Structural Studies on Oligosaccharide Transport and Sugar Scavenging in Bacteria



Micah Oliver Lee

Jesus College

University of Oxford

Supervised by Professor James H. Naismith, Professor Matthew K. Higgins

Submitted: Trinity Term 2021 (Initial), Hilary Term 2022 (Final)

This thesis is submitted to the Nuffield Department of Medicine, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Table of Contents

Abstract V

Acknowledgements VII

Abbreviations VIII

List of Figures X

List of Tables XIV

1. Introduction 1

1.a. Protein crystallography in modern structural biology 1

1.a.i. Experimental phasing by native sulfur SAD 5

1.a.ii. Antibody fragments and nanobodies 9

1.a.iii. Cell-free expression 13

1.b. Membrane proteins 14

1.b.i. Detergent-based reconstitution systems 16

1.b.ii. Lipid-based reconstitution systems 21

1.b.iii. Lipidic cubic phase 25

2. Structure determination of LPS flippase Wzx by native sulfur SAD 30

2.a. Abstract 30

2.b. Introduction 30

2.b.i. Lipopolysaccharide (LPS) biosynthesis in Gram negative bacteria 30

2.b.ii. Wzx, inner membrane scramblase in the LPS biosynthesis Pathway 33

2.b.iii. Functional characterisation of flippases 36

2.b.iv. Structural characterisation of flippases 37

2.b.v. Previous work done by laboratory members 40

2.c. Materials and methods 41

2.c.i. Preparation of chemically competent phage resistant cell lines 41

2.c.ii. Expression and purification of Wzx 41

2.c.iii. Expression and purification of nanobodies 43

2.c.iv. Preparation of Strep-tag nb10 45

2.c.v Western blot and dot blot 46

2.c.vi Crystallisation of the Wzx/nb10 complex 47

2.c.vii. Preparation of *In Meso In Situ* Crystallography (IMISX) plates 48

2.d. Results 49

2.d.i. Preparation of chemically competent phage resistant cell lines 49

2.d.ii. Purification and crystallisation of the Wzx/nb10 complex 50

2.d.iii. Preparation of Strep-tag nb10 54

2.d.iv. Double pulldown assay for ternary complex screening 57

2.d.v. Exogenous lipids as additives for LCP crystallisation 63

2.d.vi. Data collection at the I23 beamline 65

2.d.vii. Data processing for native s-SAD 69

2.d.viii. Apo-structure of *E.* coli K-12 WzxE 71

2.d.ix. Substrate complex crystallisation trials 78

2.d.x. *In Meso In Situ* Crystallography (IMISX) 81

2.d.xi. Comparison with structural features in homologues 83

2.e. Conclusion 84

2.f. Discussion and future directions 86

3. Structure determination of sialic acid oxidoreductase RgNanOx 87

3.a. Abstract 87

3.b. Introduction 88

3.b.i. Sialic acid metabolism in gut bacteria 88

3.b.ii. Canonical enzymatic mechanisms at the anomeric centre 90

3.b.iii. Non-canonical glycosidase mechanisms 94

3.b.iv. RgNanOx is a sialic acid oxidoreductase in *Ruminococcus gnavus* 96

3.c. Materials and methods 99

3.c.i. Molecular cloning and mutagenesis 99

3.c.ii. Expression and purification of RgNanOx 101

3.c.iii. Crystallisation of RgNanOx and substrate co-crystallisation 102

3.c.iv. Structure determination 103

3.c.v. Liquid chromatography-mass spectrometry (LCMS) 104

3.c.vi. Isothermal titration calorimetry (ITC) 105

3.c.vii. Nuclear magnetic resonance (NMR) spectroscopy 106

3.d. Results 106

3.d.i. Recloning and mutagenesis for RgNanOx 106

3.d.ii. Expression and purification of RgNanOx 108

3.d.iii. Determination of RgNanOx holo- structure 112

3.d.iv. Sequence comparison with homologues reveals unusual catalytic motif 116

3.d.v. Identification of putative catalytic residues using high resolution data 121

3.d.vi. Establishment of an LCMS functional assay for RgNanOx 126

3.d.vii. ITC substrate binding assay 136

3.d.viii. NMR kinetic assay 137

3.d.ix. Co-crystallisation screening 144

3.d.x. Elucidation of the dehydration step using the Neu complex structure 145

3.d.xi. Elucidation of the cyclisation step using the Anh complex structure 153

3.d.xii. Proposal of the RgNanOx catalytic mechanism 155

3.e. Conclusion 158

3.f. Discussion and future directions 159

4. Executive Summary 162

Bibliography 163

Appendix 191

A. Crystal screen compositions 191

B. XSCALE data table 192

C. Crystallographic tables 192

Abstract

This thesis reports on the structural and functional studies performed using x-ray crystallography on two proteins involved in bacterial sugar metabolism. The first protein, the inner membrane oligosaccharide scramblase Wzx, is central to capsule and lipopolysaccharide export in pathogens and is thus a potential drug target. The crystal structure was determined at 2.3 Å using sulfur phasing at the long wavelength beamline I23 at Diamond Light Source. This required us to obtain high quality crystals using lipidic cubic phase, exogenous lipids, and a nanobody. The crystal structure reveals a positively charged patch at both openings to a central cavity. The structure will guide future site directed mutagenesis and mechanistic studies. Other nanobodies were identified and these could help trap other conformational states.

The second protein studied was the sialic acid oxidoreductase RgNanOx. This protein used in sialic acid metabolism by the gut bacterium *Ruminococcus gnavus*, which converts sialic acid (Neu5Ac) into the bicyclic form (2,7-anhydro-Neu5Ac). This involves an unusual chemical mechanism where the substrate is both oxidised and reduced. The crystal structure of RgNanOx was first determined at 1.7 Å. The structure suggested a mechanism for this unusual chemical transformation which was tested by site directed mutagenesis. The enzyme was shown to be active against the sialidase inhibitor 2,3-dehydro-2-deoxy-Neu5Ac (DANA). Co-crystallisation with DANA produced an α-Neu5Ac complex. A mutant was used to obtain a 2,7-anhydro-Neu5Ac complex. These complexes allow the mechanism to be clarified and the key catalytic residues identified. It is an unresolved question whether the true substrate is the α or β form of the sialic acid. NMR data suggest that α-Neu5Ac is consumed, which would indicate the enzyme proceeds by an unusual *syn*- elimination. The work opens

opportunities to definitively assign the mechanism, and shed light on the role of the enzyme in the gut environment.

Acknowledgements

I would like to thank my primary supervisor, Professor Jim Naismith, for the opportunity to undertake this DPhil project in his laboratory and his advice over the past three years. My secondary supervisor, Professor Matt Higgins, and my thesis committee consisting of Professor Frank von Delft and Professor Yvonne Jones, also provided invaluable guidance.

I would like to thank the Naismith group for their support in the laboratory. Most notably, Dr Audrey Le Bas who acted as my direct supervisor along with Dr Haigang Song who provided generous support for the RgNanOx project. My thanks extend to other members of the Naismith lab (particularly the Sugar Team consisting of Dr Peter Harrison, Dr Phil Ward, Dr Miriam Weckner, and Dr Yun Yang), Membrane Protein Laboratory led by Dr Andrew Quigley, Protein Production UK led by Professor Ray Owens, and the team at the Rosalind Franklin Institute. I would like to thank my collaborators at the I23 beamline led by Dr Armin Wagner, Professor Martin Caffrey and his laboratory for training me in LCP and IMISX plates, Professor Nathalie Juge and Dr Andrew Bell at University of East Anglia for the RgNanOx project, Professor Ben Davis and Mr Ben Gaunt for the RgNanOx NMR study, and Dr Gemma Harris for training me in ITC. I would also like to thank the college and the university for the immensely enjoyable three years at Oxford.

Finally, I would like to express my immense gratitude to my friends, family, and my boyfriend Igor whose support I relied on through these memorable years.

Abbreviations

2D/3D	Two/three dimensional
ABC	ATP biding cassette
Anh	2,7-anhydro-N-acetylneuraminic acid
ATP	Adenosine triphosphate
CMC	Critical micelle concentration
COC/COP	Cyclic olefin (co)polymer
(Cryo)EM	Cryo-electron microscopy
CPS	Capsular polysaccharide
DANA	N-acetyl-2,3-dehydro-2-deoxyneuraminic acid
DDM	N-dodecyl β-D-maltoside
DM	N-decyl β-D-maltoside
DNA	Deoxyribonucleic acid
DSS	3-(Trimethylsilyl)-1-propanesulfonic acid
ECA	Enterobacteria common antigen
GlcNAc	N-Acetyl-Glucosamine
GPCR	G protein coupled receptor
HPLC	High performance liquid chromatography
IMAC	Immobilised metal affinity chromatography
IMISX	*In meso in situ* crystal
IPTG	Isopropylthio-β-galactoside
ITC	Isothermal titration calorimetry
KDN	Ketodeoxynonulosonic acid
LB	Luria Broth
(LC)MS	Liquid chromatography mass spectrometry

LCP	Lipid cubic phase
LPS	Lipopolysaccharide
MSP	Membrane scaffold protein
NAc	N-acetyl
NAD(P)	Nicotinamide adenine dinucleotide (phosphate)
Neu5Ac/Neu	N-acetylneuraminic acid
NMR	Nuclear magnetic resonance
OGNG	Octyl Glucose Neopentyl Glycol (2,2-dihexylpropane-1,3-bis-β-D-glucopyranoside)
O/N	Overnight
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
RMSD	Root mean standard deviation
RPM	Rotations per minute
RT	Room temperature
SAD	Single-wavelength anomalous diffraction
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TB	Terrific Broth
T_m	Melting temperature
UndPP	Undecaprenol pyrophosphate
UV	Ultraviolet
WT	Wild type

List of Figures

1.1	Fourier transform	5
1.2	Average sulfur content of bacterial proteins	8
1.3	Domain organisation of antibody fragments and nanobodies	10
1.4	PglK co-crystallised with nanobody	11
1.5	Diagram summarising membrane protein reconstitution methods	18
1.6	Molecular dynamics simulation of amphipol binding	20
1.7	Bound lipids are observed in crystalline structures	22
1.8	Detergent extraction of liposome lipids	22
1.9	Chemical structure of the SMA repeat unit	24
1.10	Membrane protein crystal packing	27
1.11	LCP forms at 20 °C at 33.3% water	28
1.12	Beamline setup for IMISX	29
1.13	Diagram of IMISX plate	29
2.1	The LPS and CPS pathways	32
2.2	Lipid III and Lipid II	35
2.3	Structurally resolved catalytic stages of flippases	39
2.4	Assembled in house IMISX plate	49
2.5	Testing wk6 for chemical competency	50
2.6	Western blot shows no expression of selenomethionine Wzx	51
2.7	Purification of the Wzx/nb10 complex	53
2.8	nb10 His-tag was swapped to Strep-tag	55
2.9	nb10-strep expression required modified protocol	56
2.10	Nanobody IMAC elution profiles	58
2.11	Dot blot suggested many nanobodies can form ternary complex	59

2.12	SDS-PAGE of pulldown fractions	61
2.13	Western blot was also inconclusive	63
2.14	Proprietary I23 pins	65
2.15	Reducing beam size improved data quality	67
2.16	Manual grid scan at I23	68
2.17	Wzx data processing workflow	68
2.18	S-SAD anomalous signal map	71
2.19	Crystalline structure of Wzx/nb10 complex	72
2.20	Wzx binding by nb10	73
2.21	Comparison of Wzx structure with homologues	75
2.22	Surface representation of Wzx/nb10 structure	76
2.23	Low sequence conservation between Wzx_{E} and Wzx_{PAO1}	77
2.24	Substrate analogue structures	78
2.25	No significant substrate density in soaked Wzx crystals	80
2.26	IMISX is more similar to standard LCP glass plate than SWSSICI	82
2.27	Wzx could not coordinate Na+ like MurJ	83
2.28	Wzx could coordinate cation like NorM	84
3.1	Sialic acid metabolism in bacteria	89
3.2	The oxocarbenium ion	90
3.3	Classification of canonical enzyme mechanism	92
3.4	Mechanisms proposed for lysozyme	93
3.5	BglT mechanism	95
3.6	Simplified NanA mechanism	96
3.7	Pfam identified two domains in RgNanOx	96
3.8	Catalytic versatility in the Gfo/Idh/MocA family	98

3.9	RgNanOx general reaction scheme	99
3.10	RgNanOx crystals	103
3.11	Linearisation of pHisTEV and RgNanOx plasmid	107
3.12	Agarose gel showing successful site directed mutagenesis	107
3.13	SDS-PAGE gels verify RgNanOx expression and IMAC purification	109
3.14	MALDI-MS spectra showed RgNanOx TEV cleavage	109
3.15	RgNanOx was homogenous after SEC	111
3.16	Western blot confirmed selenomethionine-RgNanOx expression	112
3.17	Holo-structure of RgNanOx	114
3.18	NAD^+ and citrate were bound to holo-RgNanOx	116
3.19	Multiple sequence alignment showed putative catalytic residues	119
3.20	Superposition of putative catalytic residues	120
3.21	Alignment of multi-histidine catalytic motifs in the Gfo/Idh/Mofa family	121
3.22	High resolution structure confirmed citrate confirmation	122
3.23	Citrate was positioned identically in both structures	122
3.24	Citrate conformation used to show possible substrate orientations	125
3.25	Positive controls resulted in recurrent LCMS peaks	129
3.26	Negligible MS peaks from recurrent HPLC peaks	130
3.27	Positive controls verified by LCMS	130
3.28	RgNanOx interconverts Neu and Anh	131
3.29	Histidine mutants appear inactive in MS spectra	133
3.30	R150A is prima facie inactive	134
3.31	RgNanOx showed activity against DANA	135
3.32	Initial titration curves showed no clear signal	136
3.33	Dialysed protein did not bind Neu	137

3.34	NMR showed Neu-Anh conversion	140
3.35	NMR showed Anh-Neu conversion	141
3.36	Anomers of Neu5Ac	141
3.37	Anh signal over time at varying starting Neu concentrations	142
3.38	Starting with Anh produced greater signal	142
3.39	Signal intensity traces were inconsistent between anomers	143
3.40	Equation used to fit Anh and α-Neu peak integrals	143
3.41	Equation used to fit β-Neu peak integrals	143
3.42	Graphical method for estimated K_M and k_{cat}	144
3.43	Optimisation of complex crystals	145
3.44	RgNanOx/Neu complex structure	148
3.45	Superposition of holo- and complexed RgNanOx	149
3.46	Hydrogen bonding network of Neu	149
3.47	Neu5Ac carboxylates are well-coordinated	150
3.48	Dehydration step mechanism proposed for each anomer	151
3.49	Superposition of co-crystallised Neu and citrate	152
3.50	Catalytic residues ideally positioned for dehydration	152
3.51	RgNanOx/Anh complex structure	154
3.52	Bicyclic density in Anh complex structure	154
3.53	Hydrogen bonding network of Anh	155
3.54	Proposed RgNanOx mechanism	157

List of Tables

1.1	Commercial available detergents	17
2.1	Amino acid sequences of nanobodies	44
2.2	Primers used for nb10 strep-his tag swapping	45
2.3	Screening resolution increases after lipid addition	64
3.1	Primers for mutagenesis	100
3.2	Candidate catalytic residues	126

1. Introduction

1.a. Protein crystallography in modern structural biology

In the mid-twentieth century, structure determination by x-ray crystallography was adapted from small molecules to biological macromolecules such as proteins [1].To form crystals, proteins are purified into highly homogenous solutions that are mixed with precipitants such as polyethylene glycol, 2-Methyl-2,4-pentanediol, and glycerol, which is then slowly saturated by vapour diffusion [2]. This is often a limiting step, particularly for membrane proteins (Section 1.b.). Further, unlike small molecule crystallography, obtaining the diffraction pattern resulting from x-ray exposure of crystals is insufficient for structure determination. This stems from the inability of x-ray detectors to measure phases of diffracted beams, which cannot be determined *de novo* due to the complexity of biomolecules and the similarity in electron density between atoms found in biological molecules. This 'Phase Problem' is solved by obtaining phases to calculate electron density within the repeating unit cell of the crystal, in one of three ways. Firstly, if there is a determined structure of a homologous protein, it can be used as a model to orient in the crystal unit cell from which approximate phases are derived, known as molecular replacement [3, 4]. Secondly, a derivative dataset can be obtained by soaking proteins crystals with heavy atoms, in a process known as isomorphous replacement [5]. Assuming crystal packing and space group remain unchanged, the higher electron density of heavy atoms will alter only contribution to reflection intensities. Since they bind to specific residues, the 'sub-structure' they represent can be located within the unit cell by comparing with the native diffraction pattern, and used to determine phases. Finally, x-rays can cause anomalous diffraction if their wavelength match an atom's electronic transition energy [6], allowing its sub-structure within the unit cell to be determined.

Traditionally, this requires selenomethionine labelling of proteins, but at synchrotron sources long wavelength radiation can be used to measure anomalous diffraction of sulfur in native proteins (Section 1.a.i.) [7].

The technological advances in cryo-electron microscopy (cryoEM) [8] allowed its adaptation from studying viral particle structures to becoming an increasingly popular method for macromolecular structure determination, particularly large proteins and macromolecular assemblies [9-11]. The key advance is in improving signal-to-noise ratio by sensitive direct electron detectors, and the electron beam is increasingly bright and coherent (such as by using cooler cathodes) [12]. In cryoEM, purified protein solutions are vitrified by flash freezing in liquid ethane, trapping molecules in amorphous ice. A transmission electron microscope is then used to generate 2D projections of single particles of these molecules. The projections are then classified *in silico* according to their conformational and orientational similarities, each of the resulting classes are used to generate an averaged projection called an Eigen image. Assuming that particles align with the support in random orientations, these Eigen images can be used to construct a 3D density map. The Eigen image is refined by further classification and averaging iteratively until a high-resolution density map can be produced, usually requiring thousands of single particle images [13]. cryoEM has allowed the structural characterisation of proteins and complexes that are unamenable to crystallisation. Vitrification also preserves proteins in a near-native state that is more representative than their crystalline states [14]. Currently, only large proteins of approximately 50 kDa are amenable to cryoEM [15], although the threshold is likely to decrease as technology advances and other innovations such as megabodies (Section 1.a.ii.) are developed for the technique. Additionally, structures determined by cryoEM are typically resolved at lower resolutions than by

crystallography, leading studies to take a hybrid approach where cryoEM density maps obtained of proteins or complexes are fitted with high resolution crystal structures of its constituent parts *in silico* [16].

Although Nuclear magnetic resonance (NMR) spectroscopy is often used as a complementary technique to crystallography and cryoEM [17-19], it can also be used for structural determination [20]. An electromagnet is used to align spins of odd-numbered nuclei such as ^{1}H, ^{13}C, and ^{15}N. Transition between the resulting parallel and antiparallel states relative to the applied magnetic field are induced by radio waves, from which the resonance effect is detected. The frequency of the resonance effect depends on the atom's chemical environment, allowing each signal to be assigned to a particular amino acid. These signals can be transferred between proximal nuclei in a macromolecule, by covalent linkage or proximity in real space, using two experiments. Firstly, triple resonance experiments are used to couple resonance signals between adjacent amino acids along the backbone, yielding a stretch of amino acids that is unique in the primary structure [21]. Additionally, their intensities can be used to derive torsion angles of their covalent linkage [22]. Secondly, signals are broadened by nuclear Overhauser effects that arise from spatially proximal nuclei, which can then be used to calculate a network of internuclear distances to derive the protein's 3D structure [23]. Unlike in crystallography or cryoEM, signals can be obtained from "disordered" regions that do not take up a fixed conformation, and the sample remains in solution state. However, biomolecular NMR suffers from two main disadvantages. Firstly, ^{13}C and ^{15}N labelling requires expression in minimal medium. This may result in low yield, and although it can be mitigated by cell-free expression [24] this remains challenging especially for membrane proteins (Section 1.a.iii.). Secondly, the resonance signal decreases as molecule size increases due

to slower tumbling times, restricting the methodology largely to proteins smaller than 25 kDa [25].

Protein structures are largely governed by the folding process that is intrinsically linked to the physicochemical properties of constituent amino acids. This led to the development of accurate computational tools for *de novo* prediction, such as secondary structure prediction, protein fold prediction, and non-covalent contact prediction [26]. The analysis of this relationship using experimentally derived structures has enabled *ab initio* predictions of whole protein structures. For instance, RoseTTAFold is an online server that uses an artificial neural network to relate information from the primary structure, to the a 2-dimensional distance map, and the resulting 3-dimensional coordinates to predict protein structures [27]. The accuracy of these techniques has recently been pushed by Alphabet's DeepMind, who developed an artificial neural network platform Alpha Fold trained using unpublished structures [28]. Through the CASP framework [29] for assessing prediction accuracy against experimentally derived structures, Alpha Fold was shown to successfully predict structures to 0.96 Å RMSD [30]. Although this is a promising avenue for bypassing time consuming laboratory work, it is currently unclear whether *in silico* predictions will replace experimental structural biology studies that require high resolution data to make precise hypotheses, such as our report on the RgNanOx complex structures (Sections 3.d.x., 3.d.xi.).

In this section, we highlight the innovations that allowed protein crystallography to remain accessible and relevant in light of other tools available in modern structural biology.

1.a.i. Experimental phasing by native sulfur SAD

X-ray crystallography operates on the basis that a crystal acts as a three-dimensional grating with spacings comparable to the wavelength of x-rays. Thus, when exposed to an x-ray source, a diffraction pattern arises from Bragg's law. X-ray diffraction arises from electrons in the molecule arrayed in the crystal unit cell, meaning electron density is related to the diffraction pattern by a Fourier Transform (Figure 1.1) [31].

$$p(xyz) = \frac{1}{V}\Sigma|F_{hkl}|e^{i\alpha_{hkl}}e^{(-2\pi ihx+ky+lz)}$$

Figure 1.1 Fourier transform *Relationship between the electron density (p) at one point in the unit cell (xyz) with the volume (V), and x-ray diffraction from Bragg planes of the crystal (hkl).*

The Fourier Transform shows that electron density is determined by two key factors from the summation of diffracted waves from plane (hkl) and their contribution to point (xyz); the amplitude ($|F_{hkl}|$) and the relative phases of each diffracted wave (α_{hkl}) [31]. The amplitude, which is proportional to the intensity detected in an experiment, is determined by the structure factor (F) (number of electrons along the diffracted plane). However, phases are independent of intensities and must be determined separately.

Phases only relate to one another between reflections and the resulting electron density [32]. Although this allows their *ab initio* calculation for small molecules, the required resolution of 1.2 Å [33] is unrealistic for protein crystals. Consequently, both anomalous diffraction and isomorphous replacement relies on obtaining phases and calculating the atomic model of a constituent sub-structure in the unit cell. This can then be used to calculate the phases for the remaining atoms using a Patterson function, which is a Fourier Transform of only intensities and generates relative positions of atoms as interatomic

vectors. Incidentally, Patterson functions are also used in molecular replacement to orientate a homologous protein with sequence identity of 25% or above [34] in the unit cell to obtain estimated phases [32].

Each atom has a scattering factor (f) that is dependent on its atomic number and hence number of electrons that contribute to x-ray diffraction. Therefore, x-ray diffraction by heavy atoms have distinctive peaks. The lack of heavy atoms in macromolecules is usually mitigated by isomorphous replacement, where co-crystallised
water molecules are replaced heavy atoms by soaking or co-crystallising the crystal in heavy atom solution [35]. This alters both the intensity and amplitude of the diffracted wave [32]. The difference in structure factors between the native crystal and the soaked crystals is then used to calculate the amplitudes of the heavy atoms, whose phases are determined using a Patterson function.

Anomalous diffraction utilises constructive interference when the wavelength of the incident x-ray is close to the absorption K edge of a sub-structure atom, where an inner shell electron is promoted. This results two further atomic structure factor terms, the dispersive f' and the absorptive f''. The latter is 90° out of phase with the normal atomic scattering factor f_0, leading to the breakdown of the Friedel's law: $|F_{hkl}| = |F_{-h-k-l}|$. The resulting Bijvoet difference between $|F_{hkl}|$ and $|F_{-h-k-l}|$ can then be used to calculate phases of the sub-structure [32].

Alas, where molecular replacement is unfeasible, the additional experimental manipulation in *ab initio* techniques can hinder structure determination. For instance, isomorphous replacement requires both the binding of heavy atoms to the protein, and

without significant alteration to the unit cell. Similarly, elements that diffract anomalously at conventional x-ray wavelengths of 1-2 Å such as selenium do not naturally exist in biomolecules. These elements can be incorporated into proteins by labelling, such as using selenomethionine media to grow cultures of cell lines auxotrophic for methionine. However, due to the indiscriminate replacement of sulfur atoms in all proteins, selenomethionine incorporation is cytotoxic and typically leads to lower expression levels [36].

For these reasons, the use of anomalous diffraction in low molecular weight atoms in native proteins, most notably sulfur and phosphorous, is indispensable for challenging systems. While conventional synchrotron sources have successfully used for native sulfur SAD at 2.1 Å [37], the detection of anomalous signal away from the absorption K edge at 5.0155 Å limits their application to high sulfur content proteins [38]. For crystal sizes lower than 200 μm, the increase in anomalous signal at longer wavelengths than 2.1 Å outweighs the inherent decrease in diffraction intensity [39].

Performing crystallography at longer wavelengths also presents its own problems, most notably x-ray attenuation by absorption and background scattering from non-crystal species and air, significantly reducing signal:noise ratio [39]. To mitigate this effect, an *in vacuo* beamline I23 was built at Diamond Light Source. The BL-1A beamline at KEK employs a similar approach by encasing the beamline in a helium chamber [40]. Published shortly before the completion of I23, a metanalysis showed conventional wavelengths are used in *de novo* native s-SAD for proteins of up to approximately 3% sulfur content [41]. According to published proteomic data [42] showing gaussian distribution of sulfur content between 12 Gram negative bacteria species, 3% corresponds to the steep portion (20^{th}-80^{th} percentile) (Figure 1.2). The first protein structured determined using long

wavelength sulfur phasing at I23 was in 2016, of the cyanobactin oxidase ThcOx resolved at 2.65 Å using a 3.10 Å [43]. It is noteworthy that although a higher resolution dataset at 1.54 Å was collected suing a 1.77 Å beam at I02, the anomalous signal was insufficient for structure determination [43]. With sulfur content of 0.13%, ThcOx exemplifies the repertoire of proteins now amenable to native sulfur phasing by long wavelength crystallography. The Wzx/nb10 complex, our system of interest in Chapter 2, has a sulfur content of 0.21% which is lower than the majority of protein sampled in the proteomic study [42].

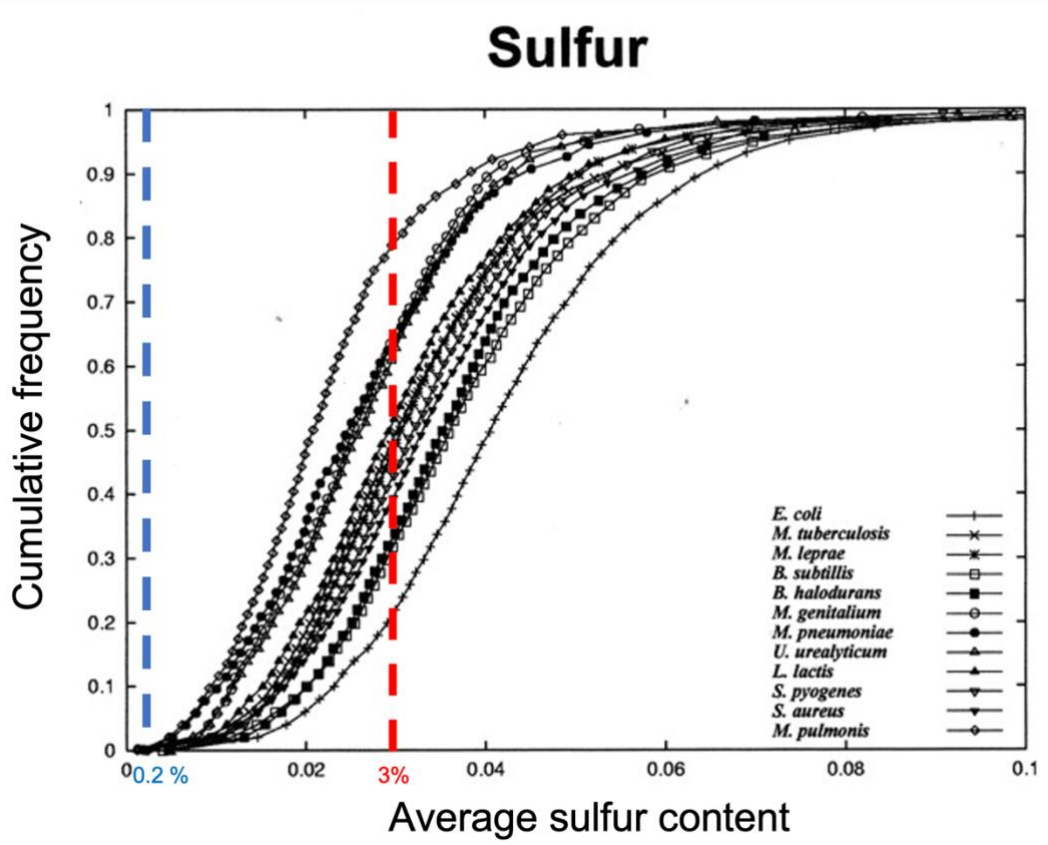


Figure 1.2 Average sulfur content of bacterial proteins *Modified from Baudouin-Cornu et al. 2004 [42], showing the cumulative frequency of protein sample plotted against sulfur content. Sulfur content threshold of 3% (structure determination achievable with conventional wavelength) and 0.2% (corresponding to Wzx/nb10 complex) are highlighted in red and blue respectively.*

1.a.ii. Antibody fragments and nanobodies

Specific binding of antibody derivatives were initially used in conjunction with fluorescent proteins for tracking [44, 45], or immobilised to use as affinity tag binders in chromatography [46, 47]. It is also used in structural studies, and extend far beyond solving structures of its complexes with their cognate antigen [48]. Small camelid antibodies can be raised against a specific antigen, such as a protein of interest, to aid crystallography and cryoEM. The antigen-binding fragment (F_{ab}) consisting of four variable domain subunits, or the heavy chain minimal epitope-binding subunit (nanobody) can then be cloned for heterologous expression (Figure 1.3). These can be raised to bind in a conformation-specific manner, which allows preparation of a homogenous sample of an unstable conformer for structure determination. For instance, the mechanism of the symporter LeuT was elucidated by crystalline structures of different conformations 'trapped' using antibody fragments, without the use of substrate for co-crystallisation [49].

The small size of nanobodies make them particularly useful for molecular biology owing to their solubility and compatibility with cloning. Most importantly, it allows access to small binding pockets such as to functionally important moieties, which are frequently flexible and hence impede structure determination. This is exemplified by the adenosine A_{2A} receptor structure published by García-Nafría et al [50]. There, the nanobody binds at the interface between the G_{α} protein and the $G_{\beta\gamma}$ subunits, which are effectors that dissociate upon receptor activation. Nanobodies can also be used for inhibitory roles, such as against prion oligomerisation of prion to derive its crystalline structure [51].

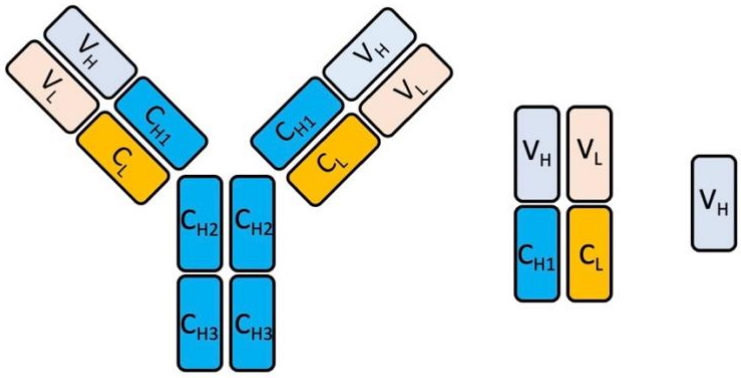


Figure 1.3 Domain organisation of antibody fragments and nanobodies *Diagram of an antibody (left), consisting of 2 heavy chains (blue) and 2 light chains (yellow), each made up of 4 and 2 domains (rectangles, labelled with domain names) respectively. The antibody fragment (middle) consists of four domains, two each from the light and heavy chains. The nanobody (right) is a single heavy chain domain. The antigen-binding variable domains (V_H and V_L) are coloured lightly.*

Conformational trapping can be subsequently used to understand structure-function relationships of proteins. For example, nanobodies were used to inhibit ATP-binding cassette (ABC) transporter ABCB1 and where conformational changes were observed using cryoEM, even at 15 Å [52]. Later on, ATPase activity assays were used to identify an inhibitory nanobody for another ABC transporter, PglK, which was used in co-crystallisation to obtain a high resolution structure [53] (Figure 1.4, Section 2.b.iv.). Compared to previous structures of a closed PglK [54], putative substrate-binding residues take up identical conformations and remain inaccessible. However, the cytoplasmic domains in the homodimer are much closer to each other, leading to the proposition that this is an intermediate, ADP-binding state. Surprisingly, only one nanobody is bound to the symmetrical homodimer, suggesting that the two independent ATPase domains are structurally coupled in the active state. Therefore, the use of nanobody in this study informed the understanding of the PglK mechanism.

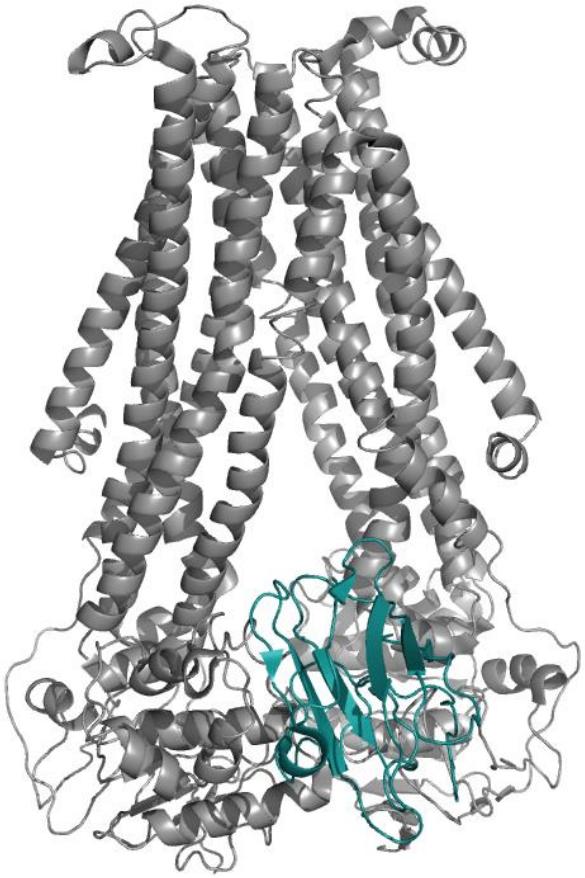


Figure 1.4 PglK co-crystallised with nanobody *Crystalline structure of PglK (grey), showing nanobody (cyan) that was co-crystallised with PglK (PDB code: 5NBD)*

For cryoEM where there is a molecular weight threshold of approximately 50 kDa [15] above which single particles are seen, specific binders are desirable to overcome the threshold for proteins of interests just below it. To this end, nanobodies can be fused with large rigid proteins, resulting in a 'megabody' [55]. The first lipid bilayer-reconstituted $GABA_A$ receptor (in nanodiscs) structure, for instance, was solved using cryoEM with a megabody [56].

Nanobody behaviour isn't simple, as they bind to target proteins with different affinities, stoichiometries, and inhibitory effects. Although their affinities for target proteins can be validated by physical methods such as isothermal calorimetry (ITC) and surface plasmon resonance (SPR), their suitability for structural characterisation can often only be determined at later stages by techniques such as SEC-MALS. However, recently Köhler et al developed a comprehensive screening strategy for nanobody screening using mass spectrometry (MS) [57]. MS is used to analyse fragments of constituting molecules of the sample, which are then separated according to their mass or charge. The resulting spectrum therefore provides information about the distribution of different species in the sample. However, traditionally in MS the detergent required to maintain solubility of membrane proteins contributes to background spectrometric signal, limiting the use of MS to studying soluble proteins. Cadene et al devised a sample preparation strategy for matrix-associated laser desorption ionisation (MALDI) MS that is compatible for membrane proteins without the need for buffer exchange and optimisation [58]. In essence, the matrix binds to proteins to form a thin film, reducing detergent concentration to just above its critical micelle formation. However, this heavily limits the choice to several non-ionic detergents, such as dodecyl maltoside. This non-disruptive technique, combined with the development of high mass detectors, now enables the biochemical characterisation of protein complexes such as glycosylation of ABC transporter Cdr1p and stoichiometry of cross-linked proteins [59]. The methodology was used to screen nanobodies by investigation the variation in mass spectra at different nanobody:protein ratios for each nanobody, to determine binding affinity and specificity[57]. The affinity for target proteins is determined by the minimum concentration needed to detect protein/nanobody complexes. Binding specificity, another crucial parameter for a homogenous sample, is determined by the distribution of complex peaks. Nonspecific

binding is indicated by broad signals, indicating many modes of binding as a result of nanobody saturation.

Interestingly, this methodology was applied to the ABC transporter PglK, which closes its lumen upon ATP binding [53]. The authors were therefore able to screen for nanobodies that bind to the lumen by identifying those that inhibit ATP binding, resulting in the structure determination of the occluded conformer that was hitherto elusive [53]. This methodology could be extended to other proteins to trap transient and unstable conformations, provided that there is functional information. Finally, spectra of samples incubated with two nanobodies can be used to identify those with competing binding sites, identified by the number of peaks. Again, this helps generating a homogenous sample, especially when multiple nanobodies are used to increase particle bulk for cryoEM.

1a.iii. Cell-free expression

Cell-free expression circumvents problems such as cytotoxicity and enzyme-mediated degradation in *in vivo* overexpression systems [60]. Preparative scale expression occurs in a vessel, where the required enzymes for target gene transcription and translation are either purified separately, or provided from cell extract. Macromolecules such as the DNA template, polymerases, and target proteins are retained in a dialysis membrane while nutrients are replenished via diffusion in a continuous feeding mechanism. Recently, cell-free expression protocols were adapted for *in vitro* expression of membrane proteins [61]. The translated membrane protein either precipitates in the aqueous solution, in which case it can be refolded, or directly reconstituted into detergent micelles. In theory, the target protein can alternatively be directly fed into a membrane-

like environment such as in liposomes, nanodiscs, or lipidic cubic phase. Unfortunately, this has largely been unsuccessful at preparative scale [61]. Nonetheless, the translocon SecYEG was shown to spontaneously reconstitute into nanodiscs *in vitro*, making the functional characterisation of the protein export/insert pathways possible [62]. This technique could one day be adapted to cell-free expression [63]. To date, only a handful of high resolution structures were yielded from cell-free expression [64-66] due to the limiting expression levels [66]. However, the open nature and versatility offered has already proven cell-free expression to be a promising avenue especially for challenging membrane proteins. For instance, it was used to generate the selenomethionine derivative required for structure determination of EmrE [64]. Traditionally, selenomethionine labelling requires expression of target proteins in an auxotroph, which may significantly reduce expression levels. Furthermore, despite precipitation of target protein in the absence of a reconstitution agent, it is possible to refold in lipidic cubic phase for crystallisation [66].

1.b. Membrane Proteins

Although the publication of the first membrane protein structure was over two decades ago [67], detailed molecular understanding of membrane proteins remains time consuming, technically challenging, and expensive. For instance, since the first high resolution GPCR structure in 2000 [68], eight years passed before the first active receptor structure emerged [69, 70], and a further three years passed before the first functional complex was solved [71]. A range of biochemical advances were made to overcome the problems in obtaining a high-quality sample needed for structural characterisation. Though alleviated by high throughput processes, the search for the right methodology is

largely empirical. Here, I review several of these biochemical advances in this area and how they are used to overcome these challenges in studying membrane proteins.

Due to the restriction in membrane volume, membrane proteins are expressed at inherently lower levels than soluble proteins. This is partially circumvented by development of cell lines with membranes of higher volumes [72], using techniques such as lipidic cubic phase (LCP) (Section 1.b.iii.) that require a lower amount of protein, and by growing large (12-24 L) batches of cell culture. The expressed protein can be extracted, reconstituted, then purified, allowing structural and functional characterisation. Protocols are devised empirically and is experiment dependent. For instance, in a study of the bacterial lipopolysaccharide synthesis pathway [73], the polymerase is reconstituted in a different detergent than one used in substrate synthesis. It is beyond scope to provide a comprehensive analysis of advantages and limitations of reconstitution methodologies, for which there is a systematic review by Skrzypek et al [74]. Heterologous overexpression is often required to produce sufficient quantities of protein for structural determination. The inevitable cytotoxicity that arise from overexpression of membrane proteins, caused by disruption of cellular integrity and inhibition of native membrane proteins, makes alternative methods such as cell-free expression (Section 1.a.iii.) highly valuable [75].

These challenges mean that some of structural biology's most important questions lie in the determination of membrane protein structures, such as Wzx (Chapter 2) and other enzymes involved in the lipopolysaccharide synthesis pathway. In this section, we summarise techniques employed to reconstitute membrane proteins for their structural and functional characterisation.

1.b.i. Detergent-based reconstitution systems

There is a wide range of commercially available detergents with different chemical properties (see systematic review by Seddon et al [76]). Screening methods were developed for high throughput selection of suitable detergent [77], as exemplified by their use in the landmark study of photosystem II [78]. Detergents differ in their capacity of micelle formation, measured by the critical micelle concentration (CMC, minimum concentration required for formation of single-layer detergent micelles). They also differ in micelle size and micelle stability at different temperatures and pH. Structures and CMCs of various commercially available detergents are shown in Table 1.1. Reconstitution with detergent can be monitored by turbidimetry [76], allowing determination of parameters such as CMC.

The characteristics of common detergents and their usage in crystallography have been reviewed by Stetsenko and Guskov [79]. To this date, detergent remains a popular choice for membrane protein extraction and reconstitution (Figure 1.5), although the detergent used for the two steps can be different. Common detergents used in membrane protein solubilisation and purification include DDM, Triton X-100, Cymal-5, and LMNG. [79] While strong detergency (ability to extract membrane proteins) is ideal for solubilisation, it can also destabilise the membrane protein. Ionic and zwitterionic detergents such as lauryldimethylamine-N-oxide (LDAO) typically have strong detergency. Therefore, milder non-ionic detergents such as *N*-Dodecyl β-D-maltoside (DDM) may be used in purification. For instance, Koehl et al [80] solubilised a heterotrimeric G protein using the ionic detergent sodium cholate, then performed on-column detergent exchange to DDM during affinity chromatography. Nonetheless, it is possible to complete a

preparation with only one mild detergent, the alternative respiratory complex III for instance was reconstituted in DDM for cryoEM [81]. It is also possible to purify membrane proteins in strong detergents, Pan et al [82] solved the structure of a voltage-gated sodium channel/regulator complex to 3.2 Å, which was purified in the ionic detergent sodium dodecyl sulfate (SDS). Detergent exchange could also be used for other reasons, such as improving sample quality. For example, Baradaran et al found that their sample formed more uniform ice when reconstituted in digitonin instead of DDM, the latter of which was used in purification [83].

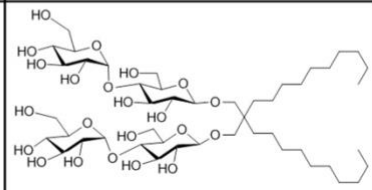
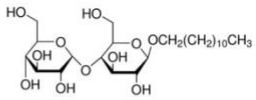
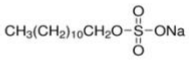
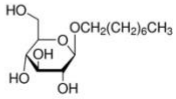
Detergent	Critical Micelle Concentration (CMC) / mM	Structure
Lauryl Maltose Neopentyl Glycol (LMNG)	0.01	
Dodecyl β-D-maltoside (DDM)	1.60	
Sodium dodecyl sulfate (SDS)	7.00 - 10.00	
Octyl β-D-glucopyranoside (OGP)	20.00 - 25.00	

Table 1.1 Commercially available detergents *Four commercially available detergents used for reconstitution of membrane proteins, showing a wide range of CMC and chemical structures*

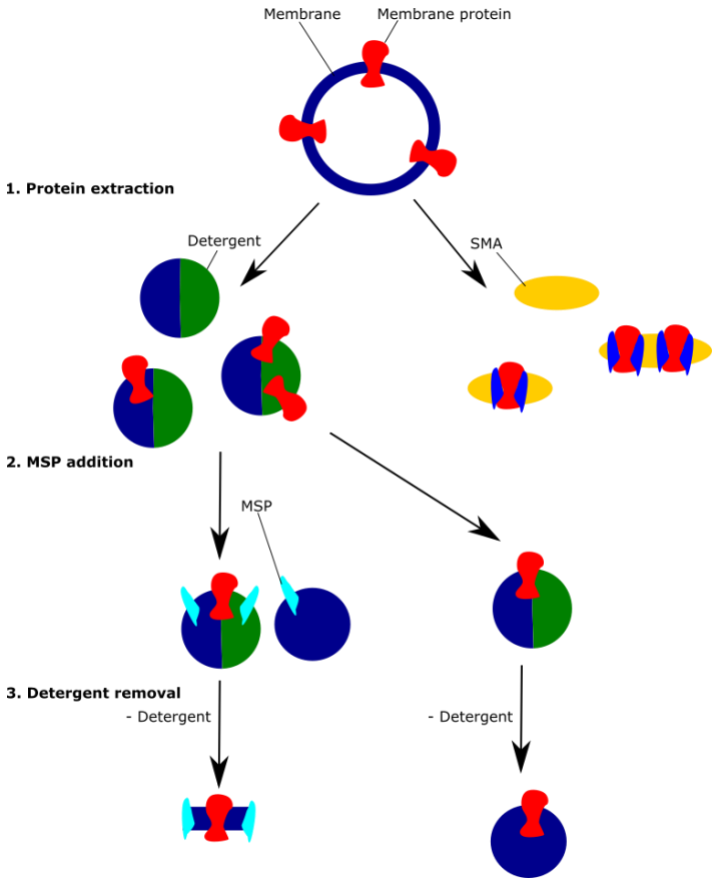
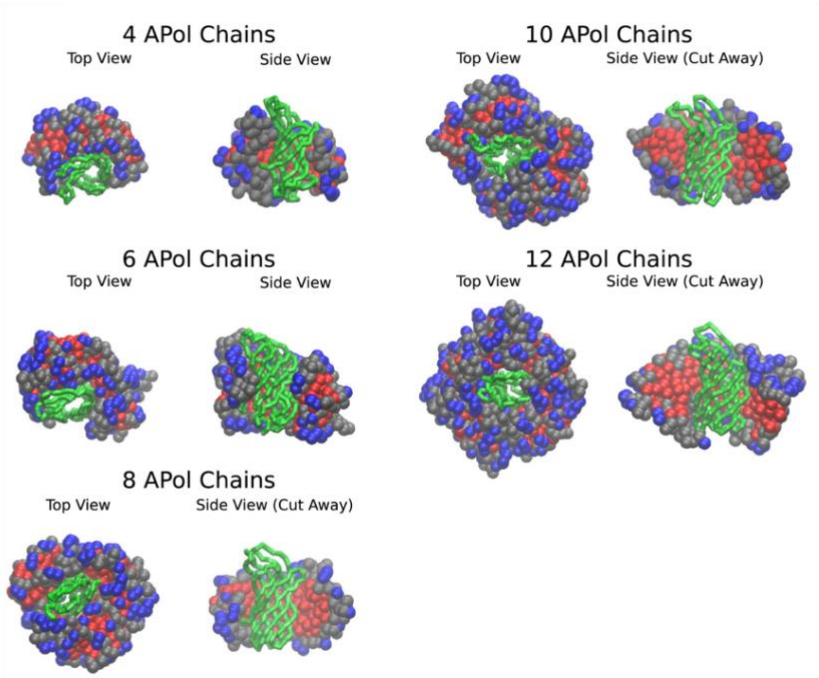


Figure 1.5 Diagram summarising membrane protein reconstitution methods *Detergents or SMA are used to extract proteins from the membrane, both of which can be used as reconstitution agents. From mixed-detergent micelles (green/blue circles), protein can be reconstituted in liposomes by detergent removal, or into nanodiscs by addition of MSP before detergent removal.*

As an alternative to exchanging to a milder detergent during purification, membrane proteins can be exchanged into polymeric detergents known as amphipathic polymers (amphipols) [84, 85]. The polymer backbone is made up of hydrophilic carboxylate groups, which interact with the solvent; with hydrophobic alkyl side chains attached, which interact with hydrophobic portions of the protein. Solvent accessibility by NMR [86] and molecular dynamics simulations [87] have shown enriched binding of amphipols at hydrophobic β strands. Although not strong enough for protein extraction, amphipols are proposed to be more potent in protein stabilisation by restricting conformation by its multiple protein binding sites per polymer molecule [88]. This hypothesis was supported

by various studies of the family of β-barrel proteins Omp. First, refolding of unfolded OmpA is inhibited kinetically, but not thermodynamically, by addition of amphipol A8-35 [89]. The sum of binding energies of intramolecular interactions in folded OmpA was proposed to be greater than that of amphipol-bound unfolded OmpA, making the former more stable [89]. However, amphipol binding makes refolding slower, presumably due to conformational restriction. (Conversely, thermodynamic inhibition would mean amphipol binding has a higher affinity than refolding.) Secondly, later NMR studies identified multiple specific interactions between A8-35 and OmpX [90]. Finally, in a molecular dynamics study of OmpX, a statistically significant increase in conformational restriction was observed [87] upon addition of amphipol molecules. More importantly, this effect is only observed up to 8 chains, at which point the hydrophobic belt is simulated to be fully surrounded (Figure 1.6). Further studies would considerably support this hypothesis if an inhibitory effect is observed to arise, or plateau, at 8 molecules as the number of amphipol molecules increase.

a


4 APol Chains
Top View
Side View
10 APol Chains
Top View
Side View (Cut Away)
6 APol Chains
Top View
Side View
12 APol Chains
Top View
Side View (Cut Away)
8 APol Chains
Top View
Side View (Cut Away)

b

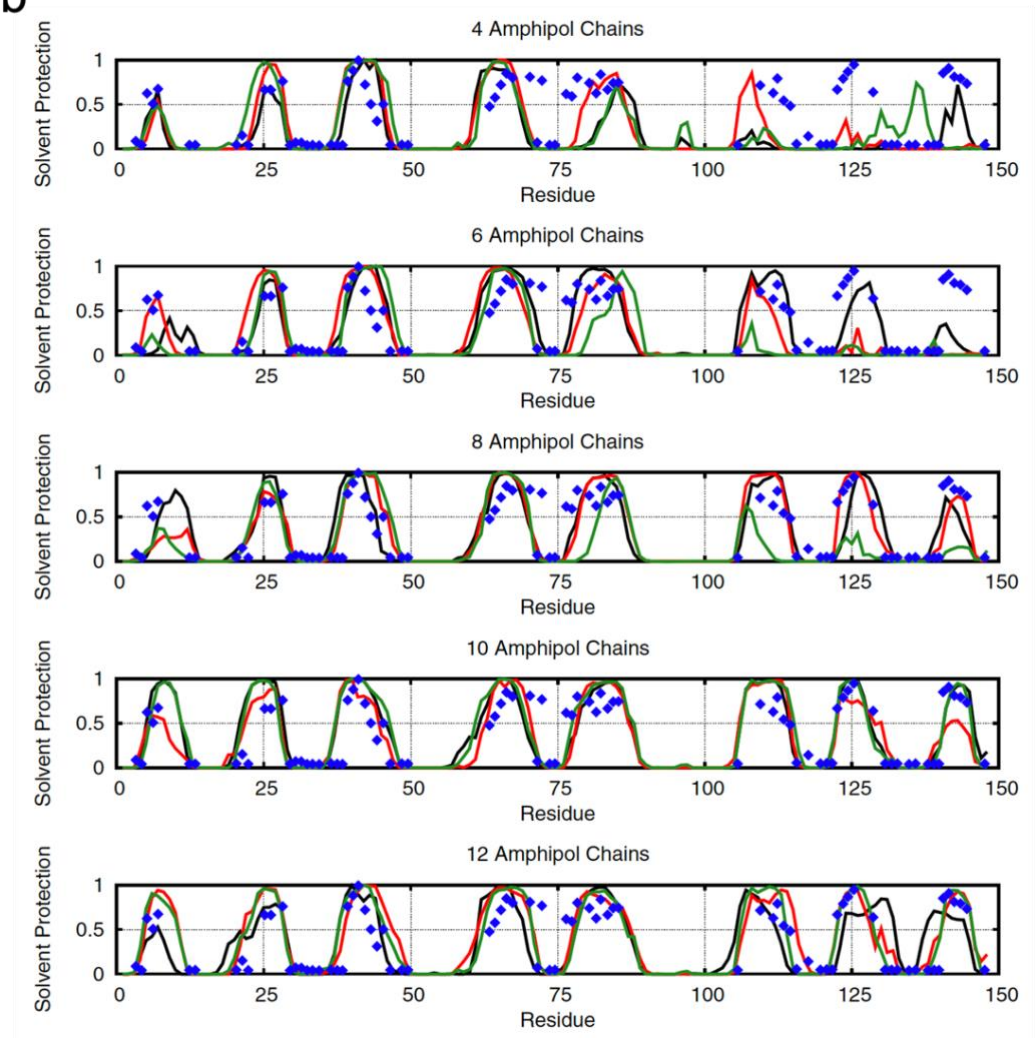

4 Amphipol Chains
6 Amphipol Chains
8 Amphipol Chains
10 Amphipol Chains
12 Amphipol Chains
Solvent Protection
Residue
0
0.5
1
0
25
50
75
100
125
150

Figure 1.6 ***Molecular dynamics simulation of amphipol binding*** Extracted from Perlmutter et al 2014 [87]. (a) Simulation of amphipol (spheres) binding bacterial outer membrane protein OmpX (ribbon), showing that 8 polymer molecules are needed to surround hydrophobic surfaces of OmpX. (b) Solvent exposure of OmpX residues from 3 independent simulations (red, green, black lines), showing higher solvent protection upon binding of 8 amphipol molecules.

1.b.ii. Lipid-based reconstitution systems

Although detergents and amphipols are widely used in structural characterisation of membrane proteins, they are poor mimics of the biological environments [74]. Therefore, it is often unclear whether the conformational and functional states of detergent-solubilised proteins are reflective of that in its native environment. Conversely, lipids can be extracted from membranes and incorporated into membrane protein-containing detergent micelles, or in lipid-based reconstitution systems (Section 1.b.ii.). Such lipids are thought to have a structural effect, as they often appear in crystalline structures of membrane proteins (Figure 1.7). The crystalline structure of melibiose permease MelB showed significant number of kinked alpha helices in the transmembrane domain, leading to the proposition that it undergoes large scale conformational changes during transport [91]. Unsurprisingly, changes in parameters such as increasing hydrophobic chain length inhibits MelB [92], specific high affinity binding between endogenous phospholipids and MelB were subsequently observed by NMR [93].

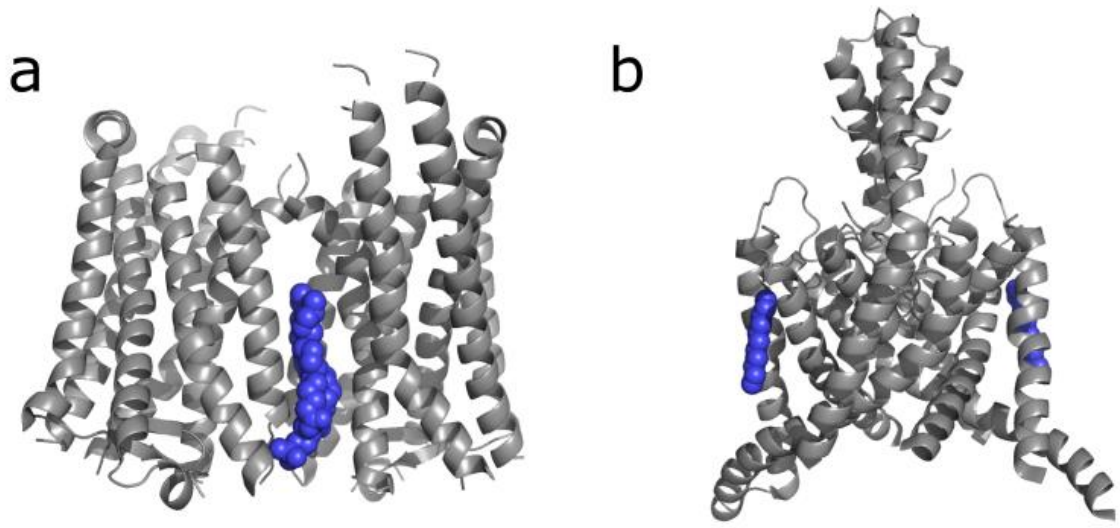


Figure 1.7 Bound lipids are observed in crystalline structures *Crystalline structures of (a) prostaglandin E receptor EP3 (PDB: 6AK3) and (b) potassium ion channel TREK2 (PDB: 4BW5) shown in grey. There is electron density detected for lipids (blue spheres), of POPC and PC respectively*

Lipids such as cholesteryl hemisuccinate (CHS) and phosphatidylcholine (PC) are often used in conjunction with mild detergents like digitonin and DDM during membrane protein purification [94-96]. The reconstitution of membrane proteins into liposomes (Figure 1.5) instead of mixed micelles, is then made possible by detergent removal (or dilution below CMC) [97-99]. This allows reconstitution of membrane proteins in a near-native environment, where phospholipids are extracted from cells of the expression organism (Figure 1.8). Additionally, unlike in micelles, the liposome bilayer exposes both extra-membrane domains of integral membrane proteins (in the case of plasma membrane proteins, the extracellular and intracellular domains), making functional characterisation more robust.

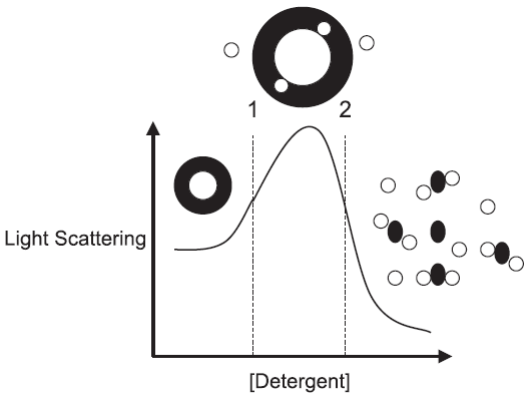


Figure 1.8 Detergent extraction of liposome lipids *Extracted from Seddon et al [76], showing extraction of phospholipid liposome (black circle) using detergent (white molecules). Upon mixing with protein samples, the process can be reversed by detergent removal, leading to protein reconstitution in liposomes.*

Commercially developed polar lipids are typically used to form liposomes (Figure 1.5). In Elferink et al.'s landmark study, thermostable archaeal lipids were shown to form liposomes and were used to reconstitute functional cytochrome c-oxidase [100]. Interestingly, they found that the less polar fractions are unable to form liposomes unless mixed with *E. coli* phospholipids, which are more polar [100]. However, proteoliposome samples can be heterogeneous with regards to size and composition. This may cause further problems in structural characterisation. Therefore, biochemical dissection of proteoliposomes remains challenging without further chemical advances. Nonetheless, embedding proteins in a membrane facilitates growth of two-dimensional crystals in the membrane plane, allowing direct structural characterisation by electron crystallography, reviewed in detail by Stahlberg et al [101]. Wang et al [102] utilised this phenomenon to grow 2D crystals of streptavidin, a tag that binds to biotin, onto lipid monolayers that are laid on EM grids. This allows specific tethering of proteoliposomes that contains biotinylated phospholipids. In addition to the increased proteoliposome homogeneity, this

methodology only requires a relatively small amount of protein, circumventing the challenge of expression.

Recombinant DNA technology and protein engineering allowed adaptation of naturally occurring proteins for reconstitution purposes. Nanodiscs containing a phospholipid layer can be made by mixing the desired lipids with membrane scaffold proteins (MSPs), which derived from helical sections of apoliprotein A-I [63]. These helices are amphipathic, allowing it to secure a homogenous bilayer in aqueous solution. Although nanodiscs are too small for crystallisation of reconstituted proteins (Figure 1.5), they can be transferred into lipidic cubic phase for crystallisation [64] and are highly homogenous, making them suitable for other methods of characterisation such as cryoEM. It has also been demonstrated that nanodiscs can be used to generate samples of sufficient quality for investigating protein function by NMR [65]. Crucially, since a polypeptide is used to secure phospholipids, this method can be optimised by genetic engineering. For instance, it was shown that by introducing repeats of helical sections in MSPs, nanodiscs as large as 17 nm can be generated [66], sufficient for reconstitution of membrane protein complexes. Sometimes, the mass of MSPs can be too similar to that of phospholipids to be for characterisation of protein complexes using mass spectrometry, although mutant MSPs with additional amino acids have been shown to circumvent this problem [67].

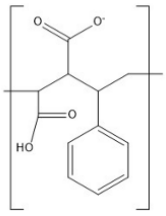


Figure 1.9 Chemical structure of the SMA repeat unit

Similarly to nanodiscs, styrene/maleic acid (SMA) polymers (Figure 1.9) can be used for direct extraction of native lipids and constituent membrane proteins (Figure 1.5), without the need of a detergent. Using NMR and isothermal titration calorimetry (ITC), Vargas et al have determined that formation of SMA-lipid particles (SMALPs) is thermodynamically more favourable than solubilisation by detergents [103]. Recently, cytochrome c oxidase in SMALPs were re-reconstituted into exogenous liposomes without the use of detergents, and were capable of forming electrochemical gradients [104]. The combination of strong 'detergency' and reconstitution in native lipids makes this methodology a promising avenue. However, it is currently limited by the insolubility of SMALPs in acidic conditions [105], under which some samples may be purified optimally.

1.b.iii. Lipidic Cubic Phase

In micelles and liposomes, shielding of hydrophobic surfaces by detergents and/or lipids limits nucleation and crystal packing, resulting in smaller crystals (Figure 1.10). The lipidic cubic phase (LCP) method was developed to improve crystal formation of membrane proteins. In essence, LCP is a biological membrane-like environment in which membrane proteins spontaneously concentrate, thereby reducing the molar amount required for crystallisation. LCP is a liquid-crystalline state with water channels, and forms when detergent-reconstituted protein is added to monoacylglycerol. This causes spontaneous reconstitution of the protein into the LCP, which is then dispensed into precipitant solutions. Based on phase diagrams, the Caffrey model [74] (Figure 1.11) proposes that monoacylglycerols are incorporated into micelles and subsequently into the LCP near the aqueous/lipidic interface. Dehydrating effects of precipitant solutions, which are added onto the LCP, causes local phase changes to the lamellar phase where

proteins may form polar contacts. Since the two local phases are continuous, nucleation is propagated by the LCP which acts as a reservoir of reconstituted proteins. It was proposed that LCP facilitates a slower rate of crystal growth and hence form more stable crystals [74]. Additionally, considerable efforts were made to make LCP compatible with other crystallographic advances, such as high throughput crystallisation by robots [75] and in situ serial crystallography [76]. The wide-ranging suitability of LCP for macromolecular crystallography is yet to be fully discovered, and recently found to be also compatible with soluble proteins [77]. Furthermore, different monoacylglycerols form LCP of different physicochemical characteristics. Recently, selenomethionine crystals of 3-phosphate acyltransferase were found to diffract to ideal resolutions only when reconstituted in 7.8monoacylglycerol than 9.9 monoacylglycerol, but the reverse is true of native crystals [78]. This methodology was used to generate bacteriorhodopsin crystals that did not form using micellar systems under the same conditions [71]. At the time of writing, 546 Protein Databank (PDB) structures were searchable under the term "lipidic cubic phase" [72]. Further details of the successes was summarised in a review of the methodology [73].

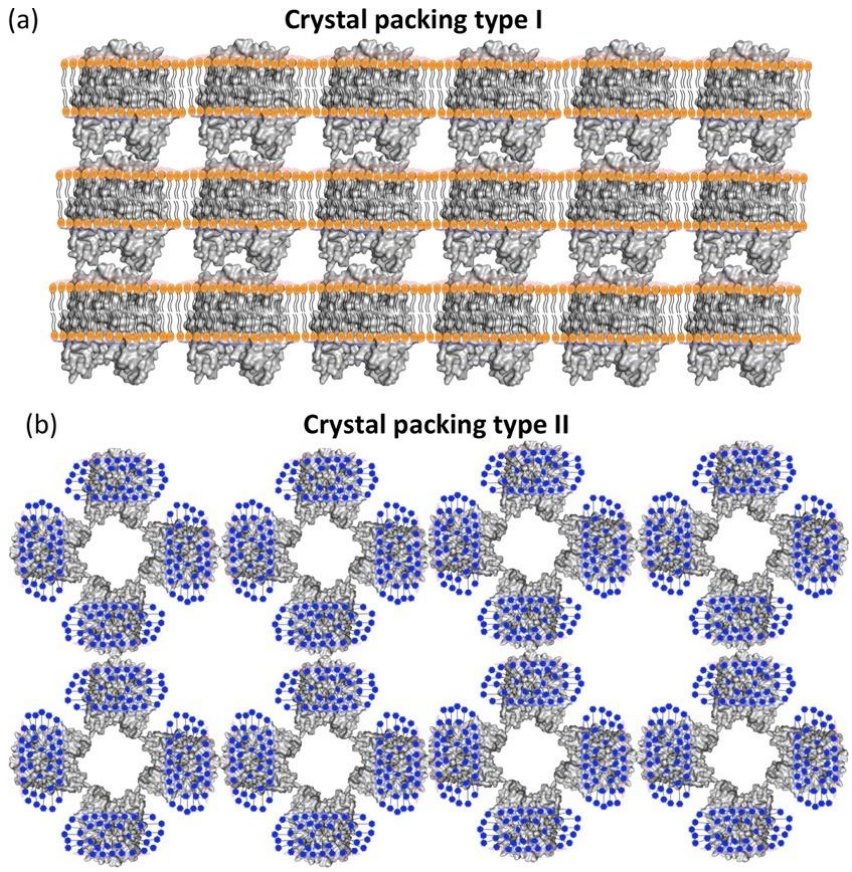


Figure 1.10 Membrane protein crystal packing *Extracted from Birch et al [106]*

While LCP has produced high-quality crystals of many challenging proteins, developments in *in situ* crystallography address difficulties with small crystals [107]. By using a bright and narrow beam, such as in microfocus beamlines [108] and with X-ray Free Electron Lasers [109], diffraction data of a large number of crystals (in the condition they are crystallised in) can be combined to generate a complete dataset. This in situ methodology, called serial crystallography, has already led to several membrane protein structures [110-112].

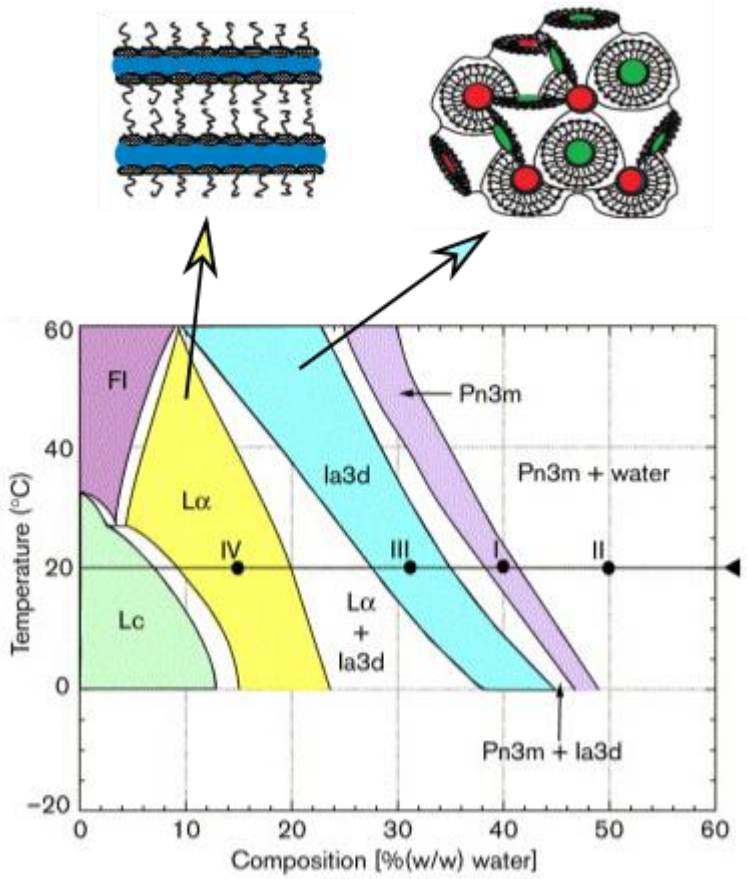


Figure 1.11 LCP forms at 20 °C at 33.3% water *Adapted from Caffrey 2000 [74], diagrammatic representations of lamellar liquid crystals (top left) and cubic mesophase (top right). Phase diagram of monoacylglycerol/water mixture is shown at the bottom.*

In situ crystallography is particularly helpful for LCP crystals for two reasons. First, the small size of LCP crystals make them difficult to harvest. Secondly, water evaporation during harvesting may destabilise the LCP (Figure 1.11) and is potentially detrimental to crystal stability. Recently, a fully *in situ* setup (Figure 1.12) is made possible by the development of specialised crystallisation plates called thin film sandwiches or *in meso in situ* crystallisation (IMISX) plates [113]. In the plate (Figure 1.13), protein is crystallised in between two films of cyclic olefin copolymer (COC), a material known for its low background x-ray scattering and biological inertness [114, 115]. This is further sandwiched by removable glass plates, to reduce water permeability. Furthermore, individual wells are removable for full 360° data collection, which is hitherto impossible

in SBS *in situ* plates. The added versatility is welcome especially for membrane proteins which may not express to sufficient levels for serial crystallography. This method has been verified to be compatible with serial crystallography of integral membrane proteins, even when phased with native sulfur SAD [116, 117]. Development of crystallographic devices remains an active field. For instance, graphene was recently incorporated in on-chip crystallography [115], where a single layer is as effective a water barrier as 100 μm of COC.

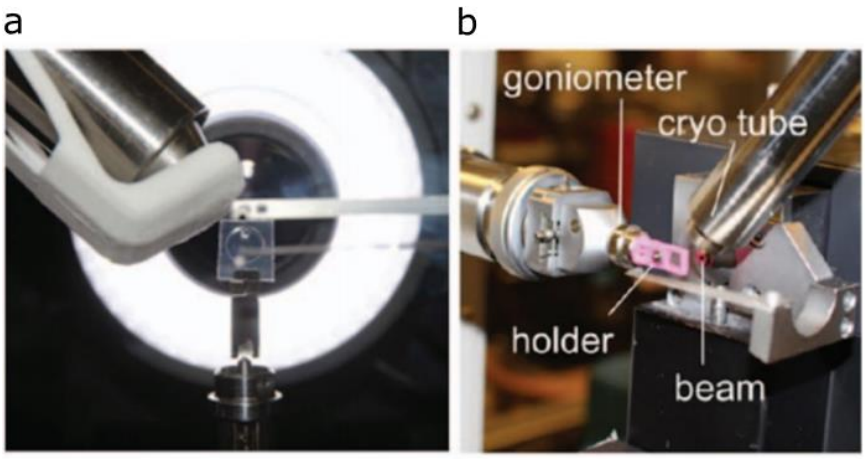


Figure 1.12 Beamline setup for IMISX *Extracted from Martiel et al 2018 [84], showing a) an individual removable well mounted on a goniometer and b) the wider setup with labelled components*

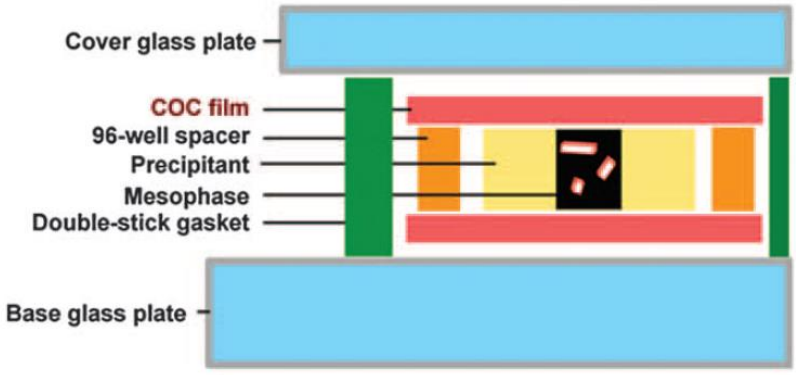


Figure 1.13 Diagram of IMISX plate *Extracted from Martiel et al 2018 [84]. The mesophase bolus/precipitant mixture is encased between two COC films, which are themselves attached to glass plates. Individual wells are removable upon detachment of glass plates.*

2. Structure determination of LPS flippase Wzx by native sulfur SAD

2.a. Abstract

The structure of Wzx_E from *E. coli* K:12 was determined by protein crystallography to a resolution of 2.3 Å. This was a challenging project that was pursued for five years prior. Previously, it was discerned that crystallisation of Wzx required a combination of detergent solubilisation, reconstitution in lipidic cubic phase (LCP) and nanobody binding. We conducted preliminary Wzx binding and epitope screening studies for eight nanobodies. We discovered that the growth of high quality Wzx crystals suitable for structure determination required addition of exogenous lipids.

It was not possible to determine the phases of the Wzx crystals by molecular placement, despite the availability of a model. The yield of selenomethionine variant protein was too low to be useful. Ultimately, Wzx required native sulfur single anomalous diffraction (s-SAD) using the long wavelength beamline I23 at Diamond Light Source. Solution of the Wzx structure showed that the I23 beamline, where x-ray diffraction is lower and background scattering is higher than conventional macromolecular crystallography beamlines, was compatible with delicate LCP crystals. We also conducted preliminary substrate co-crystallisation trials and investigated the use of *in meso in situ* crystal (IMISX) plates to facilitate further trials.

2.b. Introduction

2.b.i. Lipopolysaccharide (LPS) biosynthesis in Gram negative bacteria

Oligosaccharide synthesis and transport in Gram-negative bacteria are coordinated by overlapping pathways of soluble and membrane-bound enzymes. Being components of the peptidoglycan cell wall, lipopolysaccharide layer, and extracellular capsules, they serve biologically important functions with immense structural diversity [118].

There are two very important systems, the lipopolysaccharide (LPS) and the extracellular capsular polysaccharide (CPS) pathways (Figure 2.1). Overlap between these pathways means that there are often functional homologues of the enzymes involved. For instance, both utilise the carbohydrate polymer and therefore share the Wzx and Wzy enzymes. Their chain lengths however are regulated by functional homologues Wzz and Wzc.

LPS are antigenic molecules displayed on the outer membrane via a lipid anchor in Gram negative bacteria. Its serotype-determining moiety, the O-antigen, is synthesised in the cytoplasm by adding sugars to undecaprenol pyrophosphate (UndPP) via glycosyltransferases. This anchors the substrate to the inner membrane and is then flipped by scramblase Wzx. In the periplasm, the repeat unit is polymerised by Wzy, with chain length regulated by Wzz [73]. The polymer is then conjugated to the core oligosaccharide and lipid A anchor by the periplasmic ligase WaaL [119]. The completed LPS molecule is then transported from the inner membrane to the outer membrane via a multi-complex formed by the Lpt family [120], including the periplasmic shuttle LptA [121], and powered by its coupled ATPase LptB [122].

Unlike LPS, CPS surrounds the bacterial cell as a protective layer in the bacterial-host/-environment interface [123]. Therefore, the two pathways diverge after polymerisation by Wzy. Chain length is regulated by Wzc [124], a protein tyrosine autokinase whose structure has recently been determined [125]. Completion of the Wzc catalytic cycle relies

on the reversal of autophosphorylation by protein tyrosine phosphatase Wzb, whose structure has also been determined [126, 127]. The polymer product is then exported through the outer membrane translocon Wza [128]. Finally, the polysaccharide molecule is fixed onto the outer membrane via association with the β-barrel lectin Wzi [129].

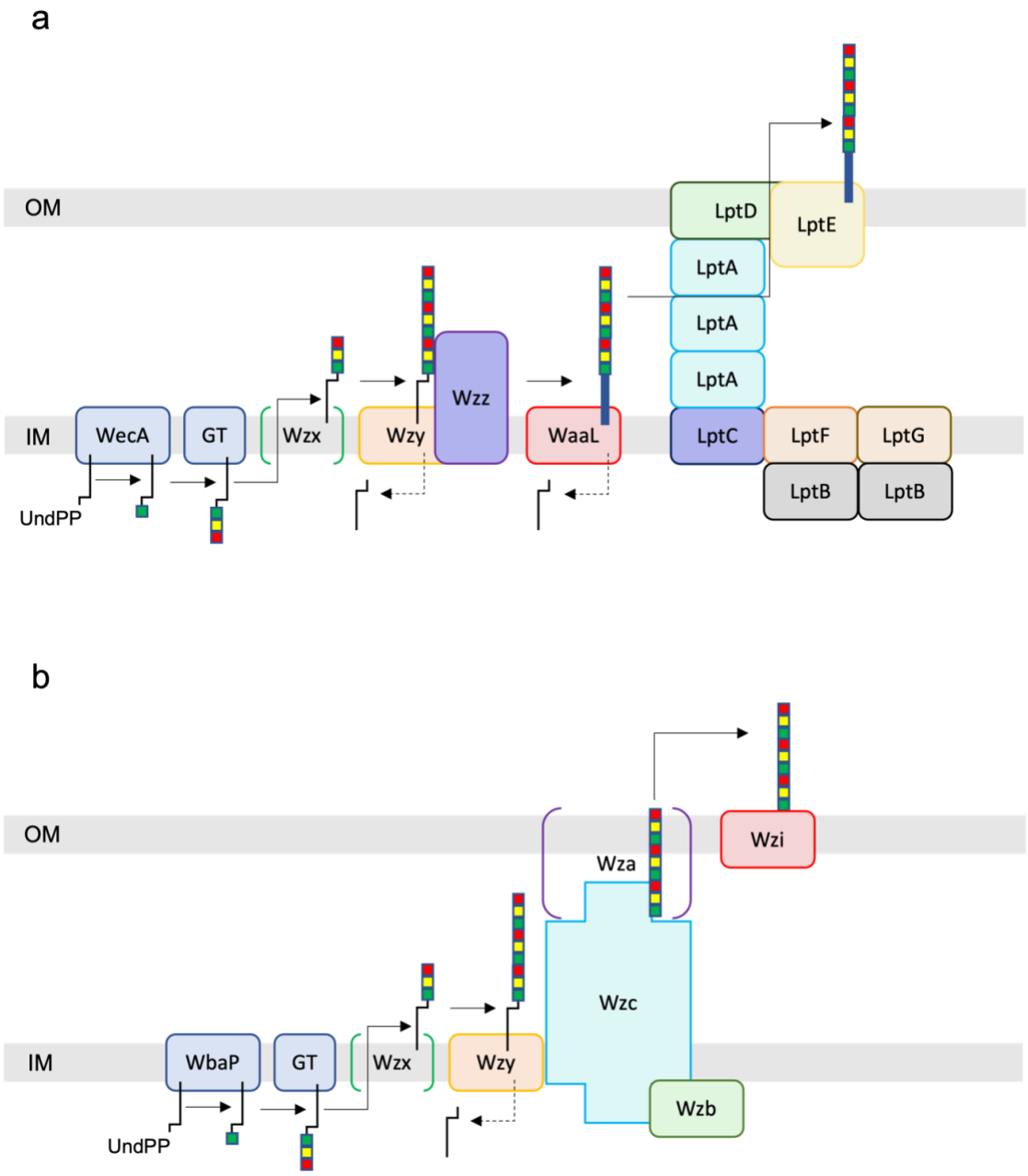


Figure 2.1 The LPS and CPS pathways *Diagrammatic representation of the synthetic pathways of LPS (a) and CPS (b). Outer membrane (OM) and Inner membrane (IM) are shown as grey bars. Both LPS and CPS pathways begin with the addition of an initiating sugar (green square) by WecA and WbaP respectively. Additional sugars (yellow and red squares) are then added to the initial sugar by a series of*

glycotransferases (GT). Both pathways use Wzx and Wzy, although Wzy is associated with different chain length regulators Wzz and Wzc/Wzb. LPS is conjugated with a lipid anchor via WaaL, then translocated to the outer membrane via the Lpt family of proteins. CPS is translated as a polysaccharide through Wza and anchored on the outer membrane via Wzi.

2.b.ii. Wzx, inner membrane scramblase in the LPS biosynthesis pathway

The flippase activity of Wzx was first identified using enzyme-linked immunosorbent assay (ELISA) assay on inverted membrane vesicles [130]. In ΔWzx mutants, UndPP-linked O-antigen was found to accumulate on the exterior of inverted membrane vesicles, when compared to when Wzx was absent. This indicates that Wzx can flip its substrate from the cytosolic to periplasmic side *in vivo* [130]. The inverted membrane vesicle system was used in further studies showing Wzx_E catalytic activity using a radio-labelling assay [131]. GlcNAc-PP-nerol, a soluble analogue of the cognate substrate with a shorter lipid chain, was labelled with tritium. Transport of radiolabel form solution into the vesicle was measured, and found to be linearly correlated with protein concentration [131]. Deletion of Wzx_E abolished transport, unlike in WT [131]. Subsequent assays in reconstituted systems showed that different Wzx subtypes are able to cross-complement deletion mutants of each other, provided that the initiating sugars (proximal to pyrophosphate) are identical. For instance, both Wzx_{O7} and Wzx_{O16} have substrates that contain GlcNAc as the initiating sugar and are able to cross complement, suggesting a prominent role of the initiating sugar in substrate recognition [132].

Wzx_E, the enzyme characterised in this study, is part of the enterobacteria common antigen (ECA) pathway. Although ECA can trigger immune response [133], it also provides other functions such as antibiotic resistance [134], maintenance of cell

membrane integrity [135], and enhancement of virulence by host effector resistance and immune suppression [136]. The common antigen repeating unit, GlcNAc-ManNAc-Fuc4NAc [137], is synthesised in different forms. It can be found on the outer membrane by conjugation to lipid A as an LPS (ECA_{LPS}) or conjugation to phospholipids (ECA_{PG}) [138]. It is also found in the periplasm as a cyclic compound (ECA_{CYC}) [138]. The substrate of Wzx_E is predicted to be UndPP-GlcNAc-ManNAc-Fuc4NAc (also known as Lipid III, Figure 2.2a) [137], which is generated from UndPP by successive glycosyltransferases WecA, WecG, and WecF [138]. Following downstream processing via the Wzy/Wzz pathway (Figure 2.1a), the ECA_{LPS} is formed. Since ECA and O-16 antigen share GlcNAc as the initiating sugar, Wzx_E can complement ΔWzx_{O16} [139]. Interestingly, this complementation only exhibits WT phenotype if the Wzy_E and Wzz_E are not expressed, indicating the three enzymes work in tandem that favours cognate counterparts [139]. Furthermore, ΔWzx_{O16} can be complemented with the ATP-dependent directional flippase PglK, the pathway of which does not contain periplasmic polymerases as it involves homopolymeric O-antigens produced in the cytoplasm [140].

Using a combination of sequence comparison and topology mapping, Islam et al reported a homology model of Wzx that retains similar surface electrostatic potential and cationic lumen as NorM [141], a sodium/drug antiporter. Although different complex structures of NorM were solved [142], only periplasm-facing conformations were captured (Figure 2.3a). Indeed, reconstituted Wzx exhibits anionic flux, leading Islam et al to propose an alternative mechanism as a proton-dependent antiporter [143]. However, due to the lack of direct observation of enzyme activity, it is unclear whether Wzx exhibits directional flippase activity or bi-directional activity (Section 2.b.iii.). Directional flippases overcome the cytosolic/periplasmic equilibrium using external energy sources such as

ATPase activity or membrane potential. Alternatively, bidirectional flippases rely on the equilibiraic shift from downstream processes to drive continual activity, which in the case of Wzx would be polymerisation by Wzy.

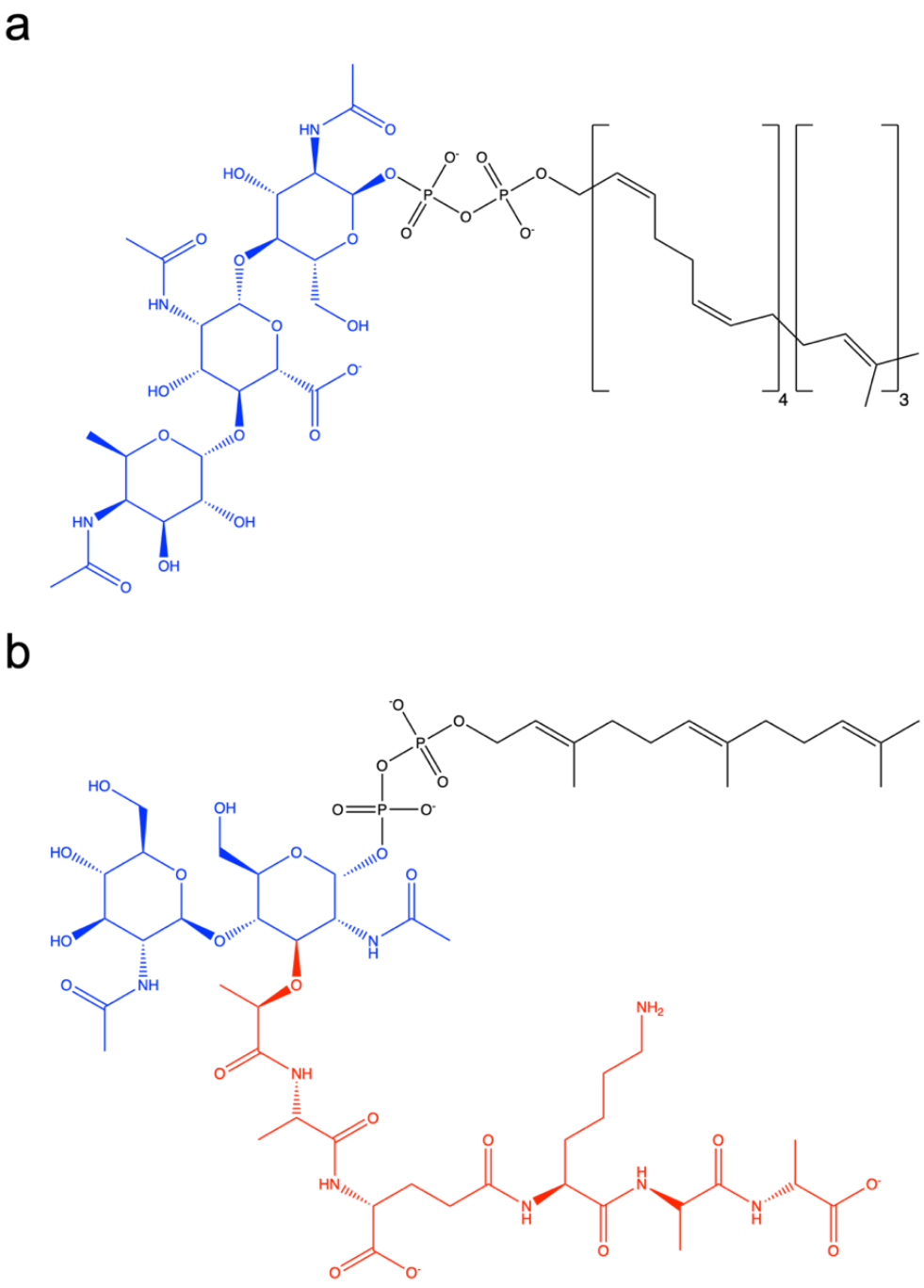


Figure 2.2. Lipid III and Lipid II *Chemical structures of Lipid III (a) and Lipid II (b). Polysaccharide groups are coloured in blue, peptide groups are coloured in red*

2.b.iii. Functional characterisation of flippases

The general term flippase is an enzyme activity term used to describe enzymes that transport lipid and lipid-linked substrates from one leaflet of a membrane to another. There are three broad classes: (i) flippases, which transport substrates into the cytoplasmic face, (ii) floppases, which transport substrates away from the cytoplasmic face, and (iii) scramblases or flip-floppases, which are bidirectional [144]. The first two allows establishment of bilayer asymmetry in biological membranes, whilst the latter catalyse the equilibration of phospholipids that would otherwise be too slow via Brownian motion or vesicular transport [145]. The lack of ATP consumption means the mechanism of scramblases are distinct from those of flippases and floppases.

Alternating access and pore are the two main mechanisms proposed for scramblases [146]. The alternating access mechanism, proposed for glycerol-3-phosphate transporter, involves a conformational change induced by substrate binding that opens the lumen to the other side of the membrane [147]. The pore model involves an aqueous lumen that binds polar head groups of lipids while the acyl chains undergoes diffusive translocation within the membrane [148].

Translocation of large substrates such as that of Wzx presents two additional challenges [146]. Firstly, the undecaprenol lipid chain adopts a non-lamellar structure to remain embedded in the membrane bilayer [149], necessitating specific binding grooves [146]. Secondly, the large polar headgroup requires specific binding partners such as arginine clusters (Section 2.b.iv.).

2.b.iv. Structural characterisation of flippases

Flippases and floppases catalyse directional transportation by consuming ATP. Of particular note are ATP-Binding Cassette (ABC) transporters, which have a wide range of biological roles. Structures of constituents solved thus far include Sav1866 multidrug floppase [150], lipid flippase MsbA [151], and the aforementioned LPS floppase PglK which is involved in protein N-linked glycosylation [152]. Although no structures have hitherto been reported of the polysaccharide transport (PST) family, to which Wzx belongs, there are structures of homologous flippases MurJ and PglK in the multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) superfamily [153].

MurJ catalyses the flipping of Lipid II in the inner membrane, a key component in the peptidoglycan synthesis pathway. Lipid II is a lipid-linked disaccharide pentapeptide, which like Lipid III contains sugar-linked UndPP (Figure 2.1b). The peptide moiety is synthesised by class A penicillin binding proteins (aPBP), which was used to develop a biotin-labelling assay that found MurJ activity relies on membrane potential [154]. The first MurJ structure solved was of an inward (cytoplasmic) facing conformation, which revealed a cationic cavity for headgroup binding and a hydrophobic groove for the hydrocarbon chain [155]. Its mechanism was proposed following the determination of further inward facing structures and an outward facing structure (Figure 2.3b) [156]. Consistent with the MOP family's canonical membrane potential-dependent activity, Na^+ binding was shown to cause conformational changes in the cytoplasmic moiety, and a 30° of rotation in transmembrane helices to expose the cavity towards the periplasm, although the cavity is too small to bind the headgroup causing its release into the periplasm [156].

PglK catalyses the flipping of GlcGalNAc5Bac-PP-undecaprenyl whilst exhibiting ATPase activity, and was determined in an inward-facing apo-state and an ADP-bound holo-state [152]. The structures show an unusual external (non-transmembrane) helix, the deletion of which caused abolition of *in vitro* activity, which forms hydrophobic grooves with the transmembrane domain [152]. The authors were able to identify an arginine cluster that are accessible only in the outward-facing state that follows ATPase activity in the catalytic cycle [152] (Figure 2.3c).

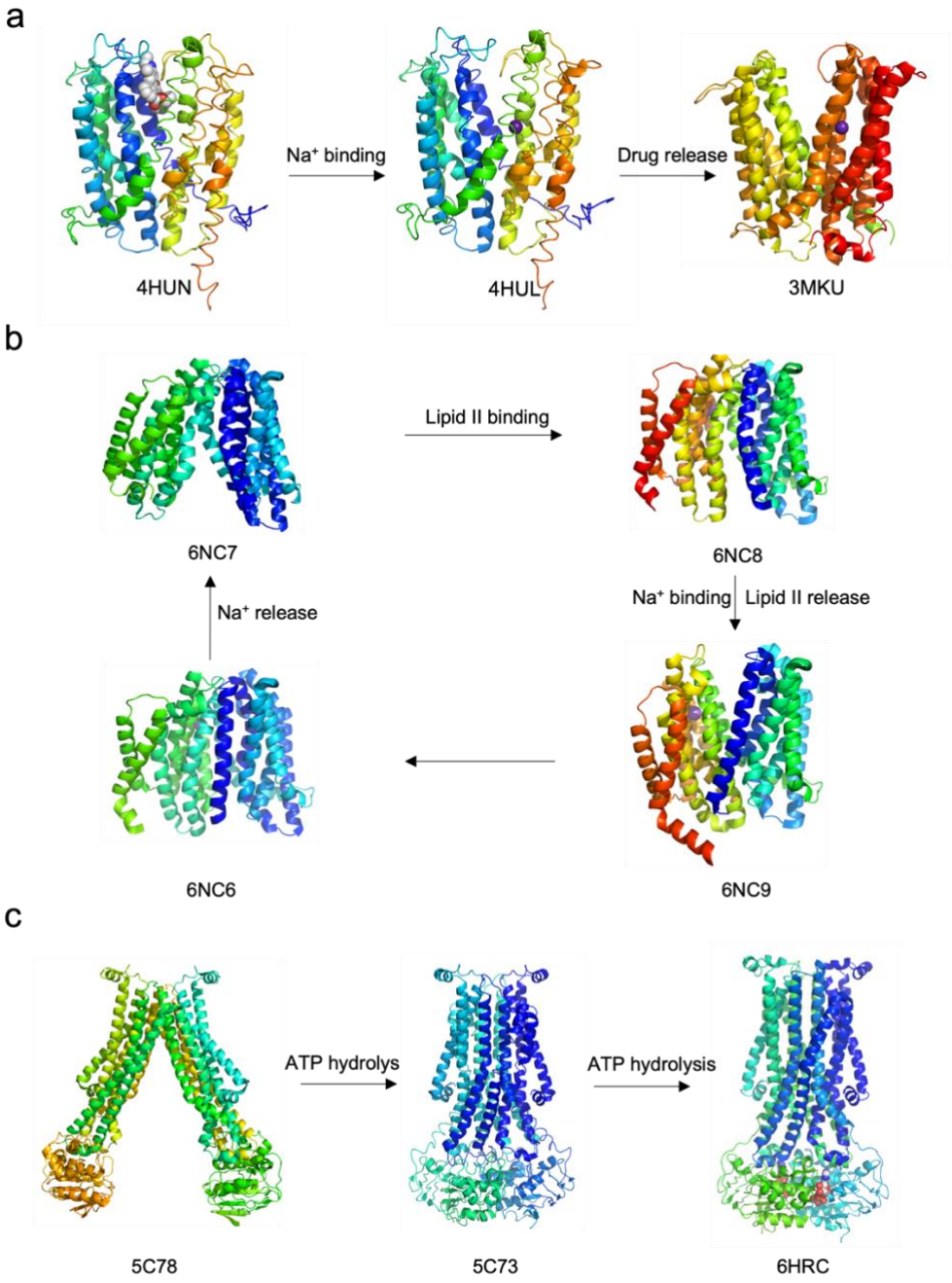


Figure 2.3 Structurally resolved catalytic stages of flippases *homologous flippase catalytic stages. (a) NorM co-crystallised with Rhodamine (left), Cs^+ (middle), and Rb^+ (right). The former two structures are of 'closed' states bound where drugs are bound, whilst the latter is of an 'open' state where drugs are released into the periplasm [142]. (b) MurJ crystallised in inward-open (top left, one of two monomers shown), inward-occluded (top right), outward (bottom right), and inward closed (bottom left, one of two monomers shown). (c) PglK crystallised in inward (left), outward-occluded, (middle), and outward semi-occluded with ATPγS (right). All structures are labelled with corresponding PDB codes, oriented with periplasmic face upwards, and where applicable only one monomer is shown.*

2.b.v. Previous work done by laboratory members

This project began in the Naismith lab in 2013 by Dr Hui Liu and Dr Audrey Le Bas.

13 clones of Wzx were obtained from 10 species of Gram negative bacteria. 3 of these were expressed to sufficient levels for crystallisation, namely *Methanocaldococcus jannaschii wzx*, *Pectobacterium atroscepticum wzxE*, and *E. coli wzxE*. 27 single mutants of conserved residues were expressed from these three genes, of which only mjWzx K196A crystallised in detergent and measured at 5.2 Å. 16 nanobodies were raised against ecWzx$_E$, 2 of which did not express to sufficient levels. Dot blots were performed on the remaining 14 to test for binding against Wzx. 9 nanobodies only exhibited weak binding, with a further nanobody (nb15) that did not show any binding. The 4 remaining nanobodies (nb1, nb4, nb5, and nb10) produced strong signals and were co-purified with Wzx by size exclusion chromatography. Only Wzx-nb10 yielded crystals, which required reconstitution in LCP.

After optimising expression, purification, and crystallisation protocols, Wzx-nb10 crystals reaching diffraction resolution of 2.5 Å were obtained. However, these results could not be replicated when the laboratory moved from the University of St Andrews to the Research Complex at Harwell. The protocols were therefore reoptimised, from which the very best crystal diffracted to 2.8 Å. Across both sites, diffraction quality of crystals were extremely variable.

2.c. Materials and Methods

2.c.i. Preparation of chemically competent phage resistant cell lines

Phage resistant cells were raised by plating infected cultures on LB Agar, from which single colonies were used to inoculate pre-cultures in 10 mL of LB at 37 °C overnight. 2 mL of pre-culture was then inoculated in 100 mL of warmed LB and grown at 37 °C until OD reached 0.6. Cells were pelleted by centrifugation at 35000 x g for 5 min. Cells were then resuspended in 16 mL of ice-cold 500 mM $CaCl_2$, 20 mM $MgSO_4$ and left on ice for 30 min. Cells were then pelleted again and resuspended in 2 mL of 500 mM $CaCl_2$, 20 mM $MgSO_4$. A quarter volume of Sterile 50% glycerol was then added to the resuspended cells and left one ice for 30 min. The resulting glycerol stock was then aliquoted in 50 μL, flash frozen with liquid nitrogen, then stored at -80 °C.

2.c.ii. Expression and purification of Wzx

The *wzxE* gene from *E. coli* K12 was cloned in a pBAD vector with a TEV-cleavable C-terminal His10 tag (gift from Dr Huanting Liu and Dr Hui Liu, University of St Andrews). The plasmid was transformed by heat shock at 42 °C for 50 seconds in C43 cells. Single colonies were inoculated in 10 mL LB preculture (100 μgmL^{-1} ampicillin) at 37 °C, and transferred to 1 L LB (100 μgmL^{-1} ampicillin). Expression was induced with 0.2% arabinose at 37 °C for 3 hr once OD reaches 1.0. Cell pellets were harvested by centrifugation at 6000 RPM for 10 min.

Selenomethionine derivatives of Wzx were expressed in methionine auxotroph strain B834 [157]. Overnight LB precultures as above were pelleted and rinsed with minimal media (NH_4Cl 1 gL^{-1}, KH_2PO_4 3 gL^{-1}, $Na_2HPO_4{\cdot}7H_2O$ 6gL^{-1}), then inoculated in 1L minimal media supplemented by 200 μm filter-sterilised Molecular Dimensions

SelenoMethionine Medium Nutrient Mix. Cultures are grown at 37 °C for 15 minutes before addition selenomethionine (60 mgL-1), induced with the 0.2% (w/v) arabinose for 24 hours.

Cell pellets were resuspended in PBS buffer and lysed by double pass high pressure cell disruption at 30 ksi. Lysate was cleared by centrifugation at 8000 x g for 30 minutes and suction filtration with 0.80 µm filters followed by 0.45 µm filters. Membrane was pelleted from cell lysate by ultracentrifugation at 40000 RPM for 75 min. Protein were then solubilised from membrane pellets homogenised with a glass homogeniser, in solubilisation buffer (20 mM tris pH 7.5, 150 mM NaCl, 5% glycerol, and 1.5% DM) at 4 °C overnight. Solubilised membrane was cleared by ultracentrifugation at 40000 RPM for 37 minutes, the supernatant from which was incubated with Ni Sepharose 6 FF beads (ABT) at 4 °C for 1 hr. IMAC is then performed on a gravity flow column, with two flow through cycles. Beads were then washed with wash buffer 1 (20 mM tris pH 7.5, 150 mM NaCl, 20 mM imidazole, 5% glycerol, 0.058% OGNG and 0.008% DDM), was buffer 2 (20 mM tris pH 7.5, 150 mM NaCl, 35 mM imidazole, 5% glycerol, 0.058% OGNG and 0.008% DDM), and elution buffer (20 mM tris pH 7.5, 150 mM NaCl, 500 mM imidazole, 5% glycerol, 0.058% OGNG and 0.008% DDM). Eluent was then desalted with wash buffer 1using CentriPure P100 column and incubated with 1 mg TEV protease at 4 °C overnight. Cleaved protein was separated from uncleaved protein and TEV protease in a further reverse IMAC (also known as difference purification [158])with one flow through cycle, wash buffer 1 and elution buffer. Flow through and wash fractions were pooled and concentrated in Vivaspin 20 MWCO 50K (GE healthcare).

2.c.iii. Expression and purification of nanobodies

Antibody was raised against Wzx in Llama and cloned into a pMESy6 vector with a non-cleavable C-terminal His6 tag via the Instruct-ERIC facility [159]. Amino acid sequences encoded by the open reading frame are shown in Table 2.1. It was transformed by heat shock at 42 °C for 50 seconds in WK6 cells. Single colonies were inoculated in 10 mL of LB preculture (100 μgmL^{-1} ampicillin). 10 mL precultures were inoculated in 1 L cultures of LB (100 μgmL^{-1} ampicillin), in which expression was induced with 1 mM IPTG at 28 °C for 24 hours once OD reaches 1.0. Cell pellets were harvested by centrifugation at 6000 RPM for 10 min. Selenomethionine derivatives were expressed identically to Wzx as above.

Cells were lysed by osmotic shock in 15 mL TES buffer (200 mM Tris pH 7.0, 0.5 mM EDTA, 500 mM sucrose) overnight at 4 °C. Buffer was the diluted with 30 mL 4x diluted TES, incubated at 4 °C for 2 hours, then diluted with TS buffer (50 mM Tris pH 7.0, 125 mM sucrose). Lysate was cleared by centrifugation at 8000 x g for 30 minutes and suction filtration with 0.80 µm filters followed by 0.45 µm filters. IMAC is then performed on a gravity flow column with Ni Sepharose 6 FF beads (ABT) as above with Wzx using wash buffer 1 (50 mM sodium phosphate pH 7.0, 1 M NaCl, 10 mM imidazole), wash buffer 2 (50 mM sodium phosphate pH 7.0, 1 M NaCl, 30 mM imidazole), and elution buffer (50 mM sodium phosphate pH 7.0, 1 M NaCl, 300 mM imidazole). The Eluent was concentrated to 5 mL with Vivaspin 20 MWCO 5K before injection in Superdex 75 SEC column (GE healthcare) with SEC buffer (50 mM sodium phosphate pH 7.0, 1 M NaCl). Purified nanobody was concentrated to 20 $mgmL^{-1}$ and frozen in 25 µL aliquots.

Nanobody	Amino Acid Sequence
1	QVQLVESGGGLVQAGGSLRLSCAASEPLSRTAMGWFRQAPGKERQFTAGISWSGDRTAYADSV KGRFTISRDNRKNTMYLQMNNLKPEDTAVYYCAAASVITNSMLYREPWQYKDWGQGTQVTVSH HHHHHEPEA
2	QVQLVESGGGLVQAGGSLRLSCAASEPLSRTAMGWFRQAPGKERQFTAGISWSGDRTAYADSV KGRFTISRDNRKNTMYLQMNNLKPEDTAVYYCAAASVTTNSMLYREPWQYKDWGQGTQVTVSS HHHHHHEPEA
3	QVQLVESGGGLVQAGGPLRLSCAASGFIGGSYGMGWFRQAPGKEREFVAAMLWSGGSTEYADS VKGRFTISRDNTKNTVYLQMNNLKPEDTSVYYCAASPIYYAGRDRQKYNAWGRGTQVTVSSHH HHHHEPEA
4	QVQLVESGGGLVQAGGSLRLSCAASGFGFSNYAMGWFRQAPGKEREFVGVIRRSGDNLYSDSV KGRFTISRDNAKKTLYLQMNSLKPEDTAVYYCARARVGDWYSSSWYREDEYVNWGQGTQVTVS SHHHHHHEPEA
5	QVQLVESGGGLVQAGGSLQLSCVASGRSYAMGWFRQAPGKEREWVGFVRWSGSSTYADSVKGR FTISRNTAKNTVYLQMNSLKPEDTAVYTCAAAPPGNPYTKYSDPDPYEYWGQGTQVTVSSHHH HHHEPEA
6	QVQLVESGGGLVQAGGSLRLSCAPSGRTLSGYAMGWFRQTPGKEREFVAAVSWSGDNIKAADS VKGRFTIARDNAKNTVYLQMNSLKPEDTAVYYCGATLKSFSYSSNYYRSSWYDYWGQGTQVTV SSHHHHHHEPEA
7	QVQLVESGGGLVQAGDSLTLSCAASGRTAYRYGMGWFRQHPGKEREFVASIWWTGTTTYYADS VKGRFTISRDDVKNMVYLQMNSLKPEDTAVYYCAAKFYGGNSKRPGDYAYWGQGTQVTVSSHH HHHHEPEA
8	QVQLVESGGGLVQTGDSLRLSCAASGRTFGKYAMGWFRQAPGKEREFVSLIDWDGSTTYYRDS VKGRFTISRDNVKNAANMVYLQMNSLKPEDTAVYYCAAKRVYYSGTYYPPAAYDYWGQGTQVT VSSHHHHHHEPEA
9	QVQLVESGGDLVQAGGSLRLSCVASGRTLSNYGVAWFRQAPGKEREFVSAISRTGSRTYYVGS VKGRFTISRDNVNNMASLQMNSLAPEDTAVYYCAAQYSNIDFAYTGGYNNWGQGTQVTVSSHH HHHHEPEA
10	QVQLVESGGGLVQAGGSLGLSCAASGRTFSNYVMAWFRQAPGKEREFVARISESRGTTDYADS VKGRFTISRDNAKNTIYLQMNSLNPGDTAVYSCAATLPAWTGIIGGRRPGNYPYWGQGTQVTV SSHHHHHHEPEA
11	QVQLVESGGGLVQPGGSLRLSCAASGRILGNSVMGWYRQAPGKQREFVARMTVRRNIGMNSDY AASVKGRFTISRDNAENTVYLQMDSLKPEDTAVYYCNYYLPGVSEDYWGRGTQVTVSSHHHHH HEPEA
12	QVQLVESGGGLVRAGGSLRLSCSASGRAVNTYAMSWFRQAPGKDREFVAGIDRSNIHTYYGDF VKGRFTISRDGAKDTVYLQMNDLNAEDTAVYYCAARRGFPAYSHITYDAWGPGTQVTVSSHHH HHHEPEA
13	QVQLVEAGGGLVQAGDSLRLSCAASVRSTYTMAWFRQAPGKDRELVAQITFGGSTNYADSVKG RFTISRDNAKNTVYLKMDSLKPEDTATYYCAAARYYGLSNSPERYTYWGQGTQVTVSSHHHHH HEPEA
14	QVQLVESGGGLVQAGGSLRLSCTASGSVFSINGMGWYRQFPGEQRDLVARITDGGITNSADFV KGRFTISRDNAKNTVYLQMNSLKPEDTAVYYCYAEIKPGPWYDYMDVWGKGTQVTVSSHHHHH HEPEA

15	QVQLVESGGGLVQAGESLRLSCTASVSLSSIKLVAWYRQVPGQPHDLIGYSTNYGYKFYVDSV KGRFNISRDNDKNMSYLQMTSLKSEDTGIYYCSTPPHDSWGQGTQVTVSSHHHHHHEPEA
16	QVQLVESGGGLVQAGGSLRLSCAASGGILRFNPMAWYRQAPGKQREVVASISGTGTTNYADSV KGRFTISRDDAKNTMSLQMNSLKPEDTAVYYCHSEVYMYRNYWGQGTQVTVSSHHHHHHEPEA

Table 2.1 Amino acid sequences of nanobodies

2.c.iv. Preparation of Strep-tag nb10

Strep-tag nb10 (nb10s) was generated by tag swapping using polymerase chain reaction (PCR) using a published protocol [160]. 50 µL PCR reactions were set up with 1x Phusion HF Buffer (NEB), 100 mM forward and reverse primers each (Table 2.2), 40 ng template (plasmid encoding His-tag nb10), 1 mM dNTP, and 0.5 µL Phusion polymerase (NEB). Reactions were initially denatured at 94 ºC for 7 min, then subjected to 12 cycles of 94 ºC for 1 min, ($T_{m(no)} – 5$) ºC for 1 min, and 72 ºC for 10 min. 3 subsequent cycles where then performed at 95 ºC for 1 min, ($T_{m(pp)} – 5$) ºC for 1 min, and 72 ºC for 10 min. Final extension was performed at 72 ºC for 20 min, before holding at 4 ºC overnight. 50 µL of PCR product was incubated with 1 µL (20 units) of DpnI (NEB) at 37 ºC for 3 hr, verified by 1% agarose gel ran at 90 V for 1 hr. 5 µL was then transformed with 100 µL of DH5α cells on which miniprep and sanger sequencing were performed. $T_{m(no)}$ and $T_{m(pp)}$ were calculated using the protocol [160] and OligoCalc [161].

Primer	**Sequence (5'-3')**
Forward	CAGGTGCTGGCAGTGGAGCGTCCGCGTGGAGTCATC CGCAATTCGAAAAATAGCACCACCATCACCATCACG
Reverse	CACTGCCAGCACCTGAACCCGCGCCTTTCTCGAA CTGAGGATGGGACCAACTGGAGACGGTGACCTG

Table 2.2 Primers used for nb10 strep-his tag swapping

Nb10s was expressed and purified identically as other nanobodies (Section 2.c.iii.) with two differences. First, expression was induced at 37 °C for 3 hr. Secondly, IMAC was performed with StrepTrap HP 5 mL column (GE Healthcare) with different buffers for wash (100 mM Tris pH 8, 150 mM NaCl, 1mM EDTA) and elution (100 mM Tris pH 8, 150 mM NaCl, 1mM EDTA, 2.5 mM desthiobiotin).

2.c.v. Western blot and dot blot

For western blots 15 μL of protein sample and 5 μL of NuPAGE LDS Sample Buffer (ThermoFisher), along with 5 μL Novex Sharp marker (ThermoFisher), were run on SDS-PAGE at 190 V for 40 min. Protein was transferred onto a Nitrocellulose membrane (ThermoFisher) using Power Blotter system (ThermoFisher) on high molecular weight setting, using transfer buffer of 20% Ethanol, 1X BioRAD running buffer (BioRAD).
For dot blots, 2 μL of protein (blotting sample) was pipetted onto nitrocellulose membrane cut to size to fit in the wells of CELLSTAR 24-well cell culture plates (Greiner). The downstream steps are identical to western blotting as outlined below, but all volumes are adjusted to 500 μL at the same concentrations.

The resulting membrane was put into a 50 mL Falcon tube and blocked with 30 mL of 5% semi-skimmed milk solution dissolved in PBST (10 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.05% Tween-20) for 1 hr at RT. It was then decanted and washed with PBST three times for 5 min each. If applicable, as in the initial tests for nanobodies that bind the Wzx/nb10s cp,[;ex (Section 2.d.iv.), membrane was incubated for 2 hr with binding protein at either 0.001 mg/mL or 5 x molar concentration of the blotted protein, whichever is higher. It was then washed with PBST three times for 5 min each. Membrane was then incubated for 2 hr with 2 μL of RHP-conjugated Anti-His antibody (prepared

in-house) or 4 μL of Anti-Strep StrepMAB-Classic conjugate (IBA). It was then washed with PBST three times for 5 min each. Excess PBST was then decanted and the membrane was imaged using chemiluminescence kit SuperSignal West Pico Plus (ThermoFisher).

2.c.vi. Crystallisation of the Wzx/nb10 complex

Purified Wzx (Section 2.c.ii.) and nb10 (Section 2.c.iii.) were incubated at 1:1.2 molar ratio for 1 hour at room temperature. The sample was then injected onto Superose 6 16/60 SEC column (GE healthcare) in SEC buffer (20 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol, 5% glucose, 0.058% OGNG, 0.008% DM), from which Wzx/nb10 complex was separated from excess nanobody. The purified complex was concentrated with Vivaspin 20 MWCO 100K (GE healthcare) to 10 $mgmL^{-1}$.

LCP bolus was formed by mixing Wzx/nb10 with the LCP mother lipid monoolein (Sigma Aldrich). Approximately 20μL of monoolein was taken using a 100 μL syringe (Hamilton, whilst 12 μL of Wzx/nb10 solution was taken using a 50 μL syringe (Hamilton). The two syringes were then coupled using a lipid mixer (Hamilton) with approximately 20 pushes in each syringe. If cubic phase has not formed, which would have been indicated by the bolus' colour change from opaque to clear, more monoolein is added to the mixture in 5 μL increments. Using the Mosquito dispenser (SPT Labtech) with the LCP module, 50 μL of LCP was dispensed onto a standard SBS format LCP glass plate (GWR), followed by 800 μL of crystallisation buffer (100 mM sodium citrate pH 6.4, 28.63% PEG400, 20mM zinc chloride, 180 mM lithium sulfate). Crystal plates were then sealed with a standard LCP cover slip (GWR) and incubated at 20 °C for 2 - 4 weeks. After crystals have stopped growing, individual wells were opened using a glass scriber. For soaking, 5 mM solutions of geranyl pyrophosphate and farnesyl

pyrophosphate and 10 mM solutions of pyrophosphate, UDP-GlcNAc, GlcNAc-1-phosphate , and GlcNAc were pipetted onto the LCP bolus. Crystals were then harvested using 10 μm cryo loops (MiTeGen) and flash frozen in liquid nitrogen.

2.c.vi. Preparation of *In Meso In Situ* Crystallography (IMISX) plates

The IMISX setup introduced in Section 1.b.ii can be assembled in house using commercially available materials (Figure 2.4). We followed a protocol provided by the Caffrey Lab (private communications).

A 2 mm wide and 140 mm thick double-sided tape gasket (Saunders), with outer dimensions of 118 x 83 mm, was placed on the corner of a LCP glass plate. The protective and filler layers were then removed, leaving only a rectangular double-stick strip on the glass plate.

A COP film (Zeon Europe) was cut out to the size of 110 mm x 75 mm, then silanised with Rain-X (ITW) and rinsed with milli Q water. A standard self-adhesive LCP spacer (GWR) was then placed onto the silanised side of the COP film. The COP film/spacer was then placed in the space enclosed by the gasket, on the glass base plate. A separate piece of unsilanised COP film was cut to 112 mm x 77 mm. This was adhered with water onto an IMISX cover slip (VWR), measured 124 mm x 84 mm x 0.15 mm. Crystal trials were set up using the glass/silanised COP plate, with identical robotic settings as standard glass LCP plates. After dispensation, the plates were sealed using the unsilanised COP/cover slip.

To extract individual wells, a glass scriber was used to extract the cover slip from the well desired. A scalpel was then used to extract the well from the rest of the plate. Mounted on the proprietary pins available as part of the IMISX kit (MiTeGen). Optionally for soaking, a scalpel was used to open the well with a diagonal cut to a corner, from which the soaking solution was then pipetted into the well. The open well was kept in a sealed box with wet tissues to prevent drying before x-ray exposure.

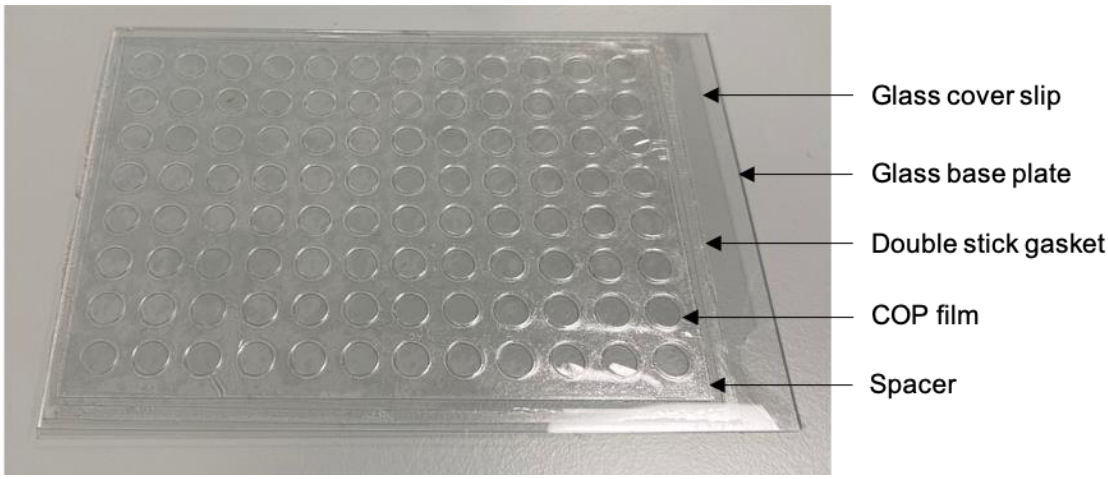


Figure 2.4 Assembled in house IMISX plate *Photograph of an assembled IMISX plate showing the commercial materials used.*

2.d. Results

2.d.i. Preparation of chemically competent phage resistant cell lines

The Research Complex at Harwell, where the laboratory was located, was affected by phage outbreak. Phage resistant strains of wk6, the nanobody expression strain, were raised by members of the laboratory. We then made phage resistant cells chemically competent (Section 2.c.i.).

Resulting cells generated were tested by transformation with pUC19 (NEB) as a positive control, and as a negative control, transformation protocol was followed but with no added plasmid. Cells were then plated on LB agar with ampicillin (100 μgmL^{-1}). Since colonies were obtained in all conditions barring the negative control (Figure 2.5), it was determined that these cells are competent and were used for expression of nanobodies used in this chapter.

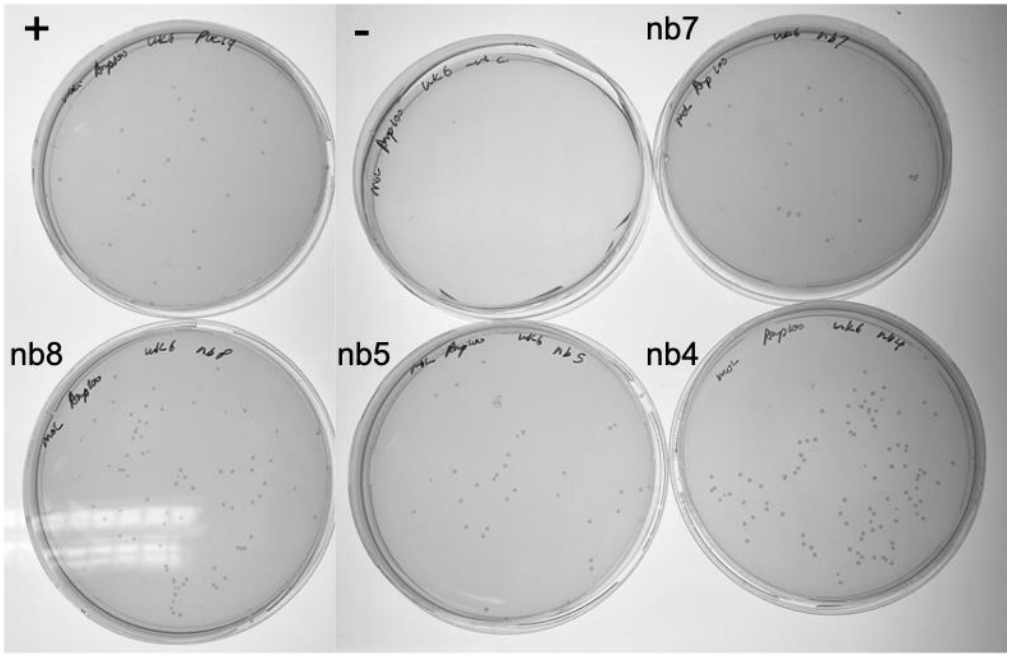


Figure 2.5 Testing wk6 for chemical competency *Plated wk6 transformed with nanobody (nb) plasmids, pUC19 is a positive control (+), and no plasmid as negative control (-) on LB agar plates with ampicillin (100 μgmL^{-1}).*

2.d.ii. Purification and crystallisation of the Wzx/nb10 complex

Following protocols outlined in Sections 2.c.i and 2.c.ii., purification of Wzx (after reverse IMAC step) and nb10 (after SEC step) were performed separately. We also screened for selenomethionine Wzx derivative expression at 16 °C, 25 °C, and 37 °C for 3 hr, 6 hr, and overnight. However, western blot performed on resulting cell lysate did

not show any expression (Figure 2.6). SDS-PAGE of IMAC (Figure 2.7a) shows separation of Wzx in the elution from other solubilised proteins in the flow through and wash fractions. SDS-PAGE of reverse IMAC shows separation of Wzx in the flow through and wash, from the His-tagged TEV protease and uncleaved Wzx in the elution (Figure 2.7b).

Similarly, SDS-PAGE of IMAC performed on nb10 shows separation of nb10 in the elution from other proteins extracted from the periplasm (Figure 2.7c). The eluted protein was then injected onto SEC where fractions appear to be homogenous (Figure 2.7d). The two samples were then pooled and incubated (Section 2.c.vi.) to form the Wzx/nb10 complex, which was injected onto SEC column. SDS-PAGE (Figure 2.7e) was used to verify that the two UV absorption peaks at 280 nm observed in the chromatogram (Figure 2.7f) corresponded to free Wzx and Wzx/nb10 complex. This protocol was optimised for LCP crystallisation by employing mild detergents DDM and OGNG (Table 1.1), in which the low number of binding sites allow subsequent reconstitution in monoolein the LCP mother lipid.

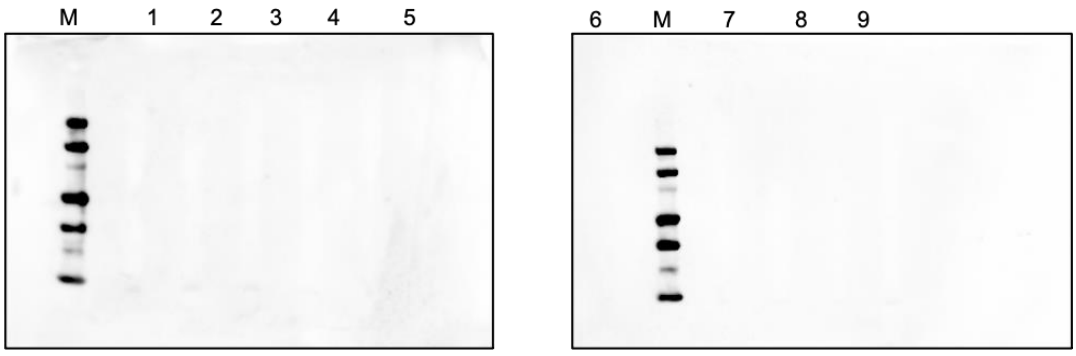


Figure 2.6 Western blot shows no expression of selenomethionine Wzx *Lysates were run on western blot along with Novex Sharp marker (M). Expression screening conditions were 16 °C for 3hr (1), 25 °C for 3hr (2), 37 °C for 3hr (3), 16 °C for 6 hr (4), 25 °C for 6 hr (5), 37 °C for 6 hr (6), 16 °C o/n (7), 25 °C o/n (8), 37 °C o/n (9).*

Crystallisation trials conducted by previous lab members yielded LCP crystals from two conditions in the JBScreen LCP (Jena Bioscience). Triangular crystals from the first condition (100 mM sodium citrate pH 6.5, 30% PEG 400, 185 mM lithium sulfate) diffracted to 6 Å. Rod-shaped crystals from a second condition (50 mM ADA pH 6.25, 25% PEGMME 550, 350 mM sodium nitrate, 50 mM disodium malonate pH 7.0) diffracted to 8 Å. The first condition was optimised by varying constituent concentrations and addition of $ZnCl_2$, as shown in its composition (Appendix A) This screen was used in henceforth production of Wzx crystals, which typically diffracted up to 3 Å.

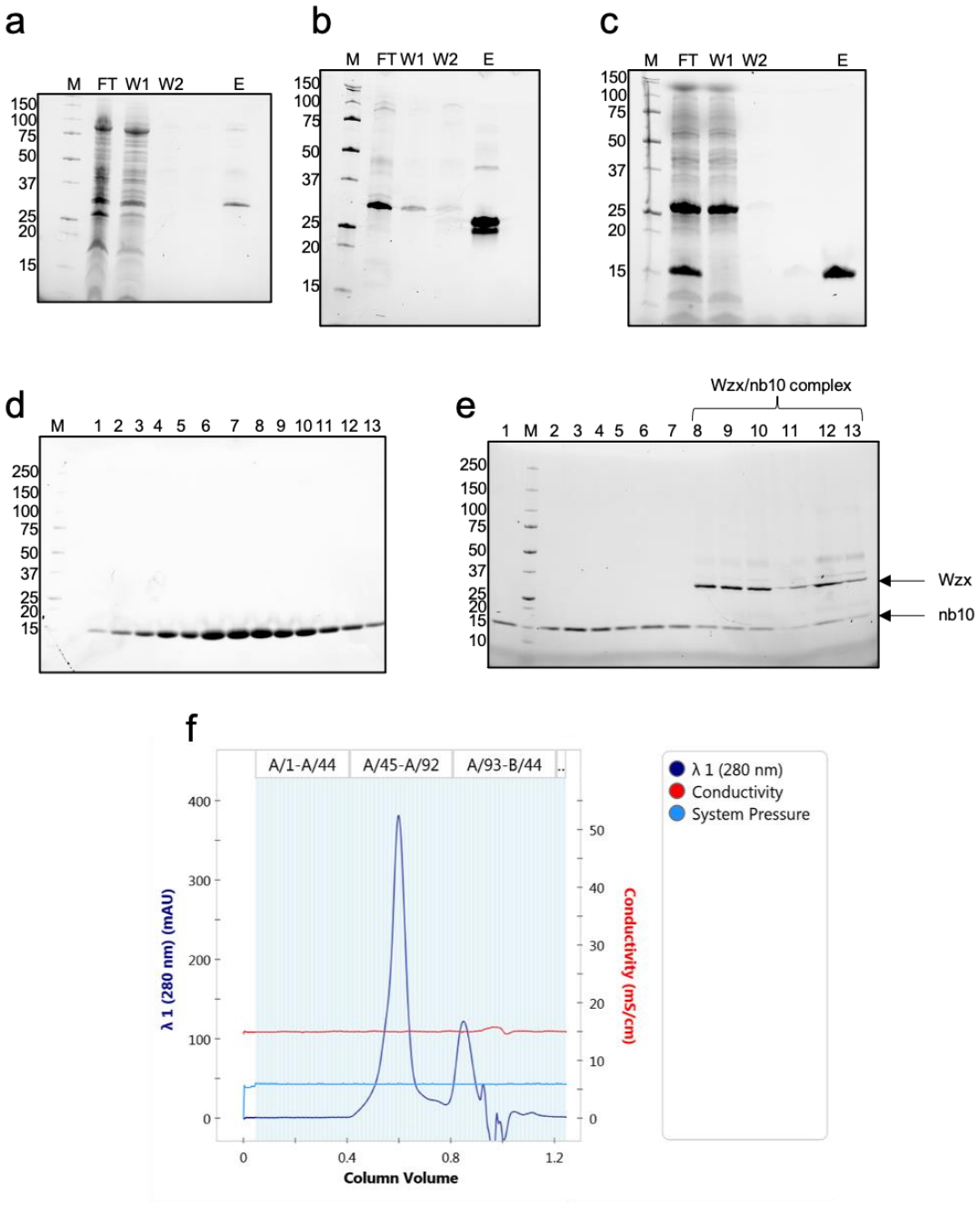


Figure 2.7 Purification of the Wzx/nb10 complex. *SDS-PAGe analysis on Wzx forward IMAC (a), Wzx reverse IMAC (b), nb10 IMAC (c), nb10 SEC (d), and Wzx/nb10 SEC (e). Lanes labelled contain BioRAD Precision Plus Proteim Unstained Standards marker (M), IMAC flow through (FT), IMAC wash (W1, W2), IMAC elution (E). For (d,e), samples (1-13) correspond to fractions of chromatogram peaks from SEC. (f) Chromatogram of SEC performed on Wzx/nb10, showing UV absorption at 280 nm (indigo). The two peaks observed were verified by SDS-PAGE as shown in (e), with first peak corresponding to lanes 8-13, and second peak corresponding to lanes 1-7.*

2.d.iii. Preparation of Strep-tag nb10

Using nb10h as a template, the tag was swapped by inserting a double Strep-tag and a stop codon between the end of the nb10 gene and the His tag (Figure 2.8a). Due to the long insertion sequence, an overlapping region in the middle was chosen for the PCR protocol [160]. The primers were therefore designed to begin with the overlapping sequence, followed by the remainder of the insertion sequence and the insertion site at the corresponding direction. The resulting open reading frame was verified by Sanger sequencing. Comparison between the amino acid sequences of nb10h and nb10s, determined using ExPASy translate tool [162], is shown in Figure 2.8b.

Since we failed to express nb10s using the nanobody protocol described in Section 2.c.iii. (Figure 2.9a), we re-screened expression conditions. We observed that nb10s expressed better when induced for 3 hr instead of overnight (Figure 2.9b), and was replicated large scale at 37 °C with 2 L cultures. Nb10s was amenable to purification by Streptactin affinity chromatography (Figure 2.9c) and SEC (Figure 2.9d), and appeared as a sharp homogenous peak in the SEC elution profile (Figure 2.9e).

a

nb10

TTTCACAAGCTTAAGGAGACAGTACATATGAAATACCTATTGCCTACGGCAGCCGCTGGATT
GTTATTACTCGCGGCCCAGCCGGCCATGGCACAGGTGCAGCTGGTGGAGTCTGGGGGAGGAT
TGGTGCAGGCTGGGGGCTCTCTTGGACTCTCCTGTGCAGCCTCTGGACGCACCTTCAGTAAC
TATGTCATGGCCTGGTTCCGCCAGGCTCCAGGGAAGGAGCGTGAGTTTGTAGCACGTATTAG
CGAGAGTCGTGGGACCACAGACTATGCGGACTCCGTGAAGGGCCGATTCACCATCTCCAGAG
ACAACGCCAAGAACACGATCTATCTCCAAATGAACAGCCTGAATCCTGGGGACACGGCCGTT
TATTCCTGTGCAGCCACCTTACCAGCCTGGACTGGCATTATAGGCGGCCGCAGACCGGGGAA
TTATCCATACTGGGGCCAGGGGACC**CAGGTCACCGTCTCCAGT**CACCACCATCACCATCACG
AACCTGAAGCCTAGTACCCGTACGACGTTCCGGACTACGGTTCCACTGTTGAAAGTTGTTTA
GCAAAACCTCATACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAAAACTTTAGATCG
TTACGCTAACTATGAGGGCTGTCTGTGGAATGCTACAGGCGTTGTCGTTTGTACTGGTGACG
AAACTCAGTGTTACGGTACATGGGTTCCTATTGGGCTTGCTATCCCTGAAAATGAGGGTGGT
GGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTACTAAACCTCCTGA
GTACGGTGATACACCTATTCCGGGCTATACTTATATCAACCCTCTCGACGGCACTTATCCGC
CTGGTACTGAGCAAAACCCCGCTAATCCTAATCCTTCTCTTGAGGAGTCTCAGCCTCTTAAT
ACTTTCATGTTTCAGAATAATAGGTTCCGAAATAGGCAGGGTGCATTAACTGTTTATACGGG
CACTGTTACTCAAGGCACTGACCCCGTTAAAACTTATTACCAGTACACTCCTGTATCATCAA
AAGCCATGTATGACGCTTACTGGAACGGTAAATTCNNAGACTGCGCTTTCCATTCTGGCTTT
AATGAGGATCCATTCGTTTGNAA

Insertion sequence

TGGTCCCATCCTCAGTTCGAGAAAGGCGCGGGTT**CAGGTGCTGGCAGTG**GAGCGTCCGCGTG
GAGTCATCCGCAATTCGAAAAATAG

Forward primer (5′-3′)

CAGGTGCTGGCAGTGGAGCGTCCGCGTGGAGTCATCCGCAATTCGAAAAATAG**CACCACCAT**
CACCATCACG

Reverse primer (5′-3′)

CACTGCCAGCACCTGAACCCGCGCCTTTCTCGAACTGAGGATGGGACCA**ACTGGAGACGGTG**
ACCTG

b

```
nb10s       MAQVQLVESGGGLVQAGGSLGLSCAASGRTFSNYVMAWFRQAPGKEREFVARISESRGTT
nb10h       MAQVQLVESGGGLVQAGGSLGLSCAASGRTFSNYVMAWFRQAPGKEREFVARISESRGTT
            ************************************************************

nb10s       DYADSVKGRFTISRDNAKNTIYLQMNSLNPGDTAVYSCAATLPAWTGIIGGRRPGNYPYW
nb10h       DYADSVKGRFTISRDNAKNTIYLQMNSLNPGDTAVYSCAATLPAWTGIIGGRRPGNYPYW
            ************************************************************

nb10s       GQGTQVTVSSWSHPQFEKGAGSGAGSGASAWSHPQFEK
nb10h       GQGTQVTVSSHHHHHHEPEA------------------
            **********
```

Figure 2.8 nb10 His-tag was swapped to Strep-tag *(a) DNA sequences of nb10 plasmid insertion sequence, and primers. The insertions sites, end of the nb10 gene and the His-tag, are highlighted in green and yellow respectively. Overlapping region of insertion sequence is highlighted in blue. (b) Amino acid sequences of nb10s and nb10h aligned with Clustal Omega [163]. Strep- and His-tags are shown in blue and red respectively.*

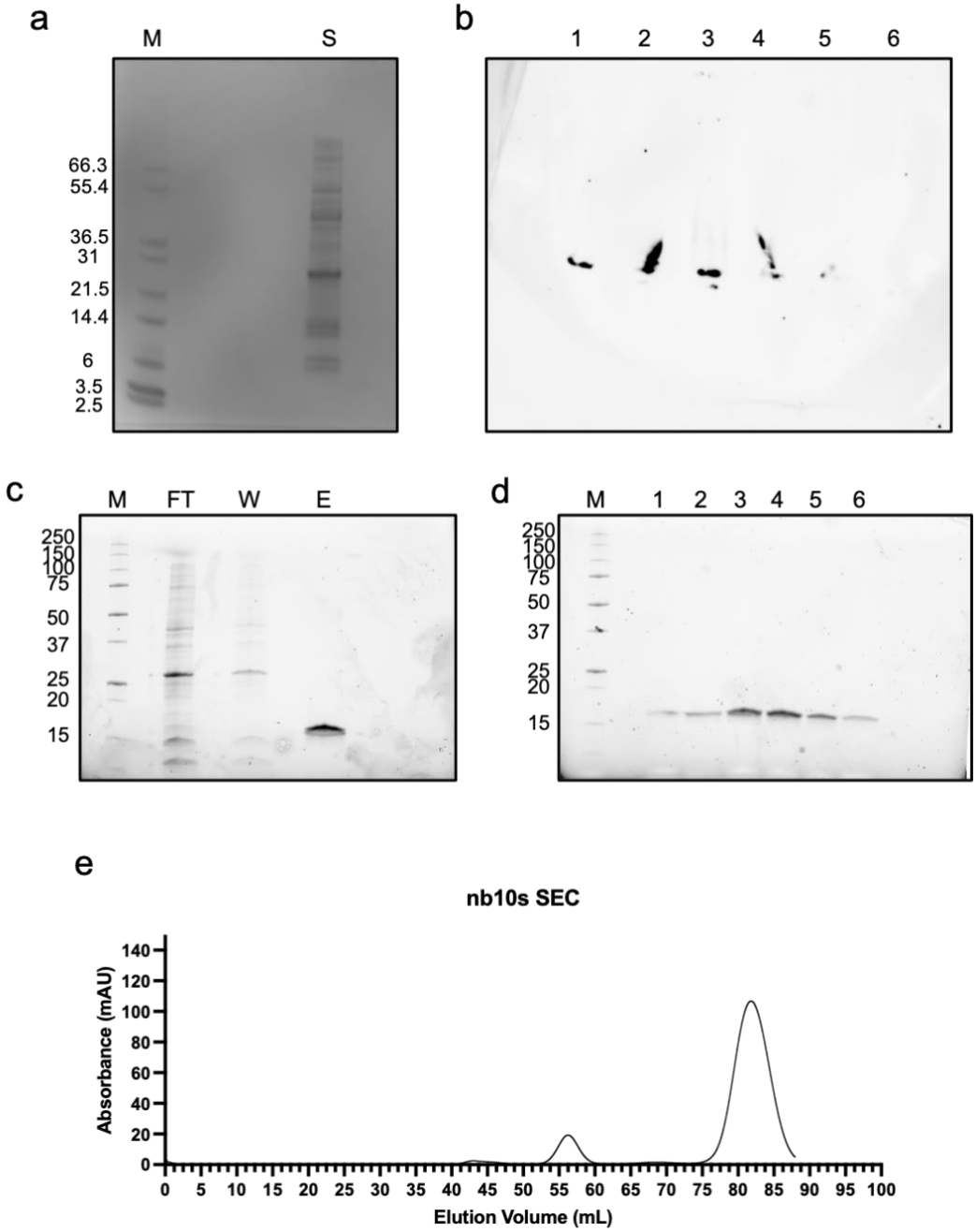


Figure 2.9 nb10-strep expression required modified protocol *(a) SDS-PAGE of Mark12 standard (M) and cleared cell lysate after expression with standard nanobody protocol (S), showing no expression.(b) Western blot performed on cell lysate of nb10s induced at 16 °C for 3hr (1), 28 °C for 3hr (2), 37 °C for 3hr (3), 16 °C o/n (4), 28 °C o/n (5), 37 °C o/n (6). (c) SDS-PAGE of BioRAD Unstained Marker (M) and Streptactin affinity chromatography flow through (FT), wash (W), and elution (E) fractions. (d) SDS-PAGE of BioRAD Unstained Marker (M) and SEC elution fractions (1-6) from the major peak eluted at 75 mL from the SEC elution profile (e).*

2.d.iv. Double pulldown assay for ternary complex screening

Since we were able to reproduce Wzx/nb10h crystals (Section 2.d.ii.), we attempted to screen for other nanobodies binding to different epitopes. This would allow us to focus efforts for crystallisation trials more likely to give rise to different conformers, which would elucidate the catalytic cycle of Wzx. We were able to express and purify eight His-tagged nanobodies using the standard nanobody protocol (Section 2.c.iii.). However, due to shift working in response to the Covid-19 pandemic, we automated the IMAC stage using the AKTA Pure system. Lysate was passed through HisTrap FF columns (GE healthcare) and washed with wash buffer until UV absorbance stabilises at the baseline, at which point elution buffer was passed. Due to the sharp elution peaks observed (Figure 2.10), samples were directly injected onto SD200 then aliquoted without verification by SDS-PAGE.

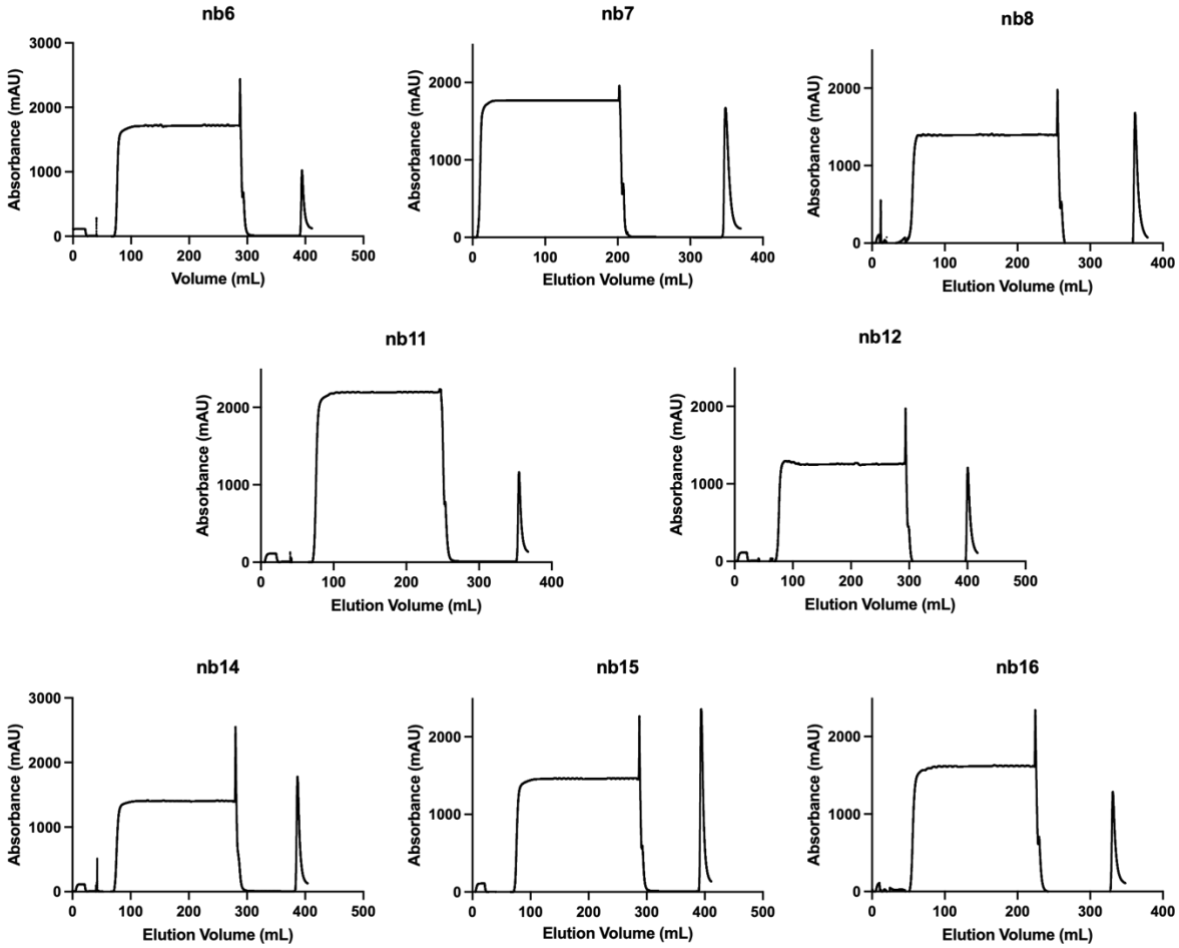


Figure 2.10 Nanobody IMAC elution profiles *Showing flow through of proteins in lysate (approximately 0 to 200-300 mL) and eluted protein (approximately 350 to 400 mL).*

We reasoned that if a nanobody was able to form a ternary complex with Wzx/nb10s, it is likely to bind to a different epitope. We attempted two methodologies to investigate this capacity from the eight nanobodies purified. Firstly, as an initial test, we performed a dot blot using Wzx as the blotting sample. The membrane was incubated with nb10s, washed, incubated with one of His-tag nanobodies. To ensure that both nanobodies remain bound, both anti-His and anti-Strep blots were performed. We observed that there was considerable background noise, which rendered results unreliable (Figure 2.11). All conditions except nb10 and nb11 yielded positive results in both anti-His and anti-Strep blots. This may represent the capture of a bona fide ternary complex, or a mixture of

Wzx/nb10 and Wzx/His-nb complexes. The nb10 condition was ambiguous, with very weak signal in the Strep blot that may be attributed to background noise. The significantly stronger signal in the His blot suggests the nb10h binds more strongly and displaces nb10s. The nb11 condition was negative in the His blot. To verify that this was due to its inability to bind the Wzx/nb10s complex, we ran additional positive control anti-His dot blots with His-tagged nanobodies as blotting sample (Figure 2.11c). Unfortunately, the positive control for nb11 failed as it only resulted in a weak signal. Finally, as an additional control, anti-Strep dot blot was performed on Wzx as blotting sample and incubated with nb10s, compared to using nb10s as blotting sample, both of which yielded positive results (Figure 2.11d).

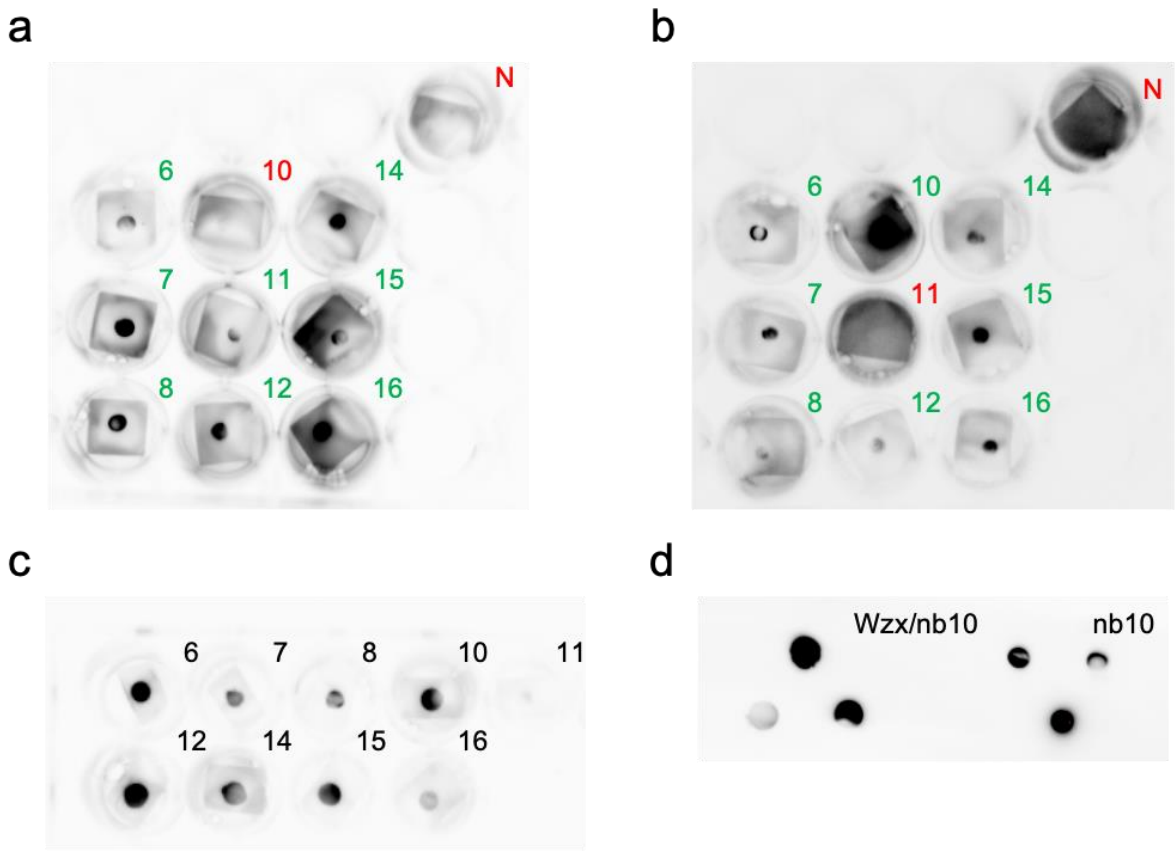


Figure 2.11 Dot blot suggested many nanobodies can form ternary complex *Anti-Strep (a) and anti-His (b) dot blots of Wzx bound with nb10s and His-tag nanobodies (numbered), with negative control (N) of Wzx without nanobody incubation. Results are labelled as either positive (green) or negative (red). (c)*

Anti-His dot blot of His-tagged nanobodies (numbered) as blotting sample. (d) Anti-Strep dot blots in triplicates of Wzx as blotting sample incubated with nb10s, and nb10s as blotting sample.

We attempted to verify these results with a double pulldown assay. 1 $mgmL^{-1}$ of Wzx was incubated in 1:1.2:1.2 molar ratio with nb10s and each His-nanobody at RT for 2 hr. Affinity chromatography (Sections 2.c.iii., 2c.iv.) was downscaled to 1mL reactions in detergent (Section 2.c.ii.), performed by pipetting relevant buffers in and out of Eppendorf tubes. Streptactin affinity chromatography was performed first, as its elution buffer is less likely to cause precipitation during bead binding for the second, His-IMAC step. We predicted that ternary complexes, where they form, will be pulled down during elution in both chromatography steps.

We analysed flow through and eluent of both chromatography steps by SDS-PAGE (Figure 2.12) Unfortunately, signal was poor and results were ambiguous. Nonetheless, gels from the first Streptactin affinity chromatography step showed two distinct low molecular weight bands, the higher of which is likely to be nb10s due to the longer, double Strep tag. In some fractions, we also observed Wzx and His-tag nanobody bands pulled down in elution fractions, which were was not present in the Wzx only control. These conditions are: nb6, nb7, nb11, nb15, and nb16. Puzzlingly, Wzx appeared to have been pulled down in the Wzx/nb10h condition despite the lack of any Strep tag protein, which was difficult to explain.

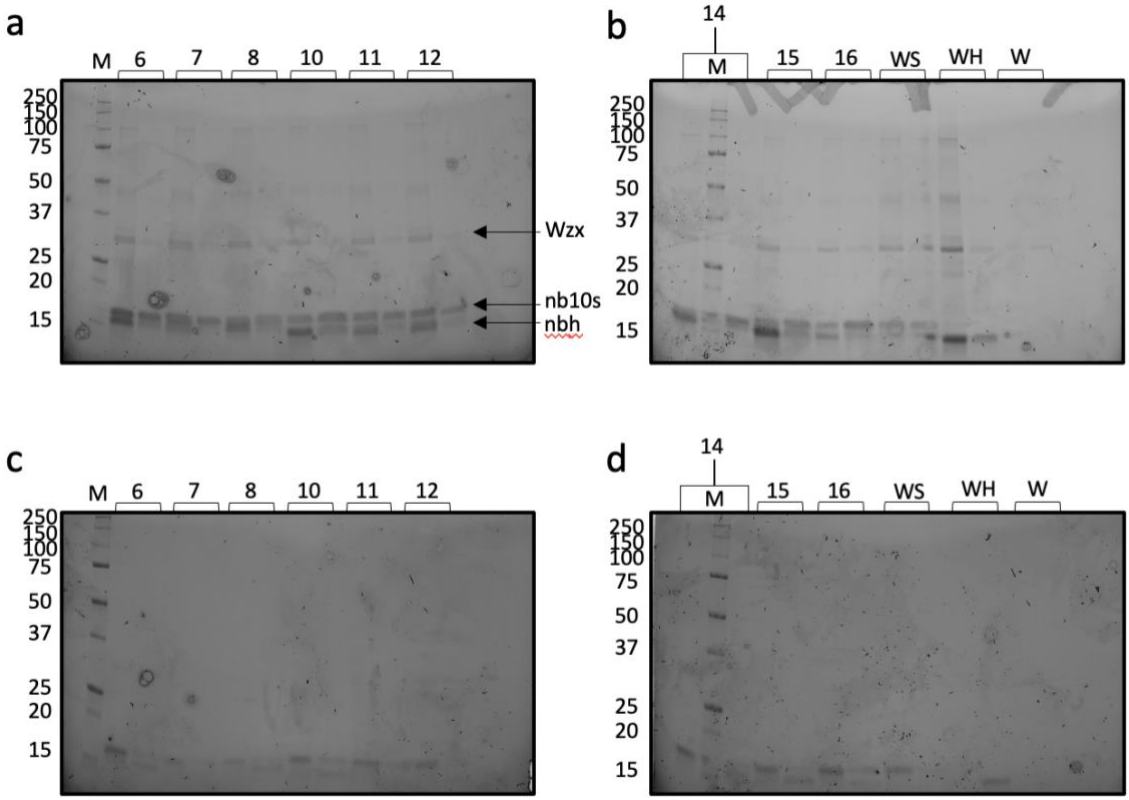


Figure 2.12 SDS-PAGE of pulldown fractions *SDS-PAGE of samples from Streptactin affinity chromatography (a,b) and His-IMAC (c,d). Lanes correspond to flow through on the left an elution on the right for each sample (numbered = nanobody, WS = Wzx + nb10s, WH = Wzx + nb10h, W = Wzx), shown alongside BioRAD marker (M). Bands corresponding to Wzx, nb10s, and His-tagged nanobody (nbh) are highlighted.*

To improve signal, we analysed eluted fractions Western blot using anti-His antibody for Streptactin affinity chromatography, and anti-Strep antibody for His-IMAC (Figure 2.13). Unfortunately, despite having used a secondary antibody to enhance signal, signal:noise ratio was poor and the results remained inconclusive. The Novex Sharp marker in anti-His blots were poorly resolved, the most intense bands from the ladder expected at >30 kDa were higher than bands that appeared in experimental conditions, the latter of which are therefore likely correspond to nanobodies (molecular weight approximately 18 kDa). Negative controls behaved as expected. Wzx incubated with nb10s only was positive only for the anti-Strep blot. Wzx incubated with nb10h only was positive, albeit with a faint band, for the anti-His blot. Wzx only was negative for both blots.

Although we did not have time to produce reliable results, nb6, nb11, nb14, and nb15 were positive in both blots and were therefore recommended for priority in crystallisation trials. nb7, nb8, nb10, nb12, and nb16 were negative for the anti-His blot despite being positive for the anti-Strep blot. If this result was reproducible, a possible explanation is that during Streptactin chromatography, Wzx is pulled down as a binary complex with nb10s, and residual His-tag nobodies remained in the eluent as the beads were washed using only pipetting. To prevent this in future studies, experiments can be scaled up to perform chromatography using columns, which would also improve the signal:noise ratio that was suboptimal for our studies.

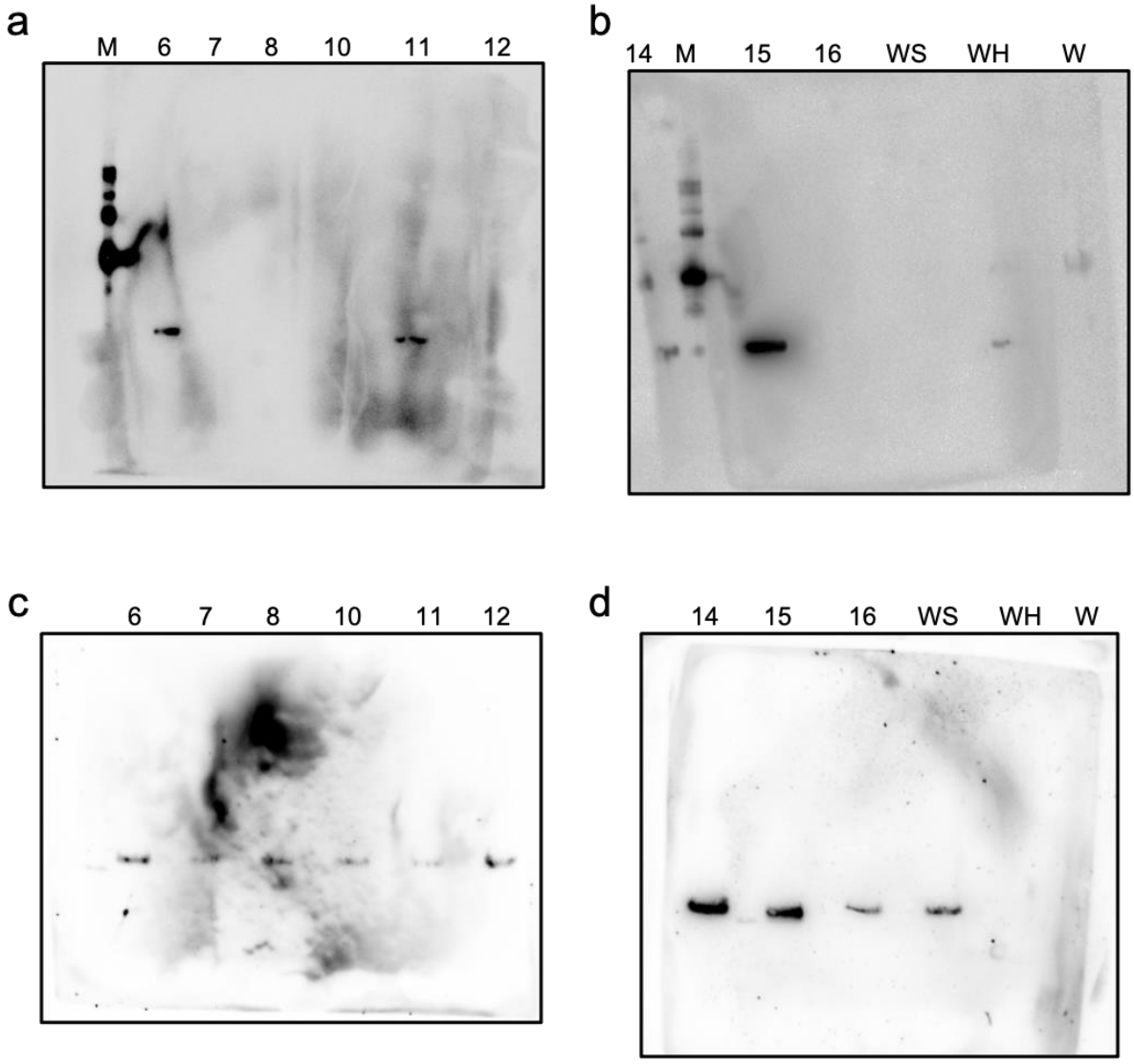


Figure 2.13 Western blot was also inconclusive *Western blots ran for Streptactin eluent using Anti-His antibody (a,b) and anti-Strep antibody (c,d). Experimental conditions were numbered according to the His-tag nanobody incubated. Controls included were Wzx/nb10s (WS), Wzx/nb10h (WH), and Wzx only (W).*

2.d.v. Exogenous lipids as additives for LCP crystallisation

Exogenous lipids were added to the LCP mixture by dissolving in monoolein at 37 °C before mixing with detergent-solubilised protein solution to form the LCP bolus. During initial trials, 1-2 mg of pre-weighed lipids were dissolved in 20 μL of monoolein, to make a final (after protein addition) concentration of 5-10%. However, it was noted that some

lipids were not dissolved, requiring addition of further monoolein of up to a further 20 μL. It was observed that when Wzx/nb10 solution and monoolein were mixed in Hamilton syringes (Section 2.c.iv.), the resulting bolus becomes significantly stiffer. However, we did not observe any changes to crystal size, morphology, or incubation time required for crystallisation. Initial trials were performed for four lipids and screened at the microfocus beamline I24, the resolution of which is shown in Table 2.3.

Lipid	Number of crystals screened	Screening resolution (Å)
*E. coli p*olar lipid mixture	9	2.0 – 4.0
CHS	7	2.3 – 3.0
POPC	10	2.7 – 5.0
DPMG	6	2.3 – 8.0
No lipid	17	3.0 – 8.0

Table 2.3 Screening resolution increases after lipid addition *Tabulated screening resolution of highest diffracting crystals containing each type of lipid tested. E. coli polar lipids extract showed the best diffraction limit at 2.0 Å*

Further optimisation was performed on addition of *E. coli* polar lipid extract (Avanti) due to its marked increase in diffraction resolution. Stocks of 20% polar lipid extract were prepared, with pre-weighed lipid first dissolved in chloroform before addition to monoolein. Residual chloroform was by evaporation using inert N_2 gas and incubation in a desiccator overnight. Using this method we were able to obtain up to 20% polar lipid in monoolein. We subsequently screened for Wzx/nb10 crystals at 5%, 10%, 15%, and 20% polar lipid, all of which resulted in abundant crystals. Since x-ray diffraction was most reproducibly obtained at the resolution limit of approximately 2.0 Å at 15% polar lipid, this was used to generate subsequent crystals.

2.d.vi. Data collection at the I23 beamline

Wzx crystals were mounted on I23 pins (Figure 2.14), which are also compatible in the I03 beamline. Crystals were pre-screened at I03 principally because the *in vacuo* robotics at I23 are considerably slower, allowing more efficient screening for diffracting crystals. Additionally, pre-screening at I03 allows utilisation of its automatic x-ray centring feature to locate crystals within the opaque LCP bolus. However we observed that after pre-screening at the I03 beamline, diffraction exhibited unintended annealing effects at I23. This was manifested in a decrease of diffraction quality, icy appearance of cryo loops, and appearance of ice rings. We deduced that this is due to thermal conductivity of I03 pin base, which was subsequently rectified by beamline scientists at I23. The disappearance of this reannealing effect was tested by loading and reloading crystals at I03. Subsequently, we observed that Wzx crystals were more robust and were able to withstand data collection for longer before x-ray induced damage.

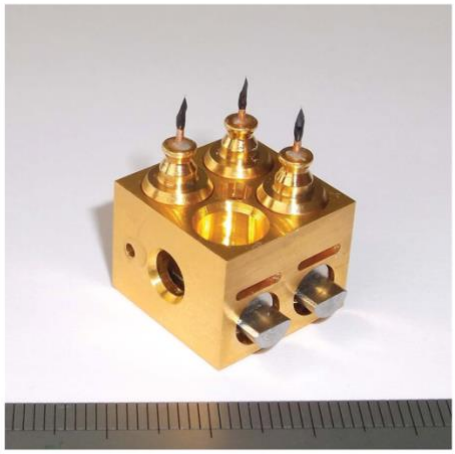


Figure 2.14 Proprietary I23 pins *Extracted from Wagner et al [39], a photograph of the copper pin and pin base system designed for in vacuo sample transfer*

Data was collected at the wavelength of 2.7552 Å. This wavelength is sufficiently high to be proximal to the sulfur K absorption edge, which increases anomalous signal

strength. Higher wavelengths increases background scattering, such as from the lipid in the LCP bolus. This background scattering led us to reduce beam size from 150 μm x 60 μm to 50 μm x 50 μm, which significantly improved data quality By reducing beam size one reduces the exposure of non-crystal containing areas of the LCP bolus, which improves signal:noise ratio (Figure 2.15a,b). This was particularly important for our LCP sample because lipid rings diffract at 3 Å, which is close to the diffraction limit of Wzx crystals due to the lower beam intensity and higher background scattering at I23 compared to I03. With smaller beam sizes, we were able to identify diffraction spots (Figure 2.15c) that would otherwise be eclipsed by the more intense lipid rings with larger beam sizes.

The reduction of beam size meant optimal beam centring was ever more important, as the smaller beam can be used to target the best diffracting parts of a crystal. The grid scan function at I23 was used to identify the optimal portions of the crystal, whilst the illumination with LED backlight available at both beamlines was used to align the I23 beam onto the optimal portion. To further optimise beam centring, we performed manual grid scans of beam centres 25 μm apart in both x and y directions (Figure 2.16). The automation of this process was identified as the beamline's next milestone in collection x-ray data from LCP crystals.

Radiation and thermal damage from X-ray exposure to crystal leads to decrease in the number of diffraction spots, and may impede data reduction. However, crystals optimised for robustness such as Wzx can withstand collection at 360° and multiple κ angles to improve data redundancy, as discussed in Section 2.d.iv. We operated a data collection workflow shown in Figure 2.17.

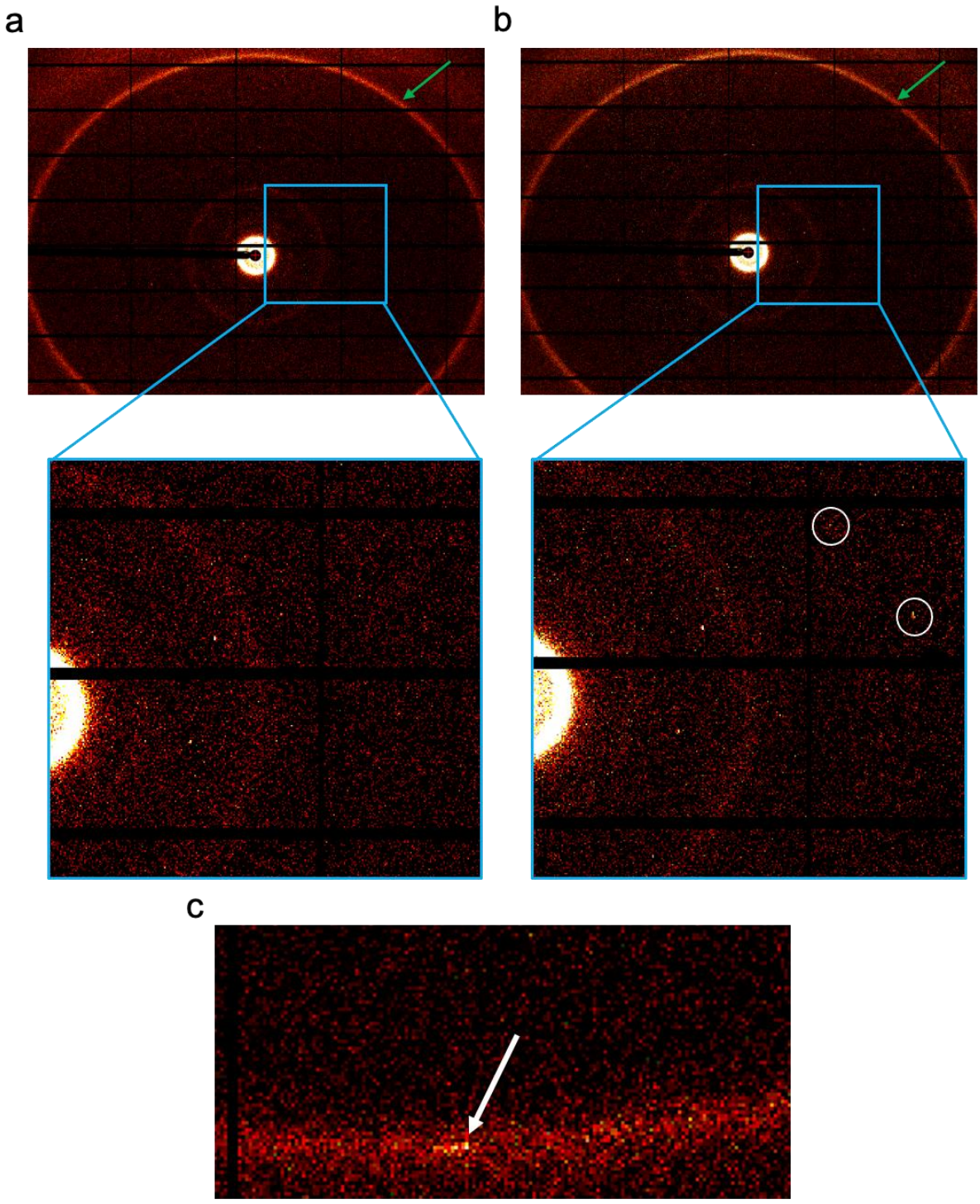


Figure 2.15 Reducing beam size improved data quality *Diffraction images at identical locations with beam sizes of 150 x 60 μm (a) and 80 x 60 μm (b). Zoomed images of areas highlighted in blue rectangles show diffraction spots that only appear with lower beam size (white circles). Lipid rings at approximately 3 Å are shown with green arrows. (c) Close up image taken with a 50 x 50 μm beam at the lipid ring, showing an overlapping diffraction spot (white arrow).*

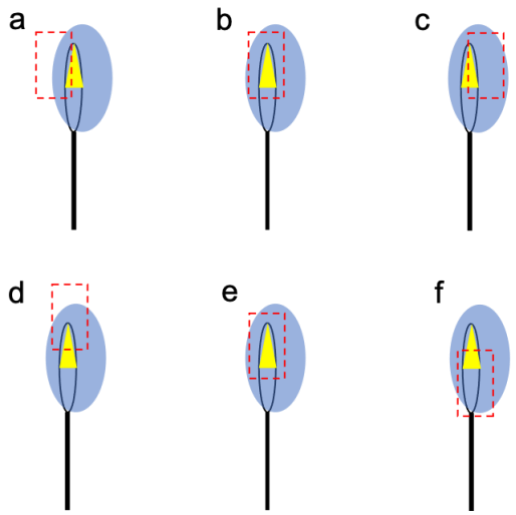


Figure 2.16 Manual grid scan at I23 *Single crystals (yellow triangle) were harvested along with the LCP bolus (blue) using the cryo loop (black). The beam area (red dashed outline) was moved along both x- (a,b,c,) and y- (d,e,f) directions, by half of beam width. The optimal positions (b,e,) have the highest possible crystal area with the least possible LCP area.*

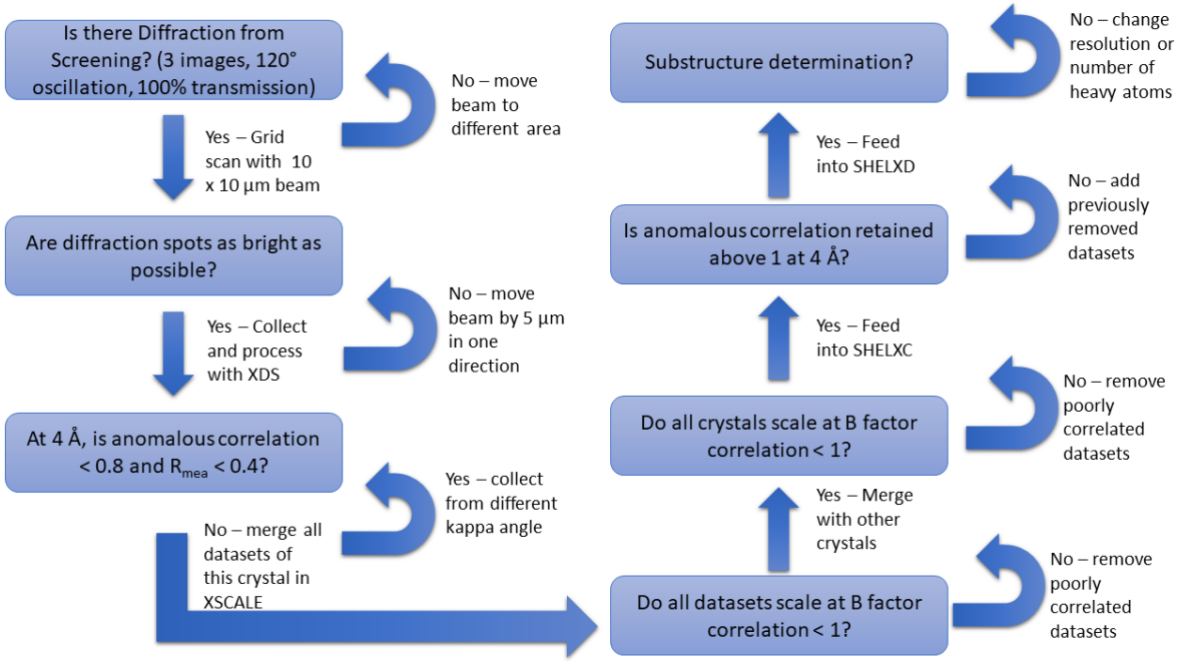


Figure 2.17 Wzx data processing workflow Flow chart for substructure (sulfur heavy atom sites) determination at the I23 beamline. Chart begins at I23 screening and grid scan and applies to every crystals at every κ angle

2.d.vii. Data processing for native s-SAD

Data were indexed using the XDS programme [164] with the space group and cell dimensions identified at I03. Data were processed to the highest resolution provided that data quality indicators R_{mea} does not exceed 0.6 and $CC_{1/2}$ does not fall below 0.5. Data quality can be improved by excluding poor images processed. For example, this can be used to omit images where there is no diffraction as the beam misses the crystal as it rotates, or exhibited x-ray damage. To take this into account, the XDS programme was set to automatically exclude any images where the percentage of diffraction spots indexed is lower than 10%. Other than data quality, XDS also provides the resolution at which anomalous diffraction data were reliable. Samples with anomalous signal resolved at lower than 6 Å were omitted from downstream processing as we found the gain in anomalous signal was minimal.

The parameters outlined above were used to determine which dataset were merged using XSCALE [165] to increase data redundancy. Merging of datasets improves anomalous signal and redundancy, at the detriment of internal consistency as indicated by R-values. To maximise the former and minimise the latter, similar datasets were merged. This is indicated by the B factor coordination, which is maintained at above 1, which indicates the variability of B factors across the unit cell. At this stage, data can also be limited to higher resolution ranges to maintain internal consistency. XSCALE generates a correlation matrix between datasets, which was used to guide the omission of datasets when the criteria outlined above were suboptimal. Following these guidelines, we merged 20 360° datasets from 4 Wzx crystals, the data table from which (Appendix B) suggests that anomalous signal can be detected above a resolution cut off at 2.7 Å. SHELXC was used to estimate sulfur structure factors (Section 1.a.i), which was used in SHELXD to

locate sulfur atoms (the substructure) in the unit cell [166]. The resulting anomalous signal map for the substructure (Figure 2.18) confirmed identification of all 18 sulfur-containing residues (14 methionines and 4 cysteines) in the Wzx/nb10 complex.

Having satisfied that the sulfur substructure is obtained, we migrated data processing to the CCP4i2 interface [167]. Pointless was used to confirm the space group by identifying symmetry-related reflection spots in the diffraction pattern [168]. Data reduction was then performed using Aimless, which averages intensities of symmetry-related reflections [169]. Using the resulting reflection intensities and phases estimated from the substructure in SHELXD, CRANK2 was used to calculate the electron density map, in which an initial model was built using the primary structure with sulfur-containing residues as anchors [170]. Alas, the resulting 2.80 Å model presented significant difficulties in refinement of the nanobody moiety due to its distorted fold. We therefore performed molecular replacement on a separate dataset collected on I24 using this structure as a model, which resulted in a 2.31 Å structure discussed below. Buccaneer used phases from I23 data to generate an electron density map, from which chains of C_α are built and if possible, assign side chains [171]. The resulting model has an R-free of 31% and completeness of 87%. The model was refined using REFMAC5 [172], which optimises electron density fit, along with: (i) molecular flexibility data obtained using TLSMD [173], which modelled the Wzx/nb10 structure as two bodies, and; (ii) geometric restraints using PDB_REDO [174] and MolProbity [175]. The final model has an R-free of 25.7% (crystallographic table in Appendix C).

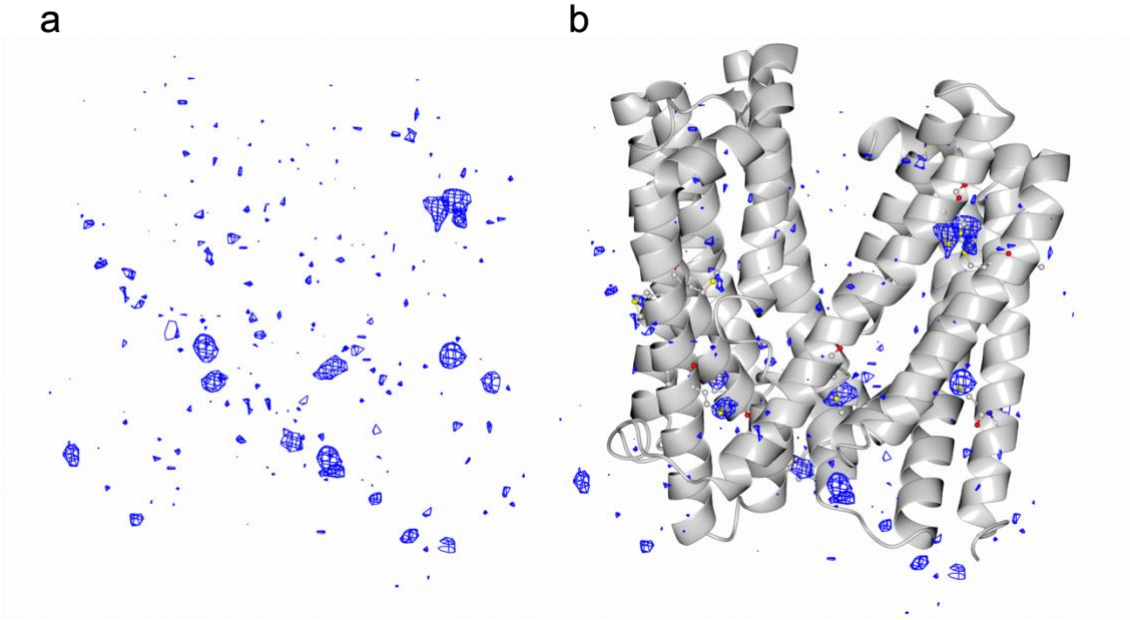


Figure 2.18 S-SAD anomalous signal map *Anomalous difference map generated from SHELXD contoured at σ = 3.0 and clipped as a 50 Å^3 cube (a) superposed on the solved Wzx structure (b), showing signals correspond to sulfur-containing residues (shown as balls and sticks).*

2.d.viii. Apo-structure of *E.* coli K-12 WzxE

Wzx has two α-helical bundles of 6 helices each, together with nb10 forms a monomer in the asymmetric unit (Figure 2.19a). We aligned Wzx with its homologue PAO1 (UniprotKb G3XD19) using EMBOSS Needle [176] to identify corresponding residues used in Islam et al.'s [177] truncation topological assay. They observed whether expression of truncated Wzx leads to a (periplasmic) alkaline phosphatase reaction and/or (cytoplasmic) β-galactosidase reaction. On this basis, we propose that nb10 binds at the periplasmic end of Wzx (Figure 2.19b).

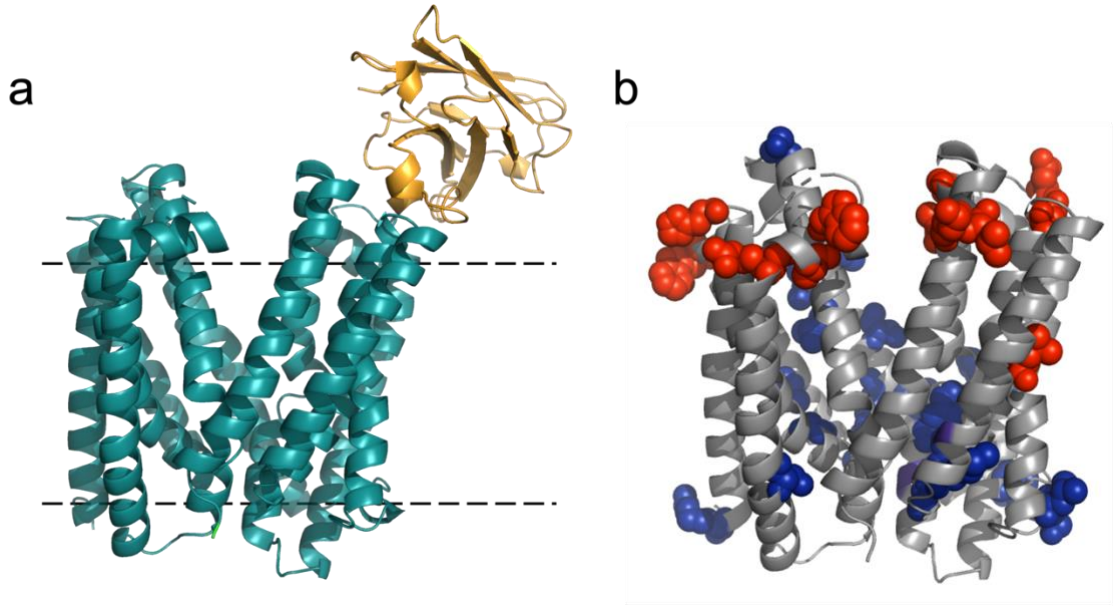
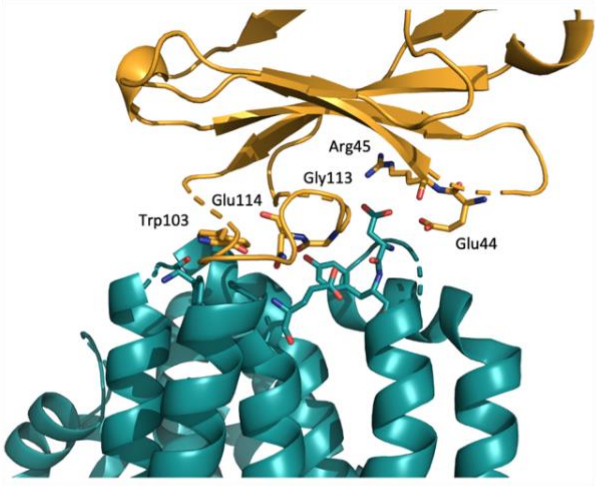


Figure 2.19 Crystalline structure of Wzx/nb10 complex *(a) apo-Wzx (green) and nb10 (yellow) shown as ribbons. Approximate outline of membrane boundaries are shown with a dotted line. (b) Wzx (grey ribbon) with spheres representing residues shown to be periplasmic (red) or cytosolic (indigo) in the Islam et al assay.*

4 Wzx residues (Phe32, Asp114, Tyr115, Glu173) and 5 nb10 residues (Glu44, Arg45, Trp103, Gly113, Asn114) participate in hydrogen bonding, involving both main chain and side chain atoms, at the Wzx/nb10 interface (Figure 2.20a). Interestingly, 2 of the 5 nb10 residues (Glu44 and Arg45), which coordinate Asp114 in Wzx, are highly conserved among other Wzx nanobodies (Figure 2.20b). This raises the worrying possibility that all nanobodies may share the same epitope, contrary to our findings in Section 2.d.iv.

a


Arg45
Gly113
Glu114
Trp103
Glu44

b

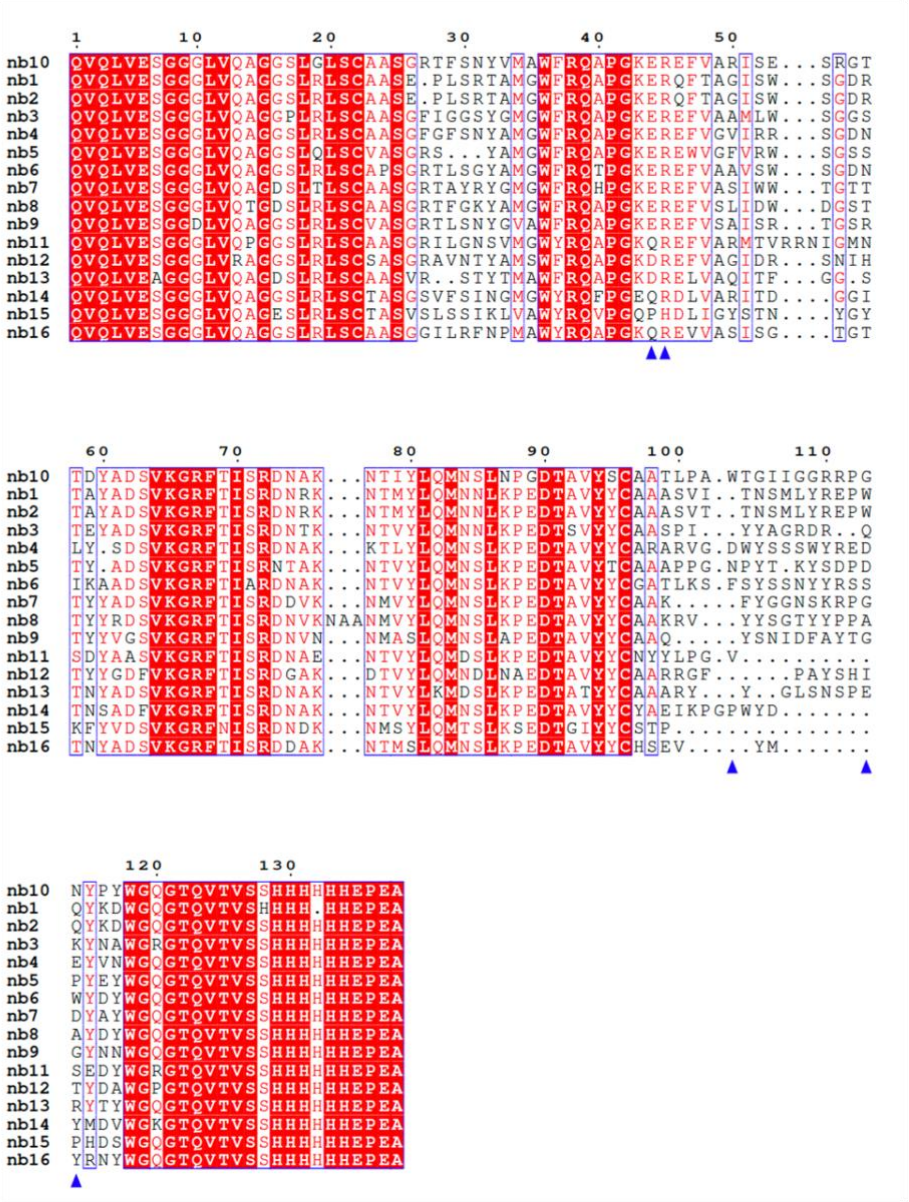

1 10 20 30 40 50
nb10 QVQLVESGGGLVQAGGSLGLSCAASGRTFSNYVMAWFRQAPGKEREFVARISE...SRGT
nb1 QVQLVESGGGLVQAGGSLRLSCAASE.PLSRTAMGWFRQAPGKERQFTAGISW...SGDR
nb2 QVQLVESGGGLVQAGGSLRLSCAASE.PLSRTAMGWFRQAPGKERQFTAGISW...SGDR
nb3 QVQLVESGGGLVQAGGPLRLSCAASGFIGGSYGMGWFRQAPGKEREFVAAMLW...SGGS
nb4 QVQLVESGGGLVQAGGSLRLSCAASGFGFSNYAMGWFRQAPGKEREFVGVIRR...SGDN
nb5 QVQLVESGGGLVQAGGSLQLSCVASGRS...YAMGWFRQAPGKEREWVGFVRW...SGSS
nb6 QVQLVESGGGLVQAGGSLRLSCAPSGRTLSGYAMGWFRQTPGKEREFVAAVSW...SGDN
nb7 QVQLVESGGGLVQAGDSLTLSCAASGRTAYRYGMGWFRQHPGKEREFVASIWW...TGTT
nb8 QVQLVESGGGLVQTGDSLRLSCAASGRTFGKYAMGWFRQAPGKEREFVSLIDW...DGST
nb9 QVQLVESGGDLVQAGGSLRLSCVASGRTLSNYGVAWFRQAPGKEREFVSAISR...TGSR
nb11 QVQLVESGGGLVQPGGSLRLSCAASGRILGNSVMGWYRQAPGKQREFVARMTVRRNIGMN
nb12 QVQLVESGGGLVRAGGSLRLSCSASGRAVNTYAMSWFRQAPGKDREFVAGIDR...SNIH
nb13 QVQLVEAGGGLVQAGDSLRLSCAASVR..STYTMAWFRQAPGKDRELVAQITF...GG.S
nb14 QVQLVESGGGLVQAGGSLRLSCTASGSVFSINGMGWYRQFPGEQRDLVARITD....GGI
nb15 QVQLVESGGGLVQAGESLRLSCTASVSLSSIKLVAWYRQVPGQPHDLIGYSTN....YGY
nb16 QVQLVESGGGLVQAGGSLRLSCAASGGILRFNPMAWYRQAPGKQREVVASISG....TGT
60 70 80 90 100 110
nb10 TDYADSVKGRFTISRDNAK...NTIYLQMNSLNPGDTAVYSCAATLPA.WTGIIGGRRPG
nb1 TAYADSVKGRFTISRDNRK...NTMYLQMNNLKPEDTAVYYCAASVI..TNSMLYREPW
nb2 TAYADSVKGRFTISRDNRK...NTMYLQMNNLKPEDTAVYYCAASVT..TNSMLYREPW
nb3 TEYADSVKGRFTISRDNTK...NTVYLQMNNLKPEDTSVYYCAASPI...YYAGRDR..Q
nb4 LY.SDSVKGRFTISRDNAK...KTLYLQMNSLKPEDTAVYYCARARVG.DWYSSSWYRED
nb5 TY.ADSVKGRFTISRNTAK...NTVYLQMNSLKPEDTAVYTCAAAPPG.NPYT.KYSDPD
nb6 IKAADSVKGRFTIARDNAK...NTVYLQMNSLKPEDTAVYYCGATLKS.FSYSSNYYRSS
nb7 TYYADSVKGRFTISRDDVK...NMVYLQMNSLKPEDTAVYYCAAK.....FYGGNSKRPG
nb8 TYYRDSVKGRFTISRDNVKNAANMVYLQMNSLKPEDTAVYYCAAKRV...YYSGTYYPPA
nb9 TYYVGSVKGRFTISRDNVN...NMASLQMNSLAPEDTAVYYCAAQ.....YSNIDFAYTG
nb11 SDYAASVKGRFTISRDNAE...NTVYLQMDSLKPEDTAVYYCNYYLPG.V.....
nb12 TYYGDFVKGRFTISRDGAK...DTVYLQMNDLNAEDTAVYYCAARRGF.....PAYSHI
nb13 TNYADSVKGRFTISRDNAK...NTVYLKMDSLKPEDTATYYCAAARY...Y..GLSNSPE
nb14 TNSADFVKGRFTISRDNAK...NTVYLQMNSLKPEDTAVYYCYAEIKPGPWYD.....
nb15 KFYVDSVKGRFNISRDNDK...NMSYLQMTSLKSEDTGIYYCSTP.....
nb16 TNYADSVKGRFTISRDDAK...NTMSLQMNSLKPEDTAVYYCHSEV.....YM.....
120 130
nb10 NYPYWGQGTQVTVSSHHHHHHEPEA
nb1 QYKDWGQGTQVTVSHHHH.HHEPEA
nb2 QYKDWGQGTQVTVSSHHHHHHEPEA
nb3 KYNAWGRGTQVTVSSHHHHHHEPEA
nb4 EYVNWGQGTQVTVSSHHHHHHEPEA
nb5 PYEYWGQGTQVTVSSHHHHHHEPEA
nb6 WYDYWGQGTQVTVSSHHHHHHEPEA
nb7 DYAYWGQGTQVTVSSHHHHHHEPEA
nb8 AYDYWGQGTQVTVSSHHHHHHEPEA
nb9 GYNNWGQGTQVTVSSHHHHHHEPEA
nb11 SEDYWGRGTQVTVSSHHHHHHEPEA
nb12 TYDAWGPGTQVTVSSHHHHHHEPEA
nb13 RYTYWGQGTQVTVSSHHHHHHEPEA
nb14 YMDVWGKGTQVTVSSHHHHHHEPEA
nb15 PHDSWGQGTQVTVSSHHHHHHEPEA
nb16 YRNYWGQGTQVTVSSHHHHHHEPEA

Figure 2.20 Wzx binding by nb10 *(a) The Wzx/nb10 interface, coloured as in Figure 2.19, with hydrogen bonding residues shown as sticks and nb10 residues labelled. (b) Multiple Sequence Alignment of Wzx nanobodies generated using Clustal Omega [163] and represented by ESPript [178]. nb10 residues involved in hydrogen bonding with Wzx residues are highlighted with blue triangles.*

Wzx superposes well with homologues NorM and MurJ (Section 2.b.iii.), which share its 12 helix fold (Figure 2.21a) where helices 3 and 8 are kinked. The top view (Figure 2.21b) shows an internal 2-fold pseudo-symmetry also found in NorM and MurJ proteins [142, 155]. Of the three NorM conformers in Figure 2.3, the cation-bound outward-closed conformer superposed best with RMSD of 4.812 Å over 2013 atoms (Figure 2.21c). Of the four MurJ conformers shown in Figure 2.3, the outward-open conformation superposed best with RMSD 3.219 Å over 2022 atoms (Figure 2.21d). Interestingly, the loops between helices 3/4 and 8/9 occlude the substrate binding site in NorM and MurJ but not Wzx, even though the helix 3/4 loop in Wzx is ordered. This favours the proposal of a pore mechanism for Wzx, but an alternating access mechanism cannot be ruled out (Section 2.b.iii.). Further analysis on structural features are discussed in Section 2.d.xi.

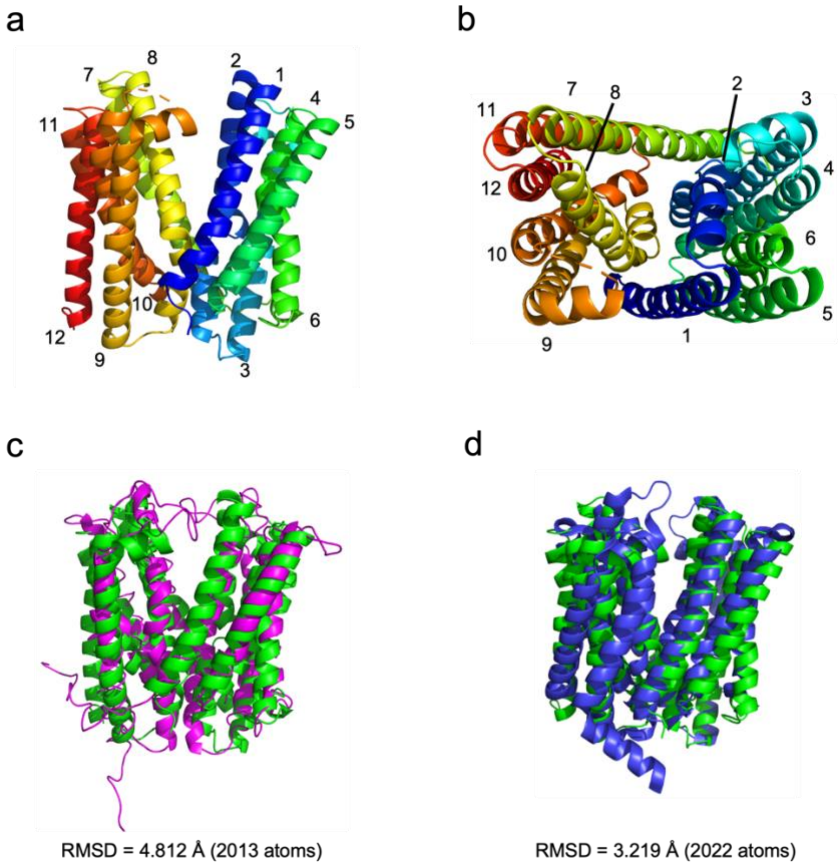


Figure 2.21 Comparison of Wzx structure with homologues *Wzx coloured as rainbow (blue = N-terminus, red = C-terminus) in side view (a) and top view (b), helices are numbered from N to C termini. Wzx (green) supposed with NorM (PDB: 4HUL) (c), MurJ (PDB: 6NC9) (d), PglK (PDB: 6HRC).*

As expected of an integral membrane protein, there is an asymmetric distribution of exposed charged residues towards the cytoplasmic and periplasmic ends of Wzx (Figure 2.22). As in previous flippase structures [152, 155], we observed a hydrophobic groove on the surface of the transmembrane domain (Figure 2.22a) as well as positively charged clusters in the lumen (Figure 2.22c). We also observed an additional cationic cluster in the cytoplasmic moiety (Figure 2.22d). The former is likely to serve as a binding site for the undecaprenyl moiety of its substrate, whilst the latter two could coordinate the pyrophosphate in the substrate. Together, these suggest an anchoring mechanism for Wzx where it binds to the sugar-phosphate moiety in its lumen while the undecaprenyl moiety

remains (partially) free in the inner membrane, although its precise sequence of conformational change remains elusive.

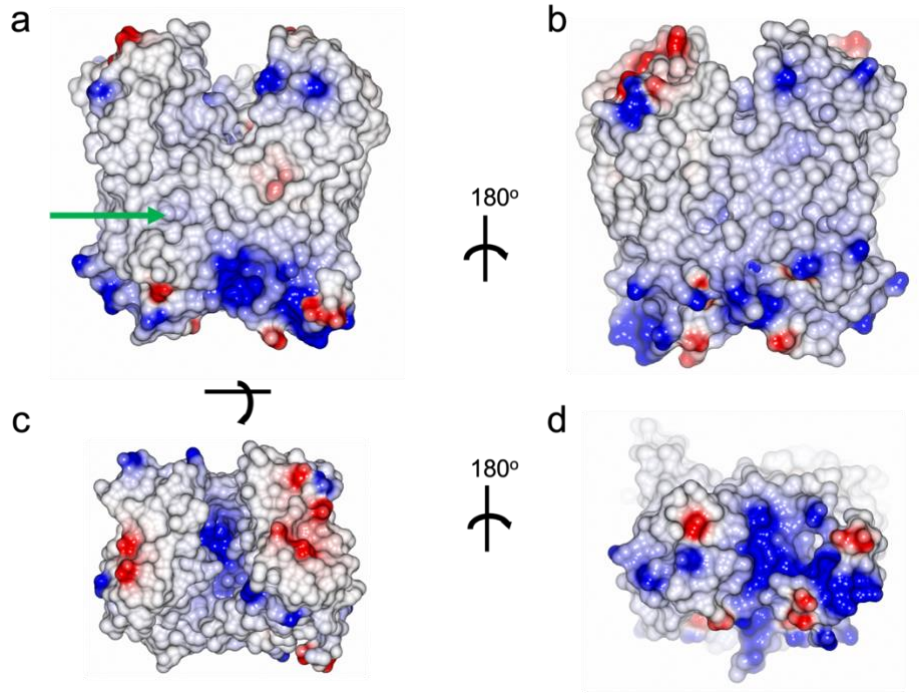


Figure 2.22 Surface representation of Wzx/nb10 structure *Generated using CCP4MG [179], surface representation of Wzx/nb10 coloured by electrostatic potential as calculated in CCP4MG [179] (positive in blue, negative in red). (a) Hydrophobic groove in the transmembrane domain shown by a green arrow, which is not observed on the reverse side (b). (c) Top view from the periplasmic side reveals a cationic lumen. (d) Bottom view from the cytoplasmic side reveals a high concentration of basic residues.*

Using EMBOSS Needle [176] to align the amino acid sequences of Wzx_E and Wzx_{PAO1}, which was used in Islam et al.'s homology model [141]. Unfortunately, sequence identity (21.5%) and similarity (36.5%) were low and many lumen-facing residues identified as functionally important were not conserved (Figure 2.23). Of these, only Phe139 (Phe143 in Wzx_E) was conserved. Arg146, Asp269, and Lys272 were aligned to physiochemically similar residues, whereas Arg59-Glu61 and Asp359 were not at all conserved.

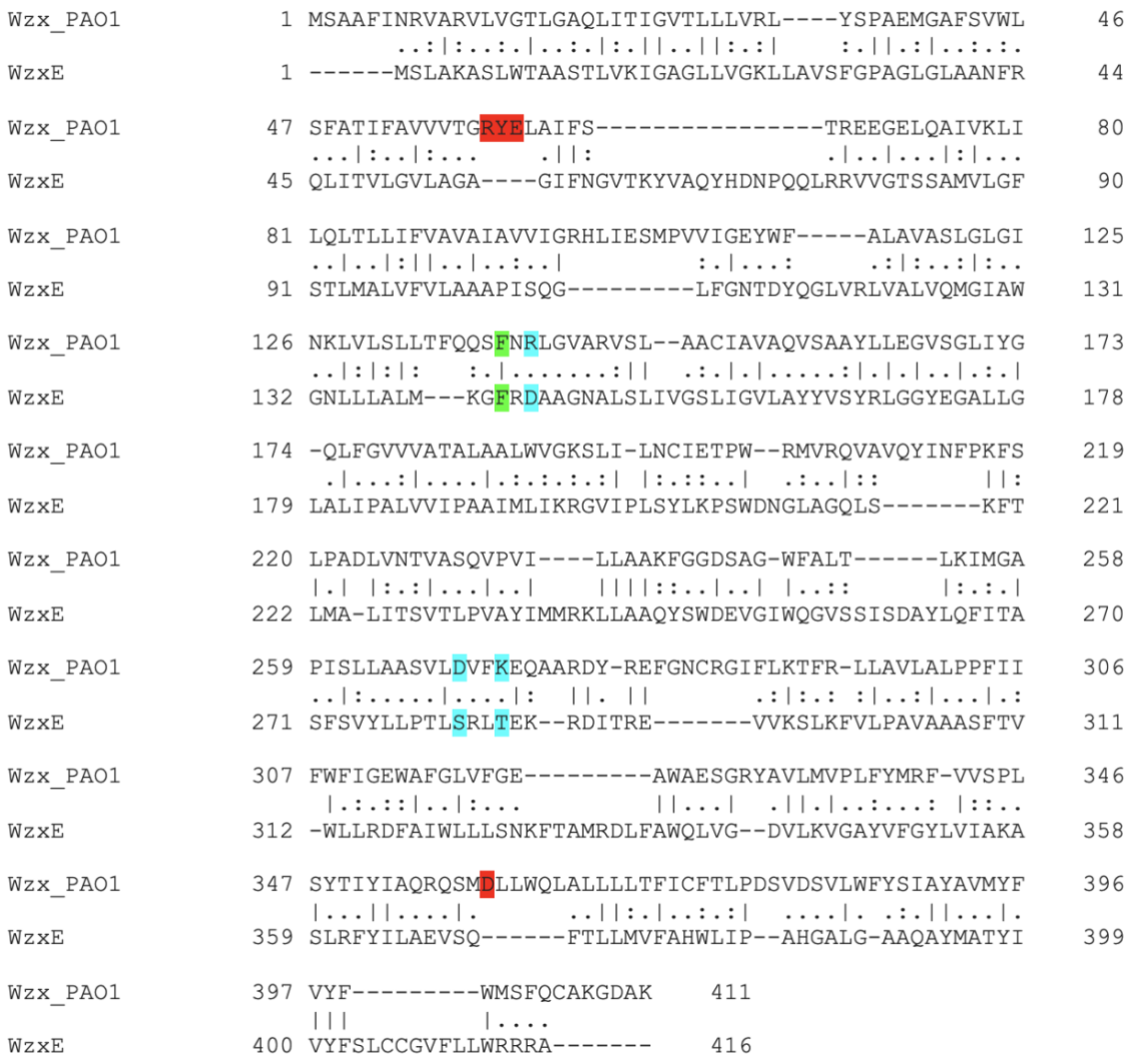


Figure 2.23 Low sequence conservation between Wzx_E and Wzx_{PAO1} *Pairwise alignment of Wzx_{PAO1} (top) and Wzx_E, with dashes indicating conserved residues pairs whilst dot represent physicochemically similar residues. Residues deemed functionally important by Islam et al's homology model [141] are highlighted in green (conserved), cyan (aligned with similar residue), and red (not conserved).*

2.d.ix. Substrate complex crystallisation trials

We attempted to obtain structures of Wzx-substrate complex by soaking Wzx-nb10 crystals. Due to the insolubility of lipid III, we used substrate analogues that mimic part of the lipid III structure. This includes geranyl pyrophosphate, farnesyl pyrophosphate, pyrophosphate, UDP-GlcNAc, GlcNAc-1-phosphate , and GlcNAc (Figure 2.24).

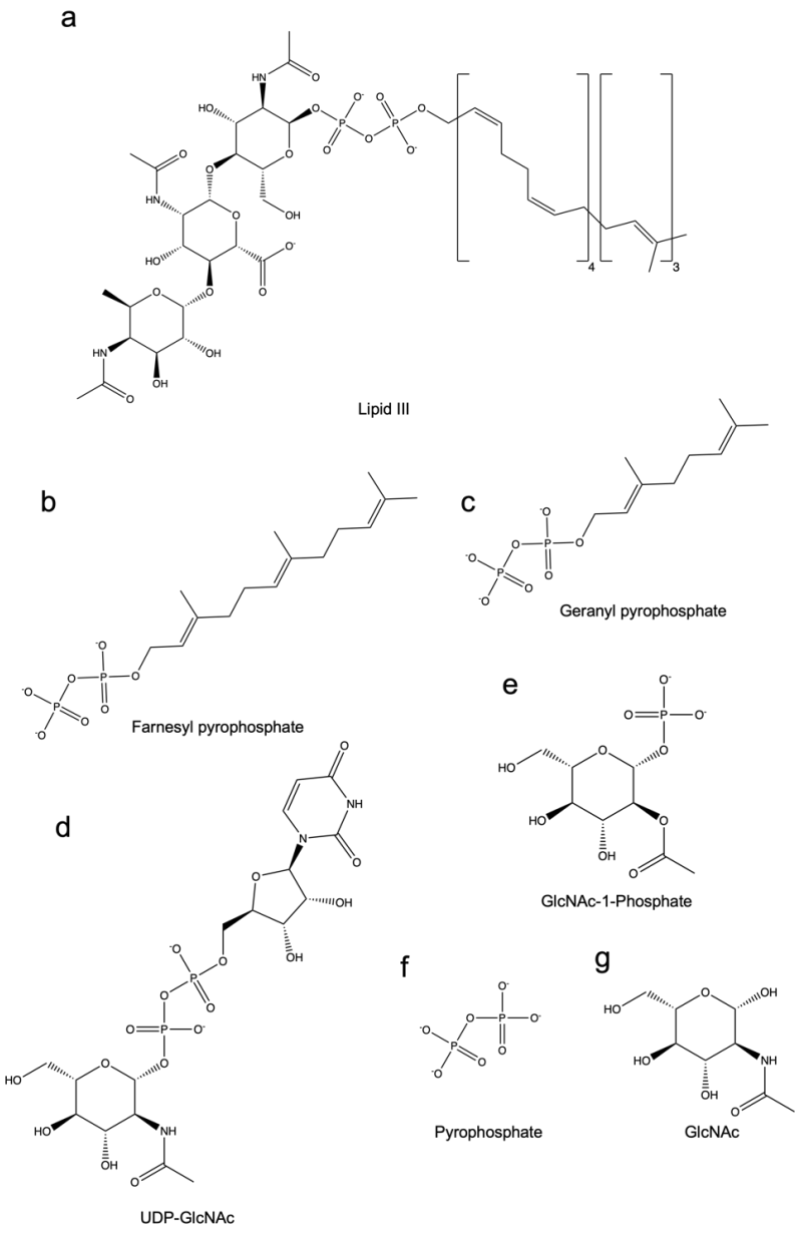


Figure 2.24 Substrate analogue structures *Chemical structures of Wzx substrate Lipid III (a) and analogues used in co-crystallisation trials (b-g).*

Unfortunately, Wzx crystals did not form when co-crystallised with any of these substrates. We therefore grew Wzx/nb10 crystals (Section 2.c.i.) and soaked in 5-10 mM substrate analogue solutions for up to 5 min during harvesting. Diffracting crystals were yielded for geranyl-pyrophosphate, UDP-GlcNAc, and GlcNAc. Crystals were then exposed at the I24 beamline, using a 20 x 20 μm beam at 20% transmission and wavelength of 0.9686 Å. Unfortunately, we did not observe significant substrate density at either the hydrophobic groove or cationic lumen (Figure 2.25).

This may indicate that nb10 and substrate binding are mutually exclusive, as substrate binding prevented crystallisation due to the lack of nanobody stabilisation and nanobody binding causes occlusion of the substrate binding site.

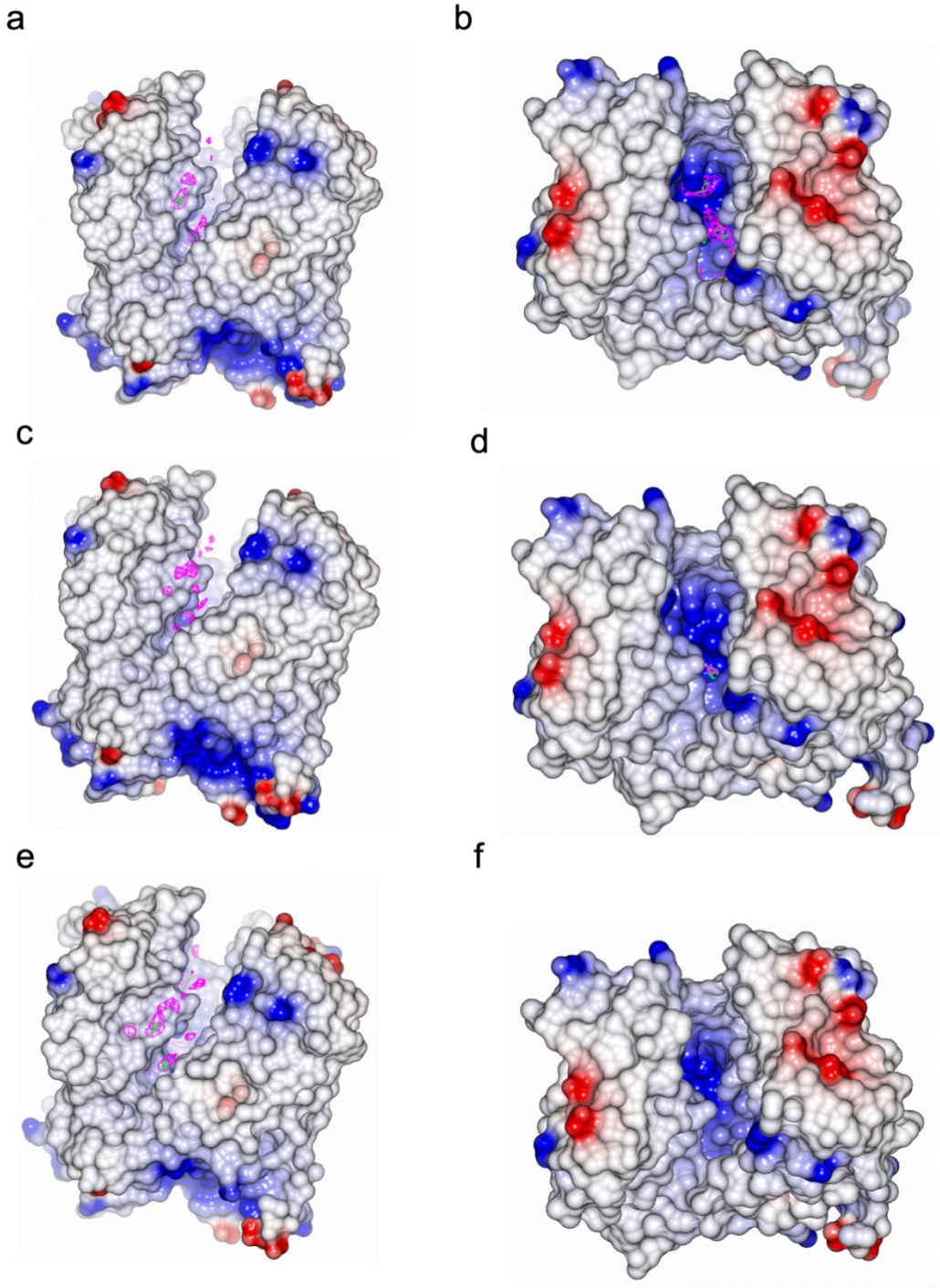


Figure 2.25 No significant substrate density in soaked Wzx crystals *$2F_O$-F_C (magenta) and F_O-F_C (green for positive and red for negative) maps, contoured at σ = 1.0 and σ = 3.0 respectively, overlaid on crystalline structures solved for soaked Wzx/nb10 crystals showing hydrophobic groove (a, c, e) and cationic lumen (b, d, f). All of crystals soaked with GlcNAc (a, b), UDP-GlcNAc (c, d), and geranyl pyrophosphate (e, f) did not show significant substrate density.*

2.d.x. *In Meso In Situ* Crystallography (IMISX)

We investigated the use of IMISX plates, introduced in Section 1.b.iii., as an alternative to long wavelength crystallography and to allow for longer substrate soaking times of up to 24 hrs. Due to mount compatibility requirements, IMISX plates were only used at the I24 microfocus beamline at Diamond light source. However, we ventured that the increased protection against radiation damage may compensate for the lower anomalous signal observed at conventional wavelengths, as we observed in the comparison of diffraction limits between I03 and I23 reported in Sections 2.d.ii. and 2.d.iii. Additionally, the increased resistance to bolus dehydration allows for longer periods of soaking when obtaining a substrate complex structure.

We collaborated with the Caffrey Lab at Trinity College, Dublin to explore options for preparing in-house plates, as a cost-effective alternative to the commercially available IMISX kit. As described in Section 2.c.iii., assembly requires standard LCP equipment with the exception of custom-ordered cyclic olefin polymer (COP) film, an oversized glass cover slip, and a double-side tape gasket to secure the film onto the standard LCP glass back plate. We were able to grow lysozyme LCP crystals within one hour using the in house plates, soak crystals in solution for an hour at room temperature without the LCP bolus drying out, and secure the film onto custom-made, I24-comptaible crystal pins. Since we obtained the Wzx structure, the plates were ultimately unused and subsequent testing was not performed. However, we note that IMISX plates would be suitable for co-crystallisation trials once a binding substrate analogue has been identified. Nonetheless, the technology was transferred to the Membrane Protein Laboratory at the Research Complex at Harwell.

In collaboration with the Membrane Protein Laboratory and VMXi beamline, we compared the IMISX setup with an alternative film plate in development by SWISSCI, which is not encased in glass and has been used for conventional *in situ* crystallography but not LCP. Since IMISX plates use the standard SBS format, LCP crystals can be set up identically to glass plates, in the standard 50:800 µL bolus:reservoir ratio. The smaller wells in the SWISSCI plate meant the crystallisation solution was pushed to the edge when sealed and would likely need to be reoptimised (Figure 2.26). This diminishes the reservoir/air interface and may affect LCP formation and subsequent crystal growth [180]. However, due to the lack of adhered glass backing, individual wells do not have to be removed before x-ray exposure for SWISSCI plates as in IMISX plates. Therefore, although crystals may need to be reoptimised for SWSSCI plate conditions, it could be more suitable for projects at the earlier stage where extensive screening is required.

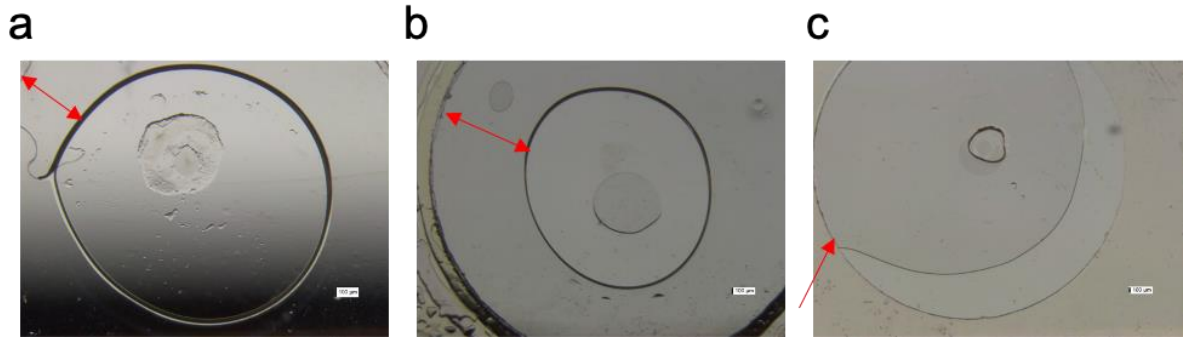


Figure 2.26 IMISX is more similar to standard LCP glass plate than SWISSCI *Bright field microscope images of an LCP drop in standard glass plate (a), IMISX plate (b), and SWISSCI film plate (c). A gap between the edge of the reservoir solution, which surrounds the LCP bolus, and the edge of the wells of glass and IMISX plates are shown with red bidirectional arrows (a,b). In contrast, the reservoir solution is spread to the edge of the well in the SWISSCI plate as shown with a red arrow (c).*

2.d.xi. Comparison with structural features in homologues

In light of the lack of an alternative conformer, we investigated whether structural features found in MurJ can be identified in Wzx due to the structural similarity between their substrates Lipid II and Lipid III (Figure 2.2). Firstly, MurJ was co-crystallised with Na^+ and is well coordinated in the outward bound structure [156] (Figure 2.27). Mutation of coordinating residues also lead to non-expression of MurJ, hinting at a role in protein folding [156]. Indeed, the Islam homology model suggested that Wzx could be an antiporter [143]. Whilst Wzx needs not be an antiporter, antiporter activity could improve its catalytic rate. However, side chains of equivalent residues in Wzx do not appear to be able to coordinate a cation. In particular, Asn374 in MurJ is replaced by Leu375 in Wzx which cannot coordinate a cation, and Asp235 in MurJ is replaced by Met239 in Wzx, the side chain of which occupies the Na^+ binding site in MurJ. Secondly, the Arg24/Asp25/Arg255 triad forms a cluster in the cytoplasm-facing but not periplasm-facing MurJ structures, which led the authors to propose that they function as an anchor point for an alternative access model [156]. Equivalent arginine residues could not be found in Wzx. Together, these suggest that Wzx does not use the alternating access model proposed for MurJ.

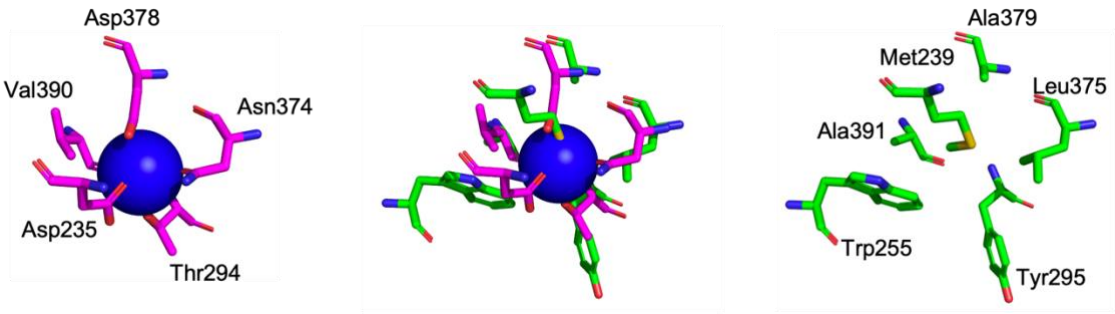


Figure 2.27 Wzx could not coordinate cation like MurJ *Superposition (middle) of cocrystallised Na^+ (blue sphere) and coordinating residues in MurJ (left) with equivalent residues in Wzx (right).*

We performed the same analysis on NorM, which was co-crystallised with caesium, though it is less conclusive as only Glu260 definitively coordinates the ion (Figure 2.28). However, we noticed that all residues within 5 Å of the ion are acidic, which is also true of Wzx when superposed (Figure 2.28). Therefore, this presents a more suitable cation binding site if Wzx is indeed an antiporter.

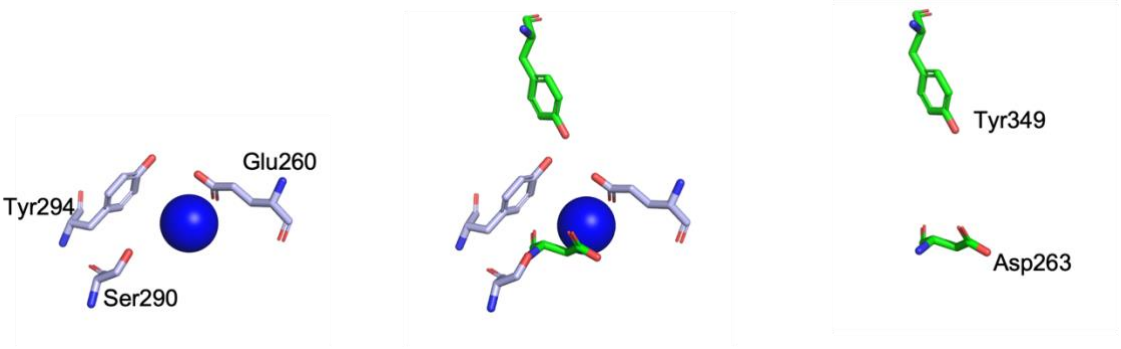


Figure 2.28 Wzx could coordinate cation like NorM *Superposition (middle) of cocrystallised Cs^+ (blue sphere) and neighbouring residues in NorM (left) with neighbouring residues in Wzx (right).*

2.e. Conclusion

Here we reported the optimisation of Wzx purification and crystallisation, most notably the addition of exogenous lipids to the LCP bolus. The resulting crystals were of sufficient diffraction resolution and robustness for structure determination using I23. Nonetheless, the background scattering and delicacy of LCP crystals necessitated optimisation of the data collection protocol, namely grid scanning and reduction of beam size. We then reported the data processing pipeline used to integrate 20 360° crystallographic datasets to generate the redundancy required for structure determination.

The Wzx structure resembles that of other determined flippases [142, 155], with 2 helical bundles, a cationic lumen, and a hydrophobic groove to accommodate the binding of

lipids linked to polar headgroups. Unfortunately, the apo-structure alone provided limited biological information on an enzyme that is expected undergo large scale conformational changes in its catalytic cycle. Furthermore, our observation that there are two positively charged clusters in the cytoplasmic face adds further complications to the mechanism of Wzx, and would likely require substrate complex structures to fully elucidate.

To obtain further biological information on Wzx, structures of alternative conformers and substrate-bound states need to be obtained. To pursue the former, we attempted to use a double pulldown assay to screen from 8 nanobodies candidates for binders to different epitopes than nb10, with which the apo-structure was obtained. Unfortunately, our results were inconclusive and would require scaling up to produce unambiguous results. There will likely be significant challenge in obtaining the crystalline structure even after such a nanobody candidate is identified, because many optimisation techniques were already employed to obtain the crystals with which the apo-structure was determined. This means each stage, from detergent solubilisation to exogenous lipid addition, will likely need to be re-optimised and would therefore require an efficient re-screening methodology. To pursue the latter, we conducted initial co-crystallisation trials which have failed. This may require additional optimisation such as through the use of IMISX plates.

2.f. Discussion and future directions

We showed that delicate LCP crystals are amenable to long wavelength native s-SAD. Although this is helpful in scenarios where conventional phasing methods are unsuccessful, determination of Wzx still required high redundancy. Long wavelength SAD may also be applied to phosphorous and may be used in conjunction with native s-SAD [181]. This also allows auxiliary advantages such as using the phosphorous substructure to aid refinement of densities from phosphorous-containing molecules like exogenous phospholipids, which ordinarily would be indistinguishable from detergent or monoolein molecules.

The Wzx structure is inconsistent with the alternating access mechanism proposed for MurJ, which has a similar substrate, on the basis that (i) substrate binding site is not occluded despite sharing the conformer (Section 2.d.viii.) and (ii) Wzx does not utilise the same arginine clusters for inducing conformational change (Section 2.d.xi.). However, with only one conformer solved the mechanism cannot be ruled out. Though Wzx could not use the same cation binding site as MurJ, it is compatible with that of NorM. The two arginine clusters identified in Wzx are key targets for site directed mutagenesis in further studies. Since arginine clusters induce conformational changes in MurJ [156], site-directed mutagenesis at these sites could be used to obtain alternative conformers. We hypothesise that if Wzx would undergo large scale conformational changes in its catalytic cycle, which may be observed by determination of a cytoplasmic-facing conformer.

3. Structure determination of sialic acid oxidoreductase RgNanOx

3.a. Abstract

We report the experimental work performed on the sialic acid oxidoreductase RgNanOx to determined its holo- and substrate complex structures. The enzyme forms part of a metabolic pathway in *R. gnavus*, a gut bacterium implicated in inflammatory bowel disease [182], where sialic acid (Neu5Ac) found in glycoproteins is converted to a bicyclic compound 2,7-anhydro-Neu5Ac [183]. We published the RgNanOx holo-structure in Bell et al 2020 [184], which was solved at 2.58 Å by molecular replacement. Using this structure and data obtained published by our collaborators [183-185], we proposed a general reaction mechanism for RgNanOx .

Using LCMS and NMR we functionally characterised mutants to confirm the importance of histidine residues in the novel HHIH catalytic motif in RgNanOx, as well as its activity against the common sialidase inhibitor N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid (DANA). This allowed the determination of two substrate complex structures. The Neu complex structure was determined by incubating WT RgNanOx with DANA, and was resolved at 2.30 Å. The unusual anomer of Neu found in this structure led us to revise our mechanism, which is now consistent with the findings of our functional studies. The H178A mutant was used to obtain the Anh complex structure, determined at 1.96 Å. This structure elucidated on the active site morphology when bound to a ring-flipped substrate, allowing us to complete our proposed RgNanOx mechanism by assigning catalytic residues.

3.b. Introduction

3.b.i. Sialic acid metabolism in gut bacteria

Gut bacteria can colonise human hosts from infancy [186] and form what is known as the microbiome, which is essential to human physiology [187]. This host-bacteria interaction means that metabolites available to bacteria in the gut environment, which affects the composition of the microbiome, can have significant health implications in the human host from obesity to cancer [188].

Sialic acids are commonly the terminal unit of vertebrate glycoproteins, of which N-acetylneuraminic acid (Neu5Ac, henceforth Neu) is the most abundant [189]. Owing to their prevalence as mucin epitopes, sialic acids are commonly catabolised by gut bacteria such as *R. gnavus* [190]. The significance of its terminal position is two-fold. First, it is the mucin interface between the co-evolving gut microbiome and host, the former of which are adapted to consume the glycans of the latter [191, 192]. Secondly, this led to the possession of *nan* clusters in bacteria that allows downstream processing of released glycans. The canonical *nanA/K/E* cluster, discovered in the model organism *E. coli*, encodes a series of enzymes that converts Neu into N-acetyleglucosamine-6-phosphate (GlcNAc-6-P) [193], allowing it to be catabolised. Alternatively in B. fragilis, the *nanL/E/T* cluster converts Neu into N-acetyl mannosamine (ManNAc) and its downstream catabolism, as knockout B. fragilis cannot grow on ManNAc as its sole source of nutrient [194] (Figure 3.1).

The gut bacteria *R. gnavus* uses a different mechanism to release sialic acid from glycan chains with the enzyme NanH, which acts on 2,3-linked sialylated glycans found in the human gut [195]. Its mechanism was structurally and functionally characterised, and was

shown to be an intramolecular-trans (IT) sialidase that releases 2,7-anhydroneuraminic acid (henceforth Anh) [196]. Anh is transported into *R. gnavus* cells using ATP-binding cassette transporters, a mechanism also used for Neu transport, in complex with a Anh-specific binding protein RgSBP [183].

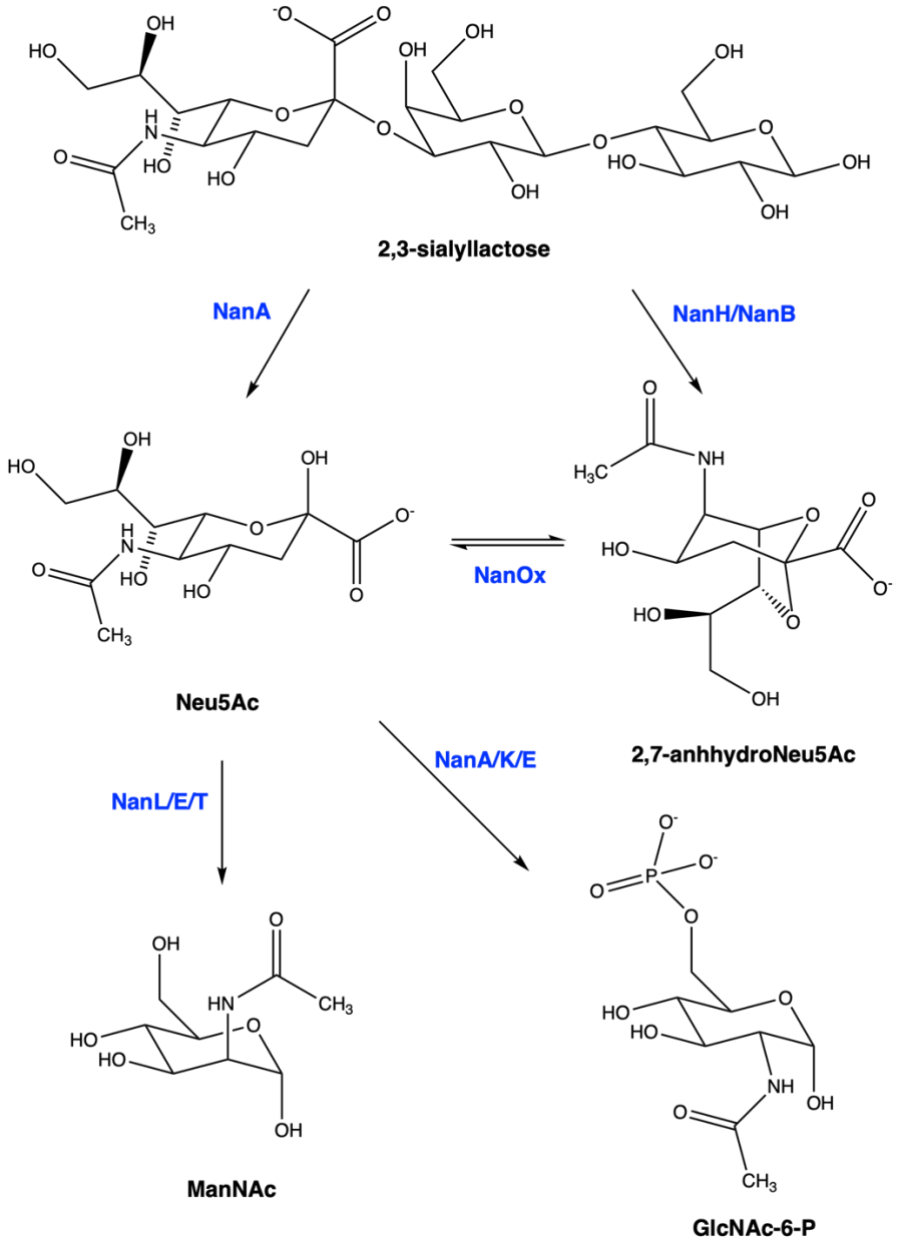


Figure 3.1 Sialic acid metabolism in bacteria *Molecular structures of products from sialic acid metabolism in bacteria. Enzymes or clusters of enzymes responsible for each reaction are shown in blue.*

3.b.ii. Canonical enzymatic mechanisms at the anomeric centre

The anomeric centre has a wider range of chemistries open to it than a normal tertiary or quaternary carbon. This stems from the fact that the ring oxygen can stabilise a positive charge at the anomeric position via an oxocarbenium cation (Figure 3.2). This means that the anomeric position is reactive under biological positions. Stereospecificity of a chemical reaction has been used to classify reaction mechanisms. In the case of nucleophilic substitutions for example, the two-step S_N1 mechanism results in formation of a cationic intermediate that can be attacked from any direction, causing a mixture of stereo centres in products. Conversely in the one-step S_N2 mechanism, nucleophilic attack from one direction is concerted with the leaving group's exit in the opposite direction, causing uniform stereochemistry in products. Such a methodology has been adapted to enzyme-catalysed reactions as early as 1950's [197].

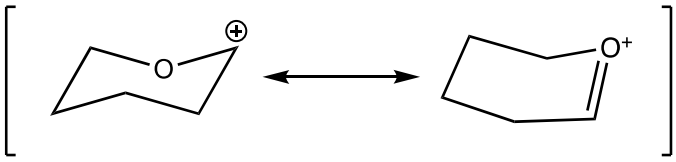


Figure 3.2 The oxocarbenium ion *Resonance structures of the oxocarbenium ion that allows stabilisation of the positive charge at the anomeric position*

Since bound substrates are surrounded in the active site, interacting with residues in multiple directions, the classification of enzymatic reactions is broadly divided between those that retain, and those that invert, stereocentres in the substrate [197]. Where there is no direct displacement of a species bound to the stereocentre via the scissile bond (Figure 3.3a), inverting enzymes go through single enzyme displacement (Figure 3.3b) whilst retaining enzymes go through double displacement (Figure 3.3c) [197]. Of particular note for this project is a mechanism proposed for retaining enzymes, whereby

the stereocentre is made susceptible to nucleophilic substitution through formation of an intra-molecular bond [197, 198] (Figure 3.3d). It was proposed that the formation of the bicyclic Anh is formed by such a mechanism, whereby 7' OH in the Neu side chain attacks the 2' position.

The re-examination of the mechanism of lysozyme [199], a retaining glycosidase, exemplifies the diversity of mechanisms within these classifications and the difficulty in distinguishing between them. Using the crystalline structure of lysozyme [200], the first protein crystalline structure to be solved, the catalytic residues Glu35 and Asp52 were identified and the canonical 'Philips Mechanism' was proposed [201]. Under this scheme, the negative charge of these residues stabilise a oxocarbenium ion that is later hydrated (Figure 3.4). However, using mass spectrometry Vocadlo et al have detected an enzyme-covalent intermediate [199]. This proved that the enzyme uses an alternative mechanism, whereby Asp52 forms a covalent linkage that is subsequently hydrolysed by nucleophilic substitution (Figure 3.4).

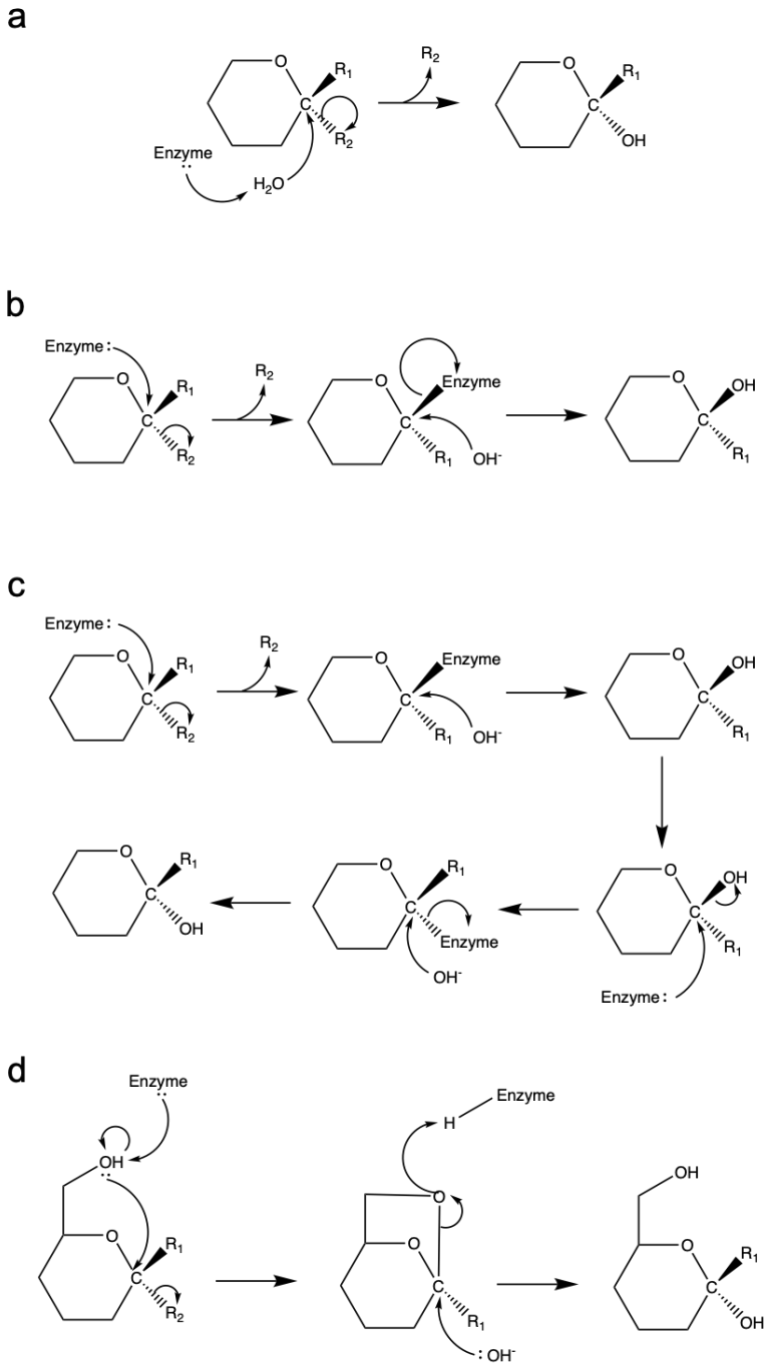


Figure 3.3 Classification of canonical enzyme mechanisms *Schematic representation of different enzyme-catalysed hydrolysis reactions, with the following mechanisms: (a) direct replacement, (b) inverting, (c) retaining, and (d) internal assistance*

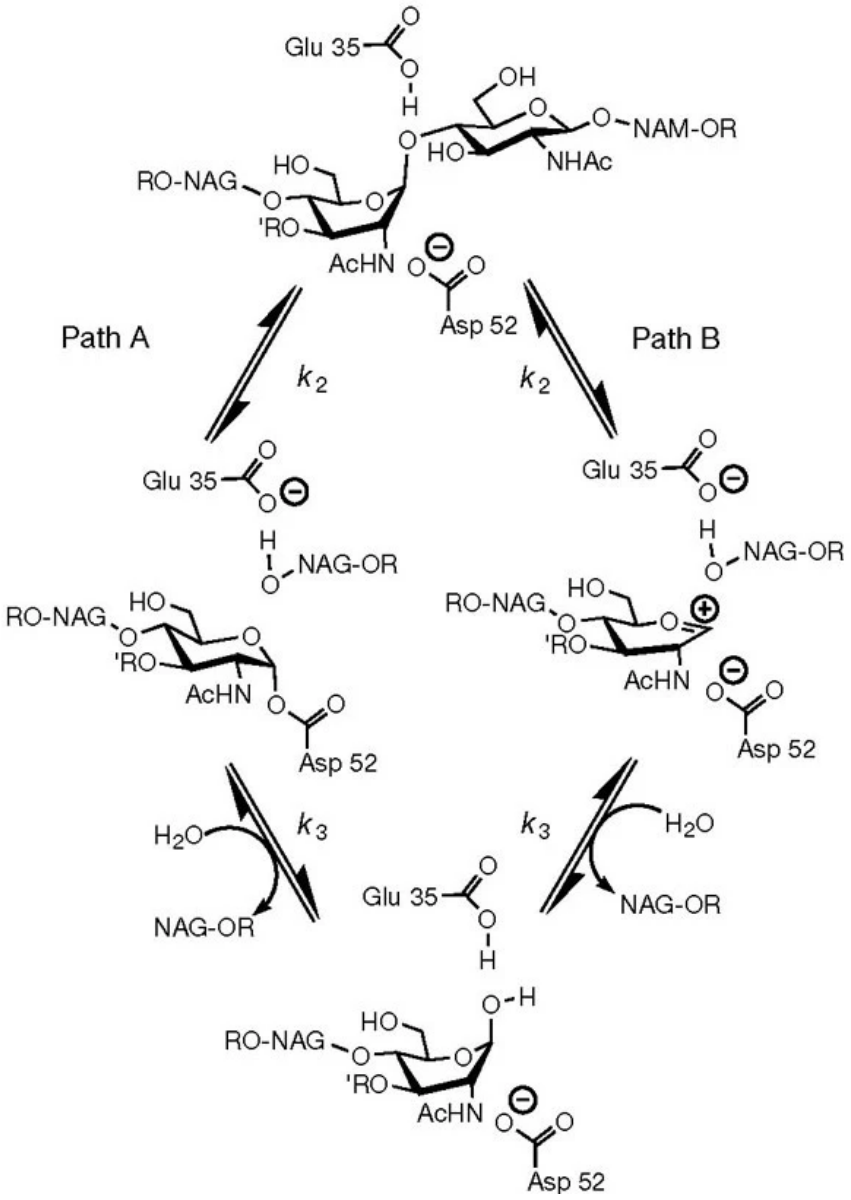


Figure 3.4 Mechanisms proposed for lysozyme *Extracted from Vocadlo et al [199]. The substrate of lysozyme, repeating unit of murin, is shown on the top. The mechanism was later proven by Vocadlo et al [199] to involve an enzyme-substrate intermediate via Asp52 (Path A). The canonical Philips mechanism initially proposed involves an oxocarbenium intermediate (Path B). Both intermediates react with water to complete the hydrolysis as shown at the bottom.*

3.b.iii. Non-canonical glycosidase mechanisms

Glycosidases that do not conform to these classical mechanisms have since been reported, two of which are particularly relevant to this project. Firstly, the GH4 and GH109 glucoside hydrolase families utilise NAD^+ as a cofactor in S_N1 mechanisms that do not involve the formation of a cationic intermediate [202]. This includes BglT, or 6-phospho-α-glucosidase, which is a GH4 glycosidase that utilises a cation as a cofactor in addition to NAD^+ [203]. In its mechanism (Figure 3.5), NAD^+ remains bound to the enzyme in the entire catalytic cycle, to oxidise 6-phospho-α-glucose at the 3' hydroxyl. This results in a 3' keto group that acidifies the adjacent 2' proton, which can then be abstracted by a base. This forms a 1'-2' double bond that mediates the exit of the O-linked 2' leaving group. The cationic cofactor, Mn^{2+} in this case, was cocrystallised with the enzyme where it coordinates the 3' and 2' hydroxyl groups [203]. This position suggests that it ensures that substrate stereocentres are retained, as NMR studies detected only the presence of β anomers at the 2' position [203]. This has interesting implications regarding the mechanisms of GH109 enzymes which does not utilise cationic cofactors [202]. Sadly, activity against α anomers by GH109 members such as α-N-acetylgalactosaminidase remains poorly characterised [204]. Similar mechanisms were later reported in other glucosidases GlvA from *B. subtilis* [205] and α-N-acetylgalactosaminidase from *E. meningosepticum* [206].

Sialidases use acidic and basic residues to hydrolyse an enzyme-linked substrate intermediate [202]. The catalytic versatility of this reaction is exemplified by NanA/B/C, a family of three sialidases discovered in the human respiratory pathogen Streptococcus pneumoniae [207]. The shared substrate α-2,3-sialyllactose is hydrolysed into Neu5Ac, 2,7-anhydroNeu5Ac, and N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid respectively

(Figure 3.1). (The former two products are substrates of RgNanOx.) Their crystalline structures reveals shared conserved catalytic residues [208, 209] that constitute a reaction mechanism (Figure 3.6) shared between the enzymes. Tyrosine forms a covalent linkage with the 2'carbon, which increases the substrate's susceptibility to nucleophilic attack by water activated by a proximal aspartate [207]. In this scheme, the site of nucleophilic attack determines the product. In NanA and NanC, the water attacks the 2' carbon and 3' proton respectively. In NanB, the water activates 7' hydroxyl that subsequently attacks the 2' carbon.

As discussed below, RgNanOx is a novel enzyme whose mechanism uses a hybrid of the two mechanisms described above. Our proposed mechanism is akin to the GH4 family Michael reaction, sharing the cofactor NAD^+ and its acidification of the adjacent proton but not the use of an anionic cofactor [203]. This allows dehydration to form a double-bonded intermediate, similar to NanC, followed by an intramolecular attack akin to NanB.

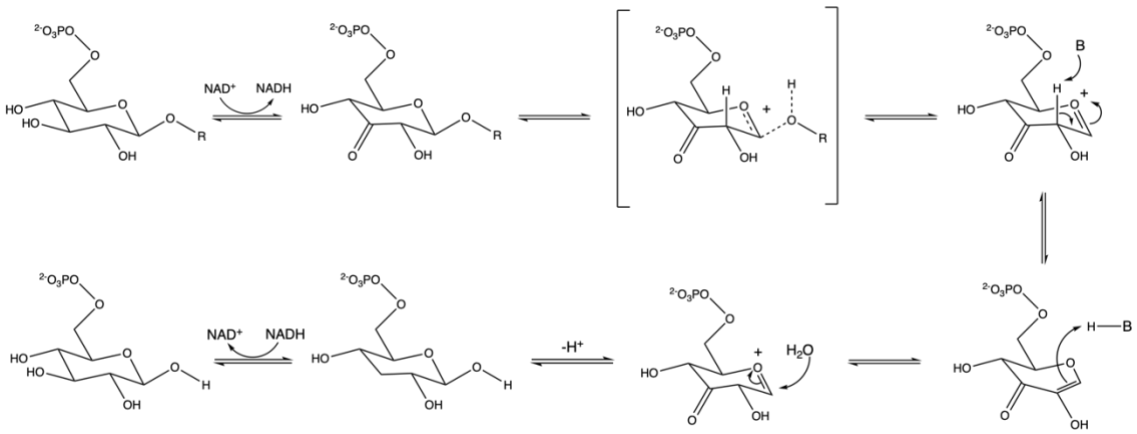


Figure 3.5 BglT mechanism *Adapted from Yip et al [203], mechanism of glycosidase BglT involving NAD^+-dependent oxidation and reduction.*

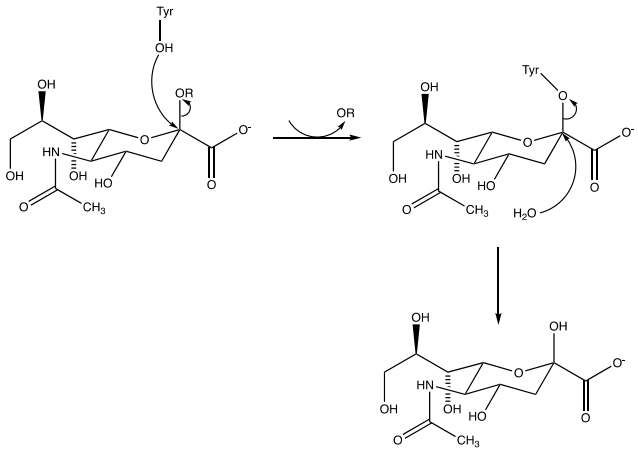


Figure 3.6 Simplified NanA mechanism *Adapted from Xu et al [207], mechanism of neuraminidase NanA involving a covalently linked intermediate.*

3.b.iv. RgNanOx is a sialic acid oxidoreductase in *Ruminococcus gnavus*

Our collaborators (A. Bell and N. Juge, University of East Anglia) identified the gene *rumgna02695*, which is coexpressed with NanH and encodes an enzyme RgNanOx that converts Anh into Neu. RgNanOx was identified as a member of the Gfo/Idh/MocA family of oxidoreductases (accession PF02894) through two significant domain matches on the Pfam database [210]. The two hits both correspond to functional domains found in the family, NAD^+ binding Rossman domain (Clan CL0063, E-value = $3.1e^{-28}$) and C-terminal alpha/beta domain (Clan CL0139, E-value = $2.7\ e^{-21}$) (Figure 3.7).

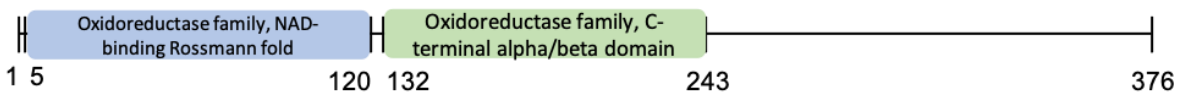


Figure 3.7 Pfam identified two domains in RgNanOx *The RgNanOx gene represented as a black line, labelled with residue numbers and domains identified through Pfam.*

The Gfo/Idh/MocA is a structural family consists of NAD(P)-dependent oxidoreductases. Its functional versatility is illustrated by its namesake; Gfo couples the oxidation of glucose and reduction of fructose to generate sorbitol and gluconolactone using a non-dissociable $NADP^+$ (Figure 3.8a) [211]; Idh catalyses the oxidation of myoinositol to form scyllo-inosose via a double-displacement reaction with NAD^+ (Figure 3.8b) [212]. Enzymes in this family can also process larger substrates, most notably α-N-acetylgalactosaminidase [204] which cleaves the blood-type determining GalNAc, and biliverdin reductase (Figure 3.8c) [213]. Interestingly, both the RgNanOx and α-N-acetylgalactosaminidase reactions are redox neutral overall. Although crystalline structures of many members have been obtained, BLAST search [214] through PDB entries resulted in only poor matches (sequence identity < 30%), except for YjhC in *E. coli*. We therefore initially pursued structure determination using selenomethionine concurrently with molecular replacement, the latter of which was ultimate successful.

Sequence similarity network analysis previously implicated the *E. coli* yjhB/C operon in the catabolism of Anh [183]. Preliminary data showed mutants of either gene can grow on Neu but not Anh. Hydrophobicity prediction and complementation assays show YjhB is likely to be Anh permease [215]. The holo-structure of YjhC has recently been determined by the Dobson group [216] and shows significant similarity, including the HHXH catalytic motif, indicating it is likely to be an orthologue of RgNanOx. They were also able to detect Neu and NAD(H) binding using differential scanning fluorimetry [216]. Interestingly, there is no known IT sialidase orthologue in *E. coli*, raising the possibility that *E. coli* utilises the Anh released by other species such as *R. gnavus*.

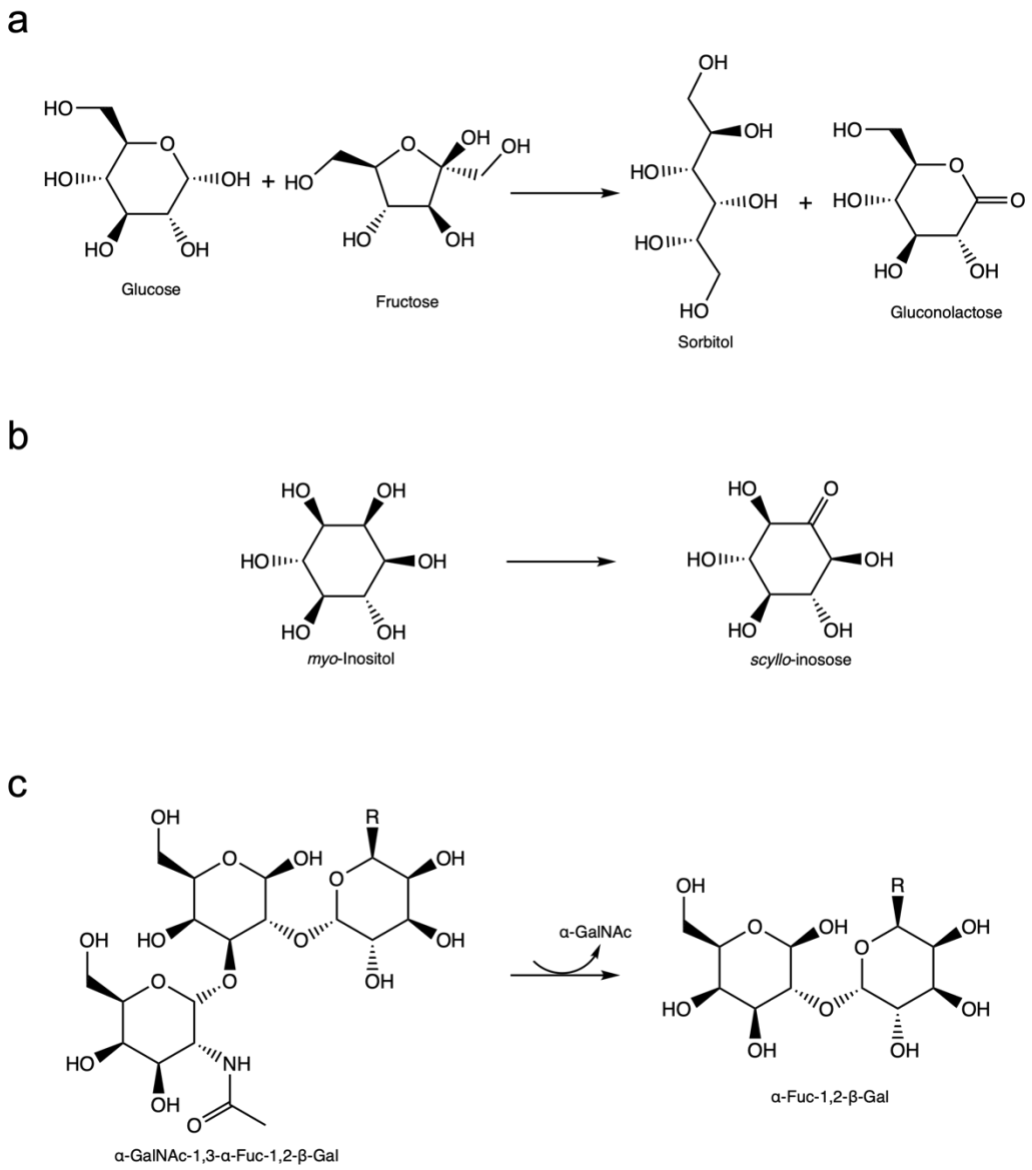


Figure 3.8 Catalytic versatility in the Gfo/Idh/MocA family *Reactions catalysed by (a) glucose-fructose oxidoreductase, (b) myo-inositol dehydrogenase, and (c) α-N-acetylgalactosaminidase*

We jointly published with collaborators in the Juge lab our findings on the molecular basis of RgNanOx catalysis in Bell et al 2020 [184]. Using NMR our collaborators observed deuterium exchange at the 3' position, which is only possible through the alpha keto effect by oxidation of the adjacent 4' hydroxyl to form a 4-keto-DANA intermediate [184]. Along with the holo-RgNanOx structure we determined that showed a typical Gfo/Idh/MocA Rossman fold (Section 3.d.iii.) and our collaborator's previous finding

that RgNanOx activity is dependent on NAD^+ [183], we propose that NAD^+ is used to mediate 4' OH oxidation. Since the Juge group's previously published showed that RgNanOx converts Neu to Anh [185], the producvt in which has lost its 2'OH, we proposed a general reaction scheme that involved elimination across 2' and 3' (Figure 3.9) [184], which is similar to the published BglT mechanism (Figure 3.5) [203]. Firstly, NAD^+-mediated oxidation of the 4'OH generates a keto group, which acidifies the adjacent 3' proton, leading to a dehydration reaction at the 2'-3' position. Ring flip then occurs, followed by 7' OH and 2' carbon cyclises by nucleophilic attack. NAD^+ then regenerates by reducing 4'OH, generating Anh.

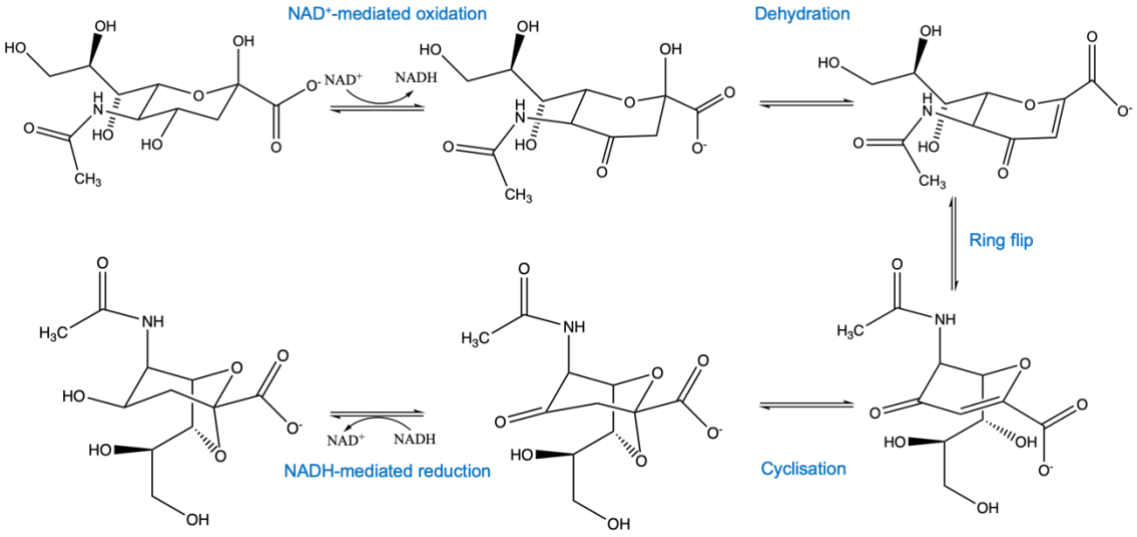


Figure 3.9 RgNanOx general reaction scheme *Reaction steps proposed in Bell et al 2020 [184]*

3.c. Materials and Methods

3.c.i. Molecular Cloning and Mutagenesis

50 µL of restriction reactions were set up, consisting of 1 µg plasmid, 1 µL (10 units) of each endonuclease (NEB), 1x CutSmart Buffer (NEB). Reactions were incubated for 1 hr at 37 ºC then ran on 1% agarose gel at 60 V for 30 min. Gel extraction was then performed using QIAquick Gel Extraction Kit (Qiagen). Ligation reactions consist of 100 ng of DNA

mixture (3:1 molar ratio of insert:vector), 1 μL (400 units) of T4 ligase (NEB), and 1x T4 DNA Ligase Buffer (NEB).

Site-directed mutagenesis was performed using a published protocol [160], primers used are listed in Table 3.1. 50 μL PCR reactions were set up with 1x Phusion HF Buffer (NEB), 100 mM forward and reverse primers each, 40 ng template, 1 mM dNTP, and 0.5 μL Phusion polymerase (NEB). Reactions were initially denatured at 94 ºC for 7 min, then subjected to 12 cycles of 94 ºC for 1 min, ($T_{m(no)} - 5$) ºC for 1 min, and 72 ºC for 10 min. 3 subsequent cycles where then performed at 95 ºC for 1 min, ($T_{m(pp)} - 5$) ºC for 1 min, and 72 ºC for 10 min. Final extension was performed at 72 ºC for 20 min, before holding at 4 ºC overnight. $T_{m(no)}$ and $T_{m(pp)}$ were calculated using the protocol [160] and OligoCalc [161].

50 μL of PCR product was incubated with 1 μL (20 units) of DpnI (NEB) at 37 ºC for 3 hr, verified by 1% agarose gel ran at 90 V for 1 hr. 5 μL was then transformed with 100 μL of DH5α cells on which miniprep and sanger sequencing were performed.

Mutation	Direction	Sequence (5' – 3')
K93A	Forward	GTGAAGCACCCATTGCACTTTCCTATCAGGATTGTGACG
	Reverse	GCAATGGGTGCTTCACAAAATACATTCACGCCATGTTCTG
R150A	Forward	TCTGCAGCAAATGGATGGGAAGAACAGCAGCCAACTATTTCCTGGA
	Reverse	CCATTTGCTGCAGAATGGCAATAGAGAACTTTTCCAATC
K163A	Forward	ACAGCAGCCAACTATTTCCTGGGCAAAAATCAGGGAAAAAT
	Reverse	ATTTTTCCCTGATTTTTGCCCAGGAAATAGTTGGCTGCTGT
Y298A	Forward	CATTATGGCAGGAAAGCCAGGAAAGAAACCGCCAATGTG
	Reverse	GGCTTTCCTGCCATAATGGCACCATCCATCTCCGTAC

Table 3.1 Primers for mutagenesis

3.c.ii. Expression and purification of RgNanOx

pET vector encoding RgNanOx with an N-terminal hexahistidine tag was obtained from our collaborators at the University of East Anglia. It was transformed into DH5α cells, and amplified using QIAprep Spin Miniprep kit (Qiagen). Following retransformation into BL21 (DE3) cells, 10 mL of LB preculture (30 μgmL^{-1} kanamycin) was inoculated with single colonies overnight at 37 °C. This was then inoculated in 1 L of auto induction media (AIM – Terrific Broth Base Including Trace Elements, Formedium). Cultures were induced at 37 °C for 3 hours, followed by 16 °C for 72 hours, and harvested by centrifugation at 8000 g.

The selenomethionine variant was expressed in methionine auxotroph strain B834. Overnight LB precultures were suspended and rinsed with minimal media (NH_4Cl $1gL^{-1}$, KH_2PO_4 $3gL^{-1}$, $Na_2HPO_4{\cdot}7H_2O$ $6gL^{-1}$), then inoculated in minimal media supplemented by Molecular Dimensions SelenoMethionine Medium Nutrient Mix. Cultures were grown at 37 °C for 15 minutes before supplemented with selenomethionine (60 mgL^{-1}), induced with the corresponding reagent for 24 hours. Expression was induced at 28 °C overnight and harvested by centrifugation at 8000 g.

Cell pellets were resuspended in lysis buffer of 25 mM Tris·HCl pH 7.5, 500 mM NaCl, 5 mM imidazole, 5% glycerol, 5 mg deoxyribonuclease, 1 tablet of cOmplete™ Protease Inhibitor Cocktail (Merck). Cells were lysed using high pressor cell disruptor at 30 ksi, from which protein was extracted from supernatant of 16000 g centrifugation. Sample was passed through HisTrap HP 5mL column (GE) for nickel IMAC, and washed with lysis buffer until absorbance at 280 nm levels to baseline using ÄKTAxpress (GE) system and associated software. Bound protein was eluted in 25 mM Tris·HCl pH 7.5, 500 mM

NaCl, 500 mM imidazole, 5% glycerol and directly injected into HiPrep™ 26/10 Desalting column. The desalted sample was incubated with TEV protease for hexahistidine tag cleavage overnight at 4 °C. Cleaved protein as separated from uncleaved protein with a further IMAC under the same conditions, the flow through and wash from which are pooled and concentrated with with 30 kDa Vivaspin 20. The sample was injected into HiLoad Superdex 200 pg preparative SEC columns (GE) with running buffer of 25 mM Tris·HCl pH 7.5, 150 mM NaCl, 5% glycerol.

3.c.iii. Crystallisation of RgNanOx and substrate co-crystallisation

Aliquots of RgNanOx, stored at 100 $mgmL^{-1}$, were diluted to 20 $mgmL^{-1}$ (469 μM) in SEC buffer with 5x molar excess NAD^+ (2.35 mM) and incubated at 37 °C for 3 hr. For co-crystallisation, Neu and DANA were purchased in powder form (Sigma Aldrich), and resuspended in 20 mM sodium phosphate buffer, and adjusted to pH 7.5 with HCl. Anh, which was received from collaborators, remained in water at pH 7.0 without further adjustment. Protein sample was incubated with 20 mM substrate and 5x molar excess NAD^+ (2.35 mM) at 37 °C for 3 hr. To set up sitting drop trials, 40 μL of crystal screen was dispensed into reservoirs of CrystalQuick™X (Greiner) plates using Hydra (Art Robbins Instruments). Mosquito® (TTP) was used to dispense 250 nL of protein solution and 250 nL reservoir into each well. Where additives were present, 200 nL of reservoir and 50 nL of additives were dispensed. Crystals were grown over fourteen days and monitored using Rock Maker® (Formulatrix) using bright field and UV imaging (Figure 3.10). The former was used to assess whether the crystals had stopped growing after fourteen days, whilst the latter was used to ensure crystals grown were proteinaceous.

Crystals were harvested using cryo-loops (MiTeGen), with which crystals are first immersed in cryoprotectant (crystal buffer + 20% glycerol) [217] then flash frozen in liquid nitrogen. Data were collected at Diamond Light Source Beamline I04 and I24. Crystals are screened with three images at 0.15° oscillation to identify resolution limits and space group of each crystal. Full 100° datasets were then collected at 100% transmission, 2.00xe^{11} flux, 0.03 s exposure time, and 0.9795Å wavelength. Data was processed automatically using the XDS 3dii pipeline, which performs indexing XDS [164] and scaling using XSCALE [165], to generate merged reflection (MTZ) files for structure determination.

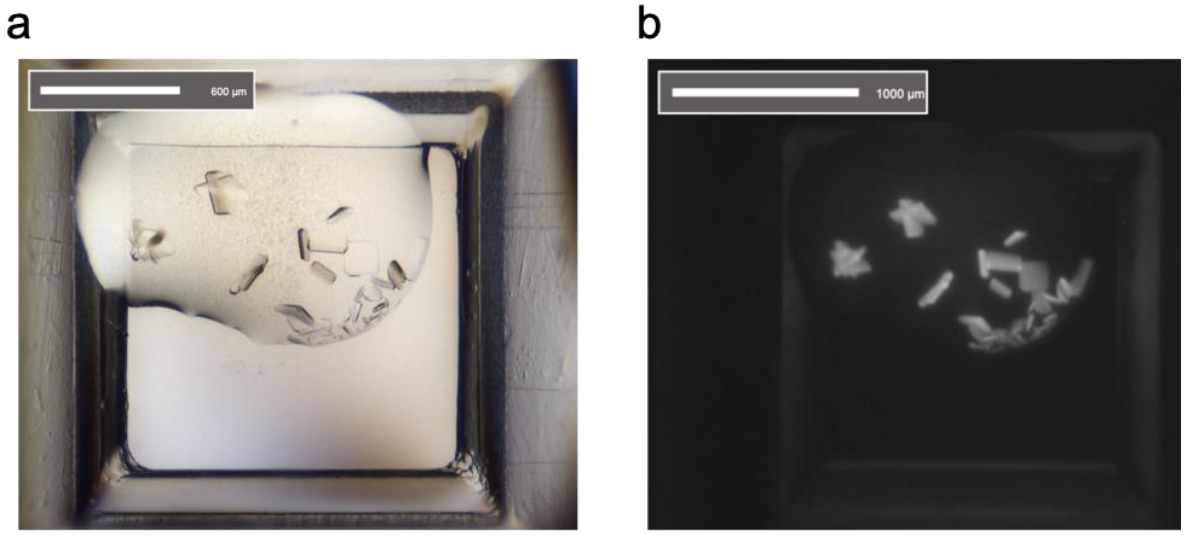


Figure 3.10 RgNanOx crystals *Images of RgNanOx crystals using Rock Maker® (Formulatrix) using (a) bright field and (b) 280 nm UV lenses.*

3.c.iv Structure Determination

Data were collected at Diamond Light Source Beamlines I03, I04, and I24. Crystals are screened with three images at 0.15° oscillation to identify resolution limits and space group of each crystal. Full 360° datasets were then collected at 100% transmission, 2.00xe^{11} flux, 0.03 s exposure time, and 0.9762Å wavelength. Data was processed

automatically using the AUTOPROC pipeline [218], which performs initial integration and space group determination, before obtaining overall statistics using XSCALE [165].

For determination of the initial holo-structure, MrBUMP [219] was used to identify A. radiobacter oxidoreductase (PDB: 5UI9) as a suitable search model for molecular replacement. Search model was truncated with Sculptor [220] and the solution was replicated on Phaser MR [221] with a TFZ score of 9.0 using the CCP4i2 suite of programmes. An initial model was generated using Phenix Autobuild [222] with R-free of 42%. Determination of the high resolution holo-structure and complex structures were performed using Phaser [221], with extracted protein coordinates of subunit A in the holo-structure as a model. Resulting datasets are screened for ligand density in the active site.

Models were refined and rebuilt using Buccaneer [171], PDB_REDO [174], TSLMD [173], REFMAC5 [172]and verified with MolProbity [175]. After refinement, models of NAD^+ and substrates were drawn using AceDRG [223], then fitted into corresponding electron density using Coot [224]. The model was then re-refined using the tools listed above. Crystallographic table and refinement statistics are shown in Appendix C.

3.c.v. Liquid chromatography-mass spectrometry (LCMS)

100 μL reactions are set up in 20 mM sodium phosphate pH 7.5, 0.1 $mgmL^{-1}$ protein, 10 mM substrate, and 5 mM NAD^+. Reactions are incubated at 37 °C for 3 hrs, after which 100 μL acetonitrile was added. The reaction was then centrifuged at 13,000 RPM, from which the supernatant was injected for LCMS.

LCMS was performed using Agilent Technologies 1200 Infinity Series fitted with a Nucleodur C18 column (Macherey-Nagel). The sample is run in the HPLC at 5% acetonitrile for 20 min then at 95% acetonitrile for 20 min. The ESI-MS mass detector is set to collect detect the m/z range of 0-500, at m/z of 1. To ensure that the mass detector is correctly calibrated, positive controls of pure Neu and Anh were tested (Section 3.d.v.). Neu was directly purchased (Sigma-Aldrich) and was verified by the vendor to be over 97% pure using an HPLC assay. Anh was synthesised by our collaborators, using membrane-enclosed multienzyme (MEME) pathway [185]. Concurrently with mass detection, UV absorption at 214 nm was also monitored. Resulting absorption peak was used to identify the retention time at which mass spectra were analysed.

3.c.vi. Isothermal titration calorimetry (ITC)

Protein sample was dialysed in 100 mL of buffer Tris-HCl 25 mM pH 7.5, 150 mM NaCl, using Slide-A-Lyzer™ casettes (Thermo Fisher) with 10 kDa cutoff, at 4 °C overnight. The resulting buffer was used in diluting protein sample to the appropriate concentration, and in preparing titrant sample by dissolving powdered Neu (Sigma Aldrich).

Using the iTC200 apparatus (GE Healthcare), syringe was filled with 60 μL of titrant and sample cell was filled with300 μL of protein. Sample cell was equilibrated against reference cell using automatic mode, at reference power of 6 $\mu cals^{-1}$. Titrant was injected at 1 injection of 0.4 μL followed by 16 injections of 2.4 μL, with 180 s spacing between each injection. Data was read and integrated using the proprietary Origin software.

3.c.vii. Nuclear magnetic resonance (NMR) spectroscopy

We provided our collaborators at the Rosalind Franklin Institute with reaction samples, who performed data collection and analysis. 200 μL reactions were set up with 20 mM sodium phosphate (pH7.5), 5 mM NAD^+, 200 μM 3-(Trimethylsilyl)-1-propanesulfonic acid (DSS), 0.7 $mgmL^{-1}$ RgNanOx. 0.2 – 15 mM of substrate (Neu/Anh) were provided separately and added to the reaction to a total volume of 200 μL before data collection. At the NMR facility at Diamond Light Source, samples were transferred to 5mM NMR tubes (Sigma Aldrich) and equilibrated to 37 °C. ^1H scans were obtained at 500 s intervals from 1.0 – 5.0 ppm. Peaks were assigned and integrated using the TopSpin programme (Bruker).

3.d. Results

3.d.i. Recloning and mutagenesis for RgNanOx

After numerous unsuccessful attempts of regenerating the original RgNanOx plasmid by miniprep and site-directed mutagenesis, we recloned the RgNanOx gene in an optimised plasmid pHisTEV [158]. NdeI and HindIII were used to excise RgNanOx from its original plasmid, and to linearise empty pHisTEV. Digestion reactions were verified by gel electrophoresis (Figure 3.11), in which linearised pHisTEV and excised gene were retrieved by gel extraction. The ligation product between the two were then transformed into DH5α cells for miniprep, and confirmed as a newly cloned plasmid by Sanger sequencing.

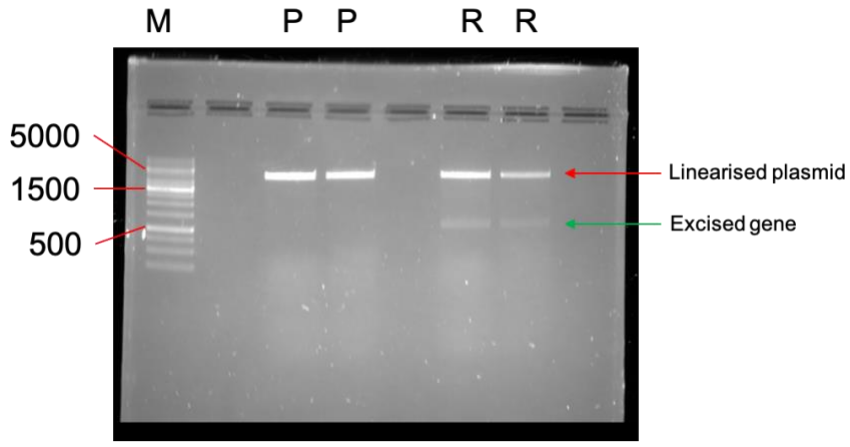


Figure 3.11 Linearisation of pHisTEV and RgNanOx plasmid *1% Agarose gel electrophoresis of GeneRuler 1kb marker (M) and digests of empty pHisTEV (P) and RgNanOx plasmid (R). Both digest have a linearised plasmid band (red arrow), whilst digest of the RgNanOx plasmid also resulted in a band corresponding to the excised gene (green arrow).*

We were successful in generating K93A, R150A, and H176A mutants by site-directed mutagenesis using the new plasmid, with a gradient to optimise annealing temperature by +/- 5 °C (Figure 3.12). H175A and H178A were obtained by miniprep of plasmids sent by our collaborators. We were able to express and purify all mutants except K93A, and purify the remaining to high homogeneity (Sections 3.c.ii. and 3.d.ii.). We attempted to generate Y13F, K163A, and Y298F mutants but were unsuccessful.

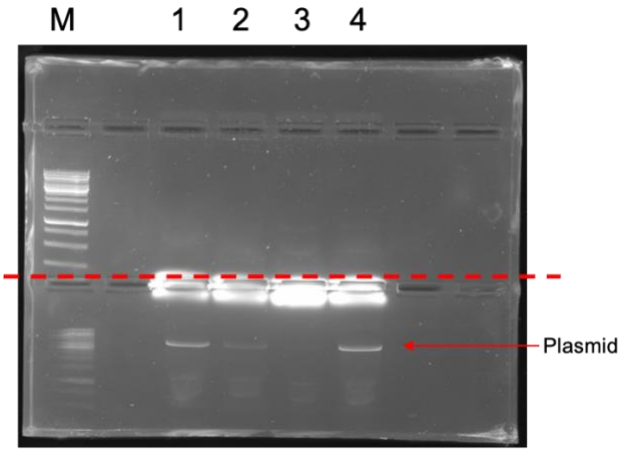


Figure 3.12 Agarose gel showing successful site directed mutagenesis *1% agarose gels of PCR products of H178A (top half, separated by red dash line) and K93A (bottom half), alongside 1 kb GeneRuler*

ladder (M). Each mutant was optimised by running at four different annealing temperatures (1- 4), resulting plasmids with K93A were highlighted with a red arrow.

3.d.ii. Expression and purification of RgNanOx

Both the original plasmid and the new construct (Section 3.d.i.) were used for expression of WT RgNanOx, which typically yields 40 mg of soluble purified protein per litre of culture as verified by SDS-PAGE (Figure 3.13). RgNanOx is approximately 42 kDa, and is expected to appear at the level between the 37 kDa and 50 kDa standards. In the first IMAC step, this band is present in every fraction (Figure 3.13a). The intense band in lysate indicates overexpression of RgNanOx, whilst the presence in flow through and wash steps may indicate the saturation of HisTrap column used. RgNanOx eluted from the column was judged to be pure.

Conversely, after IMAC 2, RgNanOx band was present only in the flowthrough and wash fractions (Figure 3.13b). This indicates that RgNanOx has been successfully cleaved by TEV protease, and thus no longer binds to HisTrap. A faint band is present in the elution fraction, which we propose to be uncleaved RgNanOx, alongside low molecular weight bands that corresponds to TEV protease (27 kDa). To ensure TEV cleavage was successful, MALDI-MS was performed via the University of St Andrews Mass Spectrometry and Proteomics Facility. SDS-PAGE bands from the beginning, middle, and tail of the elution peak were treated with trypsin. We observe the 10 residue product resulting from the digestion of cleaved RgNanOx, and did not observe the 33 residue product that would form from the digestion of uncleaved RgNanOx (Figure 3.14).

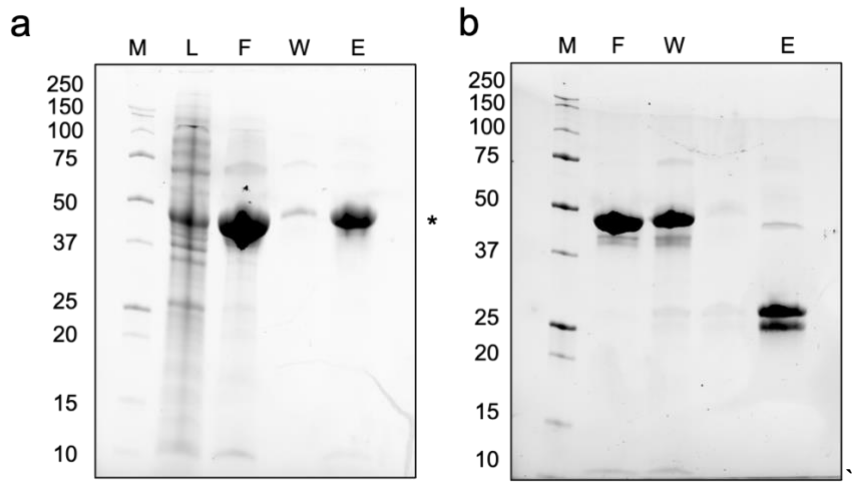


Figure 3.13 SDS-PAGE gels verify RgNanOx expression and IMAC purification *Stainless SDS-PAGE gels (BIO-RAD) with molecular weights in kDa shown adjacently, referenced using Precision Plus Protein™ Unstained Protein Standards (BIO-RAD, lane M). The level at which bands corresponding to RgNanOx is indicated with an asterisk. (a) Fractions from first IMAC with fractions corresponding to lysate (L), flow through (F), wash (W), and elution (E). (b) Fractions from second IMAC with fractions corresponding to flow through (F), wash (W), and elution (E).*

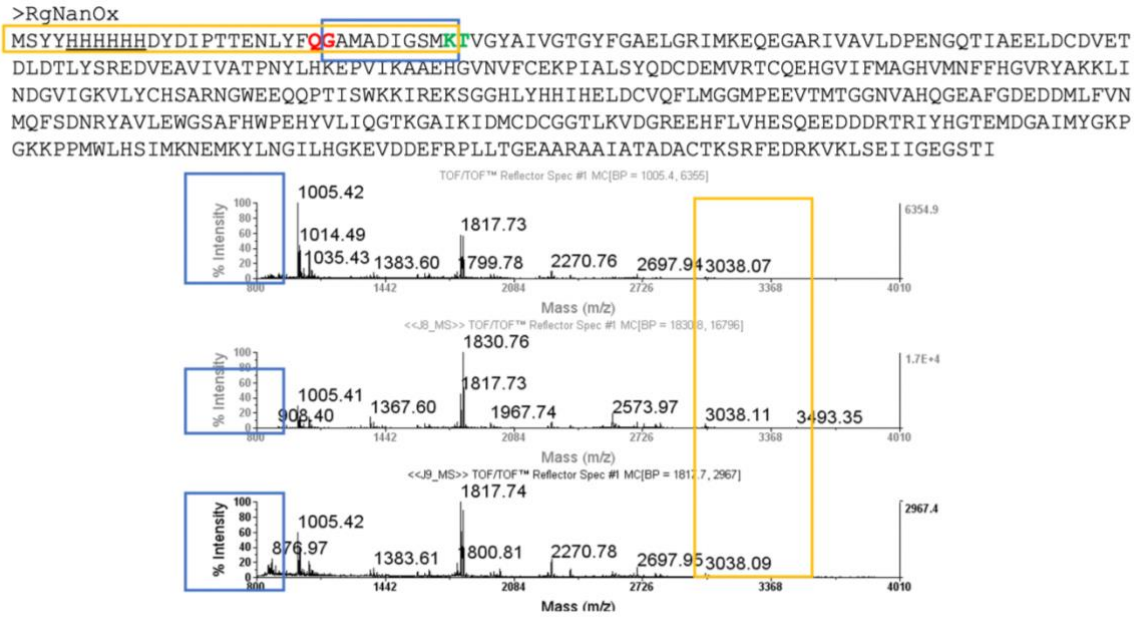


Figure 3.14 MALDI-MS spectra showed RgNanOx TEV cleavage *TEV cleavage site is shown bold in red, trypsin cleavage site is shown bold in green. We observe peaks corresponding to the mass (1000 kDa) of the fragment between the two sites, shown in blue rectangles, which would only appear if TEV*

cleavage occurred. Conversely, we did not observe any peaks in the region of 3800 kDa that would correspond to a trypsin-digested N terminus of protein uncleaved by TEV, shown in yellow triangles.

UV absorption trace in the SEC elution profile shows a highly homogenous peak that is slightly asymmetric (Figure 3.15a). This is likely due to the high amount of protein injected, rather than protein fragmentation or contamination. SDS-PAGE was used to verify this, which showed clear single bands (Figure 3.15b).

For selenomethionine derivative expression, conditions were screened with lactose (2.0 gL^{-1}) at 16, 28, and 36 degrees for 2 hr, 4 hr, and overnight. Expression from 10 mL LB cultures were assessed by western blotting with anti-His antibody, showing expression in all conditions (Figure 3.16). Large scale expression was performed 28 °C overnight to mimic expression conditions for WT RgNanOx.

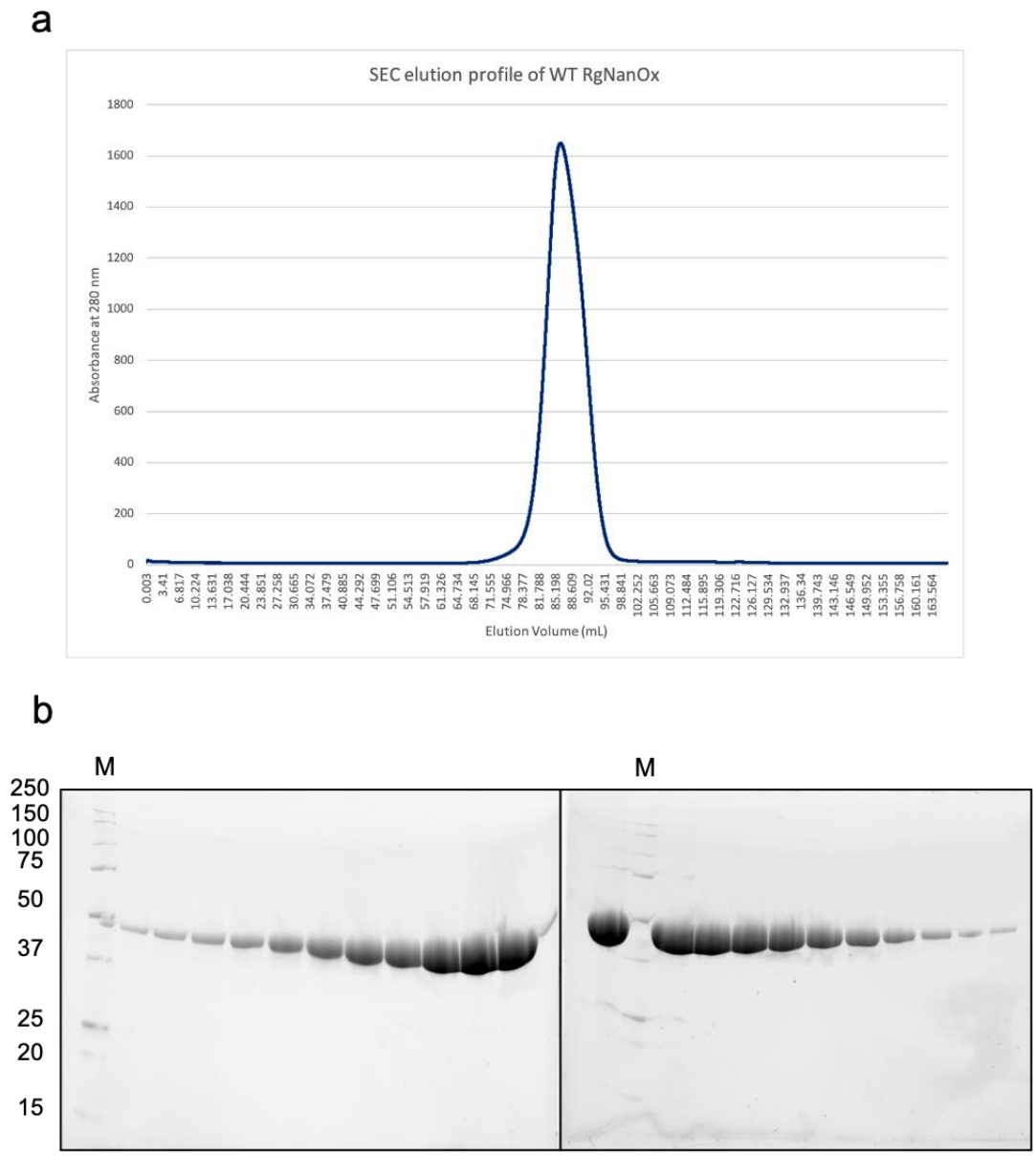


Figure 3.15 RgNanOx was homogenous after SEC *(a) 280 nm UV absorption trace of SEC elution profile. (b) SDS-PAGE of continuous fractions in the UV absorption peak, showing no noticeable change in homogeneity across fractions.*

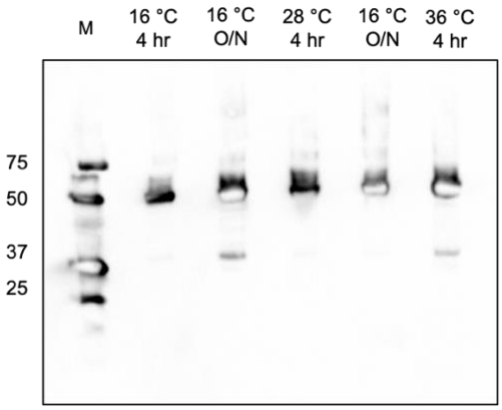


Figure 3.16 Western blot confirmed selenomethionine-RgNanOx expression *Western blot of lysed cell pellets using anti-His antibody, showing RgNanOx expression for all tested conditions. Potential contaminant of 37 kDa were detected for expression at 16 degrees overnight and 36 degrees for 4 hr.*

3.d.iii. Determination of RgNanOx holo- structure

PCT™ Pre-Crystallization Test (Hampton) was used to identify optimal protein concentration for crystallisation screening. For RgNanOx, 10 $mgmL^{-1}$ was too dilute for crystallisation, whilst 40 $mgmL^{-1}$ resulted in amorphous precipitation. Therefore, we decided to screen at 20 $mgmL^{-1}$ as suitable protein concentration for crystallisation, where samples are supplemented with 5 x molar excess NAD^{+} cofactor. Crystals formed in a wide range of conditions, albeit in poor morphologies of thin plates or needles. 32 conditions were screened at the I24 beamline, one of which resulted in a 1.7 Å dataset. This dataset could not initially be solved as Phaser [221] did not return solutions with satisfactory (> 6) translation function (TFZ) scores, indicating the solution search has too low signal:noise ratio to be reliable.

We attempted to grow larger crystals by hanging drop using VDX crystal plates (Hampton). Protein:precipitant ratios of 1:1, 2:1, and 1:2 were prepared using the condition corresponding to the 1.7 Å crystal. Drops of 1, 5, and 10 μL were set up but very few crystals grew and with no improvement in morphology. We therefore optimised

crystals by varying concentration and pH, based on two key conditions that resulted in diffracting crystals in the initial screen. These are (i) 0.1 M sodium citrate pH 5.3-5.7, 17-23 % PEG 3000 (henceforth 'citrate screen'); (ii) 0.1 M MES pH 6.0, 0.18-0.22M NaCl, 18-22 % PEG 6000, 0.10-0.14 M NaF (henceforth 'MES screen'). Complete compositions of optimised conditions can be found in Appendix A. From the H3 condition of the citrate screen, we obtained a crystals diffracting to resolution of 2.5 Å, which was used to determine the holo-structure of WT RgNanOx.

Molecular replacement was used to solve the phase problem, for which suitable truncated models were identified using the automated pipeline MrBUMP [219]. This generated an RgNanOx model with R-free of 42%, with interpretable electron density maps for refinement. $2(F_O-F_C)$ map indicates electron density surrounding the model, allowing unmodelled residues to be fitted. F_O-F_C map indicates regions of positive and negative electron density, where the model is lacking atoms (underfitted) and modelled atoms should not be present (overfitted) respectively [225]. Using B factors, which indicate the tendency of atoms to move about its modelled coordinate, the model is treated as two rigid bodies that for further automatic refinement [173]. The completed model was validated using Ramachandran plots and geometric dictionaries that detect anomalies in rotamer conformation, bond lengths, and bond angles [175]. The final R-free is 25%, complete parameters are shown in crystallographic tables provided in Appendix C.

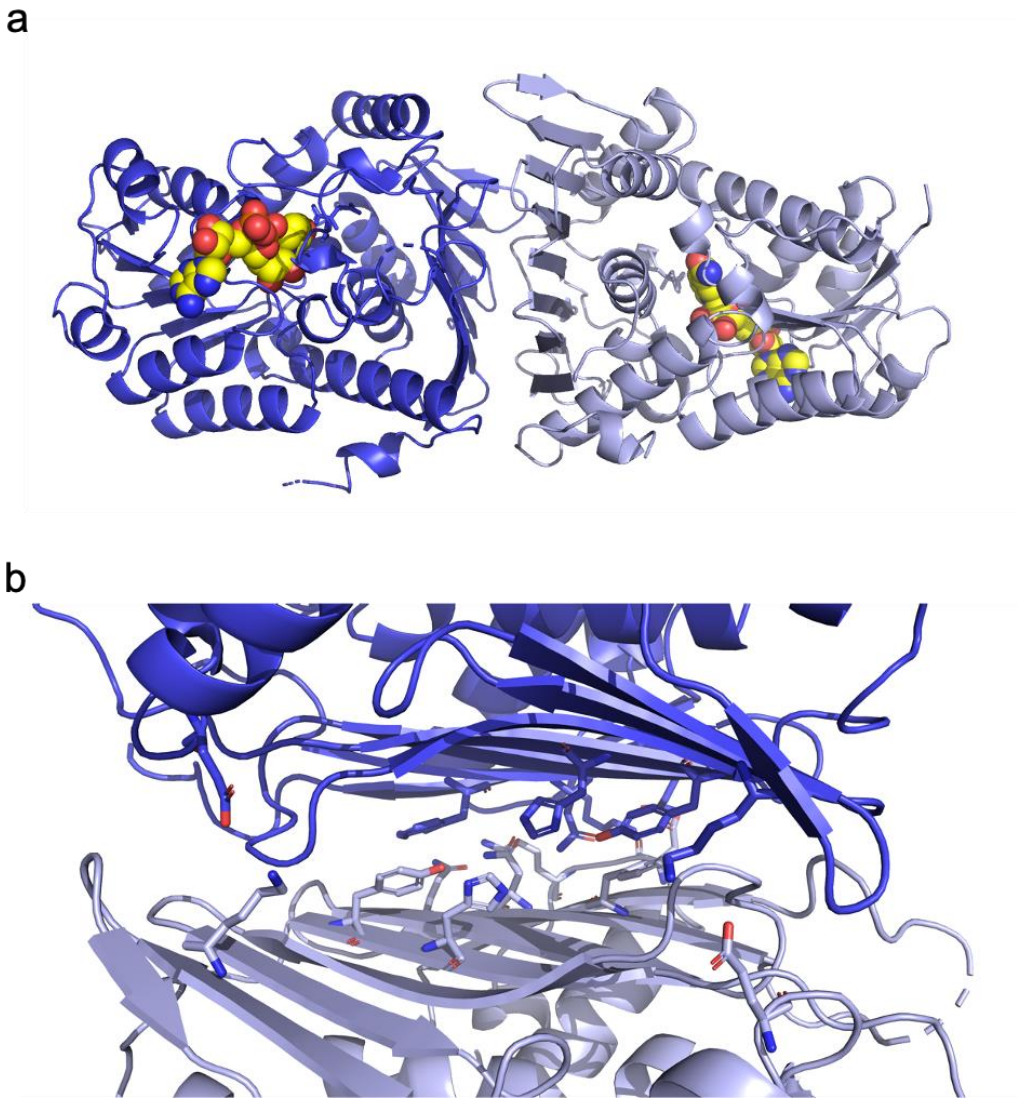


Figure 3.17 Holo-structure of RgNanOx *(a) Crystalline structure of RgNanOx dimer shown as ribbons, with NAD^+ shown as spheres. (b) Dimer interface with residues involved in hydrogen bonding and salt bridges shown as sticks*

This 2.58 Å structure is typical of the Gfo/Idh/MocA family [226] with an N-terminal nucleotide binding domain in a Rossman fold, and a C-terminal α/β substrate binding domain (Figure 3.17a). It is a dimer in the asymmetric unit where monomers are packed 'back-to-back' against each other at C-terminal β sheet. The PISA server was used to predict the biological multimeric state of RgNanOx using the crystalline structure [227]. The dimer interface has a surface area of 1835.4 $Å^2$, which constitutes approximately

6.5% of the total surface area of the assembly (28033.7 $Å^2$). In addition, the interface has a calculated CSS score of 0.505, which is calculated by the difference in solvation free energy between the dimer and monomers. This indicating that the hydrophobic residues in the interface will likely cause spontaneous dimerisation in solution, supported by specific interactions between polar or charged residues via hydrogen bonding and salt bridge (Figure 3.17b).

Two blobs of positive density were observed in the F_O-F_C map, one of which fits well with cofactor NAD^+ (Figure 3.18a), the nicotinamide ring of which forms π-π interactions with Phe14. The other density likely corresponds to the substrate binding site, as it is proximal (< 5 Å) to the catalytic C4 of NAD^+. However the density is too small to fit either Neu or Anh, and upon examination of this additional density led us to conclude that the density corresponded to citrate. Citrate was added to the model which shows a good fit and that which forms a hydrogen bonding network with surrounding basic residues in the active site (Figure 3.18b). Crystallising RgNanOx in the absence of citrate did not result in the disappearance of this density, which may indicate that citrate had been retained from the *E. coli* cell. Proximal to NAD^+ are putative catalytic residues Lys93, His178, as well as two other histidine residues His175 and His176 (Figure 3.18b). In addition, the hydrogen bonding network of citrate (Figure 3.18b) shows Tyr13, Arg150, Lys163, and Tyr298 are also proximal to the substrate binding site.

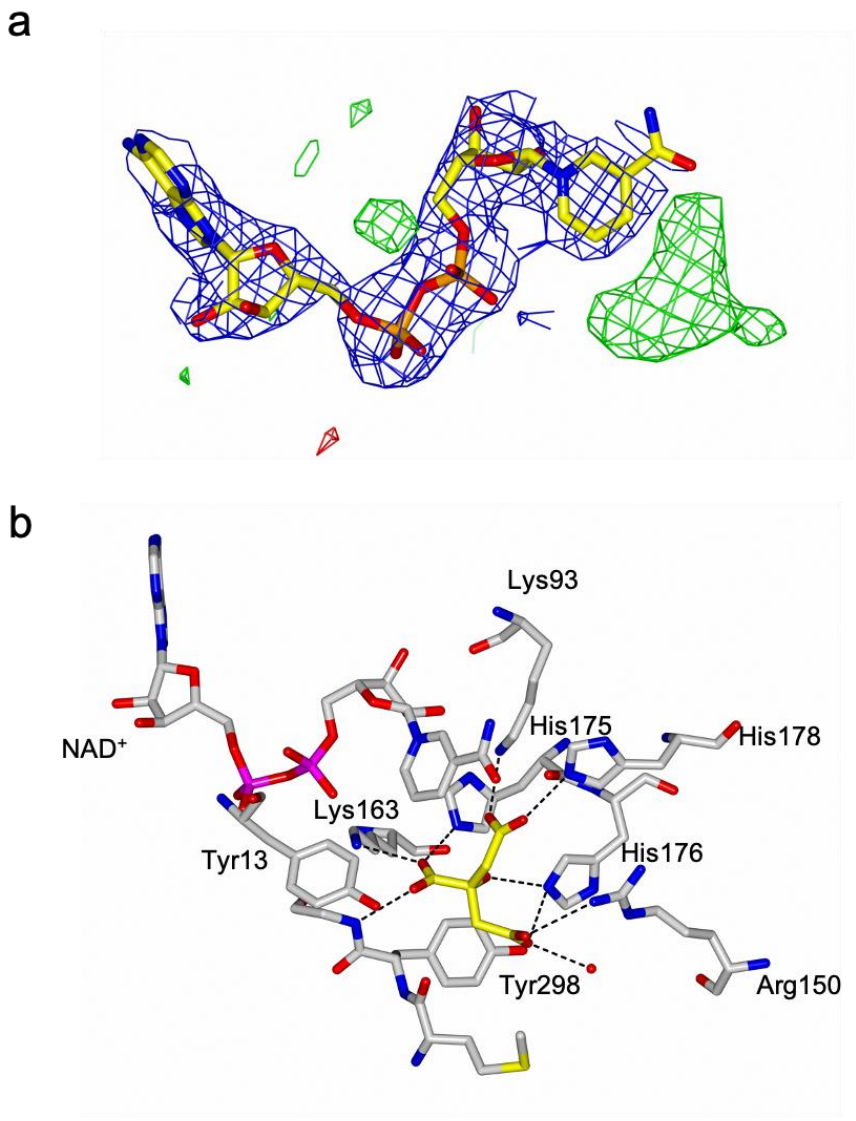


Figure 3.18 NAD$^+$ and citrate were bound to holo-RgNanOx *(a) 2F_O-F_C map in blue (σ = 1.6) surrounding NAD$^+$, showing tight fit, and surrounding F_O-F_C map in green (σ = 4.0) showing clear positive density adjacent to C4 of NAD$^+$ (b) Hydrogen bonding network of citrate (yellow), showing interactions with Lys93, Lys163, His175, His176, His178,*

3.d.iv. Sequence comparison of RgNanOx and homologues reveals unusual catalytic motif

Sequence comparison between RgNanOx and homologues was used to identify the catalytic roles of key residues identified in Section 3.d.iii. using Clustal Omega [163] was used to perform multiple sequence alignment on RgNanOx and three Gfo/Idh/MocA

enzymes with high sequence identity and known structure and function (Figure 3.19). They are 1,5-anhydro-d-fructose reductase (PDB code: 2GLX), Glucose-Fructose oxidoreductase (PDB code: 1OFG), scyllo-inositol dehydrogenase (PDB code: 5YA8).

Conserved residues in the Gfo/Idh/MocA family are characterised as follows [226]: (i) catalytic lysine (ii) catalytic histidine or tyrosine (iii) positively charged residue of unknown catalytic role, and (iv) aspartate residue proximal to the catalytic histidine or tyrosine. Their aligned positions in RgNanOx (Figure 3.19) suggest that Lys93 and His178 represent the former two residues, corroborated by their position within 4 Å from NAD^+ in our holo- structure. To confirm this, we have superposed these residues with their homologues from our alignment, showing conserved positions (Figure 3.20). Although Arg168 aligns with conserved arginines, it does not hydrogen bond with other catalytic residues or cocrystallised species in our structure, and instead faces away towards the solvent. It is possible therefore the Arg150, which hydrogen bonds with the bound citrate, takes the catalytic role. Crucially, the fourth position that corresponds to a highly conserved aspartate, which completes the DXXXH, DXXH, or HXXD motif, is missing in RgNanOx. Instead RgNanOx has a HHIH motif, of which the ultimate histidine (His178) corresponds to the second conserved site. We reasoned that His175 and/or His176 may act in lieu of the aspartate in the DXXXH motifs found in these homologues, as the two residues are also observed to be proximal to NAD^+. To explore this, we used secondary structure matching (SSM) search [228] with our RgNanOx holo-structure (Section 3.d.iii.) as query, to identify candidates for homologous motifs. Using the conserved His178 position as a reference, proximal (+/- 5) histidine residues were observed (Figure 3.21). Of the 402 matches with similarity of 70% or higher, only two entries contain the HHIH motif. These correspond to the *E. coli* homologue of RgNanOx

(YjhC, UniProtKB-P39353) and an uncharacterised *S. pneumoniae* oxidoreductase (UniProtKB-Q54728). A similar HHAAH motif is found in *Sphingobium sp* 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase (UniProtKB-Q9KWL3). Finally, two other entries contained one additional histidine in their catalytic motifs: an uncharacterised *S. fredii* oxidoreductase (UniProtKB-P55480) and putative *B. subtilis* oxidoreductase (UniProtKB-O34371). All other entries contain canonical motifs with no proximal histidines.

Since all members of the Gfo/Idh/MocA family utilise either NAD^+ or $NADP^+$, we concluded that the conserved His178 acts in tandem with RgNanOx's cofactor during the NAD^+-mediated oxidation and NADH-mediated reduction steps (Figure 3.9). It is likely that the positive charge on Lys93 assists His178 in this capacity by lowering the nearby pK_a for electron transfer, similar to other mechanisms proposed for enzymes in the family [211]. However, it remains to be addressed whether His175 and His176 in the unique HHIH motif, or Arg150, play catalytic roles.

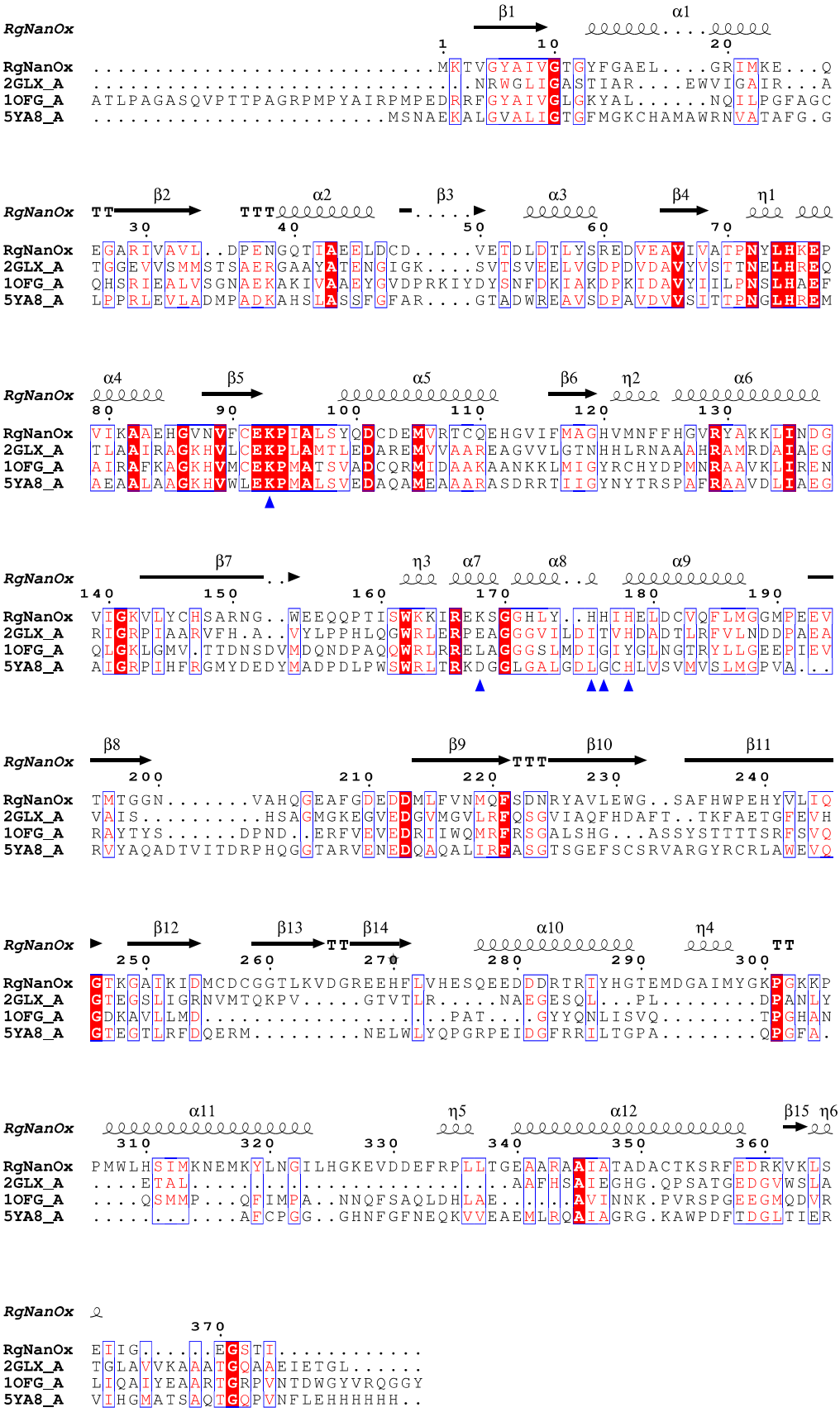


Figure 3.19 Multiple sequence alignment showed putative catalytic residues *Presented using ESPript*

[178], multiple alignment of RgNanOx with 1,5-anhydro-d-fructose reductase (2GLX_A), Glucose-Fructose oxidoreductase (1OFG_A), and scyllo-inositol dehydrogenase (5YA8_A). Secondary structures extracted from the holo- structure are shown above corresponding residues, conserved residues are shown with blue triangles below corresponding positions.

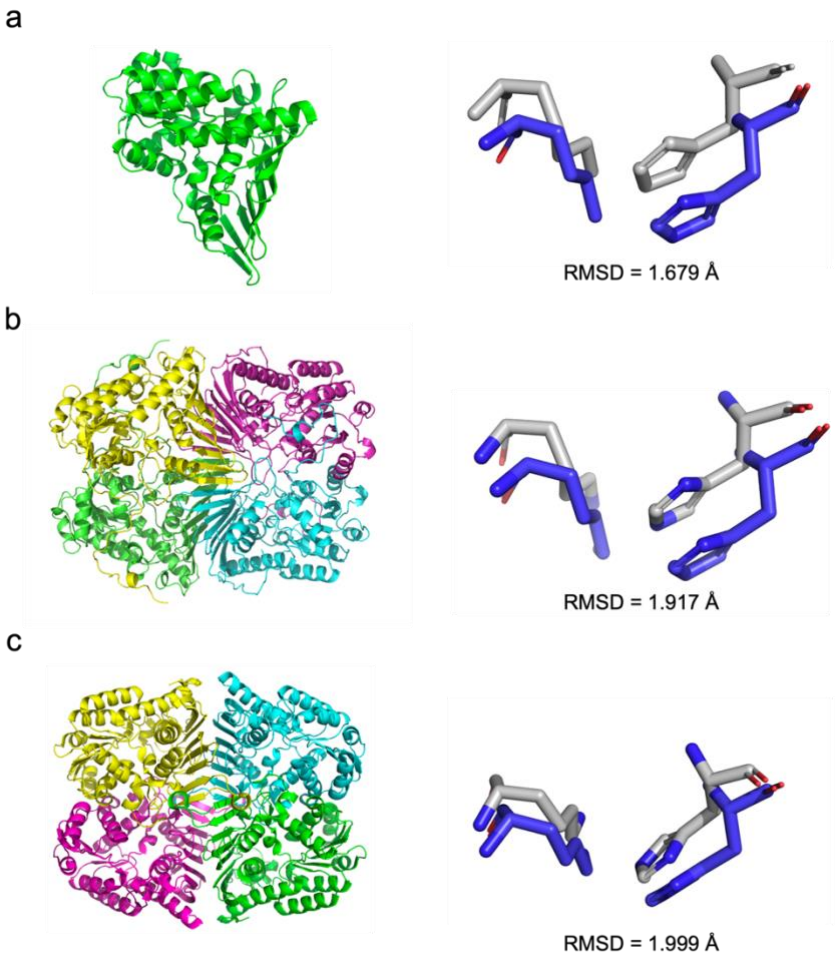


Figure 3.20 Superposition of putative catalytic residues *Shown on the left are crystalline structures of homologous proteins 2GLX (a), 1OFG (b), and 5YA8 (c) from Figure 3.19, coloured by monomer and shown as the assembly assigned on PDB. Shown on the right are superpositions of RgNanOx catalytic residues (blue) with corresponding residues in homologues (mute), along the RMSD values obtained from superposition of two monomers.*

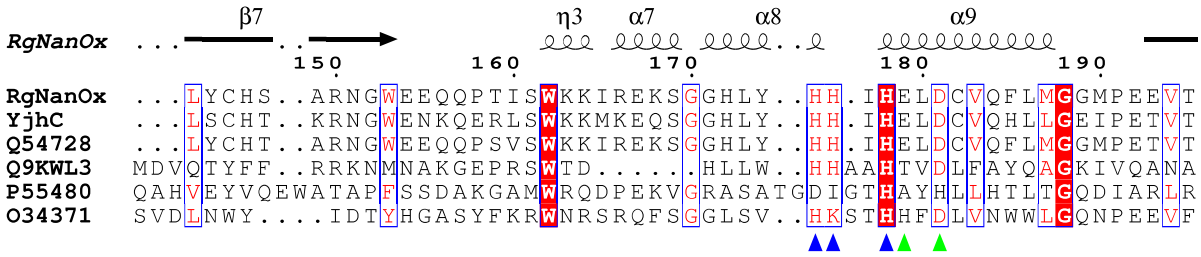


Figure 3.21 Alignment of multi-histidine catalytic motifs in the Gfo/Idh/MocA family *Presented using ESPript [178], multiple alignment of RgNanOx with homologous proteins that have multiple histidines in their catalytic motifs: YjhC (Q54728), uncharacterised S. pneumoniae oxidoreductase (Q54728), 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase (Q9KWL3), uncharacterised S. fredii oxidoreductase (P55480), and putative B. subtilis oxidoreductase (O34371). Histidines in or aligned to the RgNanOx HHIH motif were shown with blue triangles. Additional histidines proximal to the position aligned with His178 were shown with green triangles.*

3.d.v. Identification of putative catalytic residues using high resolution data

To predict the residues involved in the dehydration and cyclisation steps of RgNanOx's catalytic cycle (Figure 3.9), the 2.58 Å structure (Section 3.d.iii.) was used to solve a second dataset resolved at 1.74 Å. The higher resolution allowed unambiguous identification of the citrate ligand and its conformation (Figure 3.22). Superposition shows that each citric acid is ideally positioned (Figure 3.23). Consequently, we were able to use the three positions of the carboxylate groups in citrate to position the sole 1' carboxylate group in substrates. We therefore superposed on these loci a model of DANA extracted from the co-crystallised structure of N8 neuraminidase (PDB 2HTR) [229].

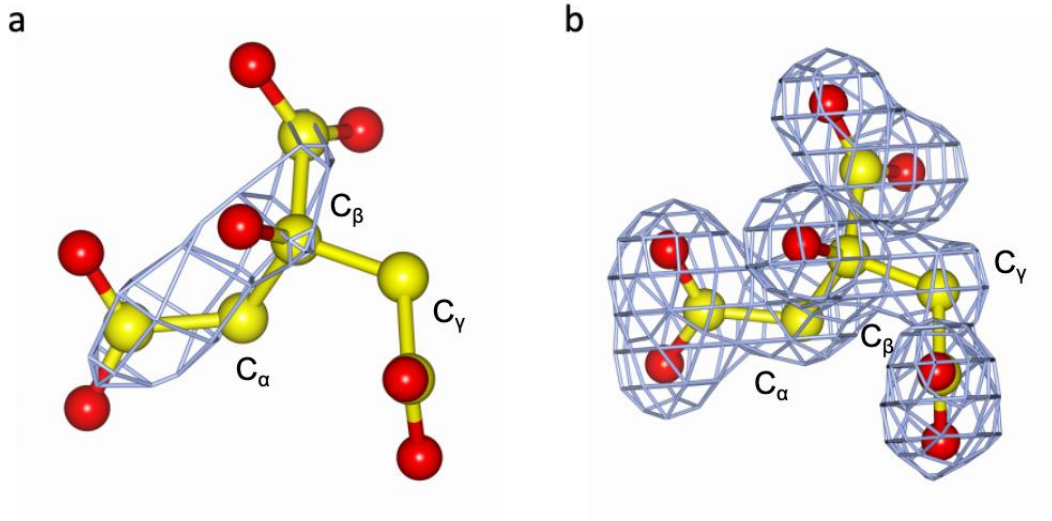


Figure 3.22 High resolution structure confirmed citrate conformation *Ball-and-Stick model and corresponding electron density of citrate in the (a) 2.58 Å structure, and; (b) 1.74 Å structure. $2F_O$-F_C maps are shown as blue chicken wire (σ = 1.2)*

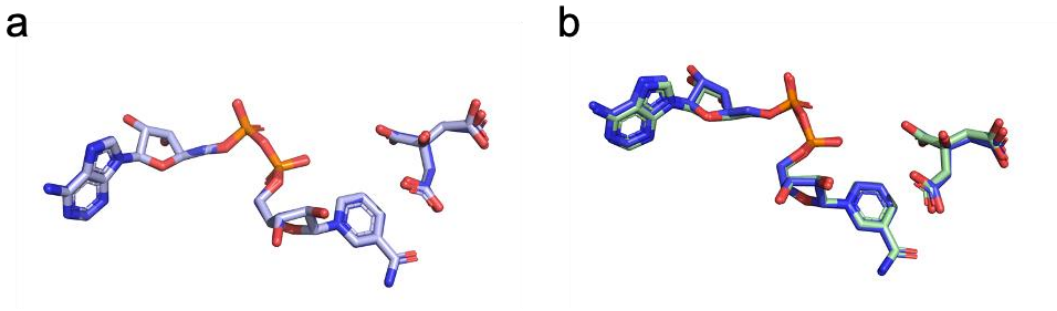


Figure 3.23 Citrate was positioned identically in both structures *Superpositions of (a) Monomer A (Blue) and monomer B (light purple), both of the lower resolutions structure, and; (b) Highest matching monomers between high resolution structure (green) and lower resolution structure (blue). Both conformations of NAD^+ and citrate are highly conserved, suggesting that they are ideally positioned.*

Hydride transfer from the C4 atom of NAD^+ requires it to be within 3.5 Å from the oxidation/reduction site 4' OH. We therefore rotated our superposed DANA model about the 1'-2' bond on this basis, whilst maintaining the nicotinamide and DANA carbon rings

approximately parallel. This adjusted DANA 'complex', consisting of the DANA model superposed on the high resolution structure with citrate removed, had its structure idealised using REFMAC5 [172]. This superposition between these refined DANA positions and the high resolution structure was shown in Figure 3.24. Although all three superpositions were possible without severe clashing, the γ position could not be adjusted to have the two rings parallel (Figure 3.24c), which led us to eliminate this position.

The remaining two superpositions were used to identify putative catalytic residues. The dehydration step (Figure 3.9) which requires a Lewis acid and a Lewis base at 2'OH and 3'C respectively. The formation of the 2'-3' double bond allows it to undergo a Michael reaction, similar to the GH4/GH109 mechanism [202]. This cyclisation step, mediated by nucleophilic attack from 7'OH on the 2' carbon, requires a further Lewis base to activate 7'OH. However, we note that since this nucleophilic attack does not occur until after the ring flip step in the Neu-to-Anh conversion, 7'OH may not be coordinated by this Lewis base in Neu or DANA complex structures. Candidate residues for these catalytic roles are summarised in Table 3.2, and were selected using two criteria. Firstly, histidines and tyrosines can act as both acid and bases, whilst basic residues will be in their conjugate acid forms and thus act as acids. Secondly, the residue must be within approximately 5 Å from the site of electron transfer. Since DANA does not have a 2' side chain position, these were taken from 1'COOH. Distances for proton abstraction are measured from the carbon attached.

We could not rule out the remaining two (C_α and C_β) superpositions as both resulted in plausible candidates for catalytic residues (Table 3.2). However, we reason that the C_β position is more likely to be representative of the in situ conformation for two reasons.

Firstly, C_β offers more suitable candidates for the base to 3'H than His178, the sole candidate in C_α. His178 is already involved in the initial 4'OH oxidation step and is therefore act as an acid, unless it donates the proton to nearby water. Secondly, it is regenerated at the end of the reaction along with NAD^+ which would be disrupted by acting as a base to 3'H in the dehydration step. Finally, His178 is more ideally positioned for 4'OH in the C_β position, whereas it is most proximal to 2' carbon in the C_α position. On this basis, we propose the following to be putative catalytic residues to be functionally characterised: Tyr13, Lys93, Lys163, His175, His176, His178.

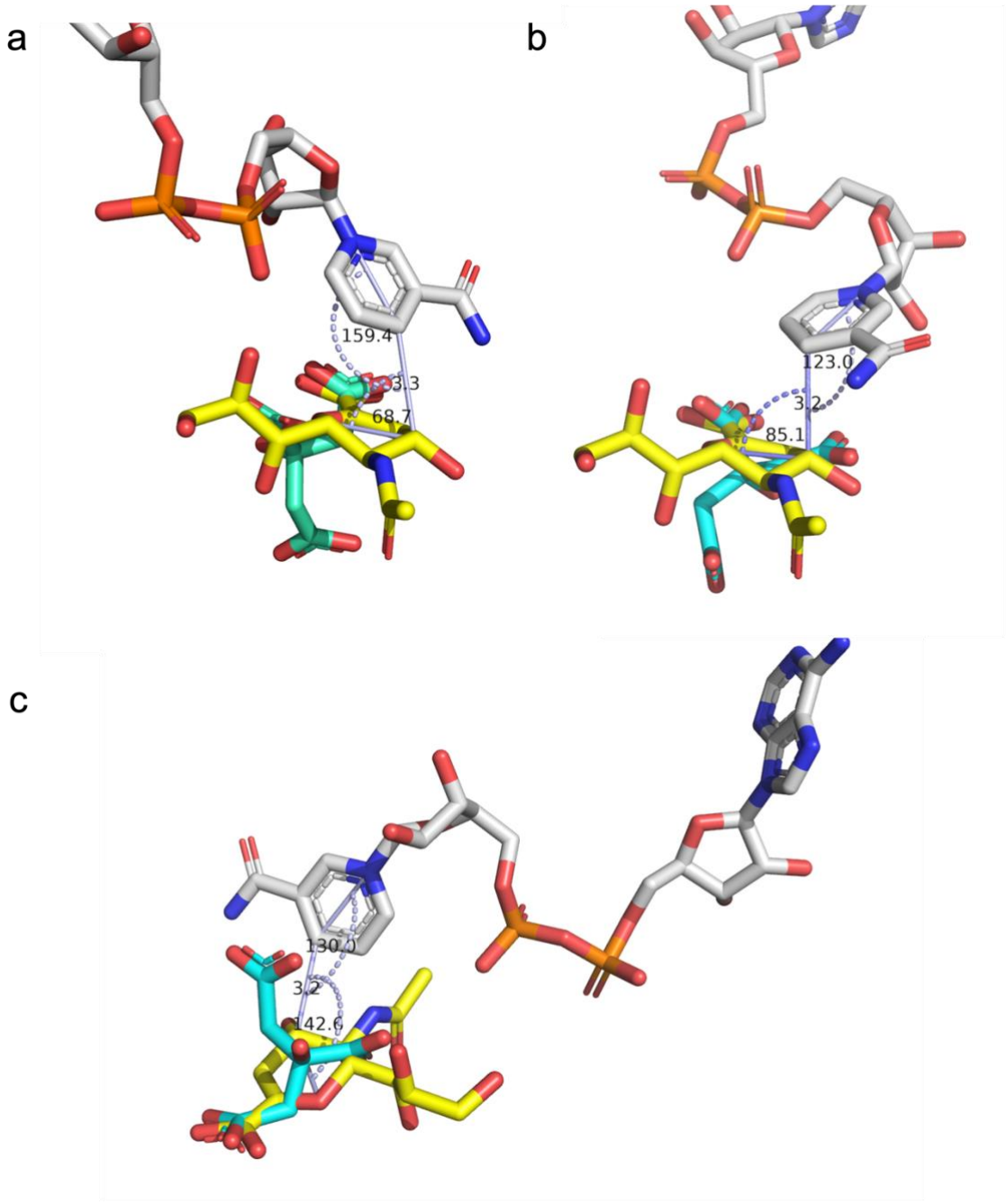


Figure 3.24 Citrate conformation used to show possible substrate orientations *Superposition of DANA (yellow sticks) on different citrate (cyan sticks) positions* C_α *(a,)* C_β *(b), and* C_γ *(c), shown alongside with* NAD^+ *(grey sticks). Distances between C4 of* NAD^+ *and 4' C of DANA are shown in angstroms. Angles between C4 and DANA sugar ring, and between 4' C and* NAD^+ *nicotinamide ring are also shown.*

Superposition on C_α in citrate (Figure 3.24a.)	
Acid to 2’ OH	Lys93 (2.7 Å) Arg150 (3.9 Å) His175 (3.1 Å) His176 (3.8 Å) His178 (2.8 Å)
Base to 3’C	His178 (5.0 Å)
Base 7’OH	His175 (5.1 Å)
Superposition on C_β in citrate (Figure 3.24b.)	
Acid to 2’ OH	Tyr13 (5.1 Å) Lys93 (5.0 Å) Lys163 (3.2 Å) His175 (2.3 Å) His176 (4.8 Å)
Base to 3’C	His175 (4.0 Å) His176 (4.3 Å) His178 (4.6 Å)
Base to 7’	None

Table 3.2 Candidate catalytic residues *Each catalytic role (left column) is listed with candidate residues (right column) and the interatomic distance between electron transfer sites. The table is split into two superposed positions, of 1’ COOH and C_a of citrate and of 1’COOH and C_b of citrate.*

.

3.d.vi. Establishment of an LCMS functional assay for RgNanOx

Our collaborators have established an Electrospray Ionisation Mass Spectrometry (ESI-MS) based assay, where samples were injected directly with a needle. Direct injection is not available on our system, which uses LCMS, adding a High Performance Liquid Chromatography (HPLC) step to separate sample into fractions, before determining the molecular weights of constituents of each fraction by ESI-MS. LCMS provides a further

dimension where the physicochemical characteristics of substrates can be used for substrate identification, and is widely used in enzymatic assays [230, 231].

Our initial efforts involved using step elution with acetonitrile in HPLC, and searching for molecular weights 291 Da (corresponding to Anh) and 309 Da (corresponding to Neu), and their mono- and di-sodium conjugates (MW + 23 and + 46). As positive controls, 2 mM Neu or Anh were injected. Under this methodology, each of the two search ranges resulted in a signal for both Neu and Anh (Figure 3.25). In the corresponding mass spectra however, peaks at the correct molecular weights are of negligible intensity (Figure 3.26). We therefore conclude that these fractions represent contaminants in the system that have been erroneously detected, and adopted another methodology.

We predicted that 1'COOH in Neu and Anh would give UV absorption at 214 nm, and decided to monitor this along in lieu of mass-specific signal detection. We repeated our positive control and observed high absorption at the 2 – 3 min region, which also overlaps with signals from both search ranges in the previous methodology. This finding is concurrent with our prediction that the polarity of Neu and Anh would lead to low affinity to the stationery phase of HPLC. Presence of Neu and Anh were verified in the MS spectra (Figure 3.27), where Neu exhibited peaks at 310, 332, and 354 Da. The 22 Da increment suggest a sodium ion (23 Da) from the buffer displaces a hydrogen, with the first likely to be at 1'COOH. Anh exhibited the expected peak at 292 Da as expected and similar sodium conjugate peaks at 314 Da and 336 Da. However, we consistently detected a contaminant peak at 274 Da alongside the 292 Da peak assigned to Anh, which could be a hydrolysed product but was not identified. This raised uncertainty to the reliability of

this assay. Nonetheless, the methodology produced characteristics peaks that are present only for Neu (310, 332, 354 Da) or Anh (292, 314 Da).

Using this methodology, we have successfully replicated the results reported in Bell et al 2020 [184]. As shown in Figure 3.28, incubation of WT with either Neu or Anh results in the appearance of the other, but only in the presence of NAD^+. Although we note the inconsistency regarding the sodium conjugation state (Figures 3.28c, 3.28d), we are nonetheless able to distinguish the Neu group of signals (310, 332, 354 Da) from the Anh group (292, 314, 336 Da). To preserve limited Anh samples, Neu was used as the substrate in all assays henceforth and peaks corresponding to Anh are assigned as reaction products.

We then investigated the activity of mutants, starting with the histidine-to-alanine mutants in the HHIH cluster (Section 3.d.iv.) (Figure 3.29). All conditions resulted in MS peaks at 310 Da, 332 Da, and 354 Da, indicating Neu. For H175A, we observed 274 Da and 292 Da peaks associated with Anh, which were absent in the Neu only control. This prima facie suggests that H175A may be catalytically active whilst H176A and H178A completely abolish catalytic activity. R150A was assayed on a different day, and while the two peaks were observed (Figure 3.30a), they were also present in the Neu-only negative control (Figure 3.30b). We conclude that these peaks resulted from non-enzymatic cleavage of Neu or data collection error during LCMS. Due to this variability, a negative control was performed alongside experimental conditions assayed on each day. We also observed that both the H175A and R150A conditions lacked the large 314 Da peak assigned to Anh in other samples (Figure 3.27b, 3.28c). These results point to the need of a more robust assay (Section 3.d.viii.).

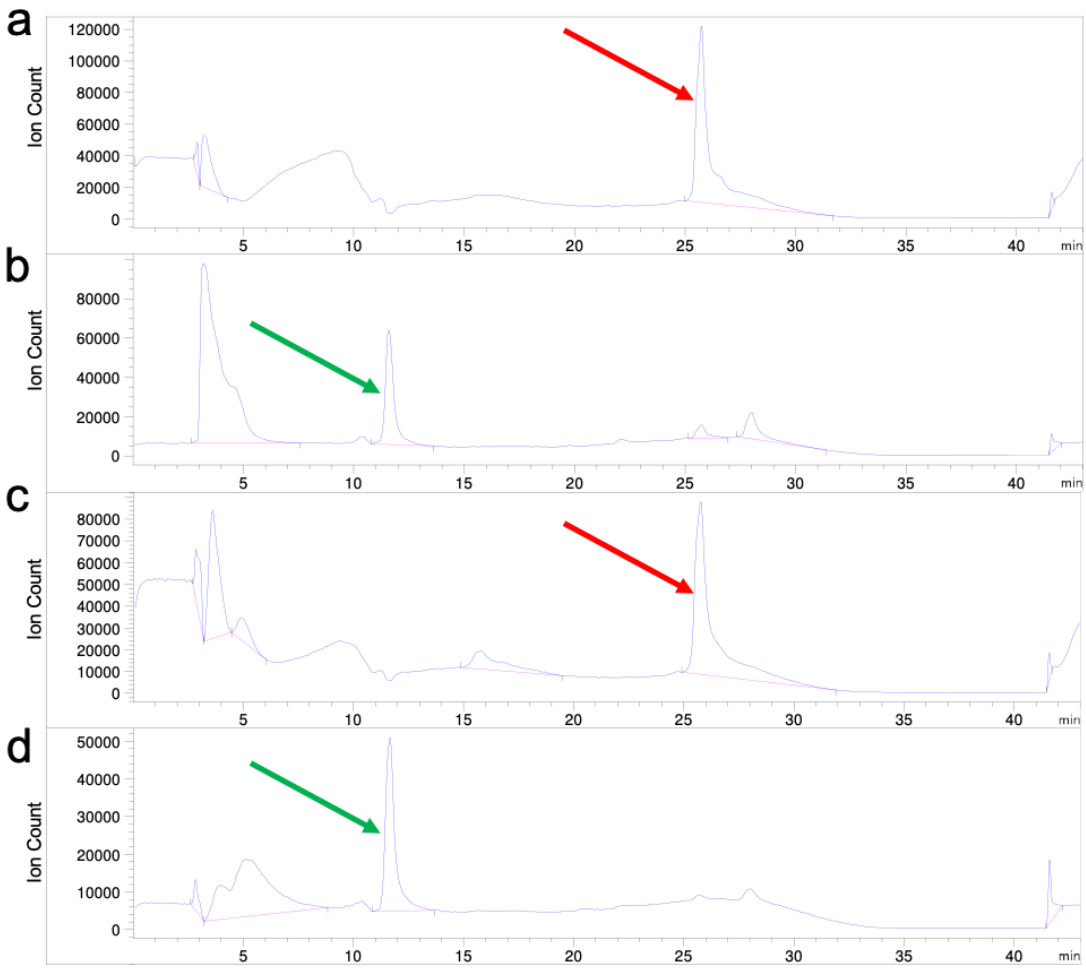


Figure 3.25 Positive controls resulted in recurrent LCMS peaks *Maximum ion count for searched range of molecular weights plotted over retention time are shown for the following samples (a) Neu injected, searching for 310, 332 Da (b) Neu injected, searching for 292, 314 Da (c) Anh injected, searching for 310, 332 Da (d) Anh injected, searching for 292, 314 Da. Recurring peaks at each search range are highlighted in red and green arrows.*

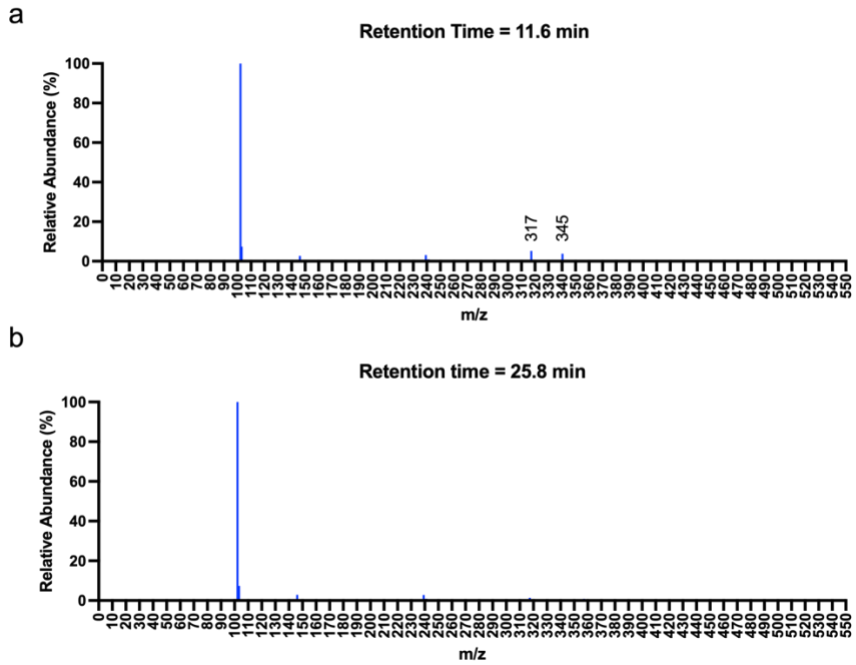


Figure 3.26 Negligible MS peaks from recurring HPLC peaks *The MS spectra from the recurring HPLC peaks with retention times of (a) 11.6 min and (b)25.8 min are shown. Peaks at relevant molecular weights are negligible, and most species at approximately 103 Da.*

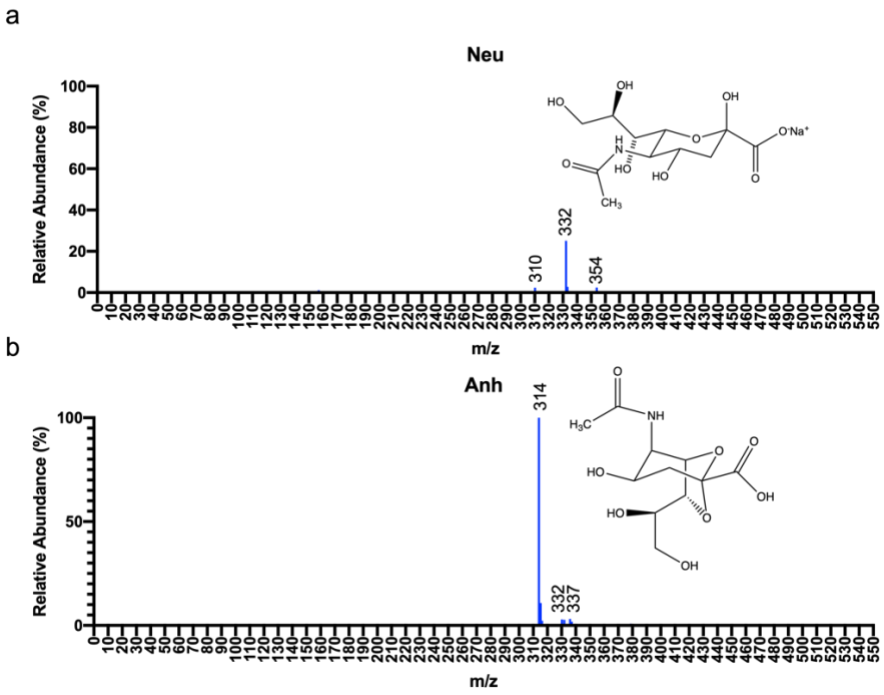


Figure 3.27 Positive controls verified by LCMS *The MS spectra of 214 nm absorption peak at 2-3 min, when either 2mM Neu or 2 mM Anh is injected. Relevant MS peaks are labelled with the corresponding molecular weights, where that of the larger of adjacent peaks are taken. The chemical structure of the proposed species corresponding to the tallest peak in each spectrum is also shown.*

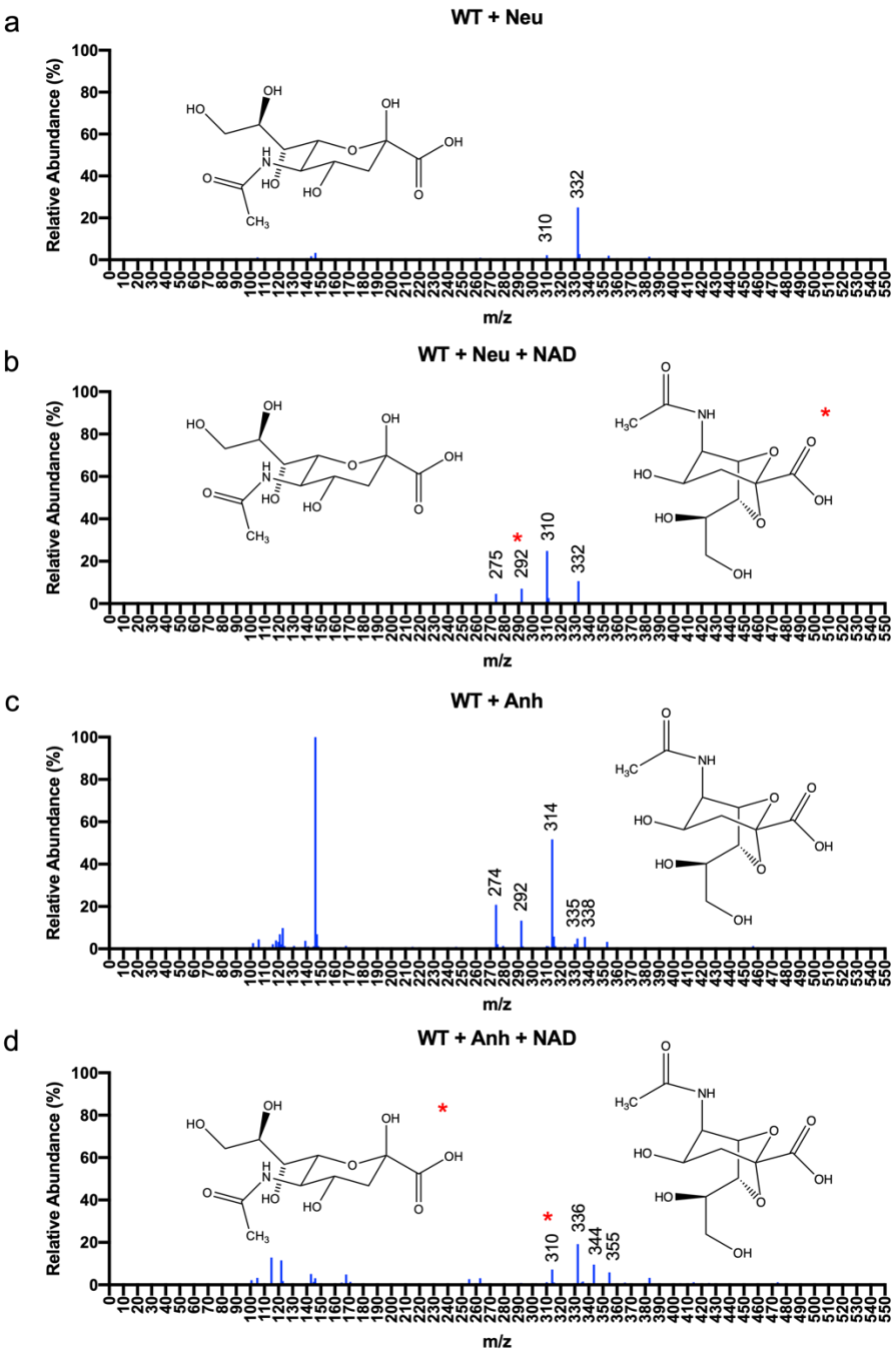


Figure 3.28 RgNanOx interconverts Neu and Anh *MS spectrum of RgNanOx incubated with (a) Neu only, (b) Neu and NAD$^+$, (c) Anh only, and (d) Anh and NAD$^+$. Comparing (a) with (b), the appearance of the 292 Da peak upon NAD$^+$ addition suggest Anh has been produced. Similarly with (c) and (d), the appearance of a 332 Da peak suggest the presence of Neu. We observe inconsistencies in sodium conjugation, such as the disappearance of the 292 Anh peak in (d) compared to (c), where there is a*

marked increase in the mono sodium-Anh signal at 332 Da. Where spectrum is a mixture of reactant and product, corresponding peaks and chemical structures are marked with a red asterisk.

We then investigated RgNanOx's activity against DANA, which was incubated with WT (Figure 3.31a) and H178A (Figure 3.31b). In the WT condition, there is a unique peak at 310 Da that is absent in the H178A condition as well as a control condition with DANA only (Figure 3.31c). This confirms that RgNanOx can catalyse the hydration of DANA into Neu, and that H178A is catalytically inactive.

Since all mutants except H175A assayed were inactive, we reasoned that some are crucial for substrate binding rather than catalysis. Obtaining kinetic data will allow us to differentiate between the two, as they will affect different kinetic parameters (K_M affected by mutation of binding residues, k_{cat} affected by mutation of catalytic residues [232]). We initially attempted to obtained this data using LCMS. WT RgNanOx was incubated with Neu, at 20 min intervals quenched with 50% acetonitrile. The summed relative abundance of peaks that correspond to Neu and Anh were then compared over time. Unfortunately, we did not observe any decrease in Neu signal nor increase in Anh signal over time. Moreover, we observed considerable fluctuation in Neu signal in the negative control condition, where no enzyme is added. No Anh signal was observed in the negative control. Overall, this indicated that the LCMS assay is best used as a qualitative assay.

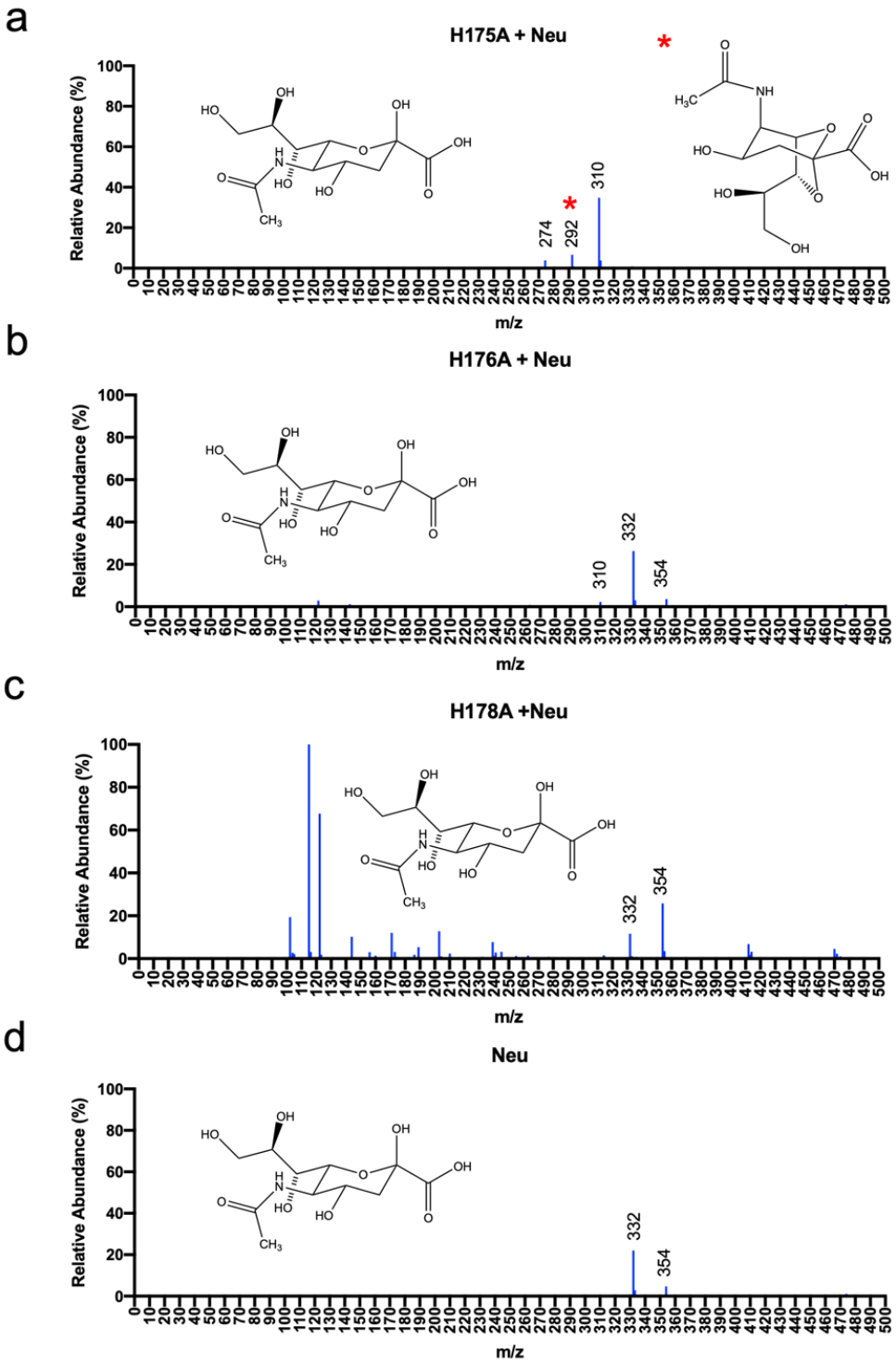


Figure 3.28 Histidine mutants appear inactive in MS spectra *of Neu incubated with (a) H175A, (b) H176A, (c) H178A, and (d) only Neu as a negative control. Only H175A showed Anh product peak 292 Da (red asterisk). Where spectrum is a mixture of reactant and product, corresponding peaks and chemical structures are marked with a red asterisk.*

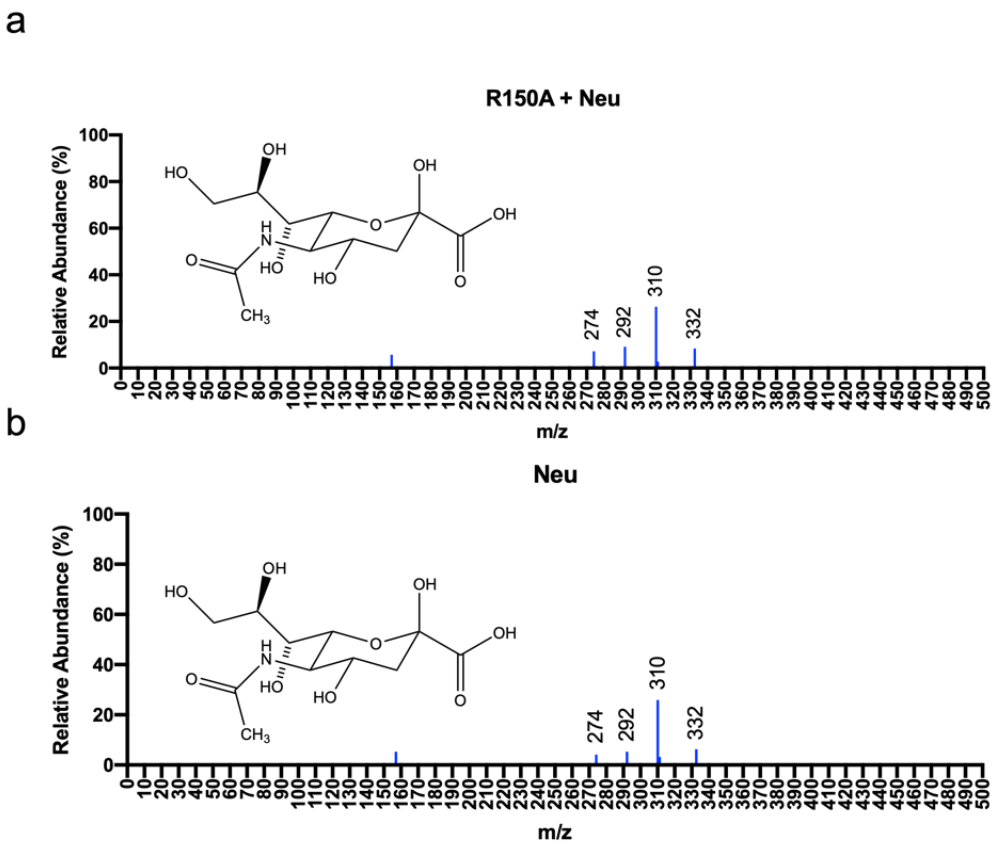


Figure 3.30 R150A is prima facie inactive *ESI-MS spectra of (a) R150A incubated with Neu and (b) Neu only as a negative control. These samples are virtually identical, with Neu peaks at 310 Da and 332 Da, as well as ambiguous peaks at 274 Da and 292 Da, were present in both conditions.*

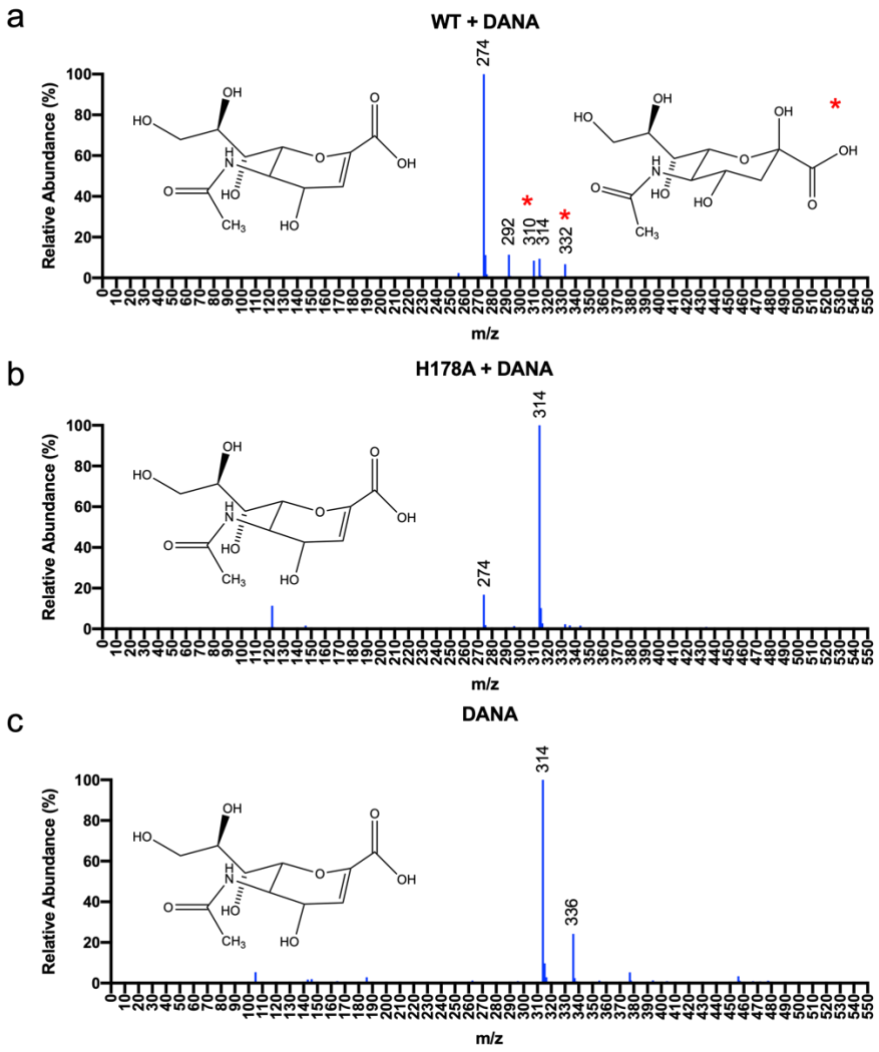


Figure 3.31 RgNanOx showed activity against DANA *ESI-MS spectra of WT RgNanOx incubated with DANA (a), showing formation of a species with molecular mass of 310 Da corresponding to Neu and its monosodium conjugate at 332 da. These product peaks (red asterisk) were not observed in the ESI-MS spectra of H178A RgNanOx incubated with DANA and NAD^+ (b), or negative control with DANA only (c). Where spectrum is a mixture of reactant and product, corresponding peaks and chemical structures are marked with a red asterisk.*

3.d.vii. ITC binding assay

Since we were unable to obtain kinetic data by LCMS, we attempted to use ITC to detect substrate binding, which would also allow us to select suitable substrates and mutants for co-crystallisation trials. Our initial experiment involved titrating Neu against protein solution. This yielded ambiguous results as there is some increase in energy with increased ligand:protein ratio (Figure 3.32a) compared to negative control of titrating Neu against buffer solution (Figure 3.32b). We attempted to increase the signal by increasing the concentrations of protein and titrant by 50%, but we were unable to interpret our results (Figure 3.32c). This suggests that our initial results were experimental artefacts and that there is weak substrate binding.

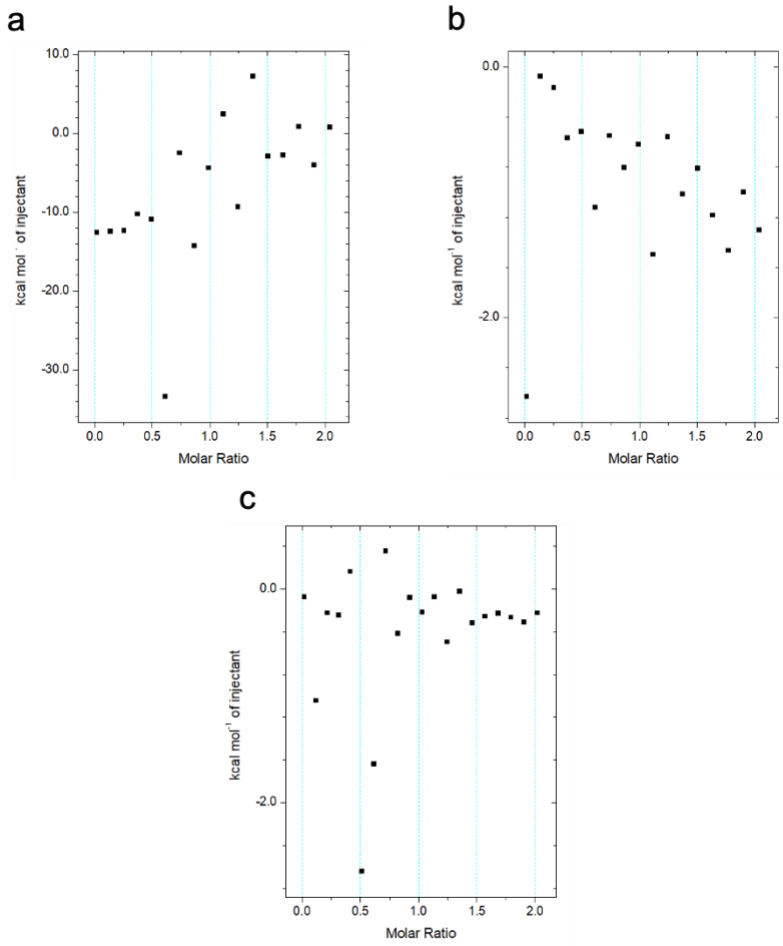


Figure 3.32 Initial titration curves showed no clear signal *Titration curves of (a) Neu against RgNanOx, (b) Neu against buffer, and (c) Neu against RgNanOx with increased concentrations*

We reasoned that citrate in the active site of the RgNanOx holo-structure (Section 3.d.iii.) could compete with substrate binding. We attempted to remove cytosolic citrate by dialysing protein before crystallisation. Since citrate binds to the active site through its acid groups, we performed three rounds of dialysis with SEC buffers containing 100 mM sodium malonate, 100 mM sodium acetate, and finally in SEC buffer only. We were still unable to detect Neu binding from dialysed RgNanOx (Figure 3.33).

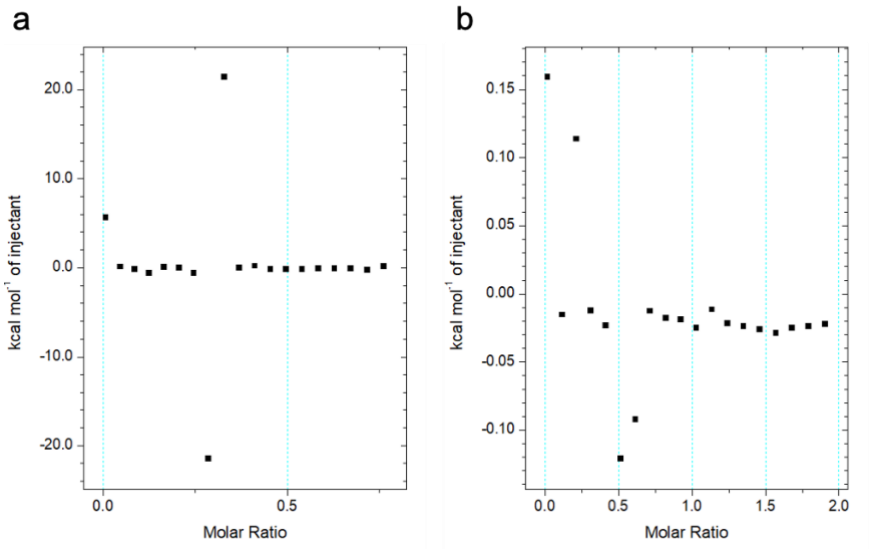


Figure 3.33 Dialysed protein did not bind Neu *Titration curves of (a) Neu against dialysed RgNanOx and (b) Neu against buffer.*

3.d.viii. NMR kinetic assay

We collaborated with Ben Gaunt and Aziz Khan of the Ben Davis Group at the Rosalind Franklin Institute to develop an NMR kinetic assay. Using pure compounds as positive controls, Neu and Anh signals were assigned from the in 2.0 – 2.10 ppm region in the 1D proton NMR spectra, corresponding to the acetyl group. We then incubated WT RgNanOx with either Neu or Anh and monitored the reaction at 500 s intervals. For both incubation with Neu (Figure 3.34) and Anh (Figure 3.35), we observed an increase in

product signal from the first timepoint at which a scan was obtained (t = 0 s) and at t = 6400 s, indicating enzymatic conversion. We concluded that the reaction has reached equilibrium as we were able to integrate corresponding peaks to replicate literature values for equilibrium between α and β anomers of Neu (92% β in solution [233]) (Figure 3.36); and between Neu and Anh (Neu:Anh = 2:1 [184]).

We optimised the starting conditions to obtain the initial rates required for calculating kinetic parameters of RgNanOx. We incubated RgNanOx with a wide range of starting Neu concentrations and found that at low concentrations, large errors incurred due to the small increase in product peak signals (Figure 3.37). This error prevented calculation of initial reaction rates at the linear portion of the graph, we therefore opted to stay within the millimolar range. We observed that using Anh instead of Neu as the reactant resulted in greater product signal when integrated against the DSS standard (Figure 3.38). This is because the chemical equilibrium between Neu and Anh lies 2:1 towards Neu [183], so starting with Anh brings about a greater change in product concentration. However, the change in reactant brought about two drawbacks. Firstly, Anh is not commercially available and was synthesised by our collaborators at the University of East Anglia [185]. The limited yield meant we were limited to a maximum starting concentration of 15 mM, and we were only able to conduct the experiment on two protein samples (WT and H175A). Secondly, although the decrease Anh peak intensity follows a similar tend to the increase in β-Neu (Figure 3.39a,b), α-Neu (Figure 3.39c) exhibited more fluctuations despite having similar overall magnitude of increase as β-Neu at equilibrium. We reasoned that RgNanOx first converts Anh into the α-anomer, which spontaneously converts into the more stable β-anomer in solution [234]. This is consistent with the sum of peak integrals, which are linearly proportional to molar concentrations, remain broadly

level at approximately 5.0 between t = 0 sand t = 4000 s (where equilibrium is assumed to be reached). Alas, the peak integrals were inacurrate as the ratio between Anh:Neu (1.0:7.3) and α:β Neu (1.0:1.4) are far from their literature values of 1:2 [183] and 1:49 [234]. Instead, we attempted to confirm this by quenching the intermediate by incubating the reaction in 10% methanol, which we reasoned would undergo nucleophilic attack. Unfortunately, we were unable to detect such an intermediate in either NMR nor LCMS. As such, the concentrations of the two Neu anomers are modelled separately (Figures 3.40 and 3.41). However, we note that the *E. coli* orthologue YjhC was recently found to behave in this way [233].

Following these changes, we incubated WT RgNanOx with starting Anh concentrations from 0.2 mM to 15 mM. Using a Python code to fit peak integrals of each species over time to three parameters (a,b,c), we described the relationship between initial rate (f(x)) and initial Anh concentration (x) (Figures 3.40 and 3.41). We estimated RgNanOx's kinetic parameters using the resulting Lineweaver-Burk plot (Figure 3.42), which has gradient of 6900 +/- 500 s, and y-intercept of 400 +/- 1000 smM^{-1}. According to the Michaelis-Menten relationship, in the Lineweaver-Burk plot the gradient corresponds to K_M/V_{max} and the y-intercept corresponds to $1/V_{max}$. This estimates the V_{max} to be 2.5 +/- 6.3 μMs^{-1} and K_M to be 17 +/- 2.5 mM. The high K_M is consistent with our difficulty in detecting binding using ITC. This suggested that a much higher substrate concentration (at least 50 mM) would be required to observe the plateau in the Michaelis-Menten curve required to calculate reliable kinetic data. Instead, we used the limited substrate to characterise H175A. Since we did not observe any Neu signal after incubation for three days, we confirmed that H175A is catalytically inactive.

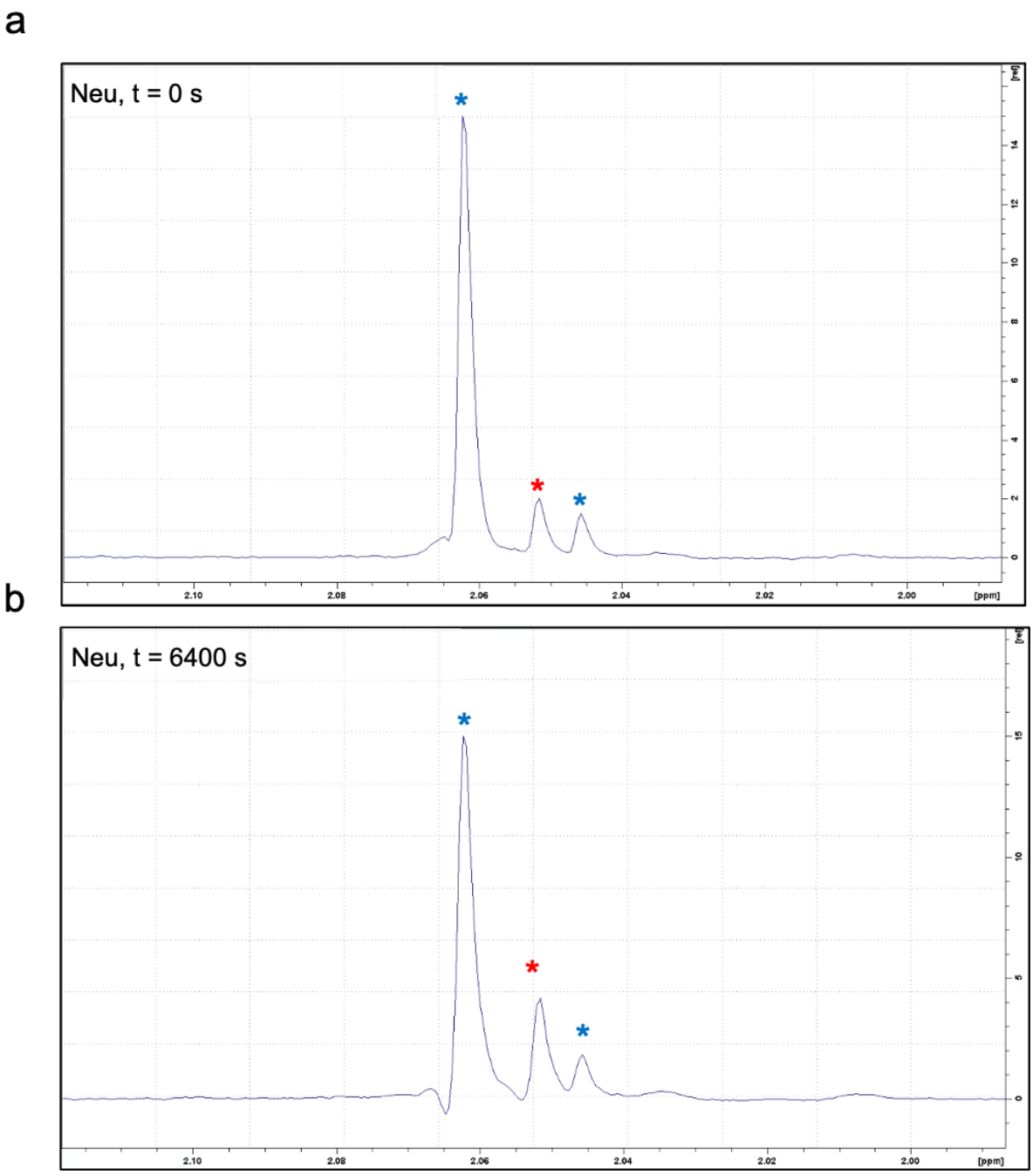


Figure 3.34 NMR showed Neu-Anh conversion *Proton NMR spectra of RgNanOx incubated with Neu at t = 0 s (a) and t = 6400 s (b). Peaks marked with blue asterisks correspond to Neu and peaks marked with red asterisks correspond to Anh. Anh peak shows a marked increase in intensity.*

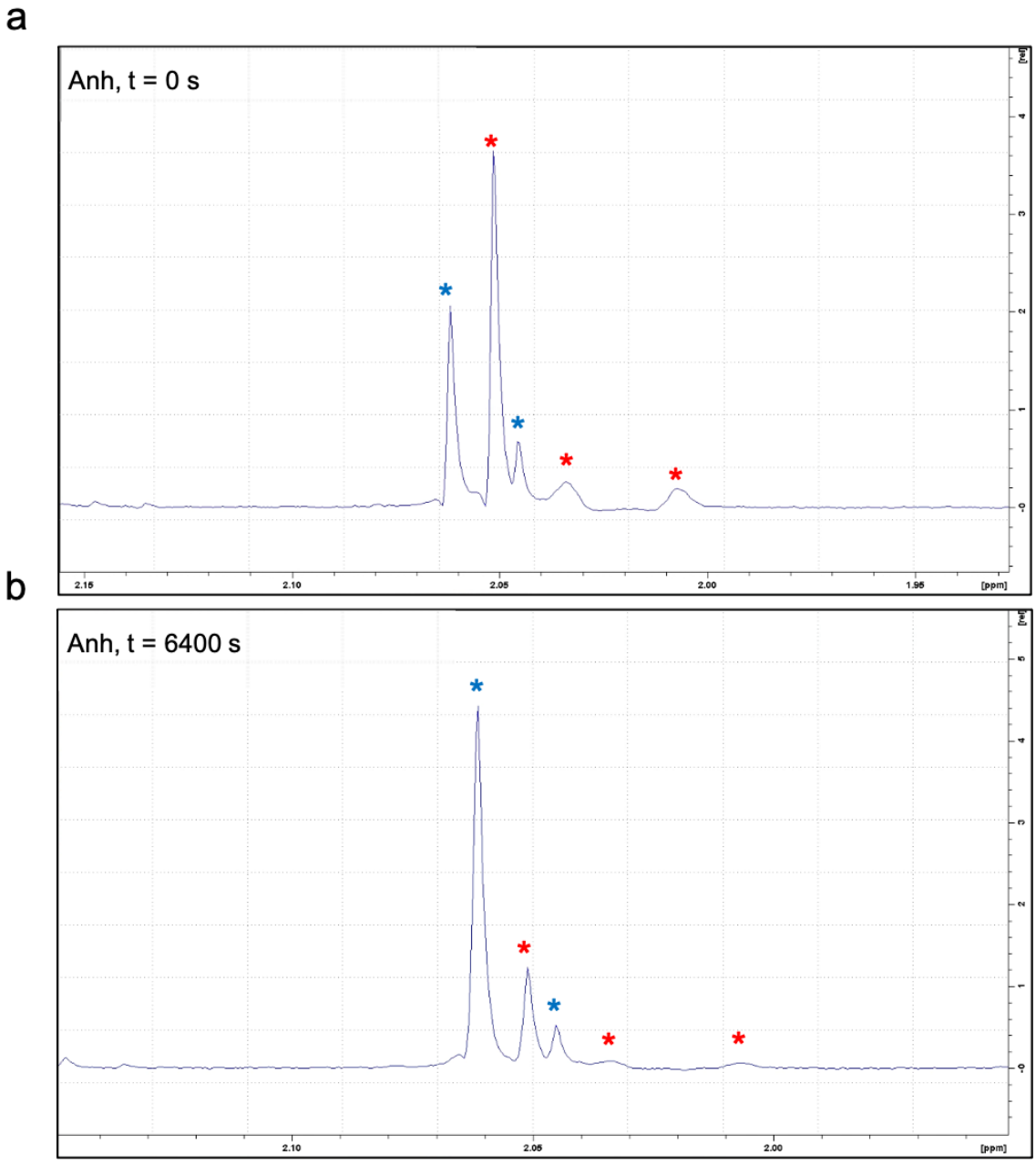


Figure 3.35 NMR showed Anh-Neu conversion *Proton NMR spectra of RgNanOx incubated with Anh at t = 0 s (a) and t = 6400 s (b). Peaks marked with blue asterisks correspond to Neu and peaks marked with red asterisks correspond to Anh. Both Neu peaks exhibited increase in intensity.*

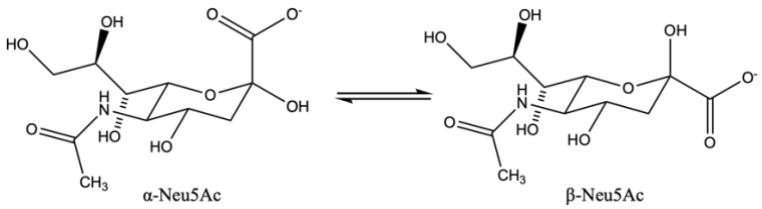


Figure 3.36 Anomers of Neu5Ac

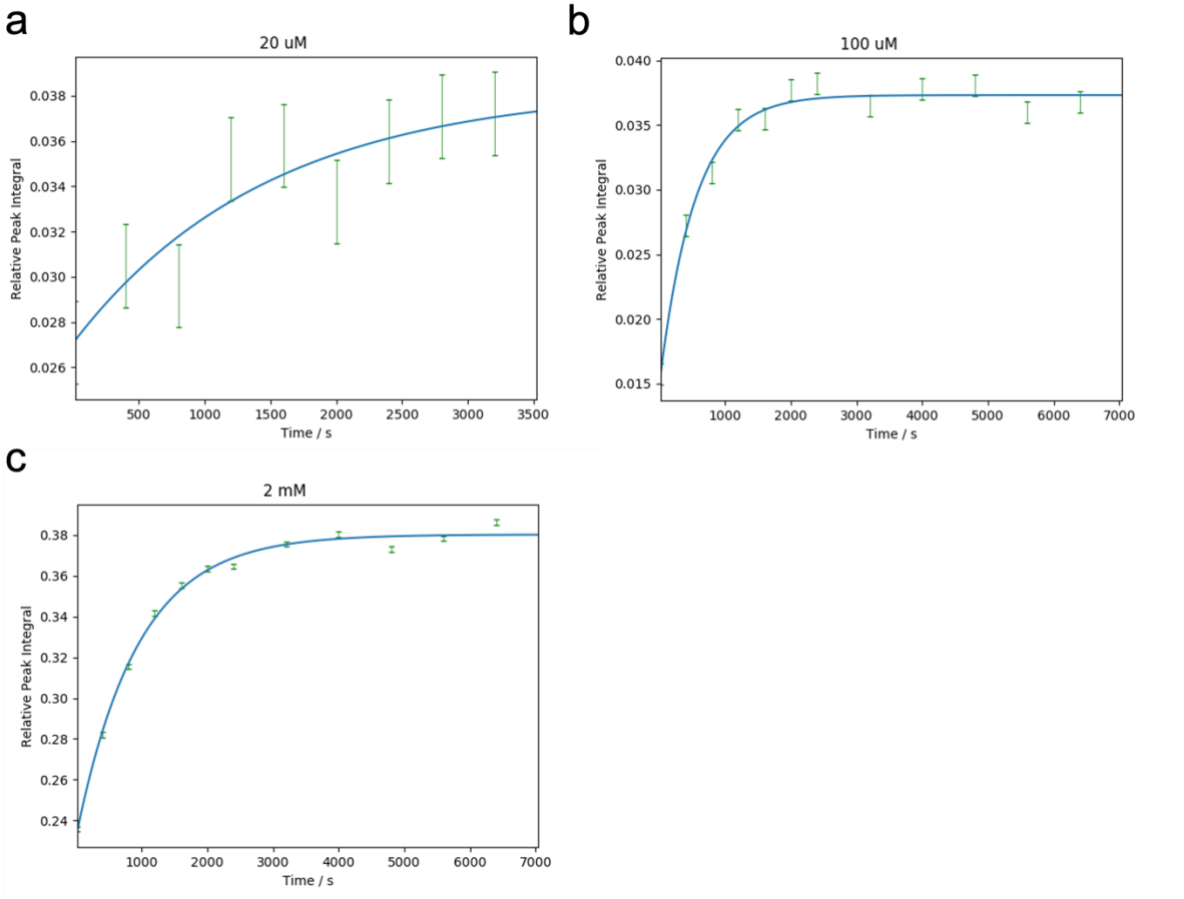


Figure 3.37 Anh signal over time at varying starting Neu concentrations *Plots of Anh concentration over time with starting Neu concentrations of (a) 20 μM, (b) 100 μM, and (c) 2 mM. Error bars, which are more significant at lower concentrations, are shown in green.*

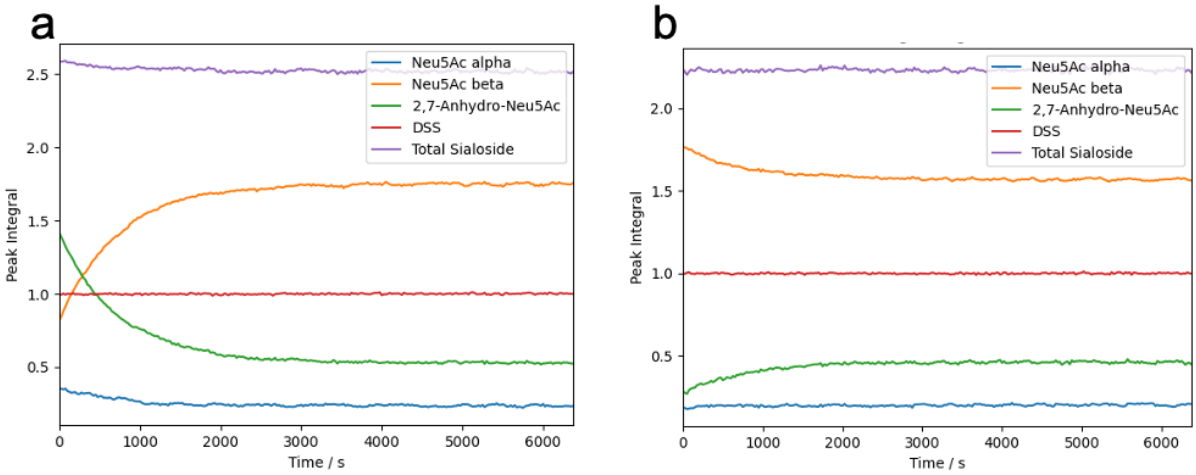


Figure 3.38 Starting with Anh produced greater signal *Signal intensities of all species following incubation of RgNanOx with Anh (a) or Neu (b). DSS is included as a standard.*

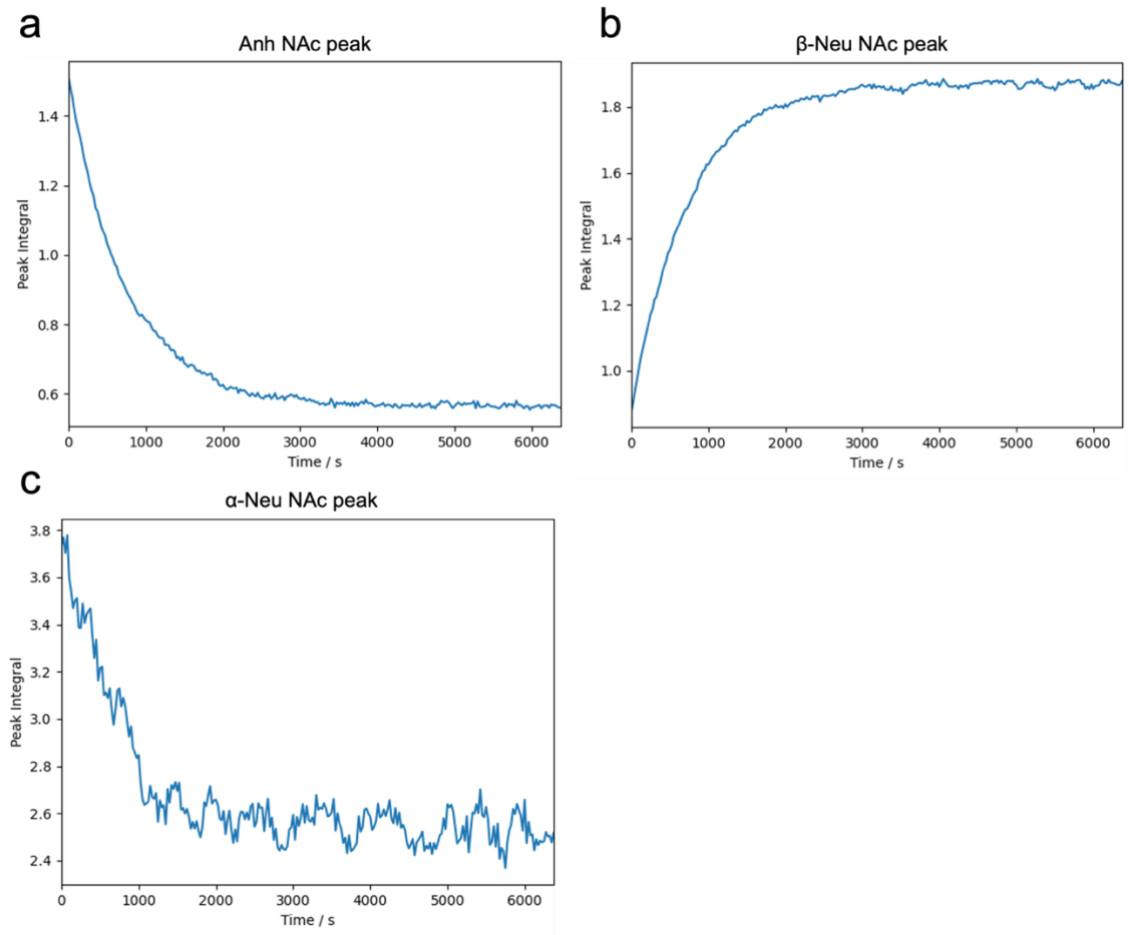


Figure 3.39 Signal intensity traces were inconsistent between anomers *Signal intensities of NAc peaks from Anh (a), β-Neu (b), and α-Neu (c) plotted over time.*

$$f(x) = a + (c \times e^{-bx})$$

Figure 3.40 Equation used to fit Anh and α-Neu peak integrals

$$f(x) = a(1 - e^{-b(x+c)})$$

Figure 3.41 Equation used to fit β-Neu peak integrals

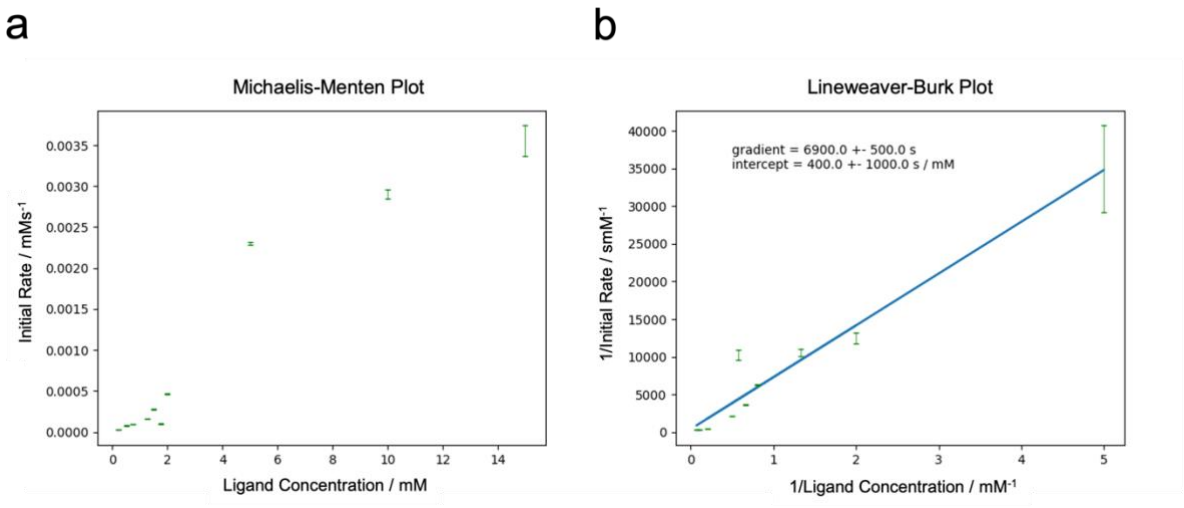


Figure 3.42 Graphical method for estimating K_M and k_{cat} *Michaelis-Menten (a) and Lineweaver-Burke (b) obtained by incubating WT RgNanOx with varying starting concentrations of Anh. A linear line of best fit (blue) was used to obtain the gradient and intercept from the Lineweaver-Burke plot. Error bars are shown in green*

3.d.ix. Co-crystallisation screening

Since we were unable to screen for substrate binding conditions through ITC, we have performed screened for co-crystallisation by incubating WT RgNanOx with 20 mM of Neu, Anh, or DANA at 37 °C for 3 hr. Crystal trials were set up to using the optimised MES and citrate screens (Section 3.d.iii.) which previously yielded crystals. However, datasets collected from crystals in these conditions yielded similar active sites with citrate-like density, which we conclude to have been pulled out of the cytosol during purification. We also performed crystal trials with dialysed protein (Section 3.d.vii.). However, crystals trials with both optimised (MES and citrate) screens and commercial screens (LMB [235], SG1 [236], Morpheus [237], JCSG [238]) yielded only crystals with poor morphology.

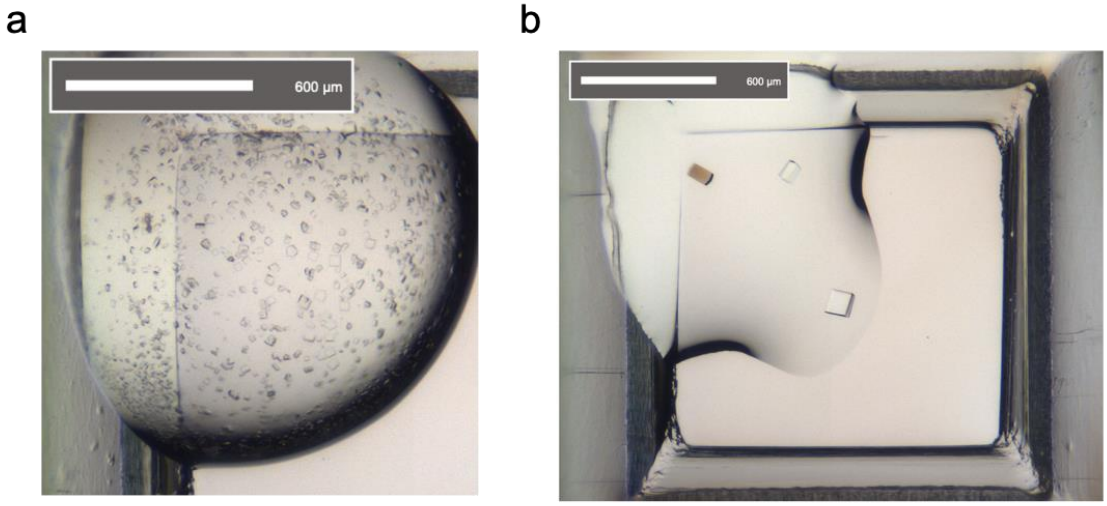


Figure 3.43 Optimisation of complex crystals *Crystals before (a) and after (b) optimisation, both containing sodium thiocynate as an additive, showing plate-like crystals appear thicker after optimisation.*

We then moved onto co-crystallisation trial with H178A, incubated with 20 mM Neu or Anh at 37 °C for 3 hr. We rescreened crystal conditions with the commercial screens listed above, which yielded numerous crystal hits. We noticed that hit conditions shared a common precipitant, PEG3350, as well as additives magnesium chloride, magnesium formate, and sodium thiocynate (Figure 3.43a). We formulated a combined optimisation screen by varying pH, precipitant concentration, and additive concentration (Appendix A), which yielded crystals with ideal morphology (Figure 3.43b).

3.d.x. Elucidation of the dehydration step using the Neu complex structure

Since RgNanOx was found to be active against DANA (Section 3.d.vi.), we co-crystallised WT RgNanOx with 20 mM DANA. The resulting 2.30 Å structure is a tetramer in the asymmetric unit (Figure 3.44a). In addition to the NAD^+ active site, there is a clear density for a cyclic compound (Figure 3.44b) which we propose corresponds to the substrate. However, we found clear electron density at the 2’ position that likely

corresponds to a hydroxyl group (Figure 3.44c). Together with the LCMS results showing conversion from DANA to Neu (Section 3.d.vi.), we propose that the bound substrate is Neu rather than DANA.

The Neu complex structure superposed well with the high resolution holo-structure (Section 3.d.iii), with RMSD of 0.513 Å over 4633 atoms. The protein retains the same fold (Figure 3.45a) and orientation for putative catalytic residues (Figure 3.45b). The hydrogen bond network of Neu reveals three clusters of residues that coordinate different positions (Figure 3.46). The Lys93/His178 cluster near 4'OH confirms our hypothesis that these conserved catalytic residues His178 and Lys93 are associated with NAD^+-mediated oxidoreductase activity. The His175/His176/Lys163 cluster coordinate the 1' and 2' positions, stabilised by the hydrogen bond between Lys163 and 1' COOH. The Tyr13/Arg150 cluster stabilises the glycerol chain and 5' NAc group respectively.

The carboxylate group in Neu cocrystallised with RgNanOx is similarly well-coordinated than Neu cocrystallised with other neuraminidases (Figure 3.47). This remains true even though *M. viridifaciens* sialidase (Figure 3.47d) was cocrystallised with β-Neu instead of α-Neu. This indicates that active site residues are key in determining substrate stereoselectivity in neuraminidases, and is further support for our hypothesis that RgNanOx uses α-Neu as its substrate.

Since DANA is the mimic of the dehydration step product 4-keto-DANA (Figure 3.48a), this structure provided crucial information on its molecular basis. We observed that the bound Neu is the α-anomer rather than the β-anomer that is predominant in solution (Section 3.d.viii). Its 1' carboxylate group is well coordinated in the active site (Figure

3.45c), and superposes well with the citrate C_β carboxylate group from the holo-structure, with RMSD = 0.513 Å over 4633 atoms (Figure 3.49). This along with our NMR data (Section 3.d.viii) suggests that α-Neu is the bona fide substrate, which could explain why we were able to obtain clearer substrate density from co-crystallising RgNanOx with DANA compared to Neu, the latter of which would reduce occupancy because only 8% of molecules will be the α anomer [234]. Nonetheless, we remain open the possibility that *in vivo* the 1' COOH and 2' OH positions may be reversed compared to our crystalline structure.

By the principle of microscopic reversibility [239], the dehydration step in the RgNanOx mechanism would follow the exact reverse of the mechanism for hydration of DANA. Since orbitals overlap is a pre-requisite for an elimination reaction, different mechanisms must be proposed for each anomer. For α-Neu, we propose a net *syn*- elimination mechanism consisting of two anti-periplanar eliminations (Figure 3.48b). Firstly, 2'OH is protonated then leaves as water after the ring oxygen forms an oxocarbenium ion. Secondly, the 3' axial proton is leaves to form the 2'-3' double bond. For β-Neu, the dehydration step would follow a simple *trans*- elimination mechanism (Figure 3.48c).

The catalytic residues were proposed on the basis of active site geometry, indicated by the dihedral angles between the Lewis acid that abstracts 2'OH, 2'C, 3'C, and Lewis base that abstracts 3'H. In the syn- mechanism, abstraction of 2'OH and 3'H occurs on the plane of the ring, leading to dihedral angle of approximately 45°. With dihedral angle of 50°, Tyr298 and His175 appear ideally positioned as the Lewis acid and Lewis base respectively (Figure 3.50a,b). For the trans mechanism, abstraction of 2'OH and 3'H occurs in opposite directions perpendicular to the 2'-3' bond, leading to dihedral angles

of approximately 180°. With dihedral angle of 130°, Lys163 and His176 are the most ideally positioned as the Lewis acid and Lewis base respectively (Figure 3.50c,d). We were unable to determine whether the syn or trans mechanism was more likely, because while the former scheme has the more optimal dihedral angle, the distance between Tyr298 and 2'OH is larger than expected, at 5.2 Å.

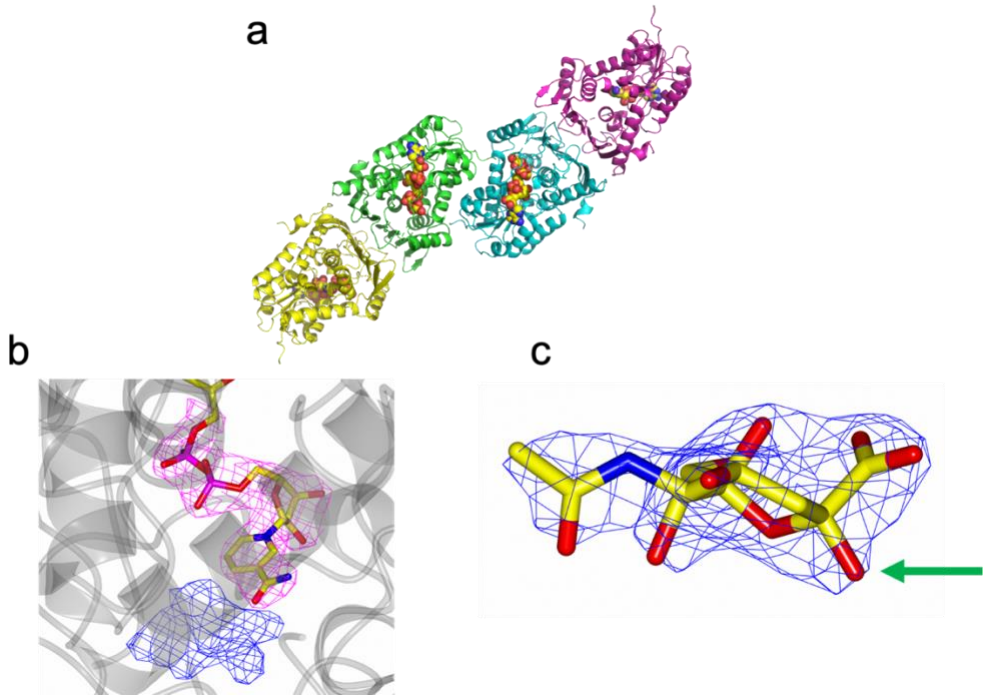


Figure 3.44 RgNanOx/Neu complex structure *(a) RgNanOx/substrate tetramer shown as ribbons, with NAD^+ shown as spheres (b) Active site structure with $2F_O$-F_C map (σ = 1.4) shown as mesh, showing electron density for a cyclic compound in blue proximal to the NAD^+ density in magenta (c) $2F_O$-F_C map (σ = 1.4) after fitting Neu, showing the 2'OH fits the electron density shown with a green arrow.*

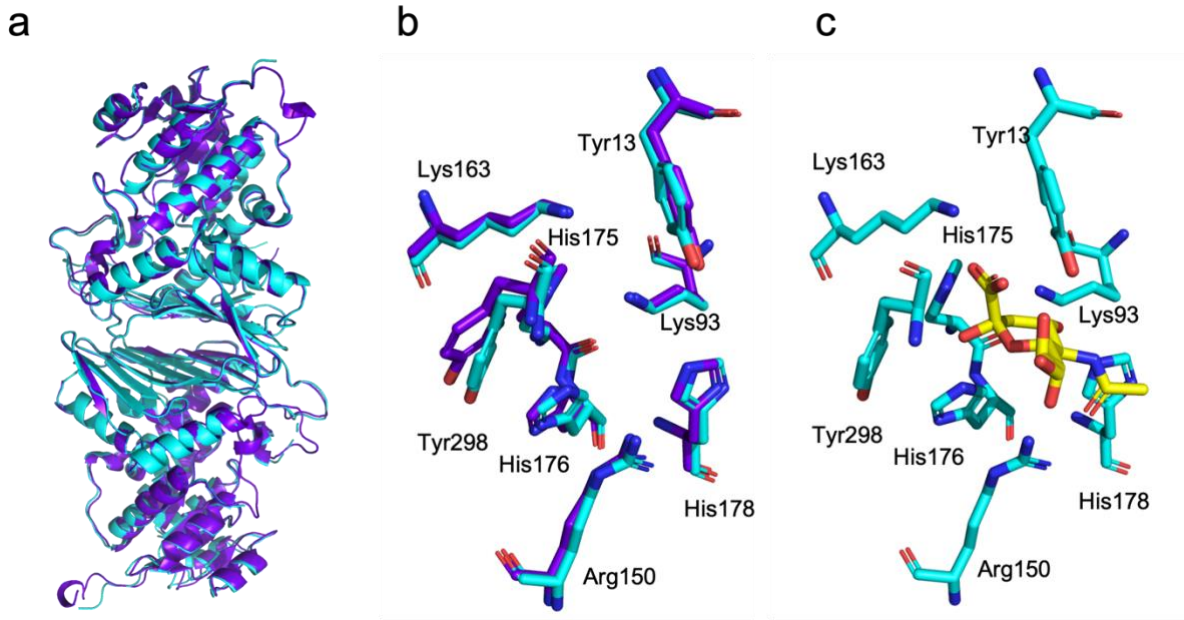


Figure 3.45 Superposition of apo- and complexed RgNanOx *(a) Superposition of the apo-structure dimer (purple) with the chain B/C dimer of complex structure (cyan) (b) Superposition of active site residues shown as sticks and in colours corresponding to (a), and Neu in the complex structure coloured according to element. (c) Complex structure active site residues coordinate α-Neu (yellow).*

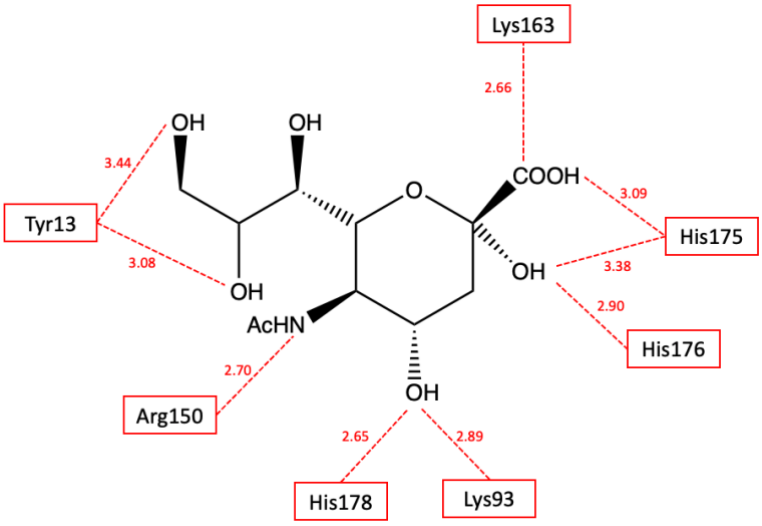


Figure 3.46 Hydrogen bonding network of Neu *Residues within 4 Å and thus capable of hydrogen bonding to each residue are represented as red squares. The distance in angstroms between the closest electronegative atoms are shown along the dotted lines that represent hydrogen bonds.*

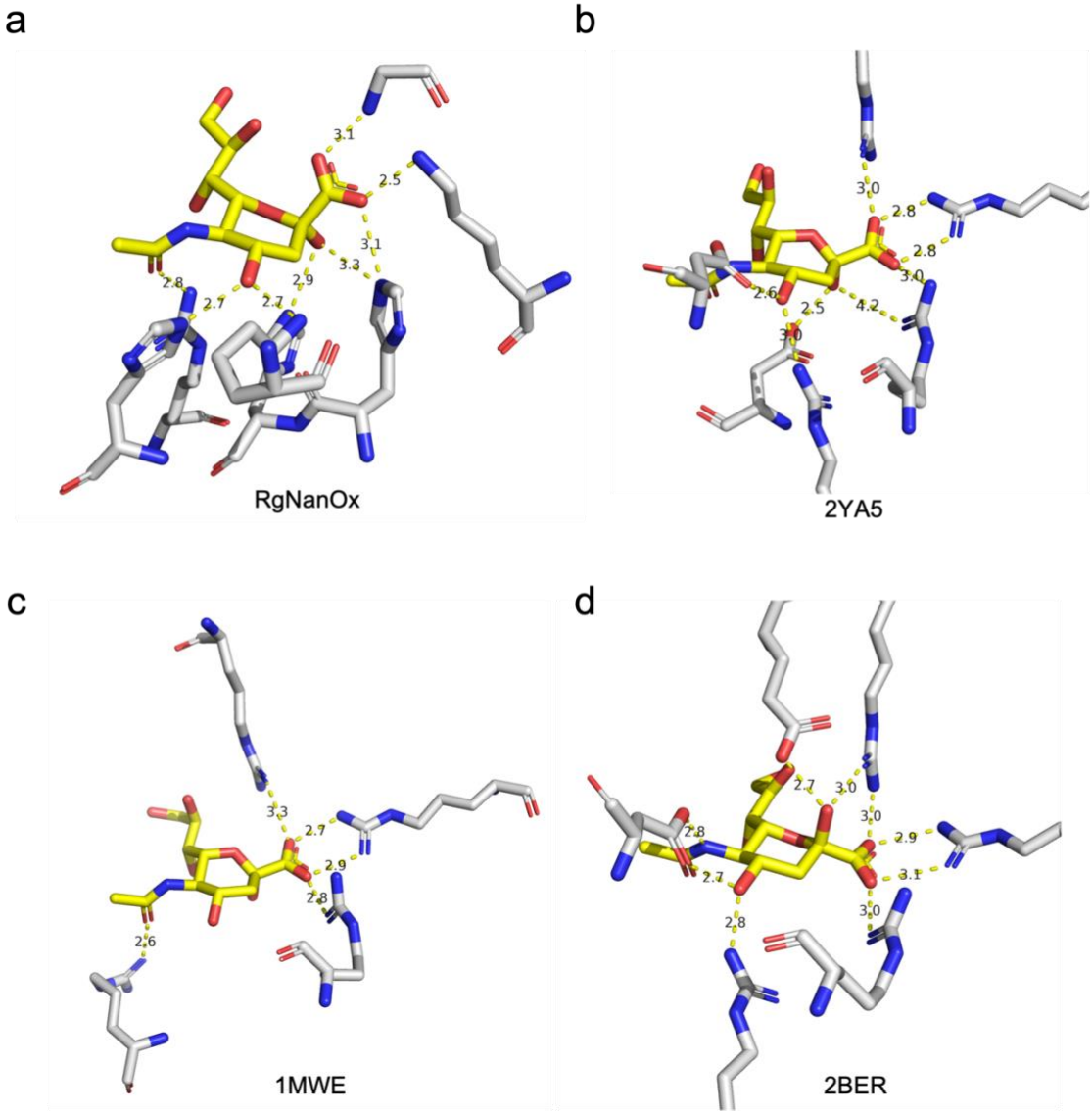


Figure 3.47 Neu5Ac carboxylates are well-coordinated *Co-crystallised Neu (yellow sticks) and hydrogen bonding residues in RgNanOx (a), spNanA (b), N9 neuraminidase (c), and M. viridifaciens sialidase (d). Distances (yellow dashes) are labelled in Angstroms.*

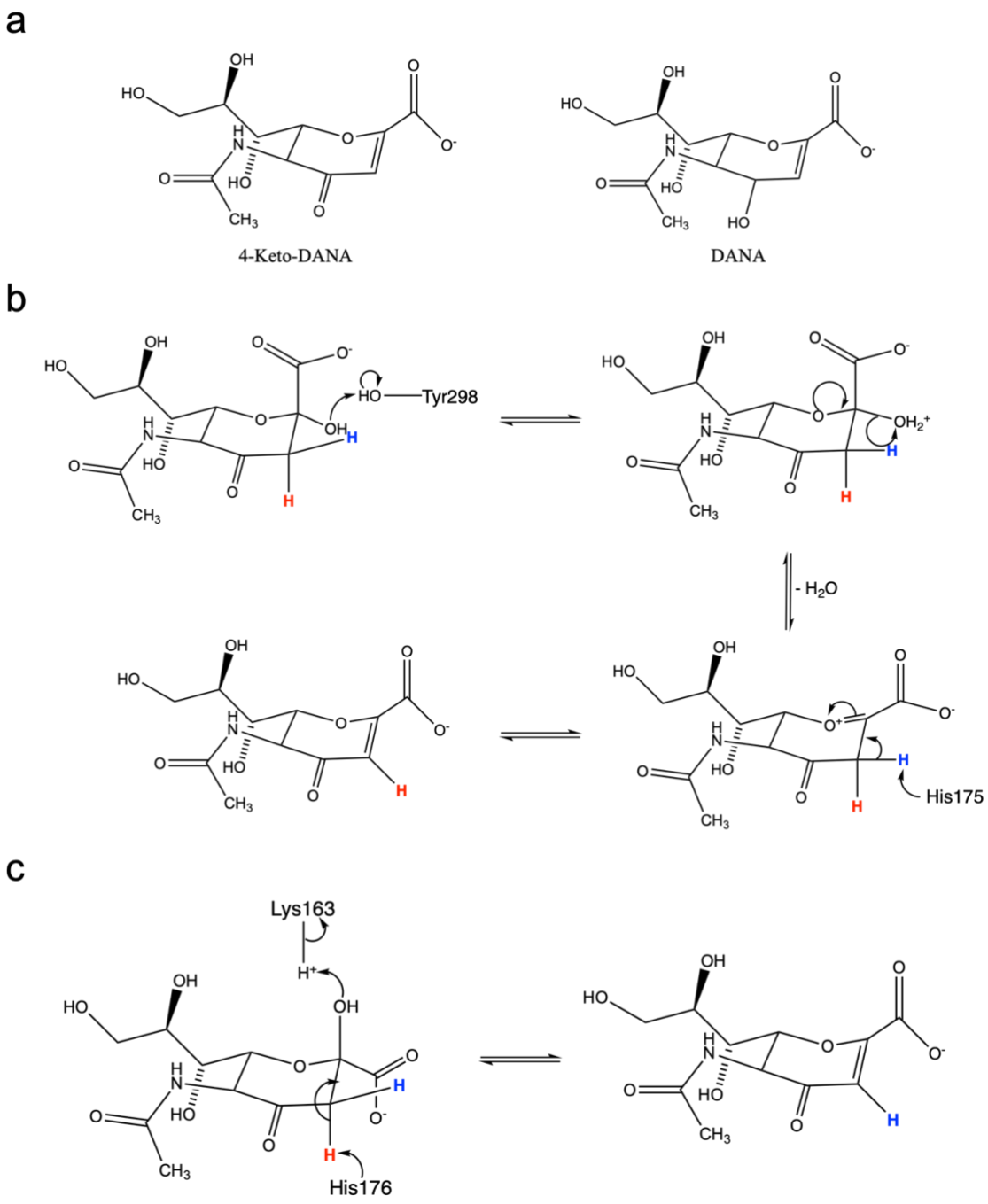


Figure 3.48 Dehydration step mechanism proposed for each anomer *(a) Structures of the dehydration step product 4-Keto-DANA and DANA. (b) Syn- elimination mechanism proposed for α anomer as substrate (c) Trans- elimination mechanism proposed for β anomer as substrate.*

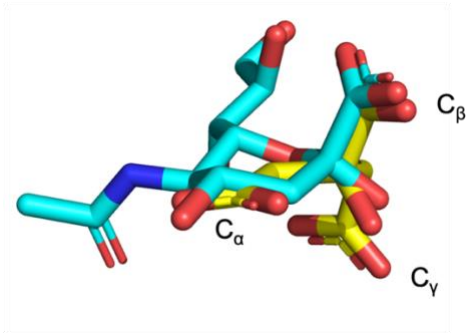


Figure 3.49 Superposition of co-crystallised Neu and citrate

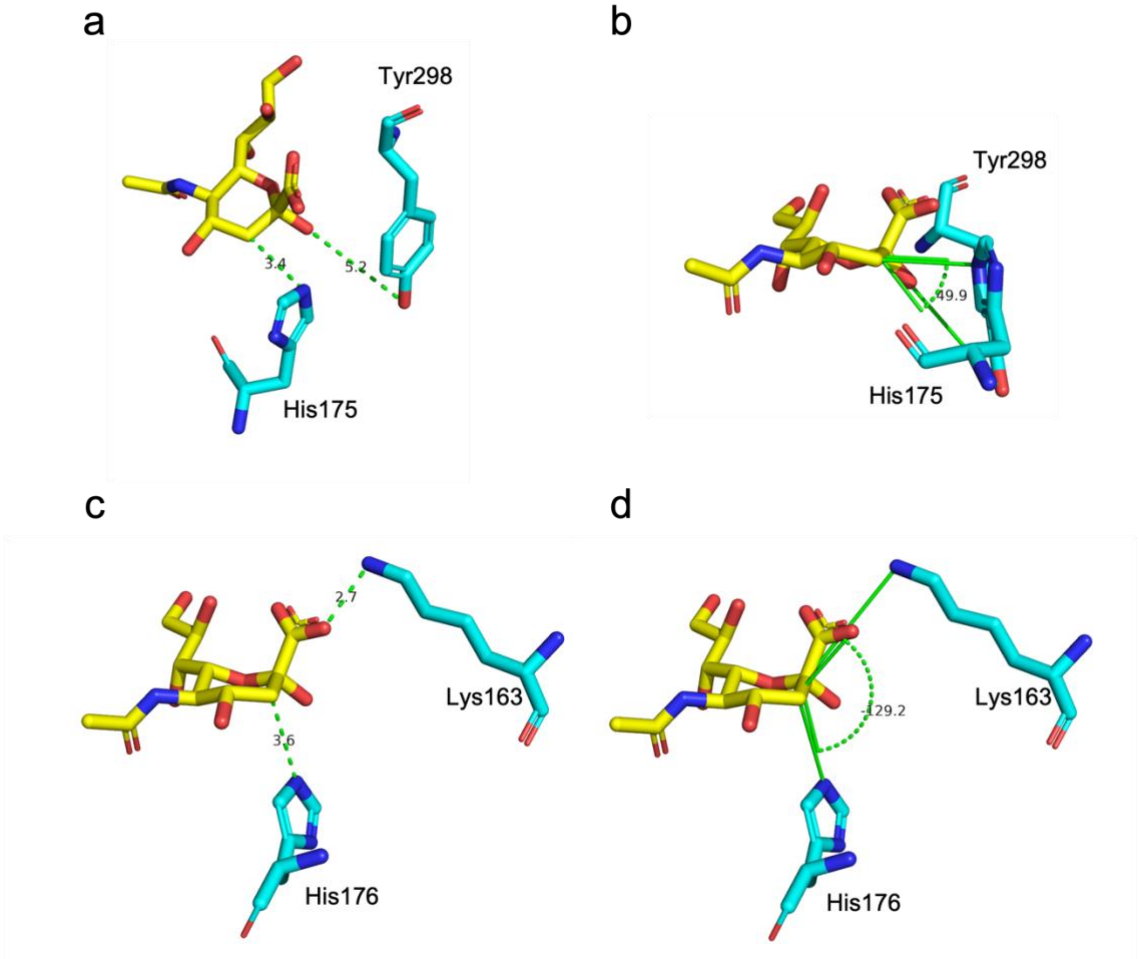


Figure 3.50 Catalytic residues ideally positioned for dehydration *Neu (yellow sticks) and putative catalytic residues (cyan sticks) from the Neu complex structure was shown. Tyr298 and His175 were proposed for the syn- mechanism as they are ideally distanced (a) and angled (b). Lys163 and His175 were proposed for the trans- mechanism as they are ideally distanced (c) and angled (d).*

3.d.xi. Elucidation of the cyclisation step using the Anh complex structure

To obtain a complex structure with the ring-flipped substrate Anh, we co-crystallised putative catalytic mutant H178A (Sections 3d.iv., 3d.vi.) RgNanOx with Anh. We obtained a 1.96 Å structure that is a dimer in the asymmetric unit (Figure 3.51a), but our efforts in fitting the substrate in its density were hindered by the weakness of its signal where electron density, even when contoured at $\sigma = 0.9$ (Figure 3.51b). Nonetheless, after refining the structure without the substrate, the resulting positive density in the difference map was clearly bicyclic (Figure 3.52a). We propose that two strong densities (Figure 3.52a, blue arrows) correspond to the carboxylate and acetyl groups, which are well coordinated by active site residues in the Neu complex structure. Therefore the plane between these two sites, which is approximately parallel to the plane of the nicotinamide ring of NAD^+, corresponds to the 6-carbon ring. This also positions 4'OH approximately 90° above the C4 atom of NAD^+, ideally positioned for its regeneration. The additional density observed on the side of this plane, opposite from the NAD^+ position (Figure 3.52a, orange arrow), we propose to be the three member ring formed between 2' and 7' carbons. We observed that after fitting Anh (Figure 3.52b) negative density that arose (Figure 3.52c), which we concluded was due to less than full occupancy.

Anh has a different hydrogen bonding network from the Neu complex structure discussed above (Figure 3.53). Here, the 2' position is coordinated only by His176 and not His175, the former of which is also within borderline distance (4.05 Å) to coordinate 9'OH. Due to the displacement of the glycerol chain after ring flip, Tyr13 no longer coordinates Anh whilst Arg150 coordinates 8'OH whilst maintaining its relative position from the 5' NAc

group. Lys93 and Lys163 both retain their positions relative to 4'OH and 1'COOH respectively.

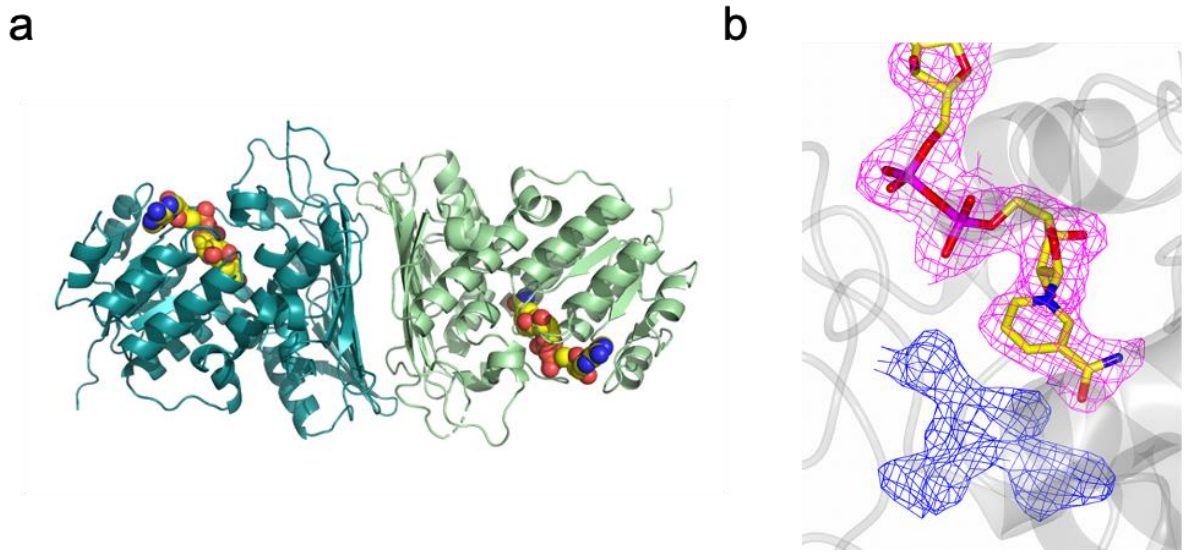


Figure 3.51 RgNanOx/Anh complex structure *(a) RgNanOx/Anh dimer shown as ribbons, with NAD^+ shown as spheres (b) Active site structure with $2F_O$-F_C map (σ = 0.9) shown as mesh, showing electron density for a substrate in blue proximal to the NAD^+ density in magenta*

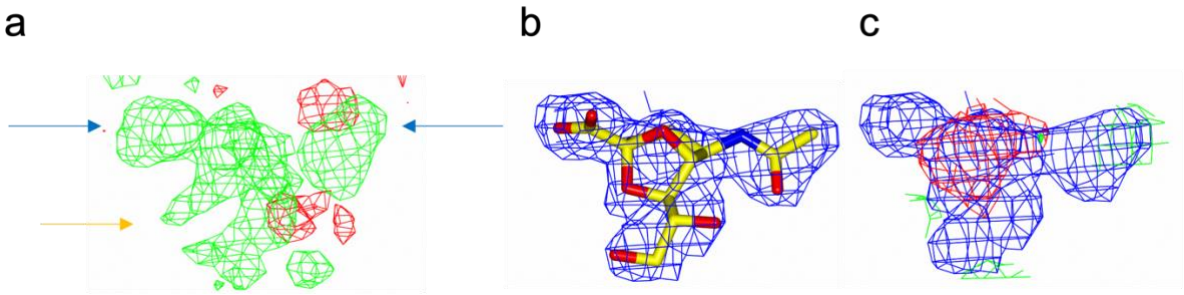


Figure 3.52 Bicyclic density in Anh complex structure *(a) F_O-F_C map (σ = +/- 3) of bicyclic substrate density. The six member ring was fitted on the plane between the blue arrows, whilst the additional density shown by the orange arrow corresponds to the three member ring. Anh fitted well in the $2F_O$-F_C map (σ = 1) (b) and F_O-F_C map (σ = +/- 3) (c), the latter showed significant negative density in the centre of the C4molecule.*

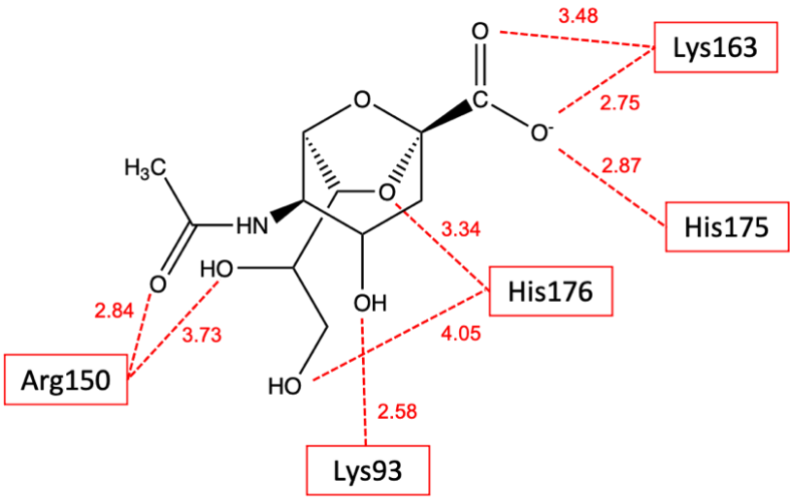


Figure 3.53 Hydrogen bonding network of Anh *Residues within approximately 4 Å and thus capable of hydrogen bonding are represented as red squares. The distance in angstroms between the closest electronegative atoms are shown along the dotted lines that represent hydrogen bonds.*

3.d.xii. Proposal of the RgNanOx Catalytic Mechanism

The data shown above allowed us to proposed a detailed RgNanOx mechanism for conversion of either α-Neu or β-Neu to Anh (Figure 3.54). Neu binds to RgNanOx with its acetyl group stabilised by Arg150 and glycerol tail by Tyr13. The reaction is initiated by NAD^+-mediated oxidation of the 4'OH in a similar manner to the canonical GFO mechanism [211], where the conserved His178 acts as Lewis base while the conserved Lys93 adjacent lowers the pK_a for electron transfer. The resulting 4' keto group acidifies the 3' proton, which is abstracted along with 2'OH to form 4-keto-DANA. If as we have shown, α-Neu is the substrate, His175 and Tyr298 are responsible for this step through *syn*- elimination. Alternatively, if β-Neu is the substrate, Lys163 and His176 are responsible for this step through *trans*- elimination (red box, Figure 3.54). 4-keto-DANA then spontaneously ring flip, with the glycerol tail previously coordinated by Tyr13 now hydrogen bonded to Arg150. His176 then acts as a Lewis base to activate 7'OH, which cyclises via a Michael addition to 2', similar to a published GH4 glycoside mechanism

[203]. This electron transfer from the 2'-3'double bond to form a 3'-4' double bond, which is in turn transferred to His176 as it now acts as a Lewis acid. Finally, NAD^+ is regenerated from NADH by reduction of the 4' keto group, producing Anh.

We propose the syn- mechanism for the dehydration step to be more likely than the trans mechanism for the following reasons. Firstly, our NMR assayed showed significant levels of α-Neu being produced by the reaction. Secondly, in the α-Neu complex structure both 1'COOH and 2'OH were well coordinated by hydrogen bonds. Finally, the mechanism predicts a catalytic role for His175, and we confirmed by NMR that H175A is catalytically inactive.

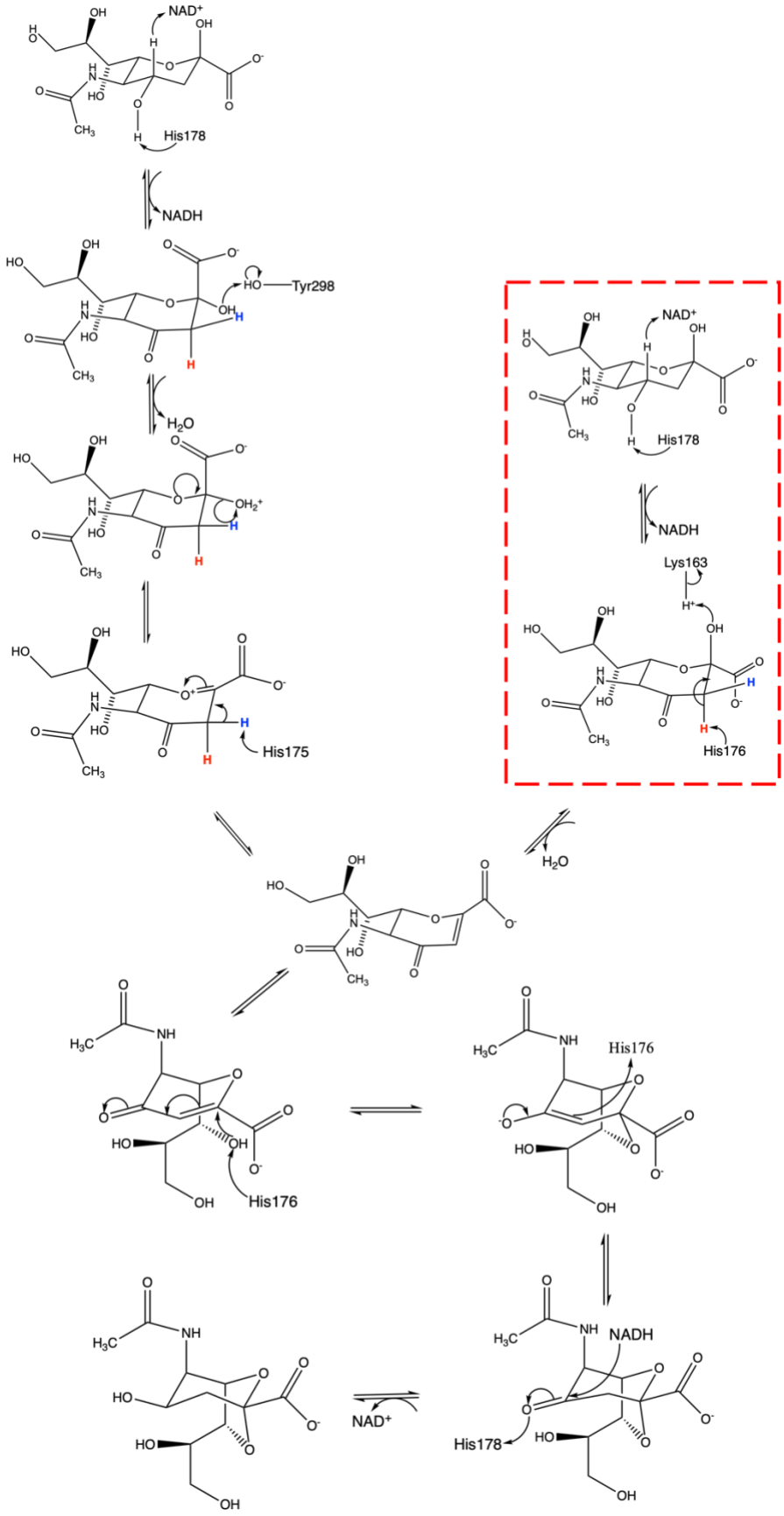


Figure 3.54 Proposed RgNanOx Mechanism *Red box shows an alternative mechanism using β-Neu*

3.e. Conclusion

The RgNanOx holo-structure, published in Bell et al 2020 [184], allowed proposal of a general mechanism that resembles those of the GH4/GH109 family of glycosidases, and is distinct from sialidase mechanisms reported to date. Using this structure we were able to identify putative catalytic residues, including the novel HHIH catalytic domain, for functional characterisation.

Successful adaptation of the direct injection ESI-MS assay to our LCMS system confirmed the published results that many mutants including R150A, H176A, and H178A were inactive, indicating catalytic roles. We also found RgNanOx was able to convert the sialidase inhibitor DANA to Neu, which we used to obtained the Neu complex structure. However, the assay was unreliable due to a contaminant peak at 274 Da, and inconsistent mass spectra that required running controls with every experiment. This also meant we were unable to obtain kinetic data, thereby unable to differentiate between binding and catalytic residues. Since our ITC binding assay failed, we turned to NMR for a functional assay for which we collaborated with the Ben Davis group. We were able to obtain a rough estimation of RgNanOx's kinetic parameters, which suggested it binds to Neu very weakly and may explain why our ITC assay was insufficiently sensitive. A more precise characterisation would require more technical optimisation and datapoints at high substrate concentration, of which we had limited amounts as it required in house synthesis by our collaborators. Using NMR we also attempted to characterise H175A due to its ambiguous result from LCMS, and found activity to be completely abolished. Optimisation of this assay is key to the expeirments proposed in Section 3.f.

We obtained the RgNanOx/Neu complex structure determined at 2.30 Å by incubating RgNanOx with DANA, which was converted into α-Neu. This led us to propose an unusual syn-elimination mechanism as discussed below. Using the H178A mutant we also obtained an Anh complex structure determined at 1.96 Å. Notwithstanding the ambiguous substrate density, we were able to observe electron density for a bicyclic compound, allowing us to fit Anh into the density. The hydrogen bonding network of the ring-flipped substrate showed His176 to be the only candidate for 7'OH activation in the cyclisation step. Together, these results allowed us to account for the catalytic roles of the histidine residues in the novel HHIH catalytic motif and propose a full reaction mechanism for RgNanOx.

3.f. Discussion and future directions

The structural studies reported herein reported have two significant implications. Firstly, α-Neu is likely the bona fide substrate for RgNanOx, and; Secondly, the dehydration step likely occurs through syn-elimination. The Neu complex structure and NMR data led us to conclude that α-Neu is the bona fide substrate. Although RgNanOx is likely to act downstream of the IT sialidase NanH in *R. gnavus* [240], this has interesting implications on sugar scavenging in gut bacteria because while β-Neu is the dominant form in solution, α-Neu is more common in biomacromolecules such as α-2,3-sialyllactose. This suggests that there is a competitive advantage in being able to scavenge α-Neu quickly, before released Neu in the gut environment spontaneously converts to the β anomer. Recent publications regarding the Anh permease YjhB and RgNanOx orthologue YjhC in *E. coli* [183, 215, 216] shows the importance of catabolism of Anh in gut bacteria species beyond *R. gnavus*. Interestingly, this is compounded by the fact that there is no known IT sialidase orthologue in *E. coli*, which generates Anh in *R. gnavus*. This raises two interesting

possibilities, that (i) YjhC (the NanOx orthologue) is expressed in *E. coli* to scavenge Anh generated by species in the microbiome that have an IT sialidase like *R. gnavus*; and that (ii) NanOx evolved from a bona fide sialidase. In the latter case, the Michael reaction mechanism of NanOx allows it to catalyse the reaction irrespective of the side chains attached the centre of reaction. In other words, it could have permitted a sialidase reaction whereby Neu is attached to glycoproteins by 2' O linkage.

Having α-Neu as the substrate means RgNanOx cannot undergo anti-periplanar *trans*-elimination, which is usually favoured by the requirement of orbital overlap. However, *syn*-elimination is rare but not impossible, and was invoked for crotonase [241], UDP-N-acetylglucosamine 2-epimerase [242], Pel10Acm sugar lyase [243], and α-1,4-glucan lyase [244]. In RgNanOx, *syn*-elimination occurs by two *trans*- elimination steps driven by the ring oxygen, first of the 2' hydroxyl then of the equatorial 3' proton.

It is a working hypothesis of this thesis that due to α-Neu being bona fide substrate, our ITC and NMR assays were inconclusive because the effective substrate concentration was at 8%. Unfortunately, we were able to identify putative catalytic residues for both the *syn*- and *trans*- mechanisms (which respectively use the α and β anomer), and we were not able to experimentally narrow down the active site residues with a catalytic role using the LCMS assay. We therefore propose the following experiments to prove the mechanism proposed for RgNanOx: (i) To generate site-selective deuterium derivatives for each 3' proton. The derivative that gives rise to a decrease in reaction rate, measured by NMR, would indicate that the corresponding position is involved in the reaction, hence determining whether RgNanOx undergoes *syn*- or *trans*-elimination. (ii) Alternatively, using time-sensitive NMR [245] to detect stereo-selective addition of water to DANA.

(iii) Characterise RgNanOx activity against ketodeoxynonulosonic acid (KDN), an Neu analogue that has a 5’ hydroxyl in place of the NAc group. If RgNanOx is active against KDN it would verify the binding, rather than catalytic, role of 5’ NAc in Neu.

4. Executive Summary

Herein we reported on structural studies performed on two enzymes involved in sugar metabolism in bacteria. Determination of the structure of the lipid III scramblase Wzx was proof of concept that long wavelength native s-SAD is compatible with delicate crystals formed in lipidic cubic phase. Though only one conformer was obtained, comparison with previously determined structures of homologues MurJ and NorM provided insight to Wzx mechanism. Most notably, our data suggest that Wzx does not employ the alternating access mechanism proposed for MurJ, the lipid II substrate of which is similar to that of Wzx. Our structure also identified two arginine clusters that we predict to be crucial for its catalytic function, and are key targets for further structural studies.

We then reported on the determination of three structures of the sialic acid oxidoreductase RgNanOx, consisting of the holo-structure as well as Neu and Anh substrate complex structures. The structures allowed us to propose two mechanisms for RgNanOx, a *trans*-mechanism for utilising β-Neu as substrate and a *syn*- mechanism for utilising α-Neu as substrate. Our data suggests that the α anomer is the bona fide substrate, which has further implications in sialic acid scavenging strategies employed in the gut microbiome.

Bibliography

1. Blow DM. Outline of crystallography for biologists / David Blow. Oxford: Oxford : Oxford University Press, 2002.; 2002.

2. Chayen NE. Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. Acta Crystallogr D Biol Crystallogr. 1998;54(Pt 1):8-15.

3. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallogr Sect D-Biol Crystallogr. 2010;66:22-5.

4. Long F, Vagin AA, Young P, Murshudov GN. BALBES: a molecular-replacement pipeline. Acta Crystallogr Sect D-Biol Crystallogr. 2008;64:125-32.

5. Ke HM. Overview of isomorphous replacement phasing. In: Carter CW, Sweet RM, editors. Macromolecular Crystallography, Pt A. Methods in Enzymology. 276. San Diego: Elsevier Academic Press Inc; 1997. p. 448-61.

6. Hendrickson WA. Determination of Macromolecular Strucutres from Anomalous Diffraction of Synchrotron Radiation. Science (New York, NY). 1991;254(5028):51-8.

7. Vasur J, Kawai R, Larsson AM, Igarashi K, Sandgren M, Samejima M, et al. X-ray crystallographic native sulfur SAD structure determination of laminarinase Lam16A from Phanerochaete chrysosporium. Acta Crystallogr Sect D-Struct Biol. 2006;62:1422-9.

8. Vinothkumar KR, Henderson R. Single particle electron cryomicroscopy: trends, issues and future perspective. Q Rev Biophys. 2016;49:1-25.

9. Ludtke SJ, Baker ML, Chen DH, Song JL, Chuang DT, Chiu W. De novo backbone trace of GroEL from single particle electron cryomicroscopy. Structure. 2008;16(3):441-8.

10. Sgro GG, Costa TRD, Cenens W, Souza DP, Cassago A, Coutinho de Oliveira L, et al. Cryo-EM structure of the bacteria-killing type IV secretion system core complex from Xanthomonas citri. Nature microbiology. 2018;3(12):1429-40.

11. Ramrath DJF, Niemann M, Leibundgut M, Bieri P, Prange C, Horn EK, et al. Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. Science (New York, NY). 2018;362(6413).

12. Henderson R. Overview and future of single particle electron cryomicroscopy. Arch Biochem Biophys. 2015;581:19-24.

13. Henderson R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q Rev Biophys. 2009;28(2):171-93.

14. Taylor KA, Glaeser RM. Electron diffraction of frozen, hydrated protein crystals. Science. 1974;186(4168):1036-7.

15. Yeates TO, Agdanowski MP, Liu Y. Development of imaging scaffolds for cryo-electron microscopy. Current Opinion in Structural Biology. 2020;60:142-9.

16. Srivastava A, Tiwari SP, Miyashita O, Tama F. Integrative/Hybrid Modeling Approaches for Studying Biomolecules. Journal of Molecular Biology. 2020;432(9):2846-60.

17. Zhang K, Keane SC, Su Z, Irobalieva RN, Chen M, Van V, et al. Structure of the 30 kDa HIV-1 RNA Dimerization Signal by a Hybrid Cryo-EM, NMR, and Molecular Dynamics Approach. Structure. 2018;26(3):490-8.e3.

18. Nielsen JT, Kulminskaya NV, Bjerring M, Linnanto JM, Ratsep M, Pedersen MO, et al. In situ high-resolution structure of the baseplate antenna complex in Chlorobaculum tepidum. Nat Commun. 2016;7:12454.

19. Demers JP, Habenstein B, Loquet A, Kumar Vasa S, Giller K, Becker S, et al. High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. Nat Commun. 2014;5:4976.

20. Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, et al. Structure of the chemokine receptor CXCR1 in phospholipid bilayers. Nature. 2012;491(7426):779-83.

21. Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc. 1999;34(2):93-158.

22. Karplus M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. Journal of the American Chemical Society. 1963;85(18):2870-1.

23. Linge JP, O'Donoghue SI, Nilges M. Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods Enzymol. 2001;339:71-90.

24. Staunton D, Schlinkert R, Zanetti G, Colebrook SA, Campbell ID. Cell-free expression and selective isotope labelling in protein NMR. Magn Reson Chem. 2006;44 Spec No:S2-9.

25. Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP. Macromolecular NMR spectroscopy for the non-spectroscopist. Febs j. 2011;278(5):687-703.

26. Kayikci M, Venkatakrishnan AJ, Scott-Brown J, Ravarani CNJ, Flock T, Babu MM. Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. Nat Struct Mol Biol. 2018;25(2):185-94.

27. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee Gyu R, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373(6557):871-6.

28. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020;577(7792):706-10.

29. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)-Round XIII. Proteins. 2019;87(12):1011-20.

30. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.

31. Taylor G. The phase problem. Acta Crystallogr D Biol Crystallogr. 2003;59(Pt 11):1881-90.

32. Taylor G. Introduction to phasing. Acta Crystallographica Section D. 2010;66(4):325-38.

33. Sheldrick G. Phase annealing in SHELX-90: direct methods for larger structures. Acta Crystallographica Section A. 1990;46(6):467-73.

34. Scapin G. Molecular replacement then and now. Acta crystallographica Section D, Biological crystallography. 2013;69(Pt 11):2266-75.

35. Perutz M. Isomorphous replacement and phase determination in non-centrosymmetric space groups. Acta Crystallographica. 1956;9(11):867-73.

36. Kivelä HM, Abrescia NG, Bamford JK, Grimes JM, Stuart DI, Bamford DH. Selenomethionine labeling of large biological macromolecular complexes: probing the structure of marine bacterial virus PM2. J Struct Biol. 2008;161(2):204-10.

37. Mueller-Dieckmann C, Panjikar S, Tucker PA, Weiss MS. On the routine use of soft X-rays in macromolecular crystallography. Part III. The optimal data-collection wavelength. Acta Crystallogr D Biol Crystallogr. 2005;61(Pt 9):1263-72.

38. Brennan S, Cowan PL. A suite of programs for calculating x-ray absorption, reflection, and diffraction performance for a variety of materials at arbitrary wavelengths. Review of Scientific Instruments. 1992;63(1):850-3.

39. Wagner A, Duman R, Henderson K, Mykhaylyk V. In-vacuum long-wavelength macromolecular crystallography. Acta Crystallogr D Struct Biol. 2016;72(Pt 3):430-9.

40. Basu S, Olieric V, Leonarski F, Matsugaki N, Kawano Y, Takashi T, et al. Long-wavelength native-SAD phasing: opportunities and challenges. IUCrJ. 2019;6(3):373-86.

41. Rose JP, Wang BC, Weiss MS. Native SAD is maturing. IUCrJ. 2015;2(Pt 4):431-40.

42. Baudouin-Cornu P, Schuerer K, Marliere P, Thomas D. Intimate evolution of proteins. Proteome atomic content correlates with genome base composition. J Biol Chem. 2004;279(7):5421-8.

43. Bent AF, Mann G, Houssen WE, Mykhaylyk V, Duman R, Thomas L, et al. Structure of the cyanobactin oxidase ThcOx from Cyanothece sp. PCC 7425, the first structure to be solved at Diamond Light Source beamline I23 by means of S-SAD. Acta Crystallogr D Struct Biol. 2016;72(Pt 11):1174-80.

44. Rothbauer U, Zolghadr K, Tillib S, Nowak D, Schermelleh L, Gahl A, et al. Targeting and tracing antigens in live cells with fluorescent nanobodies. Nat Methods. 2006;3(11):887-9.

45. Zolghadr K, Gregor J, Leonhardt H, Rothbauer U. Case study on live cell apoptosis-assay using lamin-chromobody cell-lines for high-content analysis. Methods in molecular biology (Clifton, NJ). 2012;911:569-75.

46. Klooster R, Eman MR, le Duc Q, Verheesen P, Verrips CT, Roovers RC, et al. Selection and characterization of KDEL-specific VHH antibody fragments and their

application in the study of ER resident protein expression. Journal of immunological methods. 2009;342(1-2):1-12.

47. Aubrey N, Devaux C, di Luccio E, Goyffon M, Rochat H, Billiald P. A recombinant scFv/streptavidin-binding peptide fusion protein for the quantitative determination of the scorpion venom neurotoxin Aahl. Biol Chem. 2001;382(11):1621-8.

48. Jeffrey PD, Bajorath J, Chang CY, Yelton D, Hellström I, Hellström KE, et al. The X-ray structure of an anti-tumour antibody in complex with antigen. Nature Structural Biology. 1995;2(6):466-71.

49. Krishnamurthy H, Gouaux E. X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. Nature. 2012;481:469.

50. García-Nafría J, Lee Y, Bai X, Carpenter B, Tate CG. Cryo-EM structure of the adenosine A2A receptor coupled to an engineered heterotrimeric G protein. eLife. 2018;7:e35946.

51. Abskharon RNN, Giachin G, Wohlkonig A, Soror SH, Pardon E, Legname G, et al. Probing the N-Terminal β-Sheet Conversion in the Crystal Structure of the Human Prion Protein Bound to a Nanobody. Journal of the American Chemical Society. 2014;136(3):937-44.

52. Frank GA, Shukla S, Rao P, Borgnia MJ, Bartesaghi A, Merk A, et al. Cryo-EM Analysis of the Conformational Landscape of Human P-glycoprotein (ABCB1) During its Catalytic Cycle. Molecular pharmacology. 2016;90(1):35-41.

53. Perez C, Köhler M, Janser D, Pardon E, Steyaert J, Zenobi R, et al. Structural basis of inhibition of lipid-linked oligosaccharide flippase PglK by a conformational nanobody. Scientific Reports. 2017;7:46641.

54. Perez C, Gerber S, Boilevin J, Bucher M, Darbre T, Aebi M, et al. Structure and mechanism of an active lipid-linked oligosaccharide flippase. Nature. 2015;524(7566):433-8.

55. Uchański T, Masiulis S, Fischer B, Kalichuk V, López-Sánchez U, Zarkadas E, et al. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. Nat Methods. 2021;18(1):60-8.

56. Laverty D, Desai R, Uchański T, Masiulis S, Stec WJ, Malinauskas T, et al. Cryo-EM structure of the human α1β3γ2 GABAA receptor in a lipid bilayer. Nature. 2019;565(7740):516-20.

57. Köhler M, Neff C, Perez C, Brunner C, Pardon E, Steyaert J, et al. Binding Specificities of Nanobody*Membrane Protein Complexes Obtained from Chemical Cross-Linking and High-Mass MALDI Mass Spectrometry. Analytical chemistry. 2018;90(8):5306-13.

58. Cadene M, Chait BT. A Robust, Detergent-Friendly Method for Mass Spectrometric Analysis of Integral Membrane Proteins. Analytical chemistry. 2000;72(22):5655-8.

59. Chen F, Gerber S, Heuser K, Korkhov VM, Lizak C, Mireku S, et al. High-mass matrix-assisted laser desorption ionization-mass spectrometry of integral membrane proteins and their complexes. Analytical chemistry. 2013;85(7):3483-8.

60. Watanabe M, Miyazono K, Tanokura M, Sawasaki T, Endo Y, Kobayashi I. Cell-free protein synthesis for structure determination by X-ray crystallography. Methods Mol Biol. 2010;607:149-60.

61. Schwarz D, Junge F, Durst F, Frolich N, Schneider B, Reckel S, et al. Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. Nature protocols. 2007;2(11):2945-57.

62. Alami M, Dalal K, Lelj-Garolla B, Sligar SG, Duong F. Nanodiscs unravel the interaction between the SecYEG channel and its cytosolic partner SecA. Embo j. 2007;26(8):1995-2004.

63. Katzen F, Peterson TC, Kudlicki W. Membrane protein expression: no cells required. Trends in biotechnology. 2009;27(8):455-60.

64. Chen Y-J, Pornillos O, Lieu S, Ma C, Chen AP, Chang G. X-ray structure of EmrE supports dual topology model. Proceedings of the National Academy of Sciences. 2007;104(48):18999.

65. Wada T, Shimono K, Kikukawa T, Hato M, Shinya N, Kim SY, et al. Crystal structure of the eukaryotic light-driven proton-pumping rhodopsin, Acetabularia rhodopsin II, from marine alga. J Mol Biol. 2011;411(5):986-98.

66. Boland C, Li D, Shah STA, Haberstock S, Dötsch V, Bernhard F, et al. Cell-free expression and in meso crystallisation of an integral membrane kinase for structure determination. Cellular and molecular life sciences : CMLS. 2014;71(24):4895-910.

67. Deisenhofer J, Epp O, Miki K, Huber R, Michel H. Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3Å resolution. Nature. 1985;318:618.

68. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. Science (New York, NY). 2000;289(5480):739-45.

69. Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature. 2008;454(7201):183-7.

70. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauß N, Choe H-W, et al. Crystal structure of opsin in its G-protein-interacting conformation. Nature. 2008;455:497.

71. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, et al. Structure of a nanobody-stabilized active state of the β2 adrenoceptor. Nature. 2011;469:175.

72. Miroux B, Walker JE. 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. 1996;260(3):289-98.

73. Woodward R, Yi W, Li L, Zhao G, Eguchi H, Sridhar PR, et al. In vitro bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. Nat Chem Biol. 2010;6(6):418-23.

74. Skrzypek R, Iqbal S, Callaghan R. Methods of reconstitution to investigate membrane protein function. Methods. 2018;147:126-41.

75. Klammt C, Schwarz D, Löhr F, Schneider B, Dötsch V, Bernhard F. Cell-free expression as an emerging technique for the large scale production of integral membrane protein. The FEBS Journal. 2006;273(18):4141-53.

76. Seddon AM, Curnow P, Booth PJ. Membrane proteins, lipids and detergents: not just a soap opera. Biochim Biophys Acta. 2004;1666(1-2):105-17.

77. Cudney R, Patel S, Weisgraber K, Newhouse Y, McPherson A. Screening and optimization strategies for macromolecular crystal growth. Acta Crystallogr D Biol Crystallogr. 1994;50(Pt 4):414-23.

78. Adir N. Crystallization of the oxygen-evolving reaction centre of photosystem II in nine different detergent mixtures. Acta Crystallogr D Biol Crystallogr. 1999;55(4):891-4.

79. Stetsenko A, Guskov A. An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization. Crystals. 2017;7(7):197.

80. Koehl A, Hu H, Maeda S, Zhang Y, Qu Q, Paggi JM, et al. Structure of the μ-opioid receptor–Gi protein complex. Nature. 2018;558(7711):547-52.

81. Sousa JS, Calisto F, Langer JD, Mills DJ, Refojo PN, Teixeira M, et al. Structural basis for energy transduction by respiratory alternative complex III. Nature Communications. 2018;9(1):1728.

82. Pan X, Li Z, Zhou Q, Shen H, Wu K, Huang X, et al. Structure of the human voltage-gated sodium channel $Na_v1.4$ in complex with β1. Science (New York, NY). 2018;362(6412).

83. Baradaran R, Wang C, Siliciano AF, Long SB. Cryo-EM structures of fungal and metazoan mitochondrial calcium uniporters. Nature. 2018;559(7715):580-4.

84. Arnaud C-A, Effantin G, Vivès C, Engilberge S, Bacia M, Boulanger P, et al. Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. Nature Communications. 2017;8(1):1953.

85. Huynh KW, Jiang J, Abuladze N, Tsirulnikov K, Kao L, Shao X, et al. CryoEM structure of the human SLC4A4 sodium-coupled acid-base transporter NBCe1. Nature Communications. 2018;9(1):900.

86. Catoire LJ, Zoonens M, van Heijenoort C, Giusti F, Guittet É, Popot J-L. Solution NMR mapping of water-accessible residues in the transmembrane β-barrel of OmpX. European Biophysics Journal. 2010;39(4):623-30.

87. Perlmutter JD, Popot JL, Sachs JN. Molecular dynamics simulations of a membrane protein/amphipol complex. J Membr Biol. 2014;247(9-10):883-95.

88. Picard M, Dahmane T, Garrigos M, Gauron C, Giusti F, le Maire M, et al. Protective and inhibitory effects of various types of amphipols on the Ca2+-ATPase from sarcoplasmic reticulum: a comparative study. Biochemistry. 2006;45(6):1861-9.

89. Pocanschi CL, Popot JL, Kleinschmidt JH. Folding and stability of outer membrane protein A (OmpA) from Escherichia coli in an amphipathic polymer, amphipol A8-35. European biophysics journal : EBJ. 2013;42(2-3):103-18.

90. Catoire LJ, Zoonens M, Heijenoort Cv, Giusti F, Popot J-L, Guittet E. Inter- and intramolecular contacts in a membrane protein/surfactant complex observed by heteronuclear dipole-to-dipole cross-relaxation. Journal of Magnetic Resonance. 2009;197(1):91-5.

91. Ethayathulla AS, Yousef MS, Amin A, Leblanc G, Kaback HR, Guan L. Structure-based mechanism for Na(+)/melibiose symport by MelB. Nat Commun. 2014;5:3009.

92. Dumas F, Tocanne J-F, Leblanc G, Lebrun M-C. Consequences of Hydrophobic Mismatch between Lipids and Melibiose Permease on Melibiose Transport. Biochemistry. 2000;39(16):4846-54.

93. Hariharan P, Tikhonova E, Medeiros-Silva J, Jeucken A, Bogdanov MV, Dowhan W, et al. Structural and functional characterization of protein–lipid interactions of the Salmonella typhimurium melibiose transporter MelB. BMC Biology. 2018;16(1):85.

94. Butterwick JA, del Mármol J, Kim KH, Kahlson MA, Rogow JA, Walz T, et al. Cryo-EM structure of the insect olfactory receptor Orco. Nature. 2018;560(7719):447-52.

95. Elezgarai SR, Fernández-Borges N, Eraña H, Sevillano AM, Charco JM, Harrathi C, et al. Generation of a new infectious recombinant prion: a model to understand Gerstmann–Sträussler–Scheinker syndrome. Scientific Reports. 2017;7(1):9584.

96. Singh AK, McGoldrick LL, Twomey EC, Sobolevsky AI. Mechanism of calmodulin inactivation of the calcium-selective TRP channel TRPV6. Science advances. 2018;4(8):eaau6088-eaau.

97. Rigaud J-L, Pitard B, Levy D. Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. Biochimica et Biophysica Acta (BBA) - Bioenergetics. 1995;1231(3):223-46.

98. Zhang M, Wang D, Kang Y, Wu J-X, Yao F, Pan C, et al. Structure of the mechanosensitive OSCA channels. Nature Structural & Molecular Biology. 2018;25(9):850-8.

99. Gong D, Chi X, Ren K, Huang G, Zhou G, Yan N, et al. Structure of the human plasma membrane Ca2+-ATPase 1 in complex with its obligatory subunit neuroplastin. Nature Communications. 2018;9(1):3623.

100. Elferink MG, de Wit JG, Demel R, Driessen AJ, Konings WN. Functional reconstitution of membrane proteins in monolayer liposomes from bipolar lipids of Sulfolobus acidocaldarius. J Biol Chem. 1992;267(2):1375-81.

101. Stahlberg H, Biyani N, Engel A. 3D reconstruction of two-dimensional crystals. Arch Biochem Biophys. 2015;581:68-77.

102. Wang L, Sigworth FJ. Liposomes on a Streptavidin Crystal: A System to Study Membrane Proteins by Cryo-EM. Cryo-EM Part A Sample Preparation and Data Collection. Methods in Enzymology2010. p. 147-64.

103. Vargas C, Arenas RC, Frotscher E, Keller S. Nanoparticle self-assembly in mixtures of phospholipids with styrene/maleic acid copolymers or fluorinated surfactants. Nanoscale. 2015;7(48):20685-96.

104. Smirnova IA, Ädelroth P, Brzezinski P. Extraction and liposome reconstitution of membrane proteins with their native lipids without the use of detergents. Scientific reports. 2018;8(1):14950-.

105. Sugai S, Nitta K, Ohno N. Studies on conformational transition of the maleic acid copolymer with styrene in aqueous salt solution by derivative spectroscopy. Polymer. 1982;23(2):238-42.

106. Birch J, Axford D, Foadi J, Meyer A, Eckhardt A, Thielmann Y, et al. The fine art of integral membrane protein crystallisation. Methods. 2018;147:150-62.

107. Caffrey M. A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta crystallographica Section F, Structural biology communications. 2015;71(Pt 1):3-18.

108. Axford D, Owen RL, Aishima J, Foadi J, Morgan AW, Robinson JI, et al. In situ macromolecular crystallography using microbeams. Acta Crystallogr D Biol Crystallogr. 2012;68(Pt 5):592-600.

109. Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, et al. Femtosecond X-ray protein nanocrystallography. Nature. 2011;470(7332):73-U81.

110. Nakane T, Hanashima S, Suzuki M, Saiki H, Hayashi T, Kakinouchi K, et al. Membrane protein structure determination by SAD, SIR, or SIRAS phasing in serial femtosecond crystallography using an iododetergent. Proc Natl Acad Sci U S A. 2016;113(46):13039-44.

111. Conrad CE, Basu S, James D, Wang D, Schaffer A, Roy-Chowdhury S, et al. A novel inert crystal delivery medium for serial femtosecond crystallography. IUCrJ. 2015;2(Pt 4):421-30.

112. Zhang H, Unal H, Gati C, Han GW, Liu W, Zatsepin NA, et al. Structure of the Angiotensin receptor revealed by serial femtosecond crystallography. Cell. 2015;161(4):833-44.

113. Martiel I, Olieric V, Caffrey M, Wang M. Chapter 1 Practical Approaches for In Situ X-ray Crystallography: from High-throughput Screening to Serial Data Collection.

Protein Crystallography: Challenges and Practical Solutions: The Royal Society of Chemistry; 2018. p. 1-27.

114. Nunes PS, Ohlsson PD, Ordeig O, Kutter JP. Cyclic olefin polymers: emerging materials for lab-on-a-chip applications. Microfluidics and Nanofluidics. 2010;9(2):145-61.

115. Sui S, Wang Y, Kolewe KW, Srajer V, Henning R, Schiffman JD, et al. Graphene-based microfluidics for serial crystallography. Lab on a Chip. 2016;16(16):3082-96.

116. Huang CY, Olieric V, Ma P, Panepucci E, Diederichs K, Wang M, et al. In meso in situ serial X-ray crystallography of soluble and membrane proteins. Acta Crystallogr D Biol Crystallogr. 2015;71(Pt 6):1238-56.

117. Huang CY, Olieric V, Ma P, Howe N, Vogeley L, Liu X, et al. In meso in situ serial X-ray crystallography of soluble and membrane proteins at cryogenic temperatures. Acta crystallographica Section D, Structural biology. 2016;72(Pt 1):93-112.

118. Islam ST, Lam JS. Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. Canadian Journal of Microbiology. 2014;60(11):697-716.

119. Han W, Wu B, Li L, Zhao G, Woodward R, Pettit N, et al. Defining Function of Lipopolysaccharide O-antigen Ligase WaaL Using Chemoenzymatically Synthesized Substrates*. Journal of Biological Chemistry. 2012;287(8):5357-65.

120. Woodward L, Naismith JH. Bacterial polysaccharide synthesis and export. Curr Opin Struct Biol. 2016;40:81-8.

121. Suits MDL, Sperandeo P, Dehò G, Polissi A, Jia Z. Novel Structure of the Conserved Gram-Negative Lipopolysaccharide Transport Protein A and Mutagenesis Analysis. Journal of Molecular Biology. 2008;380(3):476-88.

122. Sherman DJ, Lazarus MB, Murphy L, Liu C, Walker S, Ruiz N, et al. Decoupling catalytic activity from biological function of the ATPase that powers lipopolysaccharide transport. Proceedings of the National Academy of Sciences. 2014;111(13):4982.

123. Roberts IS. THE BIOCHEMISTRY AND GENETICS OF CAPSULAR POLYSACCHARIDE PRODUCTION IN BACTERIA. Annual Review of Microbiology. 1996;50(1):285-315.

124. Collins RF, Beis K, Clarke BR, Ford RC, Hulley M, Naismith JH, et al. 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. 2006;281(4):2144-50.

125. Yang Y, Liu J, Clarke BR, Seidel L, Bolla JR, Ward PN, et al. The molecular basis of regulation of bacterial capsule assembly by Wzc. Nature Communications. 2021;12(1):4349.

126. Lescop E, Hu Y, Xu H, Hu W, Chen J, Xia B, et al. The Solution Structure of Escherichia coli Wzb Reveals a Novel Substrate Recognition Mechanism of Prokaryotic Low Molecular Weight Protein-tyrosine Phosphatases*. Journal of Biological Chemistry. 2006;281(28):19570-7.

127. Hagelueken G, Huang H, Mainprize IL, Whitfield C, Naismith JH. Crystal structures of Wzb of Escherichia coli and CpsB of Streptococcus pneumoniae, representatives of two families of tyrosine phosphatases that regulate capsule assembly. Journal of molecular biology. 2009;392(3):678-88.

128. Dong C, Beis K, Nesper J, Brunkan-Lamontagne AL, Clarke BR, Whitfield C, et al. Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. Nature. 2006;444(7116):226-9.

129. Bushell Simon R, Mainprize Iain L, Wear Martin A, Lou H, Whitfield C, Naismith James H. Wzi Is an Outer Membrane Lectin that Underpins Group 1 Capsule Assembly in Escherichia coli. Structure. 2013;21(5):844-53.

130. Liu D, Cole RA, Reeves PR. An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. J Bacteriol. 1996;178(7):2102-7.

131. Rick PD, Barr K, Sankaran K, Kajimura J, Rush JS, Waechter CJ. Evidence That the wzxE Gene of Escherichia coli K-12 Encodes a Protein Involved in the Transbilayer Movement of a Trisaccharide-Lipid Intermediate in the Assembly of Enterobacterial Common Antigen *. Journal of Biological Chemistry. 2003;278(19):16534-42.

132. Marolda CL, Vicarioli J, Valvano MA. Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit. Microbiology (Reading). 2004;150(Pt 12):4095-105.

133. Park BS, Lee J-O. Recognition of lipopolysaccharide pattern by TLR4 complexes. Experimental & Molecular Medicine. 2013;45(12):e66-e.

134. Girgis HS, Hottes AK, Tavazoie S. Genetic architecture of intrinsic antibiotic susceptibility. PLoS One. 2009;4(5):e5629.

135. McMahon KJ, Castelli ME, García Vescovi E, Feldman MF. Biogenesis of outer membrane vesicles in Serratia marcescens is thermoregulated and can be induced by activation of the Rcs phosphorelay system. J Bacteriol. 2012;194(12):3241-9.

136. Liu B, Furevi A, Perepelov AV, Guo X, Cao H, Wang Q, et al. Structure and genetics of Escherichia coli O antigens. FEMS microbiology reviews. 2020;44(6):655-83.

137. Kajimura J, Rahman A, Rick PD. Assembly of cyclic enterobacterial common antigen in Escherichia coli K-12. J Bacteriol. 2005;187(20):6917-27.

138. Rai AK, Mitchell AM. Enterobacterial Common Antigen: Synthesis and Function of an Enigmatic Molecule. mBio. 2020;11(4).

139. Marolda CL, Tatar LD, Alaimo C, Aebi M, Valvano MA. 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. 2006;188(14):5124-35.

140. Alaimo C, Catrein I, Morf L, Marolda CL, Callewaert N, Valvano MA, et al. Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. EMBO J. 2006;25(5):967-76.

141. Islam ST, Fieldhouse RJ, Anderson EM, Taylor VL, Keates RAB, Ford RC, et al. A cationic lumen in the Wzx flippase mediates anionic O-antigen subunit translocation in Pseudomonas aeruginosa PAO1. Molecular Microbiology. 2012;84(6):1165-76.

142. Lu M, Symersky J, Radchenko M, Koide A, Guo Y, Nie R, et al. Structures of a Na+-coupled, substrate-bound MATE multidrug transporter. Proceedings of the National Academy of Sciences. 2013;110(6):2099.

143. Islam Salim T, Eckford Paul DW, Jones Michelle L, Nugent T, Bear Christine E, Vogel C, et al. Proton-Dependent Gating and Proton Uptake by Wzx Support O-Antigen-Subunit Antiport Across the Bacterial Inner Membrane. mBio. 2013;4(5):e00678-13.

144. Paulusma CC, Oude Elferink RPJ. Diseases of intramembranous lipid transport. FEBS Letters. 2006;580(23):5500-9.

145. Sprong H, van der Sluijs P, van Meer G. How proteins move lipids and lipids move proteins. Nature Reviews Molecular Cell Biology. 2001;2(7):504-13.

146. Sanyal S, Menon AK. Flipping Lipids: Why an' What's the Reason for? ACS Chemical Biology. 2009;4(11):895-909.

147. Lemieux MJ, Huang Y, Wang D-N. The structural basis of substrate translocation by the Escherichia coli glycerol-3-phosphate transporter: a member of the major facilitator superfamily. Current Opinion in Structural Biology. 2004;14(4):405-12.

148. Pomorski T, Menon AK. Lipid flippases and their biological functions. Cellular and Molecular Life Sciences CMLS. 2006;63(24):2908-21.

149. Zhou GP, Troy FA, 2nd. NMR studies on how the binding complex of polyisoprenol recognition sequence peptides and polyisoprenols can modulate membrane structure. Curr Protein Pept Sci. 2005;6(5):399-411.

150. Dawson RJP, Locher KP. Structure of a bacterial multidrug ABC transporter. Nature. 2006;443(7108):180-5.

151. Ward A, Reyes CL, Yu J, Roth CB, Chang G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proceedings of the National Academy of Sciences. 2007;104(48):19005.

152. Perez C, Gerber S, Boilevin J, Bucher M, Darbre T, Aebi M, et al. Structure and mechanism of an active lipid-linked oligosaccharide flippase. NATURE. 2015;524(7566):433-+.

153. Hong Y, Liu MA, Reeves PR. Progress in Our Understanding of Wzx Flippase for Translocation of Bacterial Membrane Lipid-Linked Oligosaccharide. J Bacteriol. 2018;200(1).

154. Rubino FA, Kumar S, Ruiz N, Walker S, Kahne DE. Membrane Potential Is Required for MurJ Function. Journal of the American Chemical Society. 2018;140(13):4481-4.

155. Kuk AC, Mashalidis EH, Lee SY. Crystal structure of the MOP flippase MurJ in an inward-facing conformation. Nat Struct Mol Biol. 2017;24(2):171-6.

156. Kuk ACY, Hao A, Guan Z, Lee S-Y. Visualizing conformation transitions of the Lipid II flippase MurJ. Nature Communications. 2019;10(1):1736.

157. Budisa N, Steipe B, Demange P, Eckerskorn C, Kellermann J, Huber R. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. Eur J Biochem. 1995;230(2):788-96.

158. Liu H, Naismith JH. A simple and efficient expression and purification system using two newly constructed vectors. Protein Expr Purif. 2009;63(2):102-11.

159. Pardon E, Laeremans T, Triest S, Rasmussen SG, Wohlkonig A, Ruf A, et al. A general protocol for the generation of Nanobodies for structural biology. Nat Protoc. 2014;9(3):674-93.

160. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnology. 2008;8(1):91.

161. Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic acids research. 2007;35(Web Server issue):W43-W6.

162. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784-8.

163. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology. 2011;7(1):539.

164. Kabsch W. XDS. Acta Crystallographica Section D. 2010;66(2):125-32.

165. Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallographica Section D. 2010;66(2):133-44.

166. Thorn A. Experimental Phasing: Substructure Solution and Density Modification as Implemented in SHELX. Methods Mol Biol. 2017;1607:357-76.

167. Potterton L, Agirre J, Ballard C, Cowtan K, Dodson E, Evans PR, et al. CCP4i2: the new graphical user interface to the CCP4 program suite. Acta Crystallographica Section D. 2018;74(2):68-84.

168. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta crystallographica Section D, Biological crystallography. 2011;67(Pt 4):282-92.

169. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallographica Section D. 2013;69(7):1204-14.

170. Ness SR, de Graaff RA, Abrahams JP, Pannu NS. CRANK: new methods for automated macromolecular crystal structure solution. Structure. 2004;12(10):1753-61.

171. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallographica Section D. 2006;62(9):1002-11.

172. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D. 2011;67(4):355-67.

173. Painter J, Merritt EA. TLSMD web server for the generation of multi-group TLS models. Journal of Applied Crystallography. 2006;39(1):109-11.

174. Joosten RP, Salzemann J, Bloch V, Stockinger H, Berglund A-C, Blanchet C, et al. PDB_REDO: automated re-refinement of X-ray structure models in the PDB. Journal of Applied Crystallography. 2009;42(3):376-84.

175. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 1):12-21.

176. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids research. 2019;47(W1):W636-W41.

177. Islam ST, Taylor VL, Qi M, Lam JS. Membrane topology mapping of the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from Pseudomonas aeruginosa PAO1 reveals novel domain architectures. mBio. 2010;1(3).

178. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research. 2014;42(W1):W320-W4.

179. McNicholas S, Potterton E, Wilson KS, Noble MEM. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallographica Section D. 2011;67(4):386-94.

180. Caffrey M. A lipid's eye view of membrane protein crystallization in mesophases. Curr Opin Struct Biol. 2000;10(4):486-97.

181. Agnarelli A, El Omari K, Duman R, Wagner A, Mancini EJ. Phosphorus and sulfur SAD phasing of the nucleic acid-bound DNA-binding domain of interferon regulatory factor 4. Acta Crystallographica Section F. 2021;77(7):202-7.

182. Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, et al. A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. Genome Medicine. 2017;9(1):103.

183. Bell A, Brunt J, Crost E, Vaux L, Nepravishta R, Owen CD, et al. Elucidation of a sialic acid metabolism pathway in mucus-foraging Ruminococcus gnavus unravels mechanisms of bacterial adaptation to the gut. Nat Microbiol. 2019;4(12):2393-404.

184. Bell A, Severi E, Lee MO, Monaco S, Latousakis D, Angulo J, et al. Uncovering a novel molecular mechanism for scavenging sialic acids in bacteria. Journal of Biological Chemistry. 2020.

185. Monestier M, Latousakis D, Bell A, Tribolo S, Tailford LE, Colquhoun IJ, et al. Membrane-enclosed multienzyme (MEME) synthesis of 2,7-anhydro-sialic acid derivatives. Carbohydrate Research. 2017;451:110-7.

186. Kraal L, Abubucker S, Kota K, Fischbach MA, Mitreva M. The prevalence of species and strains in the human microbiome: a resource for experimental efforts. PLoS One. 2014;9(5):e97279.

187. Heintz-Buschart A, Wilmes P. Human Gut Microbiome: Function Matters. Trends in Microbiology. 2018;26(7):563-74.

188. Chen MX, Wang S-Y, Kuo C-H, Tsai IL. Metabolome analysis for investigating host-gut microbiota interactions. Journal of the Formosan Medical Association. 2019;118:S10-S22.

189. Varki A, Cummings RD, Esko JD. Sialic Acids and Other Nonulosonic Acids. Essentials of Glycobiology. 3 ed. NY: Cold Spring Harbor; 2017.

190. Juge N, Tailford L, Owen C D. Sialidases from gut bacteria: a mini-review. Biochemical Society Transactions. 2016;44(1):166-75.

191. Marcobal A, Southwick AM, Earle KA, Sonnenburg JL. A refined palate: bacterial consumption of host glycans in the gut. Glycobiology. 2013;23(9):1038-46.

192. Tailford LE, Crost EH, Kavanaugh D, Juge N. Mucin glycan foraging in the human gut microbiome. Frontiers in Genetics. 2015;6(81).

193. Plumbridge J, Vimr E. Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by Escherichia coli. J Bacteriol. 1999;181(1):47-54.

194. Brigham C, Caughlan R, Gallegos R, Dallas MB, Godoy VG, Malamy MH. Sialic acid (N-acetyl neuraminic acid) utilization by Bacteroides fragilis requires a novel N-acetyl mannosamine epimerase. J Bacteriol. 2009;191(11):3629-38.

195. Crost EH, Tailford LE, Le Gall G, Fons M, Henrissat B, Juge N. Utilisation of mucin glycans by the human gut symbiont Ruminococcus gnavus is strain-dependent. PLoS One. 2013;8(10):e76341.

196. Tailford LE, Owen CD, Walshaw J, Crost EH, Hardy-Goddard J, Le Gall G, et al. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. Nat Commun. 2015;6:7624.

197. Koshland Jr DE. STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. Biological Reviews. 1953;28(4):416-36.

198. Peat S. The Chemistry of Anhydro Sugars. In: Pigman WW, Wolfrom ML, Peat S, editors. Advances in Carbohydrate Chemistry. 2: Academic Press; 1946. p. 37-77.

199. Vocadlo DJ, Davies GJ, Laine R, Withers SG. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. Nature. 2001;412(6849):835-8.

200. Blake CCF, Koenig DF, Mair GA, North ACT, Phillips DC, Sarma VR. STRUCTURE OF HEN EGG-WHITE LYSOZYME - A 3-DIMENSIONAL FOURIER SYNTHESIS AT 2A RESOLUTION. Nature. 1965;206(4986):757-&.

201. Phillips DC. THE HEN EGG-WHITE LYSOZYME MOLECULE. Proceedings of the National Academy of Sciences of the United States of America. 1967;57(3):483-95.

202. Jongkees SAK, Withers SG. Unusual Enzymatic Glycoside Cleavage Mechanisms. Accounts of Chemical Research. 2014;47(1):226-35.

203. Yip VL, Varrot A, Davies GJ, Rajan SS, Yang X, Thompson J, et al. An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from Thermotoga maritima. J Am Chem Soc. 2004;126(27):8354-5.

204. Liu QP, Sulzenbacher G, Yuan H, Bennett EP, Pietz G, Saunders K, et al. Bacterial glycosidases for the production of universal red blood cells. Nat Biotechnol. 2007;25(4):454-64.

205. Rajan SS, Yang X, Collart F, Yip VLY, Withers SG, Varrot A, et al. Novel Catalytic Mechanism of Glycoside Hydrolysis Based on the Structure of an NAD+/Mn2+-Dependent Phospho-α-Glucosidase from Bacillus subtilis. Structure. 2004;12(9):1619-29.

206. Liu QP, Sulzenbacher G, Yuan H, Bennett EP, Pietz G, Saunders K, et al. Bacterial glycosidases for the production of universal red blood cells. Nature Biotechnology. 2007;25(4):454-64.

207. Xu G, Kiefel MJ, Wilson JC, Andrew PW, Oggioni MR, Taylor GL. Three Streptococcus pneumoniae Sialidases: Three Different Products. Journal of the American Chemical Society. 2011;133(6):1718-21.

208. Xu G, Li X, Andrew PW, Taylor GL. Structure of the catalytic domain of Streptococcus pneumoniae sialidase NanA. Acta Crystallographica Section F. 2008;64(9):772-5.

209. Xu G, Potter JA, Russell RJM, Oggioni MR, Andrew PW, Taylor GL. Crystal Structure of the NanB Sialidase from Streptococcus pneumoniae. Journal of Molecular Biology. 2008;384(2):436-49.

210. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Research. 2018;47(D1):D427-D32.

211. Kingston RL, Scopes RK, Baker EN. The structure of glucose-fructose oxidoreductase from Zymomonas mobilis: an osmoprotective periplasmic enzyme containing non-dissociable NADP. Structure. 1996;4(12):1413-28.

212. Daniellou R, Zheng H, Palmer D. Kinetics of the reaction catalyzed by inositol dehydrogenase from Bacillus subtilis and inhibition by fluorinated substrate analogs. Canadian Journal of Chemistry. 2006;84:522.

213. Kikuchi A, Park SY, Miyatake H, Sun D, Sato M, Yoshida T, et al. Crystal structure of rat biliverdin reductase. Nat Struct Biol. 2001;8(3):221-5.

214. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-402.

215. Martinez J, Steenbergen S, Vimr E. Derived structure of the putative sialic acid transporter from Escherichia coli predicts a novel sugar permease domain. J Bacteriol. 1995;177(20):6005-10.

216. Horne CR, Kind L, Davies JS, Dobson RCJ. On the structure and function of Escherichia coli YjhC: An oxidoreductase involved in bacterial sialic acid metabolism. Proteins. 2020;88(5):654-68.

217. Garman EF, Owen RL. Cryocooling and radiation damage in macromolecular crystallography. Acta Crystallographica Section D. 2006;62(1):32-47.

218. Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, et al. Data processing and analysis with the autoPROC toolbox. Acta Crystallographica Section D. 2011;67(4):293-302.

219. Keegan RM, Winn MD. MrBUMP: an automated pipeline for molecular replacement. Acta Crystallogr D Biol Crystallogr. 2008;64(Pt 1):119-24.

220. Bunkóczi G, Read RJ. Improvement of molecular-replacement models with Sculptor. Acta crystallographica Section D, Biological crystallography. 2011;67(Pt 4):303-12.

221. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of Applied Crystallography. 2007;40(4):658-74.

222. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung L-W, et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallographica Section D. 2008;64(1):61-9.

223. Long F, Nicholls RA, Emsley P, Grazulis S, Merkys A, Vaitkus A, et al. AceDRG: a stereochemical description generator for ligands. Acta Crystallographica Section D. 2017;73(2):112-22.

224. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486-501.

225. Yang H, Peisach E, Westbrook JD, Young J, Berman HM, Burley SK. DCC: a Swiss army knife for structure factor analysis and validation. Journal of Applied Crystallography. 2016;49(3):1081-4.

226. Taberman H, Parkkinen T, Rouvinen J. Structural and functional features of the NAD(P) dependent Gfo/Idh/MocA protein family oxidoreductases. Protein Sci. 2016;25(4):778-86.

227. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372(3):774-97.

228. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallographica Section D. 2004;60(12):2256-68.

229. Russell RJ, Haire LF, Stevens DJ, Collins PJ, Lin YP, Blackburn GM, et al. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. Nature. 2006;443(7107):45-9.

230. Reusch D, Haberger M, Falck D, Peter B, Maier B, Gassner J, et al. Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 2: Mass spectrometric methods. mAbs. 2015;7(4):732-42.

231. Yin H, Racha J, Li SY, Olejnik N, Satoh H, Moore D. Automated high throughput human CYP isoform activity assay using SPE-LC/MS method: application in CYP inhibition evaluation. Xenobiotica. 2000;30(2):141-54.

232. Johnson KA, Goody RS. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. Biochemistry. 2011;50(39):8264-9.

233. Kentache T, Thabault L, Deumer G, Haufroid V, Frédérick R, Linster CL, et al. The metalloprotein YhcH is an anomerase providing N-acetylneuraminate aldolase with the open form of its substrate. Journal of Biological Chemistry. 2021;296.

234. Chan J, Sandhu G, Bennet AJ. A mechanistic study of sialic acid mutarotation: Implications for mutarotase enzymes. Organic & Biomolecular Chemistry. 2011;9(13):4818-22.

235. Gorrec F. Protein crystallization screens developed at the MRC Laboratory of Molecular Biology. Drug Discov Today. 2016;21(5):819-25.

236. Fazio VJ, Peat TS, Newman J. A drunken search in crystallization space. Acta Crystallogr F Struct Biol Commun. 2014;70(Pt 10):1303-11.

237. Gorrec F. The MORPHEUS protein crystallization screen. Journal of Applied Crystallography. 2009;42(6):1035-42.

238. McPherson A. A comparison of salts for the crystallization of macromolecules. Protein science : a publication of the Protein Society. 2001;10(2):418-22.

239. Casimir HBG. On Onsager's Principle of Microscopic Reversibility. Reviews of Modern Physics. 1945;17(2-3):343-50.

240. Owen CD, Tailford LE, Monaco S, Suligoj T, Vaux L, Lallement R, et al. Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont Ruminococcus gnavus. Nat Commun. 2017;8(1):2196.

241. Bahnson BJ, Anderson VE. Crotonase-catalyzed β-elimination is concerted: a double isotope effect study. Biochemistry. 1991;30(24):5894-906.

242. Morgan PM, Sala RF, Tanner ME. Eliminations in the Reactions Catalyzed by UDP-N-Acetylglucosamine 2-Epimerase. Journal of the American Chemical Society. 1997;119(43):10269-77.

243. Charnock SJ, Brown IE, Turkenburg JP, Black GW, Davies GJ. Convergent evolution sheds light on the anti-β-elimination mechanism common to family 1 and 10 polysaccharide lyases. Proceedings of the National Academy of Sciences. 2002;99(19):12067.

244. Lee SS, Yu S, Withers SG. α-1,4-Glucan Lyase Performs a Trans-Elimination via a Nucleophilic Displacement Followed by a Syn-Elimination. Journal of the American Chemical Society. 2002;124(18):4948-9.

245. Kiraly P, Nilsson M, Morris GA. Semi-real-time acquisition for fast pure shift NMR at maximum resolution. Journal of Magnetic Resonance. 2018;293:19-27.

Appendix

A. Crystal screen compositions

Wzx Optimisation Screen	
Reagent	**Concentration**
Sodium citrate (pH 6.4-6.5)	100 mM
Lithium sulfate	180 - 190 mM
PEG 400	28 - 32 %
Zinc chloride	20 - 30 mM

RgNanOx Citrate Screen	
Reagent	**Concentration**
Sodium citrate (pH 5.3 - 5.7)	100 mM
PEG 3000	17 - 23 %

RgNanOx MES Screen	
Reagent	**Concentration**
MES (pH 6.0)	100 mM
Sodium chloride	180 - 220 mM
Sodium fluoride	100 - 140 mM
PEG 6000	18 - 22 %

RgNanOx Substrate Complex Screen	
Reagent	**Concentration**
BIS-TRIS (pH 5.0 - 6.0)	10 mM
Magnesium chloride*	150 - 250 mM
Sodium thiocynate*	150 - 250 mM

Magnesium formate*	150 - 250 mM
PEG3350	17.5 - 27.5%

*Only one additive is present in each condition

B. XSCALE data table

Resolution Limit	Number of Reflections			COMPLETENESS	R-FACTOR		COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomalous Correlation	Anomalous Signal	Nano
	Observed	Unique	Possible		Observed	Expected							
11.14	31575	454	454	100.00%	5.30%	5.80%	31575	106.72	5.30%	99.8*	96*	5.09	178
7.88	55147	807	807	100.00%	5.60%	6.30%	55147	90.86	5.60%	100.0*	88*	3.151	354
6.43	71366	1038	1038	100.00%	8.70%	8.50%	71366	63.09	8.80%	100.0*	78*	2.719	473
5.57	86321	1277	1277	100.00%	12.90%	12.10%	86321	44.95	13.00%	99.9*	62*	2.033	588
4.98	95551	1438	1438	100.00%	12.70%	11.60%	95551	46.35	12.80%	99.9*	65*	1.636	671
4.55	99974	1544	1544	100.00%	11.90%	11.00%	99974	47.09	12.00%	100.0*	39*	1.579	724
4.21	112699	1743	1743	100.00%	15.30%	14.00%	112699	40.81	15.40%	99.9*	29*	1.397	821
3.94	110217	1832	1834	99.90%	20.10%	18.90%	110217	31.32	20.20%	99.8*	13*	1.184	867
3.71	115069	1969	1967	100.10%	28.80%	27.90%	115069	22.29	29.10%	99.7*	20*	1.156	937
3.52	115327	2016	2016	100.00%	38.90%	37.80%	115327	17.02	39.20%	99.3*	13*	1.11	959
3.36	117041	2227	2227	100.00%	57.00%	57.00%	117041	10.77	57.50%	98.5*	12*	1.1	1068
3.22	114379	2219	2219	100.00%	78.40%	80.70%	114379	7.59	79.20%	97.5*	11*	1.001	1059
3.09	116287	2398	2400	99.90%	98.90%	103.70%	116287	5.76	100.00%	96.1*	25*	1.047	1149
2.98	111920	2376	2376	100.00%	121.60%	130.70%	111920	4.47	122.90%	93.8*	13*	0.947	1140
2.88	107100	2659	2659	100.00%	142.60%	154.80%	107100	3.36	144.40%	90.4*	14*	0.877	1281
2.79	100886	2619	2623	99.80%	173.80%	192.60%	100886	2.59	176.10%	85.9*	11*	0.812	1258
2.7	100939	2722	2722	100.00%	207.30%	231.50%	100939	2.06	210.10%	82.0*	18*	0.805	1313
2.63	102436	2836	2840	99.90%	258.00%	295.40%	102436	1.5	261.70%	73.0*	7	0.723	1367
2.56	96977	2810	2808	100.10%	301.00%	348.80%	96977	1.2	305.60%	63.4*	8	0.683	1357
2.49	97133	3011	3033	99.30%	400.40%	467.60%	97116	0.7	407.10%	18.2*	14*	0.653	1451
Total	1958344	39995	40025	99.90%	22.00%	22.40%	1958327	17.18	22.20%	99.8*	25*	1.138	19015

C. Crystallographic tables

Wzx (I23 Data)			
PDB Code	N/A		
Space Group	C121		
Cell Dimensions	a = 271.5 Å, b = 38.5 Å, c = 56.960 Å; α = γ = 90°, β = 93°		
Resolution (highest shell)	45.19 - 2.80 (2.95 - 2.80) Å		
I/σI	32.1 (3.2)		
R-merge	0.153 (0.938)		
Wilson B	55.3 $Å^2$		
Completeness	99.9		
R-factor/R-free	0.29/0.36		
RMS Deviations Bonds Angles	 0.008 Å 1.63°		
	Overall	Inner	Outer
Number unique reflections	14932	525	2147
Half-set correlation CC(1/2)	0.999	0.999	0.908
Completeness %	100	99.3	99.9

Wzx (I24 Data)	
PDB Code	N/A
Space Group	C121
Cell Dimensions	a = 271.5 Å, b = 38.5 Å, c = 56.960 Å; α = γ = 90°, β = 93°
Resolution (highest shell)	135.02 - 2.31 (2.37 - 2.31) Å

I/σI	22.1 (7.1)		
R-merge	0.182 (1.731)		
Wilson B	44.7 Å^2 (Wzx), 65.6 Å^2 (nb10)		
Completeness	98.7 (98.2)		
R-factor/R-free	0.22/0.26		
RMS Deviations Bonds Angles	0.004 Å 1.22°		
	Overall	Inner	Outer
Number unique reflections	25893	510	2532
Half-set correlation CC(1/2)	1.0	1.0	0.4
Completeness %	99	99.1	98.4

RgNanOx Holo-Structure (Low resolution)			
PDB Code	6Z3B		
Space Group	$I4_122$		
Cell Dimensions	a = b = 150.1 Å;, c = 193.3 Å; α = β = γ = 90°		
Resolution (highest shell)	119 - 2.58 (2.7 - 2.58) Å		
I/σI	7.5 (1.3)		
R-merge	0.180 (1.717)		
Wilson B	61 Å^2		
Completeness	99.4 (99.5)		
R-factor/R-free	0.213/0.250		
RMS Deviations Bonds Angles	0.007 Å 1.53°		
	Overall	Inner	Outer
Number unique reflections	34401	1871	1651
Half-set correlation CC(1/2)	1.0	1.0	0.8
Completeness %	98.6	98.2	97.1

RgNanOx Holo-Structure (High Resolution)	
PDB Code	6Z3C
Space Group	$P2_12_12_1$
Cell Dimensions	a = 48.11 Å, b = 54.97 Å, c= 282.43 Å; α = β = γ = 90°
Resolution (highest shell)	47.47-1.74 (1.77-1.74) Å
I/σI	5.9 (1.3)
R-merge	0.158 (0.362)
Wilson B	17.45 Å^2
Completeness	99.9 (99.6)
R-factor/R-free	0.18/0.22
RMS Deviations Bonds Angles	0.010 Å 1.52°

	Overall	Inner	Outer
Number unique reflections	78162	658	4245
Half-set correlation CC(1/2)	1.0	1.0	0.9
Completeness %	92.0	100.0	84.7

RgNanOx/Neu Complex Structure			
PDB Code	N/A		
Space Group	$I4_122$		
Cell Dimensions	a = b = 150.0 Å;, c = 194.6 Å; α = β = γ = 90°		
Resolution (highest shell)	106.28-2.30 (2.36-2.30) Å		
I/σI	41.3 (13.0)		
R-merge	0.22 (8.382)		
Wilson B	42.56 $Å^2$		
Completeness	99.0 (89.1)		
R-factor/R-free	0.22/0.25		
RMS Deviations Bonds Angles	 0.008 Å 1.57°		
	Overall	Inner	Outer
Number unique reflections	79180	693	4476
Half-set correlation CC(1/2)	1.0	1.0	0.4
Completeness %	100.0	99.9	100.0

RgNanOx/Anh Complex Structure			
PDB Code	N/A		
Space Group	$I4_122$		
Cell Dimensions	a = b = 150.0 Å;, c = 194.7 Å; α = β = γ = 90°		
Resolution (highest shell)	118.81 - 1.96 (2.01 - 1.96) Å		
I/σI	20.2 (0.4)		
R-merge	0.10 (9.55)		
Wilson B	32.50 $Å^2$		
Completeness	99.9 (98.8)		
R-factor/R-free	0.20/0.22		
RMS Deviations Bonds Angles	 0.007 Å 1.39°		
	Overall	Inner	Outer
Number unique reflections	83974	712	2883
Half-set correlation CC(1/2)	1.0	1.0	0.4
Completeness %	99.9	99.9	100.0