, G., and Creuzenet, C. (2009). Elucidating the formation of 6-deoxyheptose: biochemical characterization of the GDP-D-*glycero*-D-*manno*-heptose C6 dehydratase, DmhA, and its associated C4 reductase, DmhB. Biochemistry *48*, 7764–7775.

Caffrey, M. (2015). A comprehensive review of the lipid cubic phase or *in meso* method for crystallizing membrane and soluble proteins and complexes. Acta Crystallogr. Sect. F Struct. Biol. Commun. *71*, 3–18.

Carter, J.A., Blondel, C.J., Zaldivar, M., Alvarez, S.A., Marolda, C.L., Valvano, M.A., and Contreras, I. (2007). O-antigen modal chain length in *Shigella flexneri* 2a is growth-regulated through RfaH-mediated transcriptional control of the *wzy* gene. Microbiology *153*, 3499–3507.

Carter, J.A., Jimenez, J.C., Zaldivar, M., Alvarez, S.A., Marolda, C.L., Valvano, M.A., and Contreras, I. (2009). The cellular level of O-antigen polymerase Wzy determines chain length regulation by WzzB and WzzpHS-2 in *Shigella flexneri* 2a. Microbiology *155*, 3260–3269.

Chang, C.-W., Tran, E.N.H., Ericsson, D.J., Casey, L.W., Lonhienne, T., Benning, F., Morona, R., and Kobe, B. (2015). Structural and biochemical analysis of a single amino-acid mutant of WzzBSF that alters lipopolysaccharide O-antigen chain length in *shigella flexneri*. PLoS One *10*, 1–21.

Chaptal, V., Delolme, F., Kilburg, A., Magnard, S., Montigny, C., Picard, M., Prier, C., Monticelli, L., Bornert, O., Agez, M., Ravaud, S., Orelle, C., Wagner, R., Jawhari, A., Broutin, I., Pebay-Peyroula, S., Jault, J-M., Kaback, H,R,, le Maire, M. and Falson, P. (2017). Quantification of Detergents Complexed with Membrane Proteins. Sci. Rep. *7*, 1–12.

Chen, V.B., Arendall III, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. Crystallogr. *D66*, 12–21.

Christendat, D., Saridakis, V., Dharamsi, A., Bochkarev, A., Pai, E.F., Arrowsmith, C.H., and Edwards, A.M. (2000). Crystal structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase from *Methanobacterium thermoautotrophicum* complexed with dTDP. J. Biol. Chem. *275*, 24608–24612.

Chung, B.C., Zhao, J., Gillespie, R.A., Kwon, D.-Y., Guan, Z., Hong, J., Zhou, P., and Lee, S.-Y. (2013). Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis. Science. *341*, 1012–1016.

Chung, B.C., Mashalidis, E.H., Tanino, T., Kim, M., Matsuda, A., Hong, J., Ichikawa, S., and Lee, S.-Y. (2016). Structural insights into inhibition of lipid I production in bacterial cell wall synthesis. Nature *533*, 557–560.

Clarke, B.R., Cuthbertson, L., and Whitfield, C. (2004). Nonreducing Terminal Modifications Determine the Chain Length of Polymannose O Antigens of *Escherichia coli* and Couple Chain Termination to Polymer Export via an ATP-binding Cassette Transporter. J. Biol. Chem. *279*, 35709–35718.

Clarke, B.R., Greenfield, L.K., Bouwman, C., and Whitfield, C. (2009). Coordination of Polymerization, Chain Termination, and Export in Assembly of the *Escherichia coli* Lipopolysaccharide O9a Antigen in an ATP-binding Cassette Transporter-dependent Pathway. J. Biol. Chem. *284*, 30662–30672.

Collins, R.F., Beis, K., Clarke, B.R., Ford, R.C., Hulley, M., Naismith, J.H., and Whitfield, C. (2006). Periplasmic protein-protein contacts in the inner membrane protein Wzc form a tetrameric complex required for the assembly of *Escherichia coli* group 1 capsules. J. Biol. Chem. *281*, 2144–2150.

Collins, R.F., Beis, K., Dong, C., Botting, C.H., McDonnell, C., Ford, R.C., Clarke, B.R., Whitfield, C., and Naismith, J.H. (2007). The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in *Escherichia coli*. Proc. Natl. Acad. Sci. USA *104*, 2390–2395.

Collins, R.F., Kargas, V., Clarke, B.R., Siebert, C.A., Clare, D.K., Bond, P.J., Whitfield, C., and Ford, R.C. (2017). Full-length, Oligomeric Structure of Wzz Determined by Cryoelectron Microscopy Reveals Insights into Membrane-Bound States. Structure *25*, 806–815.

Comstock, L.E., and Kasper, D.L. (2006). Bacterial Glycans: Key Mediators of Diverse Host Immune Responses. Cell *126*, 847–850.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1 . Tracing protein chains. Acta Crystallogr. Sect. D Biol. Crystallogr. *D62*, 1002–1011.

Cowtan, K. (2008). Fitting molecular fragments into electron density. Acta Crystallogr. Sect. D Biol. Crystallogr. *D64*, 83–89.

Cuthbertson, L., Powers, J., and Whitfield, C. (2005). The C-terminal Domain of the Nucleotide-binding Domain Protein Wzt Determines Substrate Specificity in the ATP-binding Cassette Transporter for the Lipopolysaccharide O-antigens in *Escherichia coli* serotypes O8 and O9a. J. Biol. Chem. *280*, 30310–30319.

Cuthbertson, L., Kimber, M.S., and Whitfield, C. (2007). Substrate binding by a bacterial ABC transporter involved in polysaccharide export. Proc. Natl. Acad. Sci. *104*, 19529–19534.

Dambacher, C.M., and Lander, G.C. (2015). Site-specific labeling of proteins for electron microscopy. J. Struct. Biol. *192*, 151–158.

Daniels, C., and Morona, R. (1999). Analysis of Shigella flexneri Wzz (Rol) function by mutagenesis and cross-linking: Wzz is able to oligomerize. Mol. Microbiol. *34*, 181–194.

Daniels, C., Vindurampulle, C., and Morona, R. (1998). Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy). Mol. Microbiol. *28*, 1211–1222.

Daniels, C., Griffiths, C., Cowles, B., and Lam, J.S. (2002). *Pseudomonas aeruginosa* O-antigen chain length is determined before ligation to lipid A core. Environ. Microbiol. *4*, 883–897.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall III, W.B., Snoeyink, J., Richardson, J.S. and Richardson D.C. (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, 375–383.

Davis Jr, M.R., and Goldberg, J.B. (2012). Purification and Visualization of Lipopolysaccharide from Gram-negative Bacteria by Hot Aqueous-phenol Extraction. J. Vis. Exp. *63*, 1–3.

Delcour, A.H. (2009). Outer Membrane Permability and Antibiotic Resistance. Biochim. Biophys. Acta. *1794*, 808-816.

Dong, C., Beis, K., Giraud, M.-F., Blankenfeldt, W., Allard, S., Major, L.L., Kerr, I.D., Whitfield, C., and Naismith, J.H. (2003a). A structural perspective on the enzymes that convert dTDP-D-glucose into dTDP-L-rhamnose. Biochem. Soc. Trans. *31*, 532–536.

Dong, C., Major, L.L., Allen, A., Blankenfeldt, W., Maskell, D., and Naismith, J.H. (2003b). High-resolution structures of RmlC from *Streptococcus suis* in complex with substrate analogs locate the active site of this class of enzyme. Structure *11*, 715–723.

Dong, C., Beis, K., Nesper, J., Brunkan-Lamontagne, A.L., Clarke, B.R., Whitfield, C., and Naismith, J.H. (2006). Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein. Nature *444*, 226–229.

Dong, C., Major, L.L., Srikannathasan, V., Errey, J.C., Giraud, M.F., Lam, J.S., Graninger, M., Messner, P., McNeil, M.R., Field, R.A., Whitfield, C., and Naismith, J.H. (2007). RmlC, a C3′ and C5′ Carbohydrate Epimerase, Appears to Operate via an Intermediate with an Unusual Twist Boat Conformation. J. Mol. Biol. *365*, 146–159.

Dong, H., Xiang, Q., Gu, Y., Wang, Z., Paterson, N.G., Stansfeld, P.J., He, C., Zhang, Y., Wang, W., and Dong, C. (2014). Structural basis for outer membrane lipopolysaccharide insertion. Nature *511*, 52–56.

Edmondson, D.G., and Dent, S.Y.R. (2001). Identification of protein interactions by far western analysis. Curr. Protoc. Protein Sci. *Chapter 19*, 1–10.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. *D66*, 486–501.

Evans, P. (2006). Scaling and assessment of data quality research papers. Acta Crystallogr. Sect. D Biol. Crystallogr. *D62*, 72–82.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. *D69*, 1204–1214.

Fan, J., Jiang, D., Zhao, Y., Liu, J., and Zhang, X.C. (2014). Crystal structure of lipid phosphatase Escherichia coli phosphatidylglycerophosphate phosphatase B. Proc. Natl. Acad. Sci. U. S. A. *111*, 7636–7640.

Feldman, M.F., Marolda, C.L., Monteiro, M.A., Perry, M.B., Parodi, A.J., and Valvano, M.A. (1999). The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in *Escherichia coli* O antigen assembly is independent of the chemical

structure of the O repeat. J. Biol. Chem. *274*, 35129–35138.

Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., and Aebi, M. (2005). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *102*, 3016–3021.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., and Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. *44*, D279–D285.

Fleming, A. (1929). On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of *B. Influenzae*. Br. J. Exp. Pathol. *10*, 226-236.

Fleming, A. (1945). Pencillin, Nobel lecture. 83-93. https://www.nobelprize.org/nobel_prizes/medicine/laureates/1945/fleming-lecture.pdf (Last accessed: 01/03/18).

Flemming, D., Thierbach, K., Stelter, P., Böttcher, B., and Hurt, E. (2010). Precise mapping of subunits in multiprotein complexes by a versatile electron microscopy label. Nat. Struct. Mol. Biol. *17*, 775–778.

Freinkman, E., Okuda, S., Ruiz, N., and Kahne, D. (2012). Regulated assembly of the transenvelope protein complex required for lipopolysaccharide export. Biochemistry *51*, 4800–4806.

Fry, B.N., Feng, S., Chen, Y.-Y., Newell, D.G., Coloe, P.J., and Korolik, V. (2000). The *galE* gene of *Campylobacter jejuni* is involved in lipopolysaccharide synthesis and virulence. Infect. Immun. *68*, 2594–2601.

Furlong, S.E., Ford, A., Albarnez-rodriguez, L., and Valvano, M.A. (2015). Topological analysis of the *Escherichia coli* WcaJ protein reveals a new hexose-1-phosphate transferase family. Sci. Rep. *5*, 1–9.

Gaynor, E.C., Cawthraw, S., Manning, G., Mackichan, J.K., Falkow, S., and Newell, D.G. (2004). The Genome-Sequenced Variant of *Campylobacter jejuni* NCTC 11168 and the Original Clonal Clinical Isolate Differ Markedly in Colonization, Gene Expression, and Virulence-Associated Phenotypes. J. Bacteriol. *186*, 503–517.

Geladopoulos, T.P., Sotiroudis, T.G., and Evangelopoulos, A.E. (1991). A malachite green colorimetric assay for protein phosphatase activity. Anal. Biochem. *192*, 112–116.

El Ghachi, M., Bouhss, A., Blanot, D., and Mengin-Lecreulx, D. (2004). The *bacA* gene of *Escherichia coli* encodes an undecaprenyl pyrophosphate phosphatase activity. J. Biol. Chem. *279*, 30106–30113.

El Ghachi, M., Derbise, A., Bouhss, A., and Mengin-Lecreulx, D. (2005). Identification of multiple genes encoding membrane proteins with undecaprenyl pyrophosphate phosphatase (UppP) activity in *Escherichia coli*. J. Biol. Chem. *280*, 18689–18695.

Gildea, R.J., Waterman, D.G., Parkhurst, J.M., Axford, D., Sutton, G., Stuart, D.I., Sauter, N.K., Evans, G., and Winter, G. (2014). New methods for indexing multi-lattice diffraction data. Acta Crystallogr. Sect. D Biol. Crystallogr. *D70*, 2652–2666.

Giraud, M., McMiken, H.J., Leonard, G.A., Messner, P., Whitfield, C., and Naismith, J.H. (1999). Overexpression, purification, crystallization and preliminary structural study of dTDP-6-deoxy-L-*lyxo*-4-hexulose reductase (RmlD), the fourth enzyme of the dTDP-L-rhamnose synthesis pathway, from *Salmonella enterica* serovar *Typhirmurium*. Acta Crystallogr. Sect. D Biol. Crystallogr. *D55*, 2043–2046.

Giraud, M., Leonard, G.A., Field, R.A., Berlind, C., and Naismith, J.H. (2000). RmlC,

the third enzyme of pathway, is a new class of epimerase. Nature *7*, 398–402.

Glover, K.J., Weerapana, E., Chen, M.M., and Imperiali, B. (2006). Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. Biochemistry *45*, 5343–5350.

Graninger, M., Nidetzky, B., Heinrichs, D.E., Whitfield, C., and Messner, P. (1999). Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* Serovar *Typhimurium* LT2. Biochemistry *274*, 25069–25077.

Greenfield, L.K., Richards, M.R., Li, J., Wakarchuk, W.W., Lowary, T.L., and Whitfield, C. (2012). Biosynthesis of the polymannose lipopolysaccharide O-antigens from *Escherichia coli* serotypes O8 and O9a requires a unique combination of single- and multiple-active site mannosyltransferases. J. Biol. Chem. *287*, 35078–35091.

Gregson, N.A., Koblar, S., and Hughes, R.A. (1993). Antibodies to gangliosides in Guillain-Barre syndrome: specificity and relationship to clinical features. Q. J. Med. *86*, 111–117.

Gu, Y., Stansfeld, P.J., Zeng, Y., Dong, H., Wang, W., and Dong, C. (2015). Lipopolysaccharide is inserted into the outer membrane through an intramembrane hole, a lumen gate, and the lateral opening of LptD. Structure *23*, 496–504.

Guerry, P., Szymanski, C.M., Prendergast, M.M., Hickey, T.E., Ewing, C.P., Pattarini, D.L., and Moran, A.P. (2002). Phase Variation of *Campylobacter jejuni* 81-176 Lipooligosaccharide Affects Ganglioside Mimicry and Invasiveness *In Vitro*. Infect. Immun. *70*, 787–793.

Guo, H., Lokko, K., Zhang, Y., Yi, W., Wu, Z., and Wang, P.G. (2006). Overexpression and characterization of Wzz of *Escherichia coli* O86:H2. Protein Expr. Purif. *48*, 49–55.

Ha, S., Walker, D., Shi, Y., and Walker, S. (2000). The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. Protein Sci. *9*, 1045–1052.

Hagelueken, G., Huang, H., Clarke, B.R., Lebl, T., Whitfield, C., and Naismith, J.H. (2012). Structure of WbdD: A bifunctional kinase and methyltransferase that regulates the chain length of the O antigen in *Escherichia coli* O9a. Mol. Microbiol. *86*, 730–742.

Hagelueken, G., Clarke, B.R., Huang, H., Tuukkanen, A., Danciu, I., Svergun, D.I., Hussain, R., Liu, H., Whitfield, C., and Naismith, J.H. (2015). A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. Nat. Struct. Mol. Biol. *22*, 50–56.

Han, W., Wu, B., Li, L., Zhao, G., Woodward, R., Pettit, N., Cai, L., Thon, V., and Wang, P.G. (2012). Defining function of lipopolysaccharide O-antigen ligase waal using chemoenzymatically synthesized substrates. J. Biol. Chem. *287*, 5357–5365.

Hanukoglu, I. (2015). Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. Biochem. Mol. Biol. Educ. *43*, 206–209.

Hartley, M.D., Morrison, M.J., Aas, F.E., Børud, B., Koomey, M., and Imperiali, B. (2011). Biochemical Characterization of the O-Linked Glycosylation Pathway in *Neisseria gonorrhoeae* responsible for biosynthesis of protein glycans containing N,N'-diacetylbacillosamine. Biochemistry *50*, 4936–4948.

He, X., Szewczyk, P., Karyakin, A., Evin, M., Hong, W.-X., Zhang, Q., and Chang, G. (2010). Structure of a cation-bound multidrug and toxic compound extrusion transporter. Nature *467*, 991–994.

Ho, N., Kondakova, A.N., Knirel, Y.A., and Creuzenet, C. (2008). The biosynthesis and biological role of 6-deoxyheptose in the lipopolysaccharide O-antigen of *Yersinia*

pseudotuberculosis. Mol. Microbiol. *68*, 424–447.

Hoa Tran, E.N., Papadopoulos, M., and Morona, R. (2014). Relationship between O-antigen chain length and resistance to colicin E2 in *Shigella flexneri*. Microbiology *160*, 589–601.

Holden, H.M., Rayment, I., and Thoden, J.B. (2003). Structure and Function of Enzymes of the Leloir Pathway for Galactose Metabolism. J. Biol. Chem. *278*, 43885–43888.

Hong, M., and Payne, S.M. (1997). Effect of mutations in *Shigella flexneri* chromosomal and plasmid-encoded lipopolysaccharide genes on invasion and serum resistance. Mol. Microbiol. *24*, 779–791.

Hong, Y., and Reeves, P.R. (2014). Diversity of O-antigen repeat unit structures can account for the substantial sequence variation of Wzx translocases. J. Bacteriol. *196*, 1713–1722.

Hong, Y., Morcilla, V.A., Liu, M.A., Russell, E.L.M., and Reeves, P.R. (2015). Three Wzy polymerases are specific for particular forms of an internal linkage in otherwise identical O units. Microbiology *161*, 1639–1647.

Hu, L., and Kopecko, D.J. (1999). *Campylobacter jejuni* 81-176 associates with microtubules and dynein during invasion of human intestinal cells. Infect. Immun. *67*, 4171–4182.

Hug, I., and Feldman, M.F. (2011). Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. Glycobiology *21*, 138–151.

Hvorup, R.N., Winnen, B., Chang, A.B., Jiang, Y., Zhou, X.F., and Saier Jr, M.H. (2003). The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. Eur. J. Biochem. *270*, 799–813.

Igura, M., Maita, N., Kamishikiryo, J., Yamada, M., Obita, T., Maenaka, K., and Kohda, D. (2008). Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. EMBO *27*, 234–243.

Ilca, S.L., Kotecha, A., Sun, X., Poranen, M.M., Stuart, D.I., and Huiskonen, J.T. (2015). Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. Nat. Comms. *6*, 1–8.

Islam, S.T., and Lam, J.S. (2014). Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. Can. J. Microbiol. *60*, 697–716.

Islam, S.T., Huszczynski, S.M., Nugent, T., Gold, A.C., and Lam, J.S. (2013). Conserved-residue mutations in Wzy affect O-antigen polymerization and Wzz-mediated chain-length regulation in *Pseudomonas aeruginosa* PAO1. Sci. Rep. *3*, 1–9.

Jin, Y., Nair, A., and Van Veen, H.W. (2014). Multidrug transport protein NorM from *Vibrio cholerae* simultaneously couples to sodium- and proton-motive force. J. Biol. Chem. *289*, 14624–14632.

Joosten, R.P., Long, F., Murshudov, G.N., and Perrakis, A. (2014). The PDB_REDO server for macromolecular structure model optimization. IUCrJ *1*, 213–220.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr. Sect. D Biol. Crystallogr. *D66*, 133–144.

Kalynych, S., Ruan, X., Valvano, M.A., and Cygler, M. (2011). Structure-guided investigation of lipopolysaccharide O-antigen chain length regulators reveals regions critical for modal length control. J. Bacteriol. *193*, 3710–3721.

Kalynych, S., Yao, D., Magee, J., and Cygler, M. (2012a). Structural characterization of closely related O-antigen lipopolysaccharide (LPS) chain length regulators. J. Biol.

Chem. *287*, 15696–15705.

Kalynych, S., Valvano, M.A., and Cygler, M. (2012b). Polysaccharide co-polymerases: The enigmatic conductors of the O-antigen assembly orchestra. Protein Eng. Des. Sel. *25*, 797–802.

Kalynych, S., Cherney, M., Bostina, M., Rouiller, I., and Cygler, M. (2015). Quaternary structure of WzzB and WzzE polysaccharide copolymerases. Protein Sci. *24*, 58–69.

Kanipes, M.I., Holder, L.C., Corcoran, A.T., Moran, A.P., and Guerry, P. (2004). A Deep-Rough Mutant of *Campylobacter jejuni* 81-176 Is Noninvasive for Intestinal Epithelial Cells. Infect. Immun. *72*, 2452–2455.

Kantardjieff, K.A., and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. *12*, 1865–1871.

Kavanagh, K.L., Jörnvall, H., Persson, B., and Oppermann, U. (2008). Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: Functional and structural diversity within a family of metabolic and regulatory enzymes. Cell. Mol. Life Sci. *65*, 3895–3906.

Keegan, R.M., and Winn, M.D. (2007). Automated search-model discovery and preparation for structure solution by molecular replacement. Acta Crystallogr. Sect. D Biol. Crystallogr. *D63*, 447–457.

Keegan, R.M., and Winn, M.D. (2008). MrBUMP: an automated pipeline for molecular replacement. Acta Crystallogr. Sect. D Biol. Crystallogr. *D64*, 119–124.

Kefala, G., Kwiatkowski, W., Esquivies, L., Maslennikov, I., and Choe, S. (2007). Application of Mistic to improving the expression and membrane integration of histidine kinase receptors from *Escherichia coli.* J. Struct. Funct. Genomics *8*, 167–172.

Kim, T.H., Sebastian, S., Pinkham, J.T., Ross, R.A., Blalock, L.T., and Kasper, D.L. (2010). Characterization of the O-antigen polymerase (Wzy) of *Francisella tularensis*. J. Biol. Chem. *285*, 27839–27849.

King, J.D., Berry, S., Clarke, B.R., Morris, R.J., and Whitfield, C. (2014). Lipopolysaccharide O antigen size distribution is determined by a chain extension complex of variable stoichiometry in *Escherichia coli* O9a. Proc. Natl. Acad. Sci. U. S. A. *111*, 6407–6412.

Kintz, E.N., and Goldberg, J.B. (2011). Site-directed mutagenesis reveals key residue for O antigen chain length regulation and protein stability in *Pseudomonas aeruginosa* Wzz2. J. Biol. Chem. *286*, 44277–44284.

Kneidinger, B., Graninger, M., Puchberger, M., Kosma, P., and Messner, P. (2001). Biosynthesis of Nucleotide-activated D-*glycero*-D-*manno*-Heptose. J. Biol. Chem. *276*, 20935–20944.

Kohanski, M.A., Dwyer, D.J., and Collins, J.J. (2010). How antibiotics kill bacteria: from targets to networks. Nat. Rev. Microbiol. *8*, 423–435.

Kohlstaedt, M., Von der Hocht, I., Hilbers, F., Thielmann, Y., and Michel, H. (2015). Development of a Thermofluor assay for stability determination of membrane proteins using the Na^+/H^+ antiporter NhaA and cytochrome c oxidase. Acta Crystallogr. Sect. D Biol. Crystallogr. *D71*, 1112–1122.

Kong, L., Harrington, L., Li, Q., Cheley, S., Davis, B.G., and Bayley, H. (2013). Single-molecule interrogation of a bacterial sugar transporter allows the discovery of an extracellular inhibitor. Nat. Chem. *5*, 651–659.

Kong, L., Almond, A., Bayley, H., and Davis, B.G. (2016). Chemical polyglycosylation and nanolitre detection enables single-molecule recapitulation of bacterial sugar export.

Nat. Chem. *8*, 461–469.

Korlath, J.A., Osterholm, M.T., Judy, L.A., Forfang, J.C., and Robinson, R.A. (1985). A Point-Source Outbreak of Campylobacteriosis Associated with Consumption of Raw Milk. J. Infect. Dis. *152*, 592–596.

Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., et al. (2006). Definition of the bacterial N-glycosylation site consensus sequence. EMBO J. *25*, 1957–1966.

Krauss, I.R., Merlino, A., Vergara, A., and Sica, F. (2013). An overview of biological macromolecule crystallization. Int. J. Mol. Sci. *14*, 11643–11691.

Krissinel, E. (2009). Crystal Contacts as Nature's Docking Solutions. J. Comput. Chem. *31*, 133–143.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. Sect. D Biol. Crystallogr. *D60*, 2256–2268.

Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. J. Mol. Biol. *372*, 774–797.

Kubiak, R.L., Phillips, R.K., Zmudka, M.W., Ahn, M.R., Maka, E.M., Pyeatt, G.L., Roggensack, S.J., and Holden, H.M. (2012). Structural and functional studies on a 3'-epimerase involved in the biosynthesis of dTDP-6-deoxy-D-allose. Biochemistry *51*, 9375–9383.

Kuhn, H.-M., Meier-Dieter, U., and Mayer, H. (1988). ECA, the enterobacterial common antigen. FEMS Microbiol. Lett. *54*, 195–222.

Kuk, A.C.Y., Mashalidis, E.H., and Lee, S.-Y. (2016). Crystal structure of the MOP

flippase MurJ in an inward-facing conformation. Nat. Struct. Mol. Biol. *24*, 171–176.

Laddomada, F., Miyachiro, M.M., and Dessen, A. (2016). Structural Insights into Protein-Protein Interactions Involved in Bacterial Cell Wall Biogenesis. Antibiotics *5*, 1–13.

Laganowsky, A., Reading, E., Hopper, J.T.S., and Robinson, C. V. (2014). Mass Spectrometry of Intact Membrane Protein Complexes. Nat. Protoc. *8*, 639–651.

Langer, G.G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. *3*, 1171–1179.

Larue, K., Kimber, M.S., Ford, R., and Whitfield, C. (2009). Biochemical and structural analysis of bacterial O-antigen chain length regulator proteins reveals a conserved quaternary structure. J. Biol. Chem. *284*, 7395–7403.

Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. *51*, 2778–2786.

Lau, S.T.B., and Tanner, M.E. (2008). Mechanism and Active Site Residues of GDP-Fucose Synthase. J. Am. Chem. Soc. *130*, 17593–17602.

Lesk, A.M. (1995). NAD-binding domains of dehydrogenases. Curr. Opin. Struct. Biol. *5*, 775–783.

Levengood, M.R., Splain, R.A., and Kiessling, L.L. (2011). Monitoring processivity and length control of a carbohydrate polymerase. J. Am. Chem. Soc. *133*, 12758–12766.

Li, H., Yang, T., Liao, T., Debowski, A.W., Nilsson, H.-O., Haslam, S.M., Dell, A., Stubbs, K.A., Marshall, B.J., and Benghezal, M. (2017). The redefinition of

Helicobacter pylori lipopolysaccharide O-antigen and core-oligosaccharide domains. PLoS Pathog. *4*, 175–178.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods *10*, 584–590.

Li, X., Gu, Y., Dong, H., Wang, W., and Dong, C. (2015). Trapped lipopolysaccharide and LptD intermediates reveal lipopolysaccharide translocation steps across the Escherichia coli outer membrane. Sci. Rep. *5*, 1–8.

Liu, H., and Naismith, J.H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. *91*, 1–10.

Liu, H., and Naismith, J.H. (2009). A simple and efficient expression and purification system using two newly constructed vectors. Protein Expr. Purif. *63*, 102–111.

Lizak, C., Gerber, S., Numao, S., Aebi, M., and Locher, K.P. (2011). X-ray structure of a bacterial oligosaccharyltransferase. Nature *474*, 350–355.

Lo, M., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal. Biochem. *332*, 153–159.

Long, F., Vagin, A.A., and Murshudov, G.N. (2008). BALBES: a molecular-replacement pipeline. Acta Crystallogr. Sect. D Biol. Crystallogr. *D64*, 125–132.

Luo, Q., Yang, X., Yu, S., Shi, H., Wang, K., Xiao, L., Zhu, G., Sun, C., Li, T., Li, D., Zhang, X., Zhou, M. and Huang, Y. (2017). Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG. Nat. Struct. Mol. Biol. *24*, 469–474.

Major, L.L., Wolucka, B.A., and Naismith, J.H. (2005). Structure and function of GDP-mannose-3′,5′-epimerase: An enzyme which performs three chemical reactions at the same active site. J. Am. Chem. Soc. *127*, 18309–18320.

Marczak, M., Dźwierzyńska, M., and Skorupska, A. (2013). Homo- and heterotypic interactions between Pss proteins involved in the exopolysaccharide transport system in *Rhizobium leguminosarum* bv. *trifolii*. Biol. Chem. *394*, 541–559.

Marolda, C.L., Tatar, L.D., Alaimo, C., Aebi, M., and Valvano, M.A. (2006). Interplay of the Wzx translocase and the corresponding polymerase and chain length regulator proteins in the translocation and periplasmic assembly of lipopolysaccharide O antigen. J. Bacteriol. *188*, 5124–5135.

Matthews, B.W. (1968). Solvent Content of Protein Crystals. J. Mol. Biol. *33*, 491–497.

McCallum, M., Shaw, G.S., and Creuzenet, C. (2011). Characterization of the dehydratase WcbK and the reductase WcaG involved in GDP-6-deoxy-*manno*-heptose biosynthesis in *Campylobacter jejuni*. Biochem. J. *248*, 235–248.

McCallum, M., Shaw, S.D., Shaw, G.S., and Creuzenet, C. (2012). Complete 6-deoxy-D-*altro*-heptose biosynthesis pathway from *Campylobacter jejuni*: More complex than anticipated. J. Biol. Chem. *287*, 29776–29788.

McCallum, M., Shaw, G.S., and Creuzenet, C. (2013). Comparison of predicted epimerases and reductases of the *campylobacter jejuni* D-*altro*- and L-*gluco*-heptose synthesis pathways. J. Biol. Chem. *288*, 19569–19580.

Mccoy, A.J., Grosse-kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M.I., Pragani, R., Boxer, M.B., Earl, L.A., Milne, J.L.S., and Subramaniam, S. (2016). Breaking Cryo-

EM Resolution Barriers to Facilitate Drug Discovery. Cell *165*, 1698–1707.

Merkel, A.B., Major, L.L., Errey, J.C., Burkart, M.D., Field, R.A., Walsh, C.T., and Naismith, J.H. (2004). The position of a key tyrosine in dTDP-4-keto-6-deoxy-D-glucose-5-epimerase (EvaD) alters the substrate profile for this RmlC-like enzyme. J. Biol. Chem. *279*, 32684–32691.

St. Michael, F., Szymanski, C.M., Li, J., Chan, K.H., Khieu, N.H., Larocque, S., Wakarchuk, W.W., Brisson, J.R., and Monteiro, M.A. (2002). The structures of the lipooligosaccharide and capsule polysaccharide of *campylobacter jejuni* genome sequenced strain NCTC 11168. Eur. J. Biochem. *269*, 5119–5136.

Miroux, B., and Walker, J.E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol. *260*, 289–298.

Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S., and Stewart, P.D.S. (2014). Membrane protein structure determination - The next generation. Biochim. Biophys. Acta - Biomembr. *1838*, 78–87.

Morona, R., Van den Bosch, L., and Manning, P.A. (1995). Molecular, genetic, and topological characterization of O-antigen chain length regulation in *Shigella flexneri*. J. Bacteriol. *177*, 1059–1068.

Morona, R., Purins, L., Tocilj, A., Matte, A., and Cygler, M. (2009). Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins. Trends Biochem. Sci. *34*, 78–84.

Mulichak, A.M., Bonin, C.P., Reiter, W.-D., and Garavito, R.M. (2002). Structure of the MUR1 GDP-Mannose 4, 6-Dehydratase from *Arabidopsis thaliana*: Implications for Ligand Binding and Specificity. Biochemistry *41*, 15578–15589.

Murray, G.L., Attridge, S.R., and Morona, R. (2003). Regulation of *Salmonella typhimurium* lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second Wzz. Mol. Microbiol. *47*, 1395–1406.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallogr. Sect. D Biol. Crystallogr. *D53*, 240–255.

Naismith, J.H. (2004). Chemical insights from structural studies of enzymes. Biochem. Soc. Trans. *32*, 647–654.

Narita, S., and Tokuda, H. (2009). Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides. FEBS Lett. *583*, 2160–2164.

Nath, P., and Morona, R. (2015a). Detection of Wzy/Wzz interaction in *Shigella flexneri*. Microbiol. *161*, 1797–1805.

Nath, P., and Morona, R. (2015b). Mutational analysis of the major periplasmic loops of *Shigella flexneri* Wzy: identification of the residues affecting O antigen modal chain length control, and Wzz-dependent polymerization activity. Microbiology *161*, 774–785.

Nath, P., Tran, E.N.H., and Morona, R. (2015). Mutational analysis of the *Shigella flexneri* O-antigen polymerase Wzy: Identification of Wzz-dependent Wzy mutants. J. Bacteriol. *197*, 108–119.

Nesper, J., Lauriano, C.M., Klose, K.E., Kapfhammer, D., Kraiß, A., and Reidl, J. (2001). Characterization of *Vibrio cholerae* O1 El Tor *galU* and *galE* Mutants : Influence on Lipopolysaccharide Structure, Colonization, and Biofilm Formation. Infect. Immun. *69*, 435–445.

Newman, J. (2004). Novel buffer systems for macromolecular crystallization. Acta Crystallogr. Sect. D Biol. Crystallogr. *60*, 610–612.

Newstead, S., Ferrandon, S., and Iwata, S. (2008). Rationalizing α-helical membrane protein crystallization. Protein Sci. *17*, 466–472.

Nickerson, N.N., Mainprize, I.L., Hampton, L., Jones, M.L., Naismith, J.H., and Whitfield, C. (2014). Trapped translocation intermediates establish the route for export of capsular polysaccharides across *Escherichia coli* outer membranes. Proc. Natl. Acad. Sci. U. S. A. *111*, 8203–8208.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205–217.

OECD (2016). Antimicrobial resistance: Policy insights. https://www.oecd.org/health/health-systems/AMR-Policy-Insights-November2016.pdf (Last accessed: 01/03/18)

Oelschlaeger, T.A., Guerry, P., and Kopecko, D.J. (1993). Unusual microtubule-dependent endocytosis mechanisms triggered by *Campylobacter jejuni* and *Citrobacter freundii.* Proc. Natl. Acad. Sci. U. S. A. *90*, 6884–6888.

Okuda, S., Freinkman, E., and Kahne, D. (2012). Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in *E. coli*. Science. *338*, 1214–1217.

Olaitan, A.O., Morand, S. and Rolain, J.-M. (2014) Mechanisms of polymyxin resistance: acquired and instrinsic resistance in bacteria. Frontiers in Microbiology. *5*, 1-18.

Omasits, U., Ahrens, C.H., Muller, S., and Wollscheid, B. (2014). Protter: Interactive

protein feature visualization and integration with experimental proteomic data. Bioinformatics *30*, 884–886.

Pacinelli, E., Wang, L., and Reeves, P.R. (2002). Relationship of *Yersinia pseudotuberculosis* O antigens IA, IIA, and IVB: The IIA gene cluster was derived from that of IVB. Infect. Immun. *70*, 3271–3276.

Paiment, A., Hocking, J., and Whitfield, C. (2002). Impact of phosphorylation of specific residues in the tyrosine autokinase, Wzc, on its activity in assembly of group 1 capsules in Escherichia coli. J. Bacteriol. *184*, 6437–6447.

Painter, J., and Merritt, E.A. (2006a). TLSMD web server for the generation of multi-group TLS models. J. Appl. Crystallogr. *39*, 109–111.

Painter, J., and Merritt, E.A. (2006b). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. Sect. D Biol. Crystallogr. *D62*, 439–450.

Papp-Szabo, E., Kanipes, M.I., Guerry, P., and Monteiro, M.A. (2005). Cell-surface alpha-glucan in *Campylobacter jejuni* 81-176. Carbohydr. Res. *340*, 2218–2221.

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S.G.F., Wohlkönig, A., Ruf, A., Muyldermans, S., Hol, W.G.J., Kobilka, B.K., and Steyaert, J. (2014). A general protocol for the generation of Nanobodies for structural biology. Nat. Protoc. *9*, 674–693.

Park, B.S., Song, D.H., Kim, H.M., Choi, B.-S., Lee, H., and Lee, J.-O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. Nature *458*, 1191–1195.

Parker, J.L., and Newstead, S. (2012). Current trends in α-helical membrane protein crystallization: An update. Protein Sci. *21*, 1358–1365.

Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A.V., Moule, S., Pallen, M.J., Penn, C.W., Quail, M.A., Rajandram, M.A., Rutherford, K.M., van Vliet, A.H., Whitehead, S., and Barrell, B.G. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature *403*, 665–668.

Penna, A. (2008). Subunit stoichiometry determination by perfluorooctanoic acid polyacrylamide gel electrophoresis (PFO-PAGE). Nat. Protoc. Exch.

Perez, C., Gerber, S., Boilevin, J., Bucher, M., Darbre, T., Aebi, M., Reymond, J.-L., and Locher, K.P. (2015). Structure and mechanism of an active lipid-linked oligosaccharide flippase. Nature *524*, 433–438.

Perez, C., Köhler, M., Janser, D., Pardon, E., Steyaert, J., Zenobi, R., and Locher, K.P. (2017). Structural basis of inhibition of lipid-linked oligosaccharide flippase PglK by a conformational nanobody. Sci. Rep. *7*, 1–9.

Pérez, J.M., McGarry, M.A., Marolda, C.L., and Valvano, M.A. (2008). Functional analysis of the large periplasmic loop of the *Escherichia coli* K-12 WaaL O-antigen ligase. Mol. Microbiol. *70*, 1424–1440.

Prendergast, M.M., Lastovica, A.J., and Moran, A.P. (1998). Lipopolysaccharides from *Campylobacter jejuni* O:41 Strains Associated with Guillain-Barre Syndrome Exhibit Mimicry of GM 1 Ganglioside. Infect. Immun. *66*, 3649–3655.

Privé, G.G. (2007). Detergents for the stabilization and crystallization of membrane proteins. Methods *41*, 388–397.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general

procedure of protein complex purification. Methods *24*, 218–229.

Qiao, S., Luo, Q., Zhao, Y., Zhang, X.C., and Huang, Y. (2014). Structural basis for lipopolysaccharide insertion in the bacterial outer membrane lipopolysaccharide. Nature *511*, 108–111.

Raetz, C.R., and Whitfield, C. (2002). Lipopolysaccharide endotoxins. Annu. Rev. Biochem. *71*, 635–700.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol. *17*, 1030–1032.

Roosild, T.P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005). NMR Structure of Mistic, a Membrane-Integrating Protein for Membrane Protein Expression. Science. *307*, 1317–1321.

Ruan, X., Loyola, D.E., Marolda, C.L., Perez-Donoso, J.M., and Valvano, M.A. (2012). The WaaL O-antigen lipopolysaccharide ligase has features in common with metal ion-independent inverting glycosyltransferases. Glycobiology *22*, 288–299.

Saldías, M.S., Patel, K., Marolda, C.L., Bittner, M., Contreras, I., and Valvano, M.A. (2008). Distinct functional domains of the Salmonella enterica WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. Microbiology *154*, 440–453.

Scheres, S.H.W. (2012a). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Scheres, S.H.W. (2012b). A bayesian view on cryo-EM structure determination. J. Mol. Biol. *415*, 406–418.

Scheres, S.H.W. (2016). Processing of Structurally Heterogeneous Cryo-EM Data in RELION. Methods in Enzymology, *579*, 125-157.

Shang, Y., Ren, F., Song, Z., Li, Q., Zhou, X., Wang, X., Xu, Z., Bao, G., Wan, T., Lei, T., Wang, N., Jiao, X.-A., and Huang, J. (2016). Insights into Campylobacter jejuni colonization and enteritis using a novel infant rabbit model. Sci. Rep. *6*, 1–12.

Sherman, D.J., Lazarus, M.B., Murphy, L., Liu, C., Walker, S., Ruiz, N., and Kahne, D. (2014). Decoupling catalytic activity from biological function of the ATPase that powers lipopolysaccharide transport. Proc. Natl. Acad. Sci. U.S.A. *111*, 4982–4987.

Silhavy, T., Kahne, D., and Walker, S. (2010). The bacterial cell envelope. Cold Spring Harb. Perspect. Biol. *2*, 1–16.

Somers, W.S., Stahl, M.L., and Sullivan, F.X. (1998). GDP-fucose synthetase from *Escherichia coli*: structure of a unique member of the short-chain dehydrogenase/reductase family that catalyzes two distinct reactions at the same active site. Structure *6*, 1601–1612.

Sternberg, M.J.E., Tamaddoni-Nezhad, A., Lesk, V.I., Kay, E., Hitchen, P.G., Cootes, A., Van Alphen, L.B., Lamoureux, M.P., Jarrell, H.C., Rawlings, C.J., Soo, E.C., Szymanski, C.M., Dell, A., Wren, B.W., and Muggleton, S.H. (2013). Gene Function Hypotheses for the *Campylobacter jejuni* Glycome Generated by a Logic-Based Approach. J. Mol. Biol. *425*, 186–197.

Strop, P., and Brunger, A.T. (2005). Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. Protein Sci. *14*, 2207–2211.

Suits, M.D.L., Sperandeo, P., Dehò, G., Polissi, A., and Jia, Z. (2008). Novel Structure of the Conserved Gram-Negative Lipopolysaccharide Transport Protein A and Mutagenesis Analysis. J. Mol. Biol. *380*, 476–488.

Szymanski, C.M., Burr, D.H., and Guerry, P. (2002). *Campylobacter* Protein Glycosylation Affects Host Cell Interactions. Infect. Immun. *70*, 2242–2244.

Tang, K.H., Guo, H., Yi, W., Tsai, M.D., and Wang, P.G. (2007). Investigation of the conformational states of Wzz and the Wzz·O-antigen complex under near-physiological conditions. Biochemistry *46*, 11744–11752.

Taylor, V.L., Udaskin, M.L., Islam, S.T., and Lam, J.S. (2013). The D3 bacteriophage α-polymerase inhibitor (Iap) peptide disrupts O-antigen biosynthesis through mimicry of the chain length regulator Wzz in *Pseudomonas aeruginosa.* J. Bacteriol. *195*, 4735–4741.

Thoden, J.B., Frey, P.A., and Holden, H.M. (1996). Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from *Escherichia coli*: implications for the catalytic mechanism. Biochemistry *35*, 5137–5144.

Thomas, D.K., Lone, A.G., Selinger, L.B., Taboada, E.N., Uwiera, R.R.E., Abbott, D.W., and Inglis, G.D. (2014). Comparative Variation within the Genome of *Campylobacter jejuni* NCTC 11168 in Human and Murine Hosts. PLoS One *9*, 1–11.

Tocilj, A., Munger, C., Proteau, A., Morona, R., Purins, L., Ajamian, E., Wagner, J., Papadopoulos, M., Van Den Bosch, L., Rubinstein, J.L., Fethiere, J., Matte, M., and Cygler, M. (2008). Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. Nat. Struct. Mol. Biol. *15*, 130–138.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.M., Taly, J.-F., and Notredame, C. (2011). T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. *39*, 13–17.

Tong, S., Lin, Y., Lu, S., Wang, M., Bogdanov, M., and Zheng, L. (2016). Structural insight into substrate selection and catalysis of lipid phosphate phosphatase PgpB in the cell membrane. J. Biol. Chem. *291*, 18342–18352.

Tran, A.X., Dong, C., and Whitfield, C. (2010). Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in *Escherichia coli*. J. Biol. Chem. *285*, 33529–33539.

Vagin, A.A., and Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement. J. Appl. Crystallogr. *30*, 1022–1025.

Varki, A., Cummings, R.D., Aebi, M., Packer, N.H., Seeberger, P.H., Esko, J.D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J.J., Schnaar, R.L., Freeze, H.H., Marth, J.D., Bertozzi, C.R., Etzler, M.E., Frank, M., Vliegenthart, J.F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K.F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld. (2015). Symbol Nomenclature for Graphical Representations of Glycans. Glycobiology *25*, 1323–1324.

Vergis, J.M., Purdy, M.D., and Wiener, M.C. (2015). A high-throughput differential filtration assay to screen and select detergents for membrane proteins. Anal. Biochem. *407*, 1–11.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. Acta Crystallogr. Sect. D Biol. Crystallogr. *D67*, 293–302.

Wang, L., Liu, D., and Reeves, P.R. (1996). C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. J. Bacteriol. *178*, 2598–2604.

Wang, Z., Xiang, Q., Zhu, X., Dong, H., He, C., Wang, H., Zhang, Y., Wang, W., and

Dong, C. (2014). Structural and functional studies of conserved nucleotide-binding protein LptB in lipopolysaccharide transport. Biochem. Biophys. Res. Commun. *452*, 443–449.

Waterman, D.G., Winter, G., Gildea, R.J., Parkhurst, J.M., Brewster, A.S., Sauter, N.K., and Evans, G. (2016). Diffraction-geometry refinement in the DIALS framework. Acta Crystallogr. Sect. D Struct. Biol. *72*, 558–575.

Webb, N.A, Mulichak, A.M., Lam, J.S., Rocchetta, H.L., and Garavito, R.M. (2004). Crystal structure of a tetrameric GDP-D-mannose 4, 6-dehydratase from a bacterial GDP-D-rhamnose biosynthetic pathway. Structure *13*, 529–539.

Whitfield, C. (2006). Biosynthesis and Assembly of Capsular Polysaccharides in *Escherichia coli.* Annu. Rev. Biochem. *75*, 39–68.

WHO (2012). The global view of campylobacteriosis: Report of an expert consultation. (Last accessed: 19/06/17)

WHO (2016). WHO | *Campylobacter*. WHO. (Last accessed: 19/06/17)

WHO (2017a). Antibiotic resistance fact sheet. http://www.who.int/mediacentre/factsheets/antibiotic-resistance/en/ (Last accessed: 01/03/2018).

WHO (2017b). Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. http://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf?ua=1 (Last accessed: 01/03/2018)

Wilhelm, M.J., Sheffield, J.B., Sharifian Gh., M., Wu, Y., Spahr, C., Gonella, G., Xu, B., and Dai, H.L. (2015). Gram's Stain Does Not Cross the Bacterial Cytoplasmic Membrane. ACS Chem. Biol. *10*, 1711–1717.

Wilson, C.G.M., Magliery, T.J., and Regan, L. (2004). Detecting protein-protein interactions with GFP-fragment reassembly. Nat. Methods *1*, 255–262.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., and Wilson, K.S. (2011). Overview of the CCP4 suite and current developments research papers. Acta Crystallogr. Sect. D Biol. Crystallogr. *D67*, 235–242.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. *43*, 186–190.

Wong, A., Lange, D., Houle, S., Arbatsky, N.P., Valvano, M.A., Knirel, Y.A., Dozois, C.M., and Creuzenet, C. (2015). Role of capsular modified heptose in the virulence of *Campylobacter jejuni.* Mol. Microbiol. *96*, 1136–1158.

Woodward, L., and Naismith, J.H. (2016). Bacterial polysaccharide synthesis and export. Curr. Opin. Struct. Biol. *40*, 81–88.

Woodward, R., Yi, W., Li, L., Zhao, G., Eguchi, H., Sridhar, P.R., Guo, H., Song, J.K., Motari, E., Cai, L., Kelleher, P., Liu, X., Han, W., Zhang, W., Ding, Y., Li, M., and Wang, P.G. (2010). *In vitro* bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. Nat. Chem. Biol. *6*, 418–423.

Wu, S., Avila-Sakar, A., Kim, J., Booth, D.S., Greenberg, C.H., Rossi, A., Liao, M., Li, X., Alian, A., Griner, S.L., Juge, N., Yu, Y., Mergel, C.M., Chaparro-Riggers, J., Strop, P., Tampe, R., Edwards, R.H., Stroud, R.M., Craik, C.S., and Cheng, Y. (2012). Fabs enable single particle cryoEM studies of small proteins. Structure *20*, 582–592.

Wugeditsch, T., Paiment, A., Hocking, J., Drummelsmith, J., Forrester, C., and Whitfield, C. (2001). Phosphorylation of Wzc, a Tyrosine Autokinase, Is Essential for

Assembly of Group 1 Capsular Polysaccharides in *Escherichia coli*. J. Biol. Chem. *276*, 2361–2371.

Young, K.T., Davis, L.M., and DiRita, V.J. (2007). *Campylobacter jejuni*: molecular biology and pathogenesis. Nat. Rev. Microbiol. *5*, 665–679.

Yuki, N., Taki, T., Inagaki, F., Kasama, T., Takahashi, M., Saito, K., Handa, S., and Miyatake, T. (1993). A Bacterium Lipopolysaccharide that elicits Guillain-Barre Syndrome has a GM1 ganglioside-like structure. J. Exp. Med. *178*, 1771–1775.

Zaman, S.B., Hussain, M.A., Nye, R., Mehta, V., Mamun, K.T., and Hossain, N. (2017). A Review on Antibiotic Resistance: Alarm Bells are Ringing. Cureus. *9*, 1-9.

8 Appendix

8.1 Buffer compositions

8.1.1 Chapter 3

Wzy_{CV} solubilisation buffer	20 mM Tris pH 8.0, 500 mM NaCl, 100 mM arginine, 20 mM imidazole, 2.5 mM TCEP, 10 % (v/v) glycerol, 1.5 % (w/v) DM or 1% (w/v) DDM
Wzy_{CV} wash buffer 1	20 mM Tris pH 8.0, 500 mM NaCl, 100 mM arginine, 20 mM imidazole, 2.5 mM TCEP, 10 % (v/v) glycerol *[1]
Wzy_{CV} wash buffer 2	20 mM Tris pH 8.0, 500 mM NaCl, 100 mM arginine, 30 mM imidazole, 2.5 mM TCEP, 10 % (v/v) glycerol *[1]
Wzy_{CV} elution buffer	20 mM Tris pH 8.0, 500 mM NaCl, 100 mM arginine, 400 mM imidazole, 2.5 mM TCEP, 10 % (v/v) glycerol *[1]
Wzy_{CV} desalt buffer	20 mM Tris pH 8.0, 500 mM NaCl, 2.5 mM TCEP, 10 % (v/v) glycerol *[1]
Wzy_{CV} SEC buffer	20 mM Tris pH 8.0, 150 mM NaCl, 5 mM TCEP, 10 % (v/v) glycerol, 2 x CMC detergent

*[1] Detergent was added to each of these buffers at 3 x CMC, apart from ODM, which was used at 2.5 x CMC.

8.1.2 Chapter 4

Solubilsation buffer	PBS pH 7.2, 1% (w/v) DDM
Glutathione sepharose 4B wash buffer	PBS pH 7.2, 0.024 % (w/v) DDM
Glutathione sepharose 4B elution buffer	50 mM Tris pH 8.0, 200 mM NaCl, 10 mM glutathione, 0.024 % (w/v) DDM
Co-purification nickel wash buffer 1	PBS pH 8.0, 20 mM imidazole, 0.024 % (w/v) DDM
Co-purification nickel wash buffer 2	PBS pH 8.0, 30 mM imidazole, 0.024 % (w/v) DDM
Co-purification nickel elution buffer	PBS pH 8.0, 400 mM imidazole, 0.024 % (w/v) DDM
Co-purification streptactin elution buffer	50 mM Tris pH 8.0, 200 mM NaCl, 2.5 mM desthiobiotin, 0.024 % (w/v) DDM
Streptactin wash buffer	PBS pH 7.2, 1 mM EDTA, 0.024 % (w/v) DDM
Streptactin elution buffer	PBS pH 8.0, 2.5 mM desthiobiotin, 0.024 % (w/v) DDM
1 x NuPAGE Transfer buffer	1.25 mM Bicine, 1.25 mM Bis-tris (free base), 0.05 mM EDTA pH 7.2

8.1.3 Chapter 5

Biosprint wash buffer	PBS, 20 mM imidazole pH 7.2
Biosprint elution buffer	PBS, 250 mM imidazole pH 7.2
Soluble wash buffer 1	10 mM phosphate pH 7.5, 400 mM NaCl, 20 mM imidazole
Soluble wash buffer 2	10 mM phosphate pH 7.5, 400 mM NaCl, 30 mM imidazole
Soluble wash buffer 3	10 mM phosphate pH 7.5, 400 mM NaCl, 50 mM imidazole
Soluble elution buffer	10 mM phosphate pH 7.5, 400 mM NaCl, 400 mM imidazole
Soluble dialysis buffer	10 mM phosphate pH 7.5, 400 mM NaCl, 20 mM Tris
Soluble TEV protease cleavage buffer	50 mM Tris pH 7.5, 400 mM NaCl
Soluble SEC buffer	10 mM Tris pH 7.5, 150 mM NaCl

8.2 Differential filtration detergent screen detergents

Well	Detergent	Detergent Abbreviation	CMC (%)
A1	n-Decyl-β-D-Maltopyranoside	DM	0.087
A2	Water	Control	N/A
A3	n-Dodecyl-β-D-Maltopyranoside	DDM	0.0087
A4	n-Tetradecyl-β-D-Maltopyranoside	TetDM	0.00054
A5	n-Tridecyl-β-D-Maltopyranoside	TDM	0.0017
A6	n-Undecyl-β-D-Maltopyranoside	UDM	0.029
A7	n-Octyl-β-D-Maltopyranoside	ODM	0.89
A8	n-Nonyl-β-D-Maltopyranoside	NDM	0.28
A9	n-Nonyl-β-D-Thiomaltopyranoside	NTM	0.15
A10	n-Decyl-β-D-Thiomaltopyranoside	DTM	0.045
A11	n-Dodecyl-β-D-Thiomaltopyranoside	LTM	0.0026
A12	n-Undecyl-β-D-Thiomaltopyranoside	UTM	0.011
B1	n-Octyl-β-D-Glucopyranoside	OG	0.53
B2	n-Nonyl-β-D-Glucopyranoside	NG	0.2
B3	n-Dodecyl-β-D-Glucopyranoside	DDG	0.0066
B4	n-Decyl-β-D-Glucopyranoside	DG	0.07
B5	n-Nonyl-β-D-Thioglucopyranoside	NTG	0.093
B6	n-Heptyl-β-D-Thioglucopyranoside	HTG	8.9
B7	n-Octyl-β-D-Thioglucopyranoside	OTG	0.28
B8	n-Decyl-β-D-Thioglucopyranoside	DTG	0.3
B9	Octyl Tetraethylene Glycol Ether	C_8E_4	0.25
B10	Polyoxyethylene(8)dodecyl Ether (ANAPOE-$C_{12}E_8$)	$C_{12}E_8$	0.0048
B11	Polyoxyethylene(9)dodecyl Ether (ANAPOE-$C_{12}E_9$)	$C_{12}E_9$	0.003
B12	5-Cyclohexyl-1-Pentyl-β-D-Maltopyranoside (Cymal-5)	cym5	0.12
C1	6-Cyclohexyl-1-Hexyl-β-D-Maltopyranoside (Cymal-6)	cym6	0.028
C2	Lauryl Maltose Neopentyl Glycol	LMNG	0.001
C3	Octyl Glucose Neopentyl Glycol	OGNG	0.058
C4	Decyl Maltose Neopentyl Glycol	DMNG	0.0034
C5	Cymal-5 Neopentyl Glycol	Cym5-NG	0.0056
C6	Cymal-6 Neopentyl Glycol	Cym-6-NG	0.018
C7	n-Tetradecyl-N,N-Dimethylamine-N-Oxide	TDAO	0.0075
C8	n-Dodecyl-N,N-Dimethylamine-N-Oxide	LDAO	0.023
C9	n-Dodecylphosphocholine (FOS-CHOLINE-12)	fos12	0.047
C10	n-Tetradecylphosphocholine (FOS-CHOLINE-14)	fos14	0.0046
C11	Decanoyl-N-Hydroxyethylglucamide	Hega10	0.26
C12	Decanoyl-N-Methylglucamide	Mega10	0.21
D1	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfate	CHAPS	0.49
D2	DM / NG / LDAO	-	-
D3	1,2-Dimyristoyl-sn-Glycero-3-[Phospho-rac-(1-Glycerol)]	DMPG	-
D4	UDM / $C_{12}E_8$ / LDAO	-	-
D5	$C_{12}E_8$ / LMNG / DM	-	-
D6	DM / OGNG / LDAO	-	-
D7	DDM / NG / cym6	-	-
D8	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol	POPG	-
D9	3POPC:1POPE:1POPG	-	-
D10	*E. coli* polar lipids	polar	-
D11	DDM / DM	-	-
D12	DDM / UDM	-	-

E1	DM / UDM	-	-
E2	DDM / OG	-	-
E3	DM / OG	-	-
E4	UDM /OG	-	-
E5	DDM / NG	-	-
E6	DM / NG	-	-
E7	UDM /NG	-	-
E8	CHAPS / NG	-	-
E9	DDM / DG	-	-
E10	DM / DG	-	-
E11	DDM / $C_{12}E_8$	-	-
E12	DM / $C_{12}E_8$	-	-
F1	UDM / $C_{12}E_8$	-	-
F2	OG / $C_{12}E_8$	-	-
F3	NG / $C_{12}E_8$	-	-
F4	DG / $C_{12}E_8$	-	-
F5	DDM / LDAO	-	-
F6	DM / LDAO	-	-
F7	UDM /LDAO	-	-
F8	CHAPS / LDAO	-	-
F9	OG / LDAO	-	-
F10	NG / LDAO	-	-
F11	DG / LDAO	-	-
F12	LDAO / $C_{12}E_8$	-	-
G1	DDM / DMNG	-	-
G2	OG / DMNG	-	-
G3	NG / LMNG	-	-
G4	DDM / Cym5	-	-
G5	UDM / Cym5	-	-
G6	DDM / OGNG	-	-
G7	NG / OGNG	-	-
G8	DM / fos12	-	-
G9	LDAO / fos12	-	-
G10	$C_{12}E_9$ / fos12	-	-
G11	DDM / 3POPC:1POPE:1POPG	-	-
G12	DM / 3POPC:1POPE:1POPG	-	-
H1	LDAO / 3POPC:1POPE:1POPG	-	-
H2	OG / 3POPC:1POPE:1POPG	-	-
H3	NG / 3POPC:1POPE:1POPG	-	-
H4	$C_{12}E_8$/3POPC:1POPE:1POPG	-	-
H5	UDM /3POPC:1POPE:1POPG	-	-
H6	DDM / *E. coli* polar lipids	-	-
H7	DM / *E. coli* polar lipids	-	-
H8	LDAO / *E. coli* polar lipids	-	-
H9	OG / *E. coli* polar lipids	-	-
H10	NG / *E. coli* polar lipids	-	-
H11	$C_{12}E_8$/ *E. coli* polar lipids	-	-
H12	UDM / *E. coli* polar lipids	-	-

8.3 RUBIC Buffer Screen (Molecular Dimensions)

Well	Concentration	Units	Reagent	Concentration 2	Units 2	Buffer	pH
A1	100	%	Ultrapure water				
A2				0.119	M	Citrate	4
A3				0.119	M	Sodium acetate	4.5
A4				0.119	M	Citrate	5
A5				0.119	M	MES	6
A6				0.119	M	Potassium phosphate	6
A7				0.119	M	Citrate	6
A8				0.119	M	Bis-Tris	6.5
A9				0.119	M	MES	6.5
A10				0.119	M	Sodium phosphate	7
A11				0.119	M	Potassium phosphate	7
A12				0.119	M	HEPES	7
B1				0.119	M	MOPS	7
B2				0.119	M	Ammonium acetate	7.3
B3				0.119	M	Tris-HCl	7.5
B4				0.119	M	Sodium phosphate	7.5
B5				0.119	M	Imidazole	7.5
B6				0.119	M	HEPES	8
B7				0.119	M	Tris-HCl	8
B8				0.119	M	Tricine	8
B9				0.119	M	BICINE	8
B10				0.119	M	BICINE	8.5
B11				0.119	M	Tris-HCl	8.5
B12				0.119	M	CHES	9
C1	0.298	M	Sodium chloride				
C2	0.298	M	Sodium chloride	0.119	M	Citrate	4
C3	0.298	M	Sodium chloride	0.119	M	Sodium acetate	4.5
C4	0.298	M	Sodium chloride	0.119	M	Citrate	5
C5	0.298	M	Sodium chloride	0.119	M	MES	6
C6	0.298	M	Sodium chloride	0.119	M	Potassium phosphate	6
C7	0.298	M	Sodium chloride	0.119	M	Citrate	6
C8	0.298	M	Sodium chloride	0.119	M	Bis-Tris	6.5
C9	0.298	M	Sodium chloride	0.119	M	MES	6.5
C10	0.298	M	Sodium chloride	0.119	M	Sodium phosphate	7
C11	0.298	M	Sodium chloride	0.119	M	Potassium phosphate	7
C12	0.298	M	Sodium chloride	0.119	M	CHES	7
D1	0.298	M	Sodium chloride	0.119	M	MOPS	7
D2	0.298	M	Sodium chloride	0.119	M	Ammonium acetate	7.3
D3	0.298	M	Sodium chloride	0.119	M	Tris-HCl	7.5
D4	0.298	M	Sodium chloride	0.119	M	Sodium phosphate	7.5
D5	0.298	M	Sodium chloride	0.119	M	Imidazole	8
D6	0.298	M	Sodium chloride	0.119	M	HEPES	8
D7	0.298	M	Sodium chloride	0.119	M	Tris-HCl	8
D8	0.298	M	Sodium chloride	0.119	M	Tricine	8
D9	0.298	M	Sodium chloride	0.119	M	BICINE	8
D10	0.298	M	Sodium chloride	0.119	M	BICINE	8.5
D11	0.298	M	Sodium chloride	0.119	M	Tris-HCl	8.5

D12	0.298	M	Sodium chloride	0.119	M	CHES	9
E1				0.119	M	SPG	4
E2				0.119	M	SPG	4.5
E3				0.119	M	SPG	5
E4				0.119	M	SPG	5.5
E5				0.119	M	SPG	6
E6				0.119	M	SPG	6.5
E7				0.119	M	SPG	7
E8				0.119	M	SPG	7.5
E9				0.119	M	SPG	8
E10				0.119	M	SPG	8.5
E11				0.119	M	SPG	9
E12				0.119	M	SPG	10
F1				0.024	M	HEPES	7.5
F2				0.06	M	HEPES	7.5
F3				0.149	M	HEPES	7.5
F4				0.298	M	HEPES	7.5
F5				0.024	M	Sodium phosphate	7.5
F6				0.06	M	Sodium phosphate	7.5
F7				0.149	M	Sodium phosphate	7.5
F8				0.298	M	Sodium phosphate	7.5
F9				0.024	M	Tris-HCl	8
F10				0.06	M	Tris-HCl	8
F11				0.149	M	Tris-HCl	8
F12				0.298	M	Tris-HCl	8
G1	0.06	M	Sodium chloride	0.06	M	HEPES	7.5
G2	0.149	M	Sodium chloride	0.06	M	HEPES	7.5
G3	0.298	M	Sodium chloride	0.06	M	HEPES	7.5
G4	0.595	M	Sodium chloride	0.06	M	HEPES	7.5
G5	0.893	M	Sodium chloride	0.06	M	HEPES	7.5
G6	1.19	M	Sodium chloride	0.06	M	HEPES	7.5
G7	0.06	M	Sodium chloride	0.06	M	Tris-HCl	8
G8	0.149	M	Sodium chloride	0.06	M	Tris-HCl	8
G9	0.298	M	Sodium chloride	0.06	M	Tris-HCl	8
G10	0.595	M	Sodium chloride	0.06	M	Tris-HCl	8
G11	0.893	M	Sodium chloride	0.06	M	Tris-HCl	8
G12	1.19	M	Sodium chloride	0.06	M	Tris-HCl	8
H1				0.06	M	MES/Bis-Tris	6
H2				0.06	M	MES/Imidazole	6.5
H3				0.06	M	Bis-Tris/PIPES	6.5
H4				0.06	M	MOPS/Bis-Tris propane	7
H5				0.06	M	Phosphate/Citrate	7.5
H6				0.06	M	MOPS/Sodium HEPES	7.5
H7				0.06	M	BICINE/Tris	8.5
H8	0.119	M	Sodium chloride	0.06	M	Imidazole	7.5
H9	0.119	M	Sodium chloride	0.149	M	Imidazole	7.5
H10	0.119	M	Sodium chloride	0.298	M	Imidazole	7.5
H11	0.119	M	Sodium chloride	0.417	M	Imidazole	7.5
H12	0.119	M	Sodium chloride	0.595	M	Imidazole	7.5

8.4 In-house stochastic crystallisation screens

These screens were produced using the Custom Screens service provided by Molecular Dimensions

8.4.1 St Andrews 1

Well	Precipitant	Buffer	Additive 1	Additive 2
A1	13.14 % (w/v) PEG 10,000	0.1 M MOPS pH 6.5	0.26 M Lithium sulfate	
A2	23.19 % (w/v) PEG 6,000	0.1 M Bicine pH 9		2.05 % (v/v) PEG DME 250
A3	16.52 % (w/v) PEG 8,000	0.1 M Sodium cacodylate pH 6.5		
A4	0.95 M Sodium tartrate	0.1 M Sodium cacodylate pH 7	0.13 M Potassiun chloride	
A5	2.13 M Sodium malate	0.1 M Bicine pH 8.5	0.1 M Ammonium acetate	
A6	27.31 % (w/v) PEG 8,000	0.1 M MES pH 6.5	0.11 M Sodium acetate	
A7	43.53 % (v/v) MPD	0.1 M Sodium cacodylate pH 7		
A8	16.09 & (w/v) PEG 4,000	0.1 M CHES pH 9	0.05 M Sodium citrate	2.51 % (v/v) DMSO
A9	20.17 % (w/v) PEG 6,000	0.1 M Sodium citrate pH 5.5	0.11 M Ammonium tartrate	
A10	20.82 % (w/v) PEG 3,350	0.1 M MOPS pH 7		0.13 % (v/v) CHAPS
A11	0.89 M Sodium tartrate	0.1 M Sodium acetate pH 5	0.09 M Potassium nitrate	
A12	2.47 M Ammonium sulfate	0.1 M Bicine pH 9.5	0.21 M Magnesium chloride	
B1	24.77 % (w/v) PEG MME 2,000	0.1 M MOPS pH 7	0.07 M Ammonium tartrate	
B2	33.49 % (w/v) PEG 1,500	0.1 M Bicine pH 9	0.1 M Magnesium chloride	
B3	24.19 % (w/v) PEG 6,000	0.1 M Sodium citrate pH 5		
B4	26 % (v/v) PEG MME 550	0.1 M Bis-Tris pH 6.5		
B5	24.16 % (w/v) PEG MME 5,000	0.1 M MOPS pH 6.5	0.09 M Zinc sulfate	
B6	26.05 % (w/v) PEG MME 550	0.1 M Sodium cacodylate pH 6.5	0.12 M Sodium-potassium-tartrate	0.18 % (w/v) LDAO
B7	1.88 M Sodium-potassium phosphate	0.1 M CHES pH 9.5	0.11 M Ammonium citrate	
B8	1.5 M Sodium acetate		0.3 M Magnesium sulfate	0.15 % (w/v) OG
B9	1.18 M Sodium tartrate			
B10	23.75 % (w/v) PEG 8,000	0.1 M CHES pH 9	0.2 M Ammonium phosphate	
B11	10.64 % (w/v) PEG 10,000		0.05 M Magnesium acetate	
B12	3.24 M Sodium chloride	0.1 M Bicine pH 9.5	0.19 M Sodium chloride	
C1	26.05 % (v/v) PEG MME 550	0.1 M MES pH 6	0.14 M Zinc sulfate	
C2	2.18 M Sodium-potassium-phosphate		0.19 M Potassium nitrate	8.28 mM EDTA
C3	30.62 % (v/v) PEG 400	0.1 M Tris-HCl pH 8	0.27 M Ammonium phosphate	2.17 % (v/v) PEG MME 350
C4	1.24 M Sodium acetate	0.1 M Bicine pH 9	0.14 M Ammonium citrate	
C5	2.45 M Ammonium sulfate	0.1 Tris-HCl pH 8	0.29 M Magnesium acetate	
C6	21.47 % (w/v) PEG 8,000	0.1 M MES pH 6		1.74 % (v/v) Glycerol
C7	16.01 % (w/v) PEG 4,000	0.1 M CHES pH 9	0.03 M Zinc sulfate	0.7 % (v/v) PEG DME 250
C8	10.06 % (v/v) Isopropanol	0.1 M Sodium acetate pH 5.5		
C9	26.93 % (w/v) PEG 6,000			
C10	20.46 % (v/v) Isopropanol	0.1 M MES pH 6	0.26 M Lithium sulfate	
C11	31.64 % (v/v) MPD	0.1 M Bis-Tris pH 6		
C12	1.48 M Sodium acetate	0.1 M Bis-Tris pH 6	0.08 M Calcium acetate	
D1	2.71 M Sodium-potassium-phosphate	0.1 M Tris-HCl pH 8		0.18 % (v/v) OG
D2	27.47 % (w/v) PEG MME 5,000	0.1 M Sodium cacodylate pH 7	0.21 M Sodium bromide	0.1 % (w/v) LDAO
D3	12.88 % (v/v) Isopropanol	0.1 M Bicine pH 9	0.09 M Calcium chloride	
D4	38.31 % (v/v) PEG 400	0.1 M MOPS pH 7	0.3 M Magnesium acetate	
D5	17.2 % (w/v) PEG 8,000	0.1 M MOPS pH 6.5	0.26 M Ammonium sulfate	
D6	2.36 M Sodium formate	0.1 M Sodium citrate pH 5.5	0.16 M Ammonium sulfate	
D7	18.76 % (w/v) PEG 4,000	0.1 M Tris-HCl pH 7.5	0.08 M Ammonium citrate	
D8	2.28 M Sodium formate	0.1 M Sodium cacodylate pH 7	0.2 M Sodium chloride	
D9	1.73 M Ammonium phosphate	0.1 M Sodium citrate pH 5.5	0.12 M Potassium chloride	
D10	29.61 % (w/v) PEG 4,000	0.1 M Sodium acetate pH 5.5	0.24 M Magnesium sulfate	
D11	8.85 % (v/v) Ethanol	0.1 M MES pH 6	0.03 M Sodium acetate	

D12	1.72 M Sodium acetate	0.1 M Tris-HCl pH 8.5	0.09 M Sodium acetate	
E1	1.94 M Sodium formate	0.1 M Sodium cacodylate pH 6.5	0.22 M Sodium-potassium phosphate	
E2	2.37 M Sodium formate	0.1 M Sodium citrate pH 5	0.08 M Ammonium sulfate	
E3	26.75 % (w/v) PEG 4,000	0.1 M Tris-HCl pH 7.5	0.14 M Sodium bromide	
E4	1.54 M Magnesium sulfate	0.1 M Tris-HCl pH 8.5	0.03 M Zinc acetate	0.06 % (w/v) OG
E5	33.55 % (v/v) PEG MME 550	0.1 M Bis-Tris pH 6		
E6	26.68 % (w/v) PEG 1,500	0.1 M Bicine pH 8.5	0.06 M Ammonium tartrate	
E7	1.94 M Sodum malate	0.1 M MES pH 6	0.14 M Ammonium phosphate	
E8	30.47 % (w/v) PEG 1,500	0.1 M MOPS pH 7	0.12 M Ammonium citrate	1.31 % (v/v) Butanediol
E9	19.31 % (w/v) PEG 1,500	0.1 M Bicine pH 9.5	0.07 M Magnesium acetate	
E10	2.31 M Ammonium sulfate	0.1 M MES pH 6		
E11	20.78 % (w/v) PEG MME 5,000	0.1 M Sodium acetate pH 5	0.25 M Magnesium acetate	
E12	28.52 % (w/v) PEG 3,350	0.1 M Sodium citrate pH 5	0.09 M Sodium sulfate	
F1	15.5 % (w/v) PEG 8,000		0.23 M Magnesium sulfate	0.03 % (w/v) LDAO
F2	2.73 M Sodium acetate	0.1 M Bicine pH 9	0.26 M Magnesium sulfate	
F3	12.55 % (v/v) Isopropanol	0.1 M Sodium acetate pH 5		0.07 % (w/v) CHAPS
F4	36.95 % (v/v) PEG 400	0.1 M HEPES pH 7.5	0.04 M Calcium chloride	
F5	1.77 M Sodium malonate	0.1 M CHES pH 9		
F6	2.59 M Sodium chloride	0.1 M Tris-HCl pH 8	0.03 M Ammonium citrate	
F7	19.9 % (w/v) PEG 10,000	0.1 M Tris-HCl pH 8	0.17 M Lithium chloride	
F8	5.83 % (v/v) Isopropanol	0.1 M MOPS pH 7		
F9	1.43 M Ammonium sulfate	0.1 M Sodium acetate pH 5.5	0.13 M Potassium nitrate	3.58 % (v/v) Glycerol
F10	1.41 M Ammonium sulfate	0.1 M Bis-Tris pH 6	0.13 M Ammonium acetate	0.02 % (w/v) OG
F11	11.62 % (v/v) Isopropanol	0.1 M Sodium citrate pH 5.5		2.84 % (v/v) PEG DME 250
F12	1.11 M Sodium citrate	0.1 M CHES pH 9	0.05 M Sodium citrate	
G1	30.77 % (v/v) MPD	0.1 M MES pH 6	0.03 M Sodium acetate	1.98 % (v/v) DMSO
G2	1.85 M Sodium acetate	0.1 M CHES pH 9.5	0.09 M Ammonium sulfate	
G3	1.09 M Sodium citrate	0.1 M MES pH 6	0.09 M Ammonium tartrate	
G4	13.48 % (v/v) Isopropanol	0.1 M MOPS pH 7	0.25 M Magnesium acetate	
G5	28.8 % (w/v) PEG 1,500	0.1 M CHES pH 9.5	0.29 M Sodium bromide	0.1 % (w/v) OG
G6	1.1 M Sodium citrate	0.1 M Tris-HCl pH 8	0.05 M Sodium citrate	0.14 % (w/v) CHAPS
G7	18.7 % (v/v) Isopropanol	0.1 M MOPS pH 7		
G8	15.65 % (w/v/) PEG MME 5,000	0.1 M CHES pH 9.5	0.09 M Sodium acetate	
G9	2.29 M Sodium acetate	0.1 M Bicine pH 9	0.06 M Calcium acetate	
G10	22.45 % (w/v) PEG 8,000	0.1M HEPES pH 8		
G11	23.04 % (w/v) PEG 6,000	0.1 M Sodium acetate pH 5	0.09 M Potassium nitrate	2.32 % (v/v) BME
G12	16.23 % (w/v) PEG 8,000	0.1 M Sodium acetate pH 5.5	0.12 M Magnesium sulfate	
H1	1.07 M Sodium tartrate	0.1 M Tris-HCl pH 8.5	0.05 M Sodium chloride	
H2	55.52 % (v/v) PEG 400	0.1 M HEPES pH 7.5	0.25 M Magnesium chloride	
H3	8.86 % (v/v) Isopropanol	0.1 M HEPES pH 8		
H4	2.46 M Sodium acetate	0.1 M CHES pH 9.5	0.09 M Ammonium acetate	
H5	20.49 % (w/v) PEG MME 2,000	0.1 M MOPS pH 7	0.13 M Ammonium acetate	
H6	1.12 M Magnesium sulfate			0.56 % (v/v) Ethylene glycol
H7	2.5 M Sodium-potassium-phosphate	0.1 M Sodium citrate pH 5.5	0.23 M Potassium chloride	
H8	36.99 % (v/v) PEG MME 550	0.1 M Sodium acetate pH 4.5	0.29 M Sodium chloride	
H9	30.96 % (w/v) PEG 4,000	0.1 M Sodium acetate pH 4.5	0.11 M Lithium sulfate	
H10	41.39 % (v/v) MPD		0.08 M Calcium chloride	
H11	40.21 % (w/v) PEG 1,500	0.1 M Bicine pH 8.5	0.07 M Sodium-potassium tartrate	
H12	14.3 % (v/v) Isopropanol	0.1 M Sodium citrate pH 5.5	0.13 M Potassium nitrate	0.49 % (v/v) PEG DME 250

8.4.2 St Andrews 2

Well	Precipitant	Buffer	Additve 1	Additive 2
A1	2.07 M Sodium malate	0.1 M HEPES pH 7.5		
A2	18.41 % (w/v) PEG 6,000	0.1 M MOPS pH 6.5	0.08 M Calcium acetate	3.19 % (v/v) PEG400
A3	1.61 M Sodium malonate	0.1 M Tris-HCl pH 8	0.06 M Ammonium citrate	
A4	41.63 % (v/v) MPD	0.1 M Tris-HCl pH 7.5	0.04 M Sodium acetate	
A5	26.8 % (w/v) PEG 1,500	0.1 M Sodium cacodylate pH 7	0.06 M Zinc chloride	0.14 % (w/v) LDAO
A6	1.52 M Sodium malate	0.1 M Bis-Tris pH 6.5	0.29 M Sodium-potassium phosphate	0.07 % (w/v) OG
A7	1.56 M Magnesium sulfate	0.1 M Tris-HCl pH 8.5	0.12 M Sodium sulfate	
A8	21.94 % (w/v) PEG 6,000	0.1 M Tris-HCl pH 8	0.06 M Sodium chloride	
A9	29.83 % (w/v) PEG 4,000	0.1 M Bicine pH 8.5	0.05 M Sodium citrate	
A10	14.96 % (v/v) Ethanol	0.1 M MES pH 6	0.12 M Magnesium formate	
A11	21.37 % (w/v) PEG 4,000	0.1 M Bicine pH 8.5	0.09 M Potassium chloride	
A12	0.86 M Sodium citrate	0.1 M Bicine pH 9		
B1	54.34 % (v/v) PEG 400	0.1 M Sodium cacodylate pH 7	0.09 M Lithium sulfate	
B2	10.38 % (v/v) Isopropanol		0.23 M Sodium bromide	
B3	1.64 M Sodium malate			
B4	2.59 M Sodium acetate	0.1 M Bis-Tris pH 6		1.62 % (v/v) PEG MME 350
B5	39.74 % (w/v) PEG MME 2,000	0.1 M CHES pH 9	0.09 M Calcium chloride	
B6	13.86 % (v/v) Isopropanol	0.1 M Sodium acetate pH 5.5	0.13 M Potassium chloride	
B7	17.06 % (w/v) PEG 3,350	0.1 M Sodium acetate pH 5.5	0.23 M Potassium chloride	
B8	1.58 M Magnesium sulfate	0.1 M HEPES pH 8	0.08 M Magnesium chloride	
B9	25.33 % (w/v) PEG MME 2,000	0.1 M Tris-HCl pH 8.5	0.23 M Sodium chloride	7.24 mM EDTA
B10	2.54 M Sodium formate		0.05 M Sodium citrate	
B11	33.78 % (v/v) PEG MME 550	0.1 M MOPS pH 6.5	0.19 M Magnesium sulfate	
B12	31.99 % (v/v) PEG 400	0.1 M Bicine pH 9.5	0.28 M Sodium sulfate	1.28 % (v/v) DMSO
C1	1.83 M Sodium acetate	0.1 M MOPS pH 7	0.07 M Calcium acetate	
C2	2.67 M Sodium malonate	0.1 M MES pH 6.5	0.1 M Sodium-potassium tartrate	
C3	18.95 M PEG MME 5,000	0.1 M Sodium cacodylate pH 6.5	0.27 M Sodium bromide	
C4	0.95 M Sodium citrate	0.1 M HEPES pH 8	0.18 M Sodium bromide	
C5	20.8 % (w/v) PEG 8,000	0.1 M Bicine pH 9	0.08 M Ammonium acetate	
C6	20.82 % (w/v) PEG 3,350	0.1 M Sodium cacodylate pH 7	0.1 M Sodium acetate	
C7	2.79 M Sodium formate			
C8	36.59 % (v/v) MPD	0.1 M Bis-Tris pH 6	0.26 M Ammonium sulfate	1.74 % (v/v) Ethylene glycol
C9	34.58 % (v/v) PEG 400	0.1 M Tris-HCl pH 8	0.13 M Sodium-potassium tartrate	
C10	2.7 M Sodium formate	0.1 M Tris-HCl pH 7.5	0.28 M Ammonium sulfate	2.65 % (v/v) Dioxane
C11	8.67 % (v/v) Isopropanol	0.1 M Sodium cacodylate pH 7		2.92 % (v/v) Hexanediol
C12	1.8 M Sodium acetate	0.1 M MOPS pH 6.5	0.07 M Lithium sulfate	
D1	2.31 M Sodium malonate	0.1 M Sodium acetate pH 5.5		7.83 % (v/v) BME
D2	23.59 % (w/v) PEG 4,000	0.1 M CHES pH 9		
D3	30.06 % (w/v) PEG 3,350	0.1 M Sodium cacodylate pH 7		
D4	37.95 % (w/v) PEG MME 550	0.1 M Sodium citrate pH 5	0.1 M Sodium citrate	2.3 % (v/v) PEG 400
D5	26.55 % (w/v) PEG 4,000	0.1 M Sodium citrate pH 5	0.09 M Ammonium sulfate	0.08 % (w/v) OG
D6	17.9 % (w/v) PEG 10,000	0.1 M CHES pH 9.5	0.07 M Sodium-potassium tartrate	3.18 % (v/v) PEG DME 250
D7	18.25 % (w/v) PEG 10,000	0.1 M Sodium acetate pH 5	0.21 M Sodium bromide	0.2 % (w/v) LDAO
D8	30.86 % (v/v) PEG MME 550	0.1 M Sodium cacodylate pH 7	0.25 M Ammonium sulfate	
D9	2.25 M Sodium malate	0.1 M Sodum acetate pH 4.5	0.08 M Potassium nitrate	
D10	30.4 % (w/v) PEG MME 2,000		0.13 M Ammonium phosphate	
D11	19.94 % (w/v) PEG 1,500	0.1 M Tris-HCl pH 8	0.06 M Sodium-potassium phosphate	

D12	1.01 M Magnesium sulfate	0.1 M CHES pH 9	0.1 M Sodium bromide	
E1	52.88 % (v/v) MPD	0.1 M Bis-Tris pH 6	0.11 M Sodium acetate	
E2	12.03 % (v/v) Ethanol	0.1 M Bis-Tris pH 6	0.13 M Sodium chloride	0.12 % (w/v) CHAPS
E3	0.61 M Sodium citrate	0.1 M Sodium citrate pH 4.5	0.14 M Sodium-potassium phosphate	2.37 % (v/v) PEG MME 350
E4	1.92 M Ammonium phosphate	0.1 M MES pH 6.5	0.04 M Ammonium acetate	
E5	18.76 % (w/v) PEG 10,000	0.1 M Tris-HCl pH 8.5	0.06 M Sodium-potassium phosphate	0.03 % (w/v) OG
E6	22.06 % (w/v) PEG MME 5,000	0.1 M MES pH 6	0.28 M Ammonium phosphate	
E7	15.14 % (w/v) PEG 6,000			
E8	2.5 M Sodium-potassium phosphate	0.1 M Bis-Tris pH 6		
E9	27.32 % (v/v) PEG MME 550	0.1 M Bicine pH 9	0.14 M Sodium-potassium tartrate	
E10	29.87 % (v/v) PEG400	0.1 M Tris-HCl pH 8.5	0.08 M Zinc acetate	
E11	2.61 M Sodium-potassium phosphate			
E12	23.29 % (w/v) PEG MME 2,000	0.1 M Sodium citrate pH 4.5	0.17 M Potassium nitrate	
F1	2.19 M Sodium-potassium phosphate	0.1 M Sodium citrate pH 5		
F2	18.45 % (w/v) PEG 1,500	0.1 M Bis-Tris pH 6	0.23 M Lithium chloride	
F3	31.89 % (w/v) PEG 1,500	0.1 M Tris-HCl pH 8.5		1.89 % (v/v) Hexanediol
F4	17.92 % (w/v) PEG 4,000	0.1 M Sodium cacodylate pH 7	0.23 M Lithium chloride	3.63 % (v/v) Butanediol
F5	1.88 M Sodium acetate	0.1 M Sodium citrate pH 5		
F6	30.85 % (w/v) PEG 1,500	0.1 M MOPS pH 6.5	0.15 M Ammonium phosphate	2.03 % (v/v) PEG DME 250
F7	17.71 % (w/v) PEG 3,350		0.15 M Sodium chloride	
F8	2.16 M Sodium chloride		0.15 M Magnesium chloride	
F9	20.22 % (w/v) PEG 10,000	0.1 M Bicine pH 8.5		2.84 % (v/v) Butanediol
F10	6.34 % (v/v) Isopropanol	0.1 M Bis-Tris pH 6	0.21 M Sodium bromide	
F11	1.44 M Ammonium sulfate		0.05 M Ammonium citrate	
F12	2.32 M Ammonium sulfate	0.1 M CHES pH 9.5		5.97 % (v/v) BME
G1	9.74 % PEG 10,000	0.1 M Sodium acetate pH 5.5	0.15 M Ammonium tartrate	1.74 % (v/v) Hexanediol
G2	0.77 M Sodium citrate	0.1 M Sodium citrate pH 5.5	0.11 M Sodium citrate	1.4 % (v/v) Glycerol
G3	1.86 M Sodium formate	0.1 M Sodium acetate pH 5.5	0.21 M Sodium-potassium phosphate	
G4	1.31 M Sodium acetate	0.1 M Bicine pH 8.5		
G5	20.42 % (w/v) PEG 10,000		0.06 M Sodium chloride	
G6	16.37 % (v/v) Isopropanol	0.1 M Sodium citrate pH 5	0.14 M Ammonium phosphate	
G7	13.49 % (w/v) PEG 8,000	0.1 M Sodium acetate pH 5.5	0.14 M Lithium sulfate	
G8	10.27 % (v/v) Isopropanol	0.1 M Sodium acetate pH 5.5	0.07 M Magnesium formate	
G9	2.32 M Sodium formate	0.1 CHES pH 9.5		
G10	14.01 % (w/v) PEG 4,000	0.1 M Tris-HCl pH 7.5	0.16 M Sodium sulfate	
G11	1.87 M Ammonium phosphate	0.1 M Sodium acetate pH 4.5	0.11 M Ammonium citrate	
G12	55.67 % (v/v) PEG 400	0.1 M Sodium acetate pH 5	0.03 M Ammonium acetate	
H1	9.44 % (v/v) Isopropanol	0.1 M Sodium citrate pH 5	0.12 M Magnesium chloride	
H2	1.05 M Sodium tartrate	0.1 M Bis-Tris pH 6		
H3	2.41 M Sodium malonate	0.1 M Bicine pH 8.5	0.1 M Sodium-potassium-tartrate	
H4	25 % (w/v) PEG 6,000	0.1 M Bicine pH 8.5	0.09 M Sodium chloride	
H5	24.15 % (w/v) PEG MME 2,000	0.1 M Sodium acetate pH 5.5		
H6	2.54 M Sodium acetate	0.1 M Bicine pH 8.5		
H7	29.45 % (w/v) PEG MME 2,000	0.1 M Sodium acetate pH 4.5	0.22 M Ammonium phosphate	0.16 % (w/v) CHAPS
H8	53.79 % (v/v) PEG 400	0.1 M Sodium citrate pH 4.5	0.15 M Lithium chloride	
H9	9.81 % (v/v) Isopropanol	0.1 M Sodium citrate pH 5.5		
H10	2.32 M Sodium formate	0.1 M HEPES pH 7.5	0.03 M Calcium chloride	
H11	15.8 % (w/v) PEG 4,000		0.07 M Zinc chloride	
H12	1.09 M Magnesium sulfate	0.1 M MOPS pH 7	0.23 M Lithium sulfate	

8.4.3 St Andrews 3

Well	Precipitant	Buffer	Additive 1	Additive 2
A1	1.44 M Sodium Malate	0.1 M Sodium cacodylate pH 7	0.07 M Sodium citrate	
A2	2.48 M Ammonium sulfate		0.09 M Ammonium citrate	
A3	12.24 % (v/v) Ethanol	0.1 M Sodium cacodylate pH 7		
A4	0.63 M Sodium citrate	0.1 M Sodium acetate pH 4.5	0.13 M Ammonium acetate	0.09 % (w/v) LDAO
A5	2.15 M Ammonium sulfate	0.1 M Tris-HCl pH 7.5	0.03 M Ammonium acetate	
A6	29.72 % (w/v) PEG 1,500	0.1 M Tris-HCl pH 8.5		
A7	49.92 % (v/v) PEG 400	0.1 M Sodium acetate pH 5	0.14 M Magnesium sulfate	3.26 % (v/v) BME
A8	36.54 % (w/v) PEG MME 2,000	0.1 M Bicine pH 9.5	0.05 M Ammonium tartrate	0.08 % (w/v) LDAO
A9	35.15 % (v/v) PEG MME 550	0.1 M Sodium citrate pH 5.5		
A10	33.32 % (w/v) PEG 1,500	0.1 M CHES pH 9.5	0.09 M Lithium chloride	
A11	21.82 % (w/v) PEG 6,000	0.1 M Bicine pH 9.5	0.08 M Zinc acetate	
A12	23.35 % (w/v) PEG 4,000	0.1 M Bis-Tris pH 6.5	0.1 M Ammonium acetate	
B1	26.2 % (w/v) PEG 1,500	0.1 M MES pH 6	0.14 M Sodium-potassum phosphate	
B2	1.48 M Sodium malate	0.1 M HEPES pH 7.5		
B3	1.52 M Ammonium sulfate	0.1 M Bis-Tris pH 6		
B4	21.2 % (w/v) PEG MME 2,000	0.1 M MES pH 6		1.6 % (v/v) PEG 400
B5	3.48 M Sodium chloride	0.1 M Tris-HCl pH 8	0.29 M Lithium chloride	
B6	28.16 % (v/v) MPD			
B7	1.52 M Magnesium sulfate	0.1 M HEPES pH 8	0.07 M Magnesium chloride	1.74 % (v/v) Glycerol
B8	20.15 % (w/v) PEG 8,000	0.1 M MES pH 6	0.27 M Magnesium acetate	
B9	18.2 % (w/v) PEG 1,500		0.11 M Ammonium citrate	
B10	21.94 % (w/v) PEG 6,000	0.1 M Tris-HCl pH 7.5		
B11	15.89 % (w/v) PEG 8,000	0.1 M Sodium cacodylate pH 7	0.27 M Lithium chloride	
B12	1.96 M Sodium acetate	0.1 M Sodium acetate pH 4.5	0.07 M Sodium bromide	0.01 % (w/v) OG
C1	15.73 % (w/v) PEG 3,350	0.1 M Bicine pH 8.5	0.12 M Calcium chloride	2.16 % (v/v) Hexanediol
C2	2.5 M Ammonium phosphate	0.1 M Sodium citrate pH 5.5	0.09 M Sodium acetate	
C3	19.45 % (w/v) PEG 1,500	0.1 M Bicine pH 9	0.14 M Ammonium sulfate	
C4	1.74 M Ammonium sulfate	0.1 M Sodium cacodylate pH 6.5	0.09 M Magnesium acetate	0.01 % (w/v) LDAO
C5	1.39 M Magnesium sulfate	0.1 M Bis-Tris pH 6	0.1 M Ammonium acetate	
C6	17.09 % (w/v) PEG 4,000	0.1 M Sodium cacodylate pH 6.5	0.07 M Ammonium citrate	
C7	1.75 M Sodium acetate	0.1 M Bicine pH 9	0.29 M Lithium chloride	0.66 % (v/v) PEG400
C8	1.69 M Sodium malonate	0.1 M MOPS pH 7	0.05 M Sodium acetate	
C9	1.7 M Ammonium sulfate	0.1 M Tris-HCl pH 8.5	0.24 M Sodium chloride	
C10	2.09 M Sodium chloride	0.1 M Sodium cacodylate pH 6.5	0.03 M Ammonium tartrate	
C11	2.16 M Ammonium phosphate		0.26 M Ammonium phosphate	
C12	2.72 M Sodium malonate	0.1 M Sodium citrate pH 5	0.06 M Sodium citrate	
D1	51.3 % (v/v) MPD	0.1 M CHES pH 9.5		5.16 % (v/v) BME
D2	21.97 % (w/v) PEG MME 5,000	0.1 M Sodium cacodylate pH 6.5	0.07 M Zinc acetate	
D3	18.81 % (w/v) PEG 6,000	0.1 M MES pH 6	0.29 M Sodium bromide	
D4	1.1 M Sodium citrate	0.1 M HEPES pH 7.5	0.15 M Sodium-potassium phosphate	
D5	13.96 % (v/v) Isopropanol	0.1 M Sodium acetate pH 4.5	0.14 M Lithium sulfate	3.51 % (v/v) DMSO
D6	24.93 % (w/v) PEG 3,350		0.12 M Zinc chloride	
D7	32.45 % (v/v) PEG400	0.1 M Sodium citrate pH 4.5	0.05 M Potassium thiocyanate	
D8	1 M Sodium tartrate	0.1 M Tris-HCl pH 8		
D9	0.97 M Sodium citrate	0.1 MOPS pH 7		
D10	16.97 % (w/v) PEG 8,000	0.1 M CHES pH 9.5	0.07 M Ammonium citrate	
D11	15.61 % (v/v) Ethanol	0.1 M Sodium acetate pH 5.5		

D12	17.63 % (w/v) PEG 6,000	0.1 M HEPES pH 7.5	0.21 M Ammonium phosphate	
E1	0.73 M Sodium citrate	0.1 M Bicine pH 8.5	0.1 M Ammonium citrate	
E2	2.05 M Sodium acetate	0.1 M Sodium cacodylate pH 7		
E3	18.06 % (w/v) PEG 3,350	0.1 M CHES pH 9	0.21 M Ammonium phosphate	
E4	2.31 M Sodium chloride	0.1 M Tris-HCl pH 8.5		
E5	2.05 M Sodium-potassium phosphate	0.1 M CHES pH 9.5		
E6	1.65 M Sodium acetate	0.1 MOPS pH 7	0.08 M Sodium acetate	
E7	2.17 M Sodium malate	0.1 M HEPES pH 8	0.13 M Sodium acetate	
E8	21.89 % (w/v) PEG MME 550		0.04 M Ammonium acetate	3.65 % (v/v) Ethylene glycol
E9	1.86 M Ammonium phosphate	0.1 M Bicine pH 8.5	0.12 M Sodium chloride	
E10	5.72 % (v/v) Isopropanol	0.1 M Tris-HCl pH 7.5		
E11	1.68 M Ammonium sulfate	0.1 M Bis-Tris pH 6	0.07 M Ammonium acetate	2.33 % (v/v) PEG 400
E12	1.95 M Ammonium sulfate	0.1 M Bis-Tris pH 6		
F1	19.64 % (w/v) PEG 8,000		0.15 M Ammonium tartrate	
F2	36.04 % (v/v) PEG 400	0.1 M Bicine pH 8.5	0.25 M Sodium bromide	
F3	2.58 M Sodium malonate			
F4	23.94 % (w/v) PEG MME 5,000	0.1 M Bis-Tris pH 6.5	0.19 M Ammonium sulfate	
F5	14.08 % (w/v) PEG 10,000		0.12 M Magnesium acetate	
F6	0.82 M Sodium citrate	0.1 M HEPES pH 8		5.89 % (v/v) BME
F7	0.6 M Sodium citrate	0.1 M Bis-Tris pH 6.5		
F8	1.65 M Ammonium phosphate	0.1 M Sodium cacodylate pH 6.5	0.03 M Sodium-potassium tartrate	2.28 % (v/v) PEG DME 250
F9	1.86 M Ammonium sulfate	0.1 M Sodium citrate pH 5	0.14 M Sodium citrate	
F10	1.73 M Sodium malate		0.09 M Potassium chloride	0.09 % (w/v) CHAPS
F11	1.03 M Sodium tartrate	0.1 M HEPES pH 7.5		1.04 mM EDTA
F12	2.12 M Sodium malate	0.1 M Tris-HCl pH 7.5	0.28 M Sodium chloride	
G1	19.82 % (w/v) PEG 8,000	0.1 M Tris-HCl pH 8	0.06 M Sodium citrate	
G2	50.97 % (v/v) PEG 400	0.1 M Sodium acetate pH 5.5	0.11 M Ammonium tartrate	0.07 % (w/v) LDAO
G3	25.52 % (w/v) PEG MME 2,000	0.1 M Sodium acetate pH 5	0.04 M Magnesium formate	1.14 % (v/v) Butanediol
G4	2.77 M Sodium formate	0.1 M Sodium citrate pH 4.5	0.17 M Lithium sulfate	
G5	51.13 % (v/v) MPD	0.1 M HEPES pH 7.5		
G6	26.5 % (w/v) PEG MME 5,000	0.1 M Bicine pH 9.5	0.1 M Magnesium chloride	
G7	20.27 % (v/v) Ethanol	0.1 M Tris-HCl pH 8.5	0.09 M Sodium-potassium tartrate	
G8	30.16 % (v/v) MPD	0.1 M Bicine pH 9.5	0.21 M Potassium thiocyanate	
G9	1.67 M Sodium-potassium phosphate	0.1 M MOPS pH 7	0.14 M Sodium bromide	
G10	37.44 % (w/v) PEG 1,500	0.1 M Sodium acetate pH 5		
G11	38.82 % (w/v) PEG MME 2,000	0.1 M Sodium citrate pH 5		
G12	16.66 % (w/v) PEG MME 5,000	0.1 M MES pH 6.5		2.52 % (v/v) Methanol
H1	2.77 M Sodium formate	0.1 M Tris-HCl pH 7.5	0.16 M Ammonium sulfate	
H2	14.07 % (v/v) Isopropanol	0.1 M Bicine pH 8.5	0.08 M Sodium-potassium phosphate	
H3	20.23 % (w/v) PEG 3,350	0.1 M Sodium acetate pH 5.5	0.09 M Ammonium sulfate	0.09 % (w/v) CHAPS
H4	1.27 M Sodium tartrate	0.1 M Bicine pH 9.5	0.12 M Sodium citrate	5.75 % (v/v) BME
H5	19.38 % (w/v) PEG 10,000	0.1 M HEPES pH 7.5	0.08 M Sodium-potassium phosphate	0.1 % (w/v) CHAPS
H6	24.9 % (w/v) PEG MME 5,000	0.1 M HEPES pH 8	0.25 M Magnesium acetate	
H7	30.58 % (w/v) PEG MME 2,000	0.1 M Tris-HCl pH 7.5	0.04 M Magnesium formate	
H8	38.97 % (w/v) PEG 1,500	0.1 M MOPS pH 7	0.26 M Ammonium phosphate	
H9	22.58 % (w/v) PEG MME 5,000	0.1 M Sodium cacodylate pH 6.5	0.13 M Potassium nitrate	
H10	27.93 % (w/v) PEG 8,000	0.1 M MES pH 6.5		
H11	50.82 % (v/v) MPD	0.1 M Tris-HCl pH 8		3.53 % (v/v) Dioxane
H12	26.25 % (w/v) PEG MME 5,000	0.1 M Tris-HCl pH 7.5	0.16 M Magnesium acetate	

8.4.4 St Andrews 16

Well	Precipitant	Buffer	Additive 1	Additive 2
A1	32.52 % (w/v) PEG MME 2,000	0.1 M Sodium citrate pH 5.5	0.05 M Sodium acetate	
A2	25.8 % (w/v) PEG 3,350	0.1 M Tris-HCl pH 8.5		0.03 % (w/v) LDAO
A3	12.01 % (v/v) Ethanol	0.1 M HEPES pH 7.5		
A4	1.84 M Sodium malonate			
A5	1.63 M Sodium-potassium phosphate	0.1 M Sodium cacodylate pH 7		4.15 mM EDTA
A6	19.43 % (w/v) PEG 10,000		0.08 M Zinc chloride	
A7	32.47 % (w/v) PEG MME 2,000	0.1 M Tris-HCl pH 7.5	0.07 M Zinc acetate	
A8	34.43 % (w/v) PEG 1,500	0.1 M CHES pH 9.5	0.14 M Sodium citrate	
A9	18.28 % (w/v) PEG 6,000	0.1 M Tris-HCl pH 7.5	0.03 M Zinc acetate	
A10	1.57 M Sodium malate		0.11 M Sodium-potassium tartrate	0.51 % (v/v) Dioxane
A11	14.61 % (v/v) Ethanol	0.1 M Sodium acetate pH 4.5	0.08 M Zinc chloride	
A12	34.92 % (w/v) PEG 1,500		0.23 M Magnesium chloride	
B1	21.29 % (w/v) PEG MME 5,000	0.1 M MES pH 6.5	0.03 M Sodium-potassium tartrate	
B2	18.17 % (w/v) PEG 1,500	0.1 M Sodium acetate pH 5.5	0.08 M Zinc chloride	
B3	2.58 M Ammonium phosphate	0.1 M Bicine pH 9		
B4	24.78 % (w/v) PEG 4,000	0.1 M Bicine pH 8.5	0.14 M Sodium-potassium phosphate	
B5	24/68 % (w/v) PEG 3,350	0.1 M HEPES pH 8	0.14 M Calcium acetate	
B6	2.33 M Sodium-potassium phosphate	0.1 M Bis-Tris pH 6	0.19 M Ammonium sulfate	
B7	22.42 % (w/v) PEG 4,000	0.1 M Sodium citrate pH 5		
B8	2.73 M Sodium malonate	0.1 M CHES pH 9		
B9	12.52 % (v/v) Ethanol	0.1 M HEPES pH 8	0.05 M Magnesium acetate	
B10	2.57 M Sodium malonate	0.1 M MOPS pH 7	0.06 M Sodium citrate	
B11	33.05 %(v/v) PEGMME550	0.1 M Sodium cacodylate pH 6.5	0.15 M Sodium bromide	
B12	2.18 M Sodium malate	0.1 M CHES pH 9.5	0.11 M Ammonium acetate	2.5 % (v/v) Butanediol
C1	24.11 % (w/v) PEG 8,000	0.1 M CHES pH 9	0.23 M Lithium sulfate	
C2	26.57 % (w/v) PEG 6,000	0.1 M Sodium cacodylate pH 7	0.25 M Magnesium chloride	
C3	22.06 % (w/v) PEG MME 5,000	0.1 M MES pH 6	0.13 M Calcium chloride	
C4	18/72 % (w/v) PEG 8,000	0.1 M Bicine pH 8.5	0.24 M Magnesium chloride	
C5	11.83 % (v/v) Isopropanol	0.1 M Sodium citrate pH 5	0.21 M Lithium sulfate	
C6	2.42 M Sodium formate	0.1 M Bicine pH 9	0.09 M Sodium chloride	
C7	17.76 % (w/v) PEG 4,000	0.1 M Sodium citrate pH 4.5	0.14 M Ammonium acetate	
C8	3.46 M Sodium chloride	0.1 M Bicine pH 9	0.23 M Potassium chloride	
C9	14/83 % (w/v) PEG 10,000	0.1 M Sodium citrate pH 5	0.17 M Magnesium acetate	
C10	2.31 M Sodium malonate	0.1 M Sodium citrate pH 4.5	0.15 M Sodium-potassium phosphate	
C11	49.89 % (v/v) PEG 400	0.1 M Sodium citrate pH 5	0.08 M Sodium chloride	
C12	24.33 % (v/v) PEG MME 550			
D1	1.78 M Sodium malonate	0.1 M Tris-HCl pH 8	0.29 M Ammonium phosphate	
D2	2.03 M Sodium-potassium phosphate	0.1 M HEPES pH 7.5		1.62 % (v/v) PEG 400
D3	2.34 M Sodium formate	0.1 M Sodium citrate pH 5.5	0.09 M Ammonium tartrate	0.08 % (v/v) LDAO
D4	23.44 % (w/v) PEG 3,350	0.1 M Bis-Tris pH 6	0.06 M Calcium acetate	
D5	2.01 M Sodium formate	0.1 M Sodium acetate pH 4.5	0.18 M Potassium thiocyanate	
D6	50.74 % (v/v) MPD		0.12 M Sodium acetate	3.05 % (v/v) PEG DME 250
D7	35.6 % (w/v) PEG MME 2,000			
D8	21.86 % (w/v) PEG 8,000	0.1 M MES pH 6.5	0.08 M Zinc acetate	2.01 % (v/v) BME
D9	33.67 % (w/v) PEG 1,500	0.1 M Tris-HCl pH 8	0.15 M Sodium acetate	
D10	21.34 % (w/v) PEG 3,350	0.1 M CHES pH 9.5		
D11	2.12 M Ammonium sulfate	0.1 M Tris-HCl pH 8		1.81 % (v/v) Hexanediol

D12	22.92 % (w/v) PEG 1,500	0.1 M MOPS pH 7	0.08 M Calcium chloride	2.38 % (v/v) PEG MME 350
E1	3.21 M Sodium chloride	0.1 M Sodium acetate pH 5	0.12 M Ammonium phosphate	
E2	33.73 % (w/v) PEG MME 2,000	0.1 M Sodium citrate pH 5.5	0.06 M Sodium-potassium phosphate	
E3	32.75 % (w/v) PEG 3,350		0.12 M Sodium chloride	
E4	27.8 % (w/v) PEG MME 5,000	0.1 M Sodium citrate pH 5.5		
E5	35.89 % (v/v) MPD		0.03 M Zinc chloride	
E6	29.41 % (w/v) PEG MME 2,000	0.1 M MES pH 6	0.17 M Magnesium sulfate	3.99 % (v/v) DMSO
E7	2.15 M Sodium malonate	0.1 M Bis-Tris pH 6	0.08 M Potassium chloride	0.48 % (v/v) MPD
E8	2.25 M Sodium formate	0.1 M Bicine pH 9.5		
E9	0.82 M Sodium citrate	0.1 M Bicine pH 8.5	0.09 M Magnesium acetate	
E10	2.2 M Sodium formate	0.1 M Tris-HCl pH 7.5	0.12 M Ammonium sulfate	1.2 % (v/v) PEG MME 350
E11	18.84 % (w/v) PEG 10,000	0.1 M CHES pH 9		
E12	48.9 % (w/v) PEG MME 550	0.1 M Sodium cacodylate pH 7	0.1 M Calcium acetate	0.57 % (v/v) MPD
F1	2.16 M Ammonium phosphate	0.1 M Sodium acetate pH 5.5	0.14 M Ammonium sulfate	
F2	1.84 M Sodium malonate	0.1 M Sodium acetate pH 5	0.19 M Potassium thiocyanate	
F3	12.74 % (w/v) PEG 6,000	0.1 M CHES pH 9	0.08 M Ammonium tartrate	
F4	2.23 M Ammonium phosphate	0.1 M CHES pH 9.5		
F5	53.38 % (v/v) PEG400	0.1 M HEPES pH 7.5	0.08 M Ammonium citrate	
F6	6.68 % (v/v) Isopropanol	0.1 M HEPES pH 8		2.15 % (v/v) BME
F7	1.97 M Ammonium sulfate	0.1 M Sodium cacodylate pH 6.5	0.11 M Ammonium citrate	
F8	39.18 % (v/v) PEG 400	0.1 M HEPES pH 8	0.13 M Calcium chloride	
F9	1.6 M Sodium malate	0.1 M MES pH 6		
F10	17.94 % (w/v) PEG 8,000	0.1 M Sodium acetate pH 5	0.09 M Magnesium acetate	0.02 % (v/v) OG
F11	0.93 M Sodium tartrate	0.1 M MES pH 6		
F12	2.67 M Sodium-potassium phosphate	0.1 M Bis-Tris pH 6	0.11 M Potassium nitrate	
G1	26.9 % (w/v) PEG MME 5,000	0.1 M Sodium citrate pH 4.5		
G2	12.68 % (v/v) Isopropanol	0.1 M MES pH 6.5	0.09 M Calcium chloride	
G3	27.57 % (w/v) PEG 8,000	0.1 M Bicine pH 8.5	0.06 M Magnesium formate	1.48 % (v/v) PEG400
G4	7.73 % (v/v) Isopropanol	0.1 M CHES pH 9	0.07 M Potassium chloride	
G5	1.64 M Magnesium sulfate	0.1 M Sodium citrate pH 5.5	0.06 M Sodium bromide	
G6	38.07 % (w/v) PEG MME 2,000	0.1 M CHES pH 9	0.11 M Sodium-potassium phosphate	
G7	31.49 % (w/v) PEG MME 550	0.1 M Sodium citrate pH 4.5	0.07 M Ammonium tartrate	
G8	0.96 M Sodium citrate		0.11 M Magnesium sulfate	1.12 % (v/v) DMSO
G9	23.48 % (w/v) PEG 4,000	0.1 M Sodium acetate pH 4.5	0.12 M Zinc acetate	
G10	43.89 % (v/v) PEG 400	0.1 M HEPES pH 8	0.15 M Magnesium chloride	
G11	20 % (w/v) PEG 10,000	0.1 M Tris-HCl pH 8.5	0.04 M Calcium acetate	
G12	20.66 % (v/v) Ethanol	0.1 M Sodium citrate pH 4.5	0.04 M Ammonium citrate	
H1	40.41 % (v/v) MPD	0.1 M Bicine pH 9.5	0.1 M Potassium nitrate	
H2	1.68 M Sodium malonate	0.1 M Bicine pH 9.5		
H3	2.37 M Ammonium phosphate	0.1 M Sodium acetate pH 4.5	0.07 M Sodium acetate	
H4	27.51 % (w/v) PEG MME 2,000	0.1 M HEPES pH 7.5	0.13 M Lithium chloride	
H5	0.97 M Sodium tartrate	0.1 M Bicine pH 9.5		
H6	2.77 M Sodium acetate		0.22 M Sodium chloride	
H7	44.16 % (v/v) MPD	0.1 M Sodium citrate pH 5.5	0.06 M Lithium chloride	
H8	24.92 % (w/v) PEG 3,350	0.1 M Sodium acetate pH 5	0.05 M Calcium acetate	
H9	2.36 M Ammonium phosphate			1.03 % (v/v) PEG MME 350
H10	1.89 M Sodium malate	0.1 M MES pH 6.5	0.18 M Potassium chloride	1.42 % (v/v) BME
H11	21.16 % (w/v) PEG 4,000	0.1 M MOPS pH 6.5	0.13 M Ammonium citrate	2.24 % (v/v) Methanol
H12	2.38 M Sodium formate	0.1 M MES pH 6	0.03 M Sodium-potassium tartrate	3.12 % (v/v) Ethylene glycol

8.4.5 St Andrews PEG 1

Well	Precipitant	Buffer	Additive 1	Additive 2
A1	20.58 % (w/v) PEG 10,000	0.1 M HEPES pH 8	0.08 M Ammonium acetate	
A2	21.44 % (w/v) PEG MME 5,000	0.1 M Bicine pH 9.5	0.19 M Potassium thiocyanate	
A3	18.95 % (w/v) PEG 8,000	0.1 M Sodium cacodylate pH 6.5	0.07 M Potassium nitrate	
A4	19.29 % (w/v) PEG 6,000	0.1 M Tris-HCl pH 8	0.25 M Potassium thiocyanate	
A5	17.77 % (w/v) PEG MME 5,000	0.1 M Sodium acetate pH 5	0.1 M Sodium bromide	3.06 % (v/v) DMSO
A6	20.29 % (w/v) PEG 4,000	0.1 M MOPS pH 7	0.11 M Magnesium formate	
A7	20.64 % (w/v) PEG MME 5,000	0.1 M Bicine pH 8.5		5.72 % (v/v) BME
A8	12.6 % (w/v) PEG 8,000	0.1 M HEPES pH 8	0.19 M Potassium chloride	
A9	34.19 % (v/v) PEG 400	0.1 M MOPS pH 6.5	0.12 M Calcium acetate	0.02 % (w/v) OG
A10	37.52 % (w/v) PEG MME 550	0.1 M CHES pH 9	0.23 M Magnesium acetate	
A11	12.66 % (w/v) PEG 8,000	0.1 M MOPS pH 6.5	0.14 M Calcium chloride	
A12	46.79 % (w/v) PEG 400		0.11 M Zinc chloride	
B1	19.3 % (w/v) PEG 10,000	0.1 M Sodium citrate pH 5.5	0.18 M Sodium chloride	
B2	20.02 % (w/v) PEG MME 2,000	0.1 M Sodium cacodylate pH 7	0.14 M Sodium acetate	
B3	50.02 % (v/v) PEG 400	0.1 M Bicine pH 8.5	0.09 M Ammonium sulfate	1.56 % (v/v) PEG 400
B4	22.07 % (w/v) PEGMME2k	0.1 M Sodium cacodylate pH 7	0.11 M Zinc acetate	
B5	25.96 % (w/v) PEG 1,500	0.1 M MES pH 6	0.07 M Sodium-potassium tartrate	0.04 % (w/v) OG
B6	16.22 % (w/v) PEG 6,000	0.1 M Sodium citrate pH 4.5	0.04 M Magnesium formate	
B7	24.04 % (w/v) PEG 3,350	0.1 M MES pH 6	0.3 M Ammonium sulfate	
B8	18 % (w/v) PEG 6,000	0.1 M Sodium acetate pH 5	0.07 M Zinc chloride	0.09 % (w/v) LDAO
B9	23.35 % (w/v) PEG 4,000	0.1 M Bis-Tris pH 6	0.07 M Magnesium sulfate	
B10	24.63 % (w/v) PEG 8,000	0.1 M Bicine pH 8.5	0.12 M Sodium citrate	0.05 % (w/v) LDAO
B11	29.77 % (w/v) PEG MME 550	0.1 M Sodium acetate pH 5	0.12 M Calcium acetate	
B12	33.69 % (w/v) PEG 3,350	0.1 M Bicine pH 9.5	0.03 M Ammonium tartrate	
C1	27.65 % (w/v) PEG MME 5,000	0.1 M Sodium citrate pH 5		
C2	35.93 % (w/v) PEG 1,500	0.1 M Sodium citrate pH 4.5	0.25 M Sodium chloride	0.56 % (v/v) PEG MME 350
C3	45.62 % (w/v) PEG MME 550	0.1 M CHES pH 9.5	0.09 M Ammonium phosphate	
C4	31.4 % (w/v) PEG MME 2,000	0.1 M Sodium citrate pH 5.5	0.08 M Sodium acetate	
C5	22.28 % (w/v) PEG 8,000	0.1 M Sodium citrate pH 5		
C6	11.01 % (w/v) PEG 10,000	0.1 M CHES pH 9	0.24 M Potassium chloride	8.55 mM EDTA
C7	28.28 % (w/v) PEG 1,500	0.1 M Bicine pH 9.5	0.13 M Sodium-potassium tartrate	
C8	29.75 % (w/v) PEG 4,000	0.1 M Sodium citrate pH 5		
C9	16.88 % (w/v) PEG MME 5,000	0.1 M Bicine pH 9	0.11 M Magnesium chloride	
C10	30.84 % (w/v) PEG 3,350	0.1 M Bis-Tris pH 6.5	0.03 M Calcium chloride	
C11	23.63 % (w/v) PEG MME 550	0.1 M Sodium cacodylate pH 7	0.12 M Sodium-potassium phosphate	2.99 % (v/v) Ethylene glycol
C12	54.24 % (v/v) PEG 400	0.1 M Tris-HCl pH 8.5		9.04 mM EDTA
D1	23.27 % (w/v) PEG MME 2,000	0.1 M Bicine pH 8.5	0.1 M Ammonium tartrate	3.6 % (v/v) Methanol
D2	32.21 % (w/v) PEG 1,500	0.1 M MES pH 6		
D3	32.84 % (w/v) PEG 3,350	0.1 M MOPS pH 7	0.11 M Ammonium citrate	
D4	21.32 % (w/v) PEG MME 2,000	0.1 M Bicine pH 9.5	0.05 M Magnesium chloride	0.09 % (w/v) LDAO
D5	40.13 % (w/v) PEG 1,500	0.1 M Sodium acetate pH 5.5	0.13 M Sodium citrate	
D6	25.1 % (w/v) PEG MME 5,000		0.11 M Calcium chloride	
D7	37.73 % (w/v) PEG MME 550	0.1 M HEPES pH 7.5	0.12 M Magnesium chloride	
D8	22.73 % (w/v) PEG 8,000		0.1 M Potassium nitrate	2.41 % (v/v) MPD
D9	13.13 % (w/v) PEG 10,000	0.1 M Tris-HCl pH 7.5		1.08 % (v/v) Ethylene glycol
D10	18.03 % (w/v) PEG 4,000	0.1 M Sodium acetate pH 4.5		2.79 % (v/v) Butanediol
D11	15.59 % (w/v) PEG 3,350	0.1 M MOPS pH 7		

D12	14.02 % (w/v) PEG 4,000	0.1 M Sodium acetate pH 5.5		
E1	20.75 % (w/v) PEG 6,000	0.1 M MES pH 6		
E2	15.15 % (w/v) PEG 8,000	0.1 M Tris-HCl pH 8.5	0.04 M Sodium acetate	
E3	31.05 % (w/v) PEG 1,500		0.25 M Sodium-potassium phosphate	3.83 % (v/v) Dioxane
E4	13.11 % (w/v) PEG 8,000	0.1 M MES pH 6.5	0.1 M Lithium chloride	
E5	44.73 % (v/v) PEG 400	0.1 M MOPS pH 6.5	0.11 M Ammonium tartrate	
E6	27.66 % (w/v) PEG 8,000	0.1 M Sodium citrate pH 5		
E7	19.49 % (w/v) PEG MME 5,000	0.1 M CHES pH 9.5	0.1 M Ammonium citrate	
E8	14.68 % (w/v) PEG 10,000	0.1 M CHES pH 9.5		2.46 % (v/v) Hexanediol
E9	12.51 % (w/v) PEG 6,000	0.1 M MES pH 6	0.12 M Ammonium acetate	
E10	44.9 % (v/v) PEG 400	0.1 M Tris-HCl pH 7.5	0.13 M Calcium acetate	2.44 % (v/v) Methanol
E11	39.87 % (w/v) PEG MME 2,000	0.1 M Bis-Tris pH 6	0.17 M Potassium chloride	
E12	13.18 % (w/v) PEG 10,000	0.1 M HEPES pH 7.5		
F1	15.59 % (w/v) PEG 4,000	0.1 M Sodium citrate	0.16 M Lithium sulfate	3.39 % (v/v) PEG 400
F2	30.78 % (w/v) PEG 1,500		0.09 M Zinc chloride	
F3	16.58 % (w/v) PEG 6,000	0.1 M Sodium acetate	0.2 M Lithium chloride	
F4	20.22 % (w/v) PEG 4,000	0.1 M HEPES pH 8	0.12 M Lithium sulfate	
F5	24.31 % (w/v) PEG 1,500	0.1 M Tris-HCl pH 7.5		
F6	33.83 % (w/v) PEG 3,350	0.1 M Sodium acetatepH 4.5		
F7	32.33 % (w/v) PEG MME 2,000	0.1 M Tris-HCl pH 7.5	0.14 M Zinc chloride	
F8	19.58 % (w/v) PEG 10,000	0.1 M Tris-HCl pH 8	0.14 M Sodium-potassium-tartrate	
F9	19.42 % (w/v) PEG MME 2,000		0.14 M Calcium chloride	
F10	54.58 % (v/v) PEG 400	0.1 M HEPES pH 8		
F11	40.72 % (w/v) PEG MME 550	0.1 M Sodium acetate pH 4.5	0.21 M Lithium chloride	
F12	25.94 % (w/v) PEG MME 2,000	0.1 M HEPES pH 8	0.08 M Zinc acetate	
G1	12.22 % (w/v) PEG 10,000	0.1 M BICINE pH 8.5		2.07 % (v/v) Glycerol
G2	25.18 % (w/v) PEG 3,350	0.1 M MOPS pH 7	0.15 M Magnesium sulfate	3.47 % (v/v) PEG DME 250
G3	26.34 % (w/v) PEG 8,000		0.09 M Ammonium citrate	
G4	25.81 % (w/v) PEG MME 2,000	0.1 M Sodium citrate pH 5.5		
G5	28 % (w/v) PEG 4,000	0.1 M Tris-HCl pH 8	0.08 M Sodium-potassium tartrate	
G6	45.42 % (w/v) PEG MME 550	0.1 M Tris-HCl pH 8.5	0.04 M Calcium chloride	
G7	28.52 % (w/v) PEG MME 2,000	0.1 M MES pH 6.5		
G8	12.41 % (w/v) PEG 6,000	0.1 M Bicine pH 9		
G9	32.32 % (v/v) PEG 400	0.1 M Bicine pH 9	0.06 M Sodium bromide	1.73 % (v/v) Dioxane
G10	16.13 % (w/v) PEG 8,000	0.1 M MOPS pH 7	0.05 M Calcium chloride	
G11	18.09 % (w/v) PEG 3,350	0.1 M Bicine pH 9		
G12	22.83 % (w/v) PEG MME 550	0.1 M Bicine pH 9	0.21 M Potassium chloride	
H1	29.58 % (w/v) PEG 1,500	0.1 M CHES pH 9.5	0.24 M Magnesium sulfate	0.07 % (v/v) OG
H2	20.47 % (w/v) PEG 6,000	0.1 M Tris-HCl pH 7.5		
H3	19.51 % (w/v) PEG 8,000	0.1 M HEPES pH 8		
H4	24.43 % (w/v) PEG 6,000	0.1 M Sodium citrate pH 5.5	0.06 M Ammonium phosphate	
H5	45.33 % (w/v) PEG MME 550	0.1 M Sodium acetate pH 5	0.18 M Magnesium sulfate	
H6	19.33 % (w/v) PEG MME 5,000	0.1 M Sodium acetate pH 5.5	0.26 M Potassium thiocyanate	2.99 % (v/v) PEG DME 250
H7	22.82 % (w/v) PEG 6,000	0.1 M Sodium citrate pH 4.5	0.06 M Magnesium acetate	0.47 % (v/v) Butanediol
H8	17.88 % (w/v) PEG 10,000	0.1 M Sodium citrate pH 4.5	0.06 M Sodium bromide	
H9	21.24 % (w/v) PEG MME 550		0.15 M Magnesium chloride	
H10	25.87 % (v/v) PEG 400	0.1 M Bicine pH 9	0.22 M Ammonium sulfate	
H11	27.53 % (w/v) PEG 6,000		0.14 M Lithium chloride	
H12	25.89 % (w/v) PEG8k	0.1 M Sodium acetate pH 5.5	0.1 M Zinc acetate	

8.5 StockOptions Salt (Hampton Research)

Well	Salt
1	1.0 M Ammonium acetate
2	5.0 M Ammonium chloride
3	2.5 M Ammonium phosphate monobasic
4	10.0 M Ammonium fluoride
5	10.0 M Ammonium formate
6	2.5 M Ammonium citrate dibasic
7	10.0 M Ammonium phosphate dibasic
8	10.0 M Ammonium nitrate
9	3.5 M Ammonium sulfate
10	2.0 M Ammonium tartrate dibasic
11	1.0 M Calcium acetate hydrate
12	2.0 M Calcium chloride dihydrate
13	5.0 M Lithium acetate dihydrate
14	10.0 M Lithium chloride
15	1.5 M Lithium citrate tribasic tetrahydrate
16	8.0 M Lithium nitrate
17	2.0 M Lithium sulfate monohydrate
18	1.0 M Magnesium acetate tetrahydrate
19	2.0 M Magnesium chloride hexahydrate
20	1.0 M Magnesium formate dihydrate
21	3.0 M Magnesium nitrate hexahydrate
22	2.5 M Magnesium sulfate hydrate
23	4.0 M Nickel (II) chloride hexahydrate
24	5.0 M Potassium acetate
25	4.0 M Potassium chloride
26	2.5 M Potassium citrate tribasic monohydrate
27	1.5 M Potassium phosphate monobasic
28	6.0 M Potassium fluoride
29	14.0 M Potassium formate
30	3.0 M Potassium phosphate dibasic
31	2.0 M Potassium nitrate
32	1.5 M Potassium sodium tartrate tetrahydrate
33	0.5 M Potassium sulfate
34	8.0 M Potassium thiocyanate
35	3.0 M Sodium acetate trihydrate
36	5.0 M Sodium chloride
37	1.6 M Sodium citrate tribasic dihydrate
38	5.0 M Sodium phosphate monobasic
39	0.8 M Sodium fluoride
40	7.0 M Sodium formate
41	1.0 M Sodium phosphate dibasic dihydrate
42	3.4 M Sodium malonate pH 7.0
43	7.0 M Sodium nitrate
44	1.0 M Sodium sulfate decahydrate
45	1.5 M Sodium tartrate dibasic dihydrate
46	8.0 M Sodium thiocyanate
47	1.2 M Succinic acid pH 7.0
48	1.0 M Zinc acetate dihydrate
49	2.0 M Zinc sulfate heptahydrate