The University Of Sheffield.

Engineering improved specificity and activity into oral bacterial sialidases for glycan biotechnology applications

Marianne Satur

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

The University of Sheffield

Faculty of Medicine, Dentistry and Health

School of Clinical Dentistry

July 2019

Abstract

A wide range of pathogenic and commensal bacteria secrete sialidases which function in the hydrolysis of terminal sialic acid residues from glycoconjugates and can have roles in nutrition or virulence. Sialic acid is connected to the underlying sugar in the glycan chain via α2,3-, α2,6- or α2,8-glycosidic linkages. The literature has reported bacterial sialidases that can hydrolyse all of these linkages and sialidases that are specific to α2,3-linked sialic acid. However, there are currently no known α2,6 specific sialidases. Glycosylation of biopharmaceuticals can affect their efficacy and so must be managed. This often requires the use of recombinant sialidases and there is an increasing demand for linkage specific sialidases in the glycan biotechnology industry.

In this study, biochemical assays and X-ray crystallography were used to understand the biochemical and structural properties of the NanH sialidase from *Tannerella forsythia* before site-directed mutagenesis was employed to generate point mutants in attempt to alter the specificity of NanH.

Biochemical analysis showed that wild-type (WT) NanH is a dual specific sialidase as it hydrolyses α2,3- and α2,6-linked sialic acid. A crystal structure was solved for the apo form of NanH, along with a structure for NanH in complex with the sialidase inhibitor oseltamivir. Alongside this, the crystal structure for an inactive NanH variant (NanH-D237A) was solved in complex with sialidase substrates 3- and 6-sialyllactose. These models enabled rational selection of residues for point mutation in attempt to alter the specificity. Overall two regions in the active site of NanH have been highlighted as potentially having importance in the specificity of the enzyme, which could be probed further in the future.

Acknowledgements

I would like to thank Dr Graham Stafford and Dr John Rafferty for giving me the opportunity to undertake this PhD and for providing expert laboratory and written help when necessary. I would also like to thank Dr Daniel Spencer at Ludger Ltd for his supervision and giving me the opportunity to visit the company to analyse the enzyme discussed in this thesis. However, for this visit to be possible, I must also thank Paulina Urbanowicz who designed the experiments and patiently assisted me in the lab. Paulina also performed some further glycan analysis using my enzyme, which I am very grateful for.

Special thanks go to Dr Andrew Frey for providing me with sialidase knowledge and teaching me 'everything he knows' in my initial few months as a PhD student. I would also like to thank Brenka McCabe and Jason Heath for being excellent lab technicians and for always being welcoming and friendly. I must also thank the great PhD community in the Dental School for being a pleasure to work amongst, especially my friends Amy Harding and Hannah Armes who I have enjoyed many laughs with.

I am especially grateful to my mum and dad who have been very supportive throughout my studies and also our family dog, Lucky, as without her interesting character to cheer me up, PhD life would have been a lot harder! I am forever grateful my boyfriend, Dr Will Wood, for being patient enough to help analyse my kinetics data, constantly answer several of my scientific questions and for proofreading and discussing this thesis with me.

Finally, I would like to acknowledge the financial support from the BBSRC-CASE award that made this project possible.

Publications in support of this thesis

Frey, A. M., Satur, M. J., Phansopa, C., Parker, J. L., Bradshaw, D., Pratten, J. and Stafford, G. P. (2018) 'Evidence for a carbohydrate-binding module (CBM) of Tannerella forsythia NanH sialidase, key to interactions at the host–pathogen interface.' *Biochemical Journal,* 475(6), pp. 1159–1176.

Contents

List of figures

List of tables

Abbreviations

2AB	2-aminobenzamide
3D	3-dimensional
3-SL	3’-sialyllactose
3-SLN	3’-sialyllactosamine
4-MU	4-methylumbelliferyl
6-SL	6’-sialyllactose
6-SLN	6’-sialyllactosamine
Asp-box	Aspartate box
BCA assay	bicinchoninic acid assay
CAZy	carbohydrate-active enzymes
CBM	carbohydrate-binding motif
CHO	Chinese hamster ovary
DANA	N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid
DNA	deoxyribonucleic acid
EPO	erythropoietin
ESI	electrospray ionisation
FT	freeze thaw
GalNAc	N-acteylgalactosamine
GlcNAc	N-acetylglucosamine
GH	glycoside hydrolase
HA	hemagglutinin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Ig	immunoglobulin
IPTG	isopropyl-β-D-1-thiogalactopyranoside
IT-sialidase	intramolecular trans-sialidase
LB	Luria-Bertani
LPS	lipopolysaccharide
mABs	monoclonal antibodies

MODY	maturity onset diabetes of the young
MS	mass spectrometry
(MS/MS)	tandem mass spectrometry
MU-NANA	2’-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid
MW	molecular weight
m/z	mass-to-charge
NA	neuraminidase
NEU	neuraminidase
Neu5Ac	N-acetylneuraminic acid
Neu5Gc	N-glycolylneuraminic acid
O/N	overnight
PAGE	polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PEG	polyethylene glycol
rER	rough endoplasmic reticulum
rHuEPO	recombinant human EPO
RMS	root mean square
RMSD	root-mean-square deviation
RT	room temperature
SDM	site-directed mutagenesis
SDS	sodium dodecyl sulphate
SEC	size-exclusion chromatography
SEC-MALLS	size-exclusion chromatography-multiangle laser light scattering
SLe^A	3-Sialyl-Lewis A
SLe^X	3-Sialyl-Lewis X
TBA assay	thiobarbituric acid assay
TLR	toll-like receptor
(U)HPLC	ultrahigh-performance liquid chromatography
UV	ultraviolet
VNTR	variable number of tandem repeats
WT	wild-type

Chapter 1

Introduction

1.1 Glycosylation

Glycosylation is the most abundant and diverse protein modification in nature and it involves the addition of a carbohydrate to a protein, lipid or organic molecule (Bertozzi and Rabuka, 2009). Alone, the carbohydrate portion is referred to as a glycan, but when attached to a protein or lipid the complex is referred to as a glycoprotein or glycolipid respectively (Bertozzi and Rabuka, 2009). Glycan diversity arises because several features of glycosylation can be modified such as the glycosidic linkages between sugar residues, the composition of sugar residues in the glycan and the structure and length of the glycan (Bertozzi and Rabuka, 2009). Glycosylation is crucial for a variety of biological functions such as protein folding, mediating cell interactions with the extracellular matrix and by providing stability to proteins that may be exposed to the extracellular environment. The size of a glycan can alter protein behaviour and larger glycans are considered to protect proteins from proteolysis, hide antigenic sites and increase protein solubility (Wormald *et al*., 2002).

1.1.1 Eukaryotic glycosylation

Glycans are composed of several sugar residues that link together to form an oligosaccharide and it is common for multiple oligosaccharides to be covalently linked to proteins or lipids in plasma cell membranes or proteins that are free in secretions such as saliva (Colley *et al*., 2017; Cross and Ruhl, 2018). Eukaryotic glycosylation of proteins occurs in the rough endoplasmic reticulum (rER) and Golgi body before the protein reaches its destination of the cell membrane or extracellular environment. The oligosaccharide attaches to the protein via the amide nitrogen of an asparagine (Asn), to produce an N-glycan (section 1.1.2) (Stanley *et al*., 2009), or to the hydroxyl group of a serine (Ser) or threonine (Thr) residue to produce an O-glycan (section 1.1.3) (Brockhausen *et al*., 2009). C-mannosylation is a less common type of glycosylation, differing to N and O-glycosylation because carbon-carbon links are formed instead of carbon-nitrogen or carbon-oxygen links (Furmanek and Hofsteenge, 2000). This carbon-carbon link is formed between an α-mannopyranosyl residue to the indole C2 carbon atom of a tryptophan on the polypeptide chain and was first discovered in RNase 2 from human urine (Hofsteenge *et al*., 1994).

1.1.2 N-glycosylation in eukaryotes

N-glycosylation refers to the linkage of N-acetylglucosamine (GlcNAc) to the amide nitrogen (hence N-glycan) of an Asn residue in the protein chain and was first recognised in hen egg albumin (Johansen *et al*., 1961). The GlcNAc-β-Asn link has later been discovered in many eukaryotic proteins, together with O-glycans (section 1.1.3). There are three classes of N-glycans; oligomannose, complex and hybrid oligosaccharides. These classes share a Manα1-3(Manα1-6) Manβ1-4GlcNAcβ1-4 GlcNAcβ1-Asn core, as all N-glycans are originally composed of the same 14 residue core, $Glc_3Man_9GlcNAc_2Asn$ (Glucose, mannose and GlcNAc respectively), before they are processed into one of the three classes (Hubbard and Robbins, 1980).

Oligomannose oligosaccharides, initially discovered in calf thyroglobulin, only consist of mannose residues after the core (Figure 1.1a) (Ito *et al*., 1977). Complex oligosaccharides consist of a variety of different sugars, apart from mannose, which tends to be absent apart from within the core. Galactose, fucose and nearly always a terminal sialic acid are present in complex oligosaccharides (Figure 1.1b). Hybrid oligosaccharide only consist of mannose residues on the Manα1-6 arm, whilst the Manα1,3 arm develops into a complex chain (Figure 1.1c). The GlcNAc residues added to the highest point on the core mannose residues are referred to as antennae and when there are 2-3 GlcNac residues attached to these core mannoses, they are called bi- and tri-antennary oligosaccharides (Kornfeld and Kornfeld, 1985; Stanley *et al*., 2009).

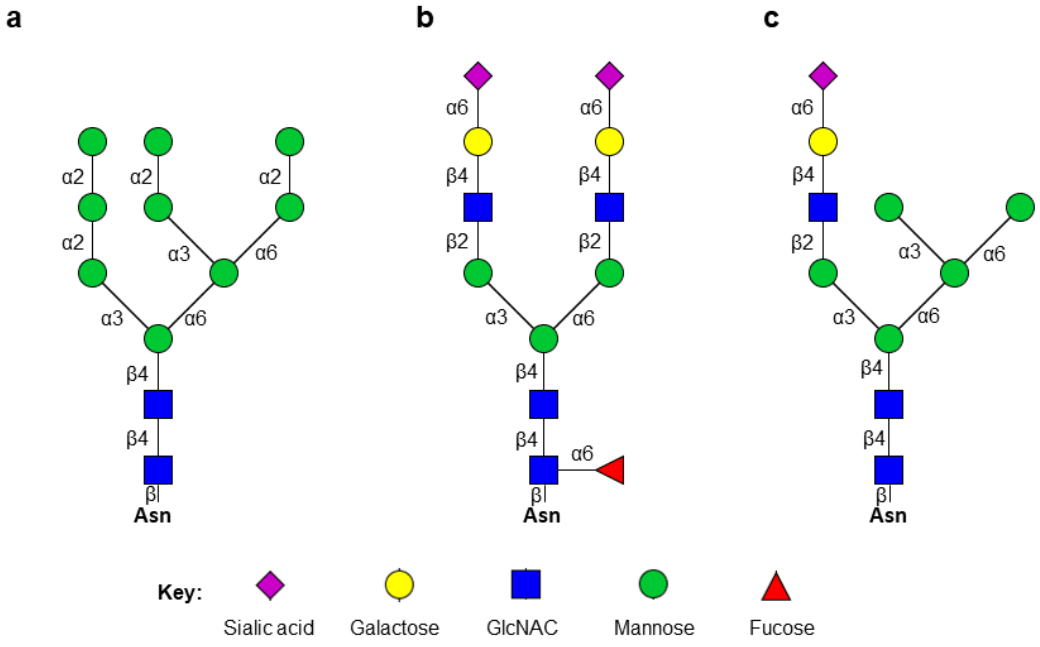


Figure 1.1: N-glycans. Examples of **(a)** oligomannose, **(b)** complex and **(c)** hybrid N-glycans. This figure was produced using GlycoWorkbench (Ceroni *et al*., 2008) and is based upon a figure seen in Essentials of Glycobiology (Stanley *et al*., 2009).

N-glycan biosynthesis begins with the production of a 14-residue precursor on a hydrophobic lipid dolichol phosphate, which is located on the cytosolic side of the rER. GlcNAc is added to the lipid via a pyrophosphate bond but the sequential sugars are attached to one another, through the action of glycotransferases, via glycosidic bonds. Once $Man_5GlcNAc_2$-PP-dolichol has formed, the precursor is flipped to the luminal side of the rER by a flippase enzyme (Stanley *et al*., 2009). The initial 7 sugars in the chain are donated from sugar nucleotides in the cytoplasm and the final four mannose and 3 glucose residues are added in the rER lumen. Oligosaccharyltransferase is responsible for the recognition of an Asn in the consensus sequence N-X-S/T (where X can be any amino acid barring proline) and the transfer of the oligosaccharide to the polypeptide (Aebi, 2013). This precursor glycan is subject to trimming and processing reactions in the rER, followed by further processing in the Golgi body, where modifications include the addition and removal of sugar residues, catalysed by glycosyltransferases and glycosidases (Stanley *et al*., 2009).

1.1.3 O-glycosylation in eukaryotes

O-linked oligosaccharides are connected to the hydroxyl groups on threonine and serine residues in a polypeptide chain via N-acetylgalactosamine (GalNAc), however there is no known consensus sequence in the protein chain where this attachment occurs (Brockhausen *et al*., 2009). The number of sugar residues in O-glycans can vary between as few as one to more than ten, with the smallest being the Tn-antigen, which is simply GalNAcα-Ser/Thr (Van den Steen *et al*., 1998). O-glycosylation is initiated in the Golgi body, where GalNAc is transferred to the polypeptide. A variety of transferases elongate the glycan in a stepwise manner, which results in the formation of eight different core O-glycan structures. The core structures are extended by the addition of fucose and sialic acid, with sialylation again being common at the terminus (Van den Steen *et al*., 1998). A major group of mammalian proteins that are heavily O-glycosylated are mucins, which are discussed in section 1.1.4. As well as this, O-glycans serve as antigens on the surface of red blood cells, which allows the human blood groups (ABO) to be identified (Brockhausen *et al*., 2009).

1.1.4 Mucins

Mucins are found in mucus secretions and the cell surface glycocalyx and are high molecular weight glycoproteins that can fight microbial infections in the oral cavity, gastrointestinal tract and urinary tract, as well as functioning in cell signalling and adhesion events when membrane bound (McGuckin *et al.,* 2011). In the oral cavity, the mucus membrane is composed of a squamous epithelial layer with underlying glands that secrete a fluid that contains mucin glycoproteins and antimicrobials, commonly referred to as saliva. The gastrointestinal tract is formed by a single layer of columnar epithelial cells that are surrounded by secreted mucus, which is composed of mucins, nonspecific antimicrobials and specific immunoglobulins and provides spatial separation between the commensal microbiota and the epithelium (McGuckin *et al*., 2011). The human MUC genes code for 20 known mucins, which are separated into secreted and membrane associated forms, both of which are heavily O-glycosylated (Corfield, 2015).

Secreted mucins contain repeated peptide stretches referred to as variable number of tandem repeats (VNTR) in the centre of their structures. The VNTRs are rich in serine and threonine residues, permitting O-glycosylation in this region, with glycans comprising about 85 % of the molecule by weight in salivary mucins (Figure 1.2) (Wu *et al*., 1994). N and C-terminals are cysteine-rich and are homologous to the C and D domains of von Willebrand factor. The cysteine-rich region allows mucins to oligomerise into homo-oligomers (gel-forming mucins), using inter-molecular disulphide bonds. A small amount of mucin N-glycosylation occurs in the ER, along with dimerization of two mucin molecules at their C-termini. The dimer then passes into the Golgi body to be O-glycosylated (at the VNTR region) and multimerization occurs at the N-terminus before the mucins are secreted in granules (Brockhausen *et al*., 2009; Moran *et al*., 2011). This complex array of O-glycosylation provides ligands for microbial adhesions and can be utilised as an energy source; however, the mucin glycoproteins can also have direct antimicrobial properties to limit microbial growth in the mucus (McGuckin *et al*., 2011).

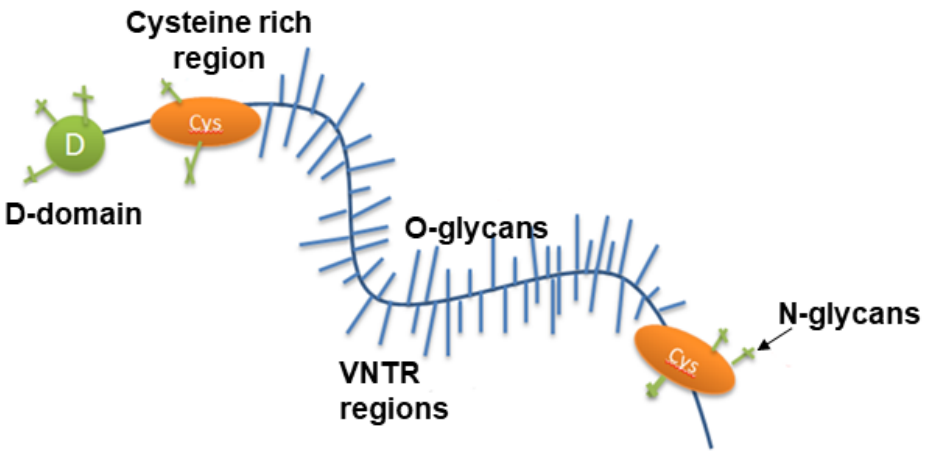


Figure 1.2: A typical secreted monomeric mucin. The VNTR region, with a cluster of O-linked oligosaccharides (sticks) are shown in blue, the cysteine rich terminals in orange, the von Willebrand domain in green and the few N-glycans are shown as green crosses.

Bacteria that breach the mucus layer reach the cell surface, which contains membrane associated mucins which have a single membrane-spanning peptide domain, a short cytoplasmic tail and an extensive extracellular O-glycosylated portion that extends above the cell surface, which provides adhesion sites for microbes (McGuckin *et al*., 2011). The dense O-glycosylation in the VNTR region creates a hydrophilic environment, lubricating the epithelia and therefore preventing access to this epithelium (Hattrup and Gendler, 2008). The cell surface mucin MUC1 is found in the gut and when bacteria adhere to its extensive network of O-glycans, the extracellular portion of MUC1 is shed, therefore acting as a decoy by limiting bacterial adhesion to the cell surface, something that has been demonstrated for *Helicobacter pylori* adhesion to the gastric mucosa (Linden *et al*., 2009). In addition to obstructing bacterial adhesion, cell surface mucins can modulate inflammatory signalling in epithelial cells through their cytoplasmic domains. For example, MUC1 extracellular domain has a specific binding site for *Pseudomonas aeruginosa* flagellin and a binding event causes MUC1 to suppress inflammatory responses that have been induced by this bacterium (Ueno *et al*., 2008).

1.1.5 Prokaryotic glycosylation

Protein glycosylation was believed to only occur in eukaryotic organisms before being discovered in prokaryotic and archaeal proteins. Our knowledge of prokaryotic glycosylation lags behind that of eukaryotic glycosylation because there is greater diversity, as a wider variety of monosaccharide building blocks can be used and several glycosylation processes occur (Esko *et al*., 2009). Glycoproteins were initially identified in the 1970s on the surface of the bacterium *Clostridium* and the archaeon

Halobacterium (Mescher and Strominger, 1976; Sleytr and Thorne, 1976), with the first complete bacterial N-glycosylation pathway being discovered in *Campylobacter jejuni,* which contains enzymes and transporters that promote an N-glycosylation pathway that resembles that in Eukarya (Szymanski *et al*., 1999). The pili of *Neisseria meningitides* are glycosylated, with PilE having three unoccupied N-glycosylation sites and an O-linked trisaccharide containing a 1-4 linked digalactose covalently linked to a 2,4-diacetamido-2,4,6-trideoxyhexose sugar (DATDH) (Stimson *et al.,* 1995). Bacterial flagella are also often glycosylated, with glycosylation having roles in flagellin filament assembly and stability, host mimicry, adhesion and host recognition (Nothaft and Szymanski, 2010). The study of bacterial glycosylation pathways is important, as the manipulation of certain components could lead to therapies against host infections and improvements in glycan synthesis and humanisation of biologics in recombinant strains.

1.1.6 Glycan-binding proteins – lectins

Lectins, also known as hemagglutinins (HA) or adhesins, are a type of carbohydrate-binding protein that have high specificity towards terminal aspects of N-glycans, O-glycans and glycosphingo-lipids, functioning as recognition molecules in cell-cell and cell-molecule interactions. The term lectin tends to be used when the sugar specificity of the protein is known, whereas HA is used when the sugar specificity is unknown (Lam and Ng, 2011). Lectins are commonly expressed by animals, plants, bacteria and viruses and can bind to oligosaccharides attached to cells or free floating glycans and monosaccharides. Lectins from these diverse sources lack primary sequence similarity but share similarities in their tertiary structures which often fold into 'jelly rolls', which are composed of β-strands, with examples being concanavalin A, human galectin 7 and ricin (Sharon and Lis, 2004).

In mammals, lectins control the protein levels in the blood and can recognise carbohydrates on the surface of pathogens to initiate an innate immune response. For example, mannose-binding lectin is a pattern recognition molecule that functions in innate immunity by targeting glycoconjugates on microbial surfaces, including lipopolysaccharide (LPS), which results in the activation the complement pathway (Turner, 2004). In addition, immune cells can also have cell surface bound sialic acid-binding immunoglobulin-like lectins (siglecs), which bind sialic acid and function in innate and adaptive immune responses. There are approximately 15 different mammalian siglecs (Lewis and Lewis, 2012), each of which differ in function depending on the cell-surface receptor ligand interaction, with a large fraction of siglecs being an important source of attenuating the immune response and dampening inflammation (Pillai *et al*., 2012). On a different note, lectins from different organisms, for example *Vicia cracia*, have been found bind to various terminal sugar residues on the surface of erythrocytes,

which specify the different blood groups and are used clinically in blood group typing or subgrouping (Khan *et al*., 2002).

Bacterial lectins behave as adhesins that enable attachment to and colonisation of host cells. *Escherichia coli* use a variety of pili to adhere to host surfaces, for example, the terminus of type 1 pili is composed of the FimH lectin, which binds to mannose on the surface of host cells in both the large intestine and urinary tract (Sokurenko *et al*., 1998). *H. pylori* employs two lectins, BabA and SabA, to bind to the heavily glycosylated epithelial lining of the human gastric mucosa. BabA has specificity for fucosylated glycoconjugates like Lewis B, whereas SabB is specific for the sialylated Lewis blood group antigens, Sialyl-Lewis X and Sialyl-Lewis A (Ilver *et al*., 1998; Mahdavi *et al*., 2002). Some secreted bacterial toxins can bind glycans and shiga toxin is one of the most potent bacterial toxins to date and is secreted by *Shigella dysenteriae* 1 (Stx) and by some serogroups of *Escherichia coli* (Stx1 and Stx2) (Ling *et al*., 1998). The Stx's are composed of two subunits (A and B), which together is referred to as AB_5. Subunit B must first identify the Gb3 (α-D-Gal(1,4)β-D-Gal(1,4)β-D-Glc(1,O-ceramide)) glycosphingolipid receptor on mammalian cells, which then enables subunit A to injure the eukaryotic ribosome and therefore halt protein synthesis in target cells (Jacewicz *et al*., 1986; Ling *et al*., 1998). Detailed studies of microbial lectins have led to the identification and production of therapeutic agents that act as inhibitors and therefore provide protection against infection. Methyl α-D-mannopyranoside and type-1-fimbriated *E. coli* were together injected into the bladder of mice, which showed a significantly reduced rate of urinary tract infection, which suggested that Methyl α-D-mannopyranoside is a competitive inhibitor of the binding of mannose by *E. coli* (Aronson *et al*., 1979).

The most well-known viral glycan-binding protein is the trimeric, influenza viral membrane protein hemagglutinin (HA). HA is used in the first step of entry to the host cell, where HA recognises and binds sialic acid receptors, eventually leading to internalisation of the virus by endocytosis and the release of viral RNA into the cytosol (Edinger *et al*, 2014). Sialic acid can be linked to the underlying galactose by an α2,3 or an α2,6 glycosidic linkage (see section 1.2.1). There are different specificities for host glycan-HA interactions depending on the subtype of influenza. HA from human strains of influenza A and B have a preference of binding to α2,6-linked sialic acid, whereas avian strains of influenza prefer α2,3-linked sialic acid (Rogers and Paulson, 1983). Several methods have been proposed for preventing sialic acid-HA interactions and therefore the inhibition of influenza infection. One example is the use of monoclonal antibodies (mABs) that interact with the globular head of HA, or mABs that block host sialic acid receptors, both of which prevent sialic acid-HA interactions. Soluble sialic acid analogues also have potential for inhibiting attachment of HA to sialic acid receptors by blocking the receptor-binding pocket of HA (Edinger *et al*., 2014).

Lectins have many practical applications when produced recombinantly, such as their use in affinity chromatography, which involves immobilising lectins on gel beads to which glycoproteins can be adsorbed and then eluted with a specific carbohydrate (Lam and Ng, 2011). For example, lectin-sepharose columns can be used to differentiate between the glycosylation patterns on commercial and endogenous forms of the glycoproteins which may be used as drugs, such as erythropoietin (EPO) (section 1.4.3) (Fraguas *et al*., 2008).

1.2 Sialic acid

1.2.1 Structure and diversity

The term sialic acid originates from the Greek word for saliva, sialon (Vimr *et al*., 2004). Sialic acid is a generic term for a monosaccharide with a 9-carbon backbone (also referred to as nonulosonic acids) that is located at the terminus of N and O-glycans and is a derivative of neuraminic acid (Zunk and Kiefel, 2014). There are over 50 variations of sialic acid, with N-acetylneuraminic (Neu5Ac) acid being the most abundant in humans and eukaryotes (Figure 1.3a) (Cohen and Varki, 2010). Sialic acid diversity arises from additional substitutions that occur at positions C-4, C-7, C-8 and C-9 (Figure 1.3a), where one or more of the hydroxyls could be replaced by O-acetyl, O-sulphate and O-methyl groups, which hinder sialic acid hydrolysis (Varki and Schauer, 2009) and also affect sialic acid recognition by glycan binding proteins (Cohen and Varki, 2010). A hydroxylated version of Neu5Ac, N-glycolylneuraminic acid (Neu5Gc), is a mammalian sialic acid however this form is not synthesised in humans (section 1.2.5) (Cohen and Varki, 2010). In addition, bacterial sialic acid variants, pseudaminic acid and legionaminic acid, have been discovered on bacterial cell surface glycans such as flagellins, pili and LPS, the function of which is not yet fully understood, but it is thought that they could be involved in host-bacterial interactions (Zunk and Kiefel, 2014).

Terminal sialic acids are linked to the underlying residue in the glycan by an α-glycosidic linkage. This can be via an α2,3- or α2,6-linkage to galactose or an α2,6-linkage to GalNAc (Figure 1.3b). Occasionally, an α2,8- or α2,9-linkage connects sialic acid to another sialic acid (forming polysialic acid) or GlcNAc (Cohen and Varki, 2010). In eukaryotes the exposed location and negative charge of sialic acids on cells, allows sialic acids to play roles in masking recognition sites or mediating cell recognition and adhesion, for example in the immune response via the action of proteins such as Siglecs that suppress several immune signalling pathways (Buschiazzo and Alzari, 2008; Lewis and Lewis 2012). In this thesis, sialic acid refers to Neu5Ac unless stated otherwise.

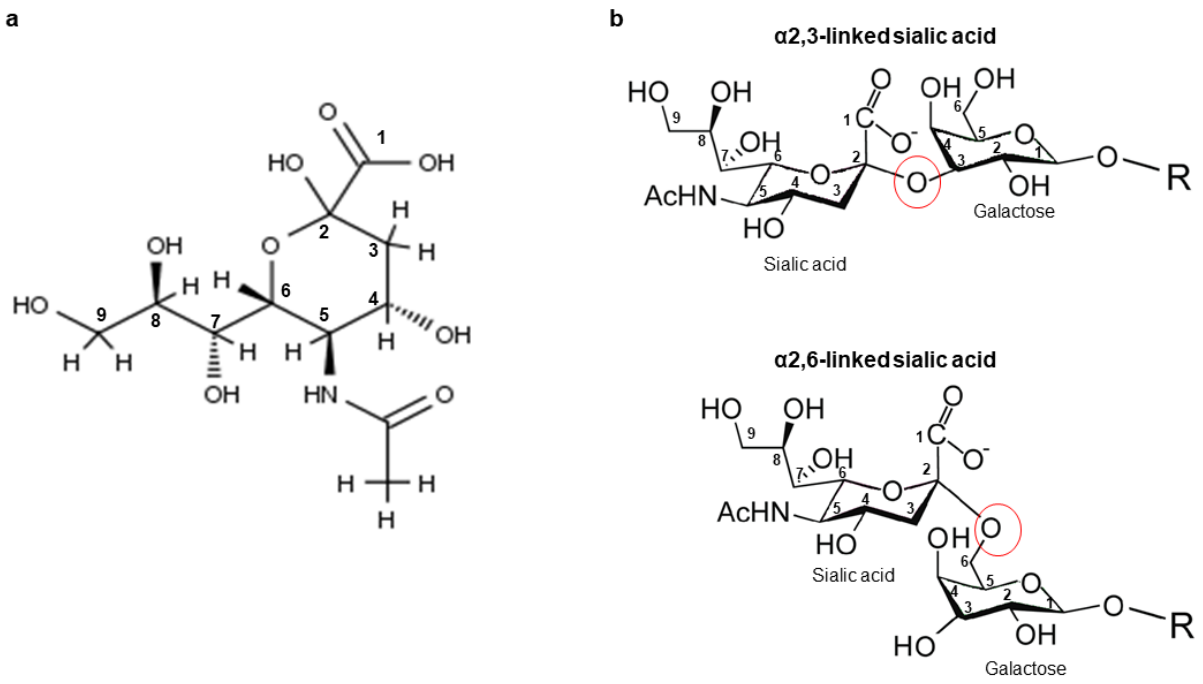


Figure 1.3: Chemical structures of (a) Neu5Ac and (b) α2,3- and α2,6-linked sialic acid. (a) was provided by C. Phansopa, Stafford research group and (b) was adapted from (Byrd-Leotis *et al*., 2017).

Secreted proteins that are found in mucosal secretions of the oral cavity, gut and respiratory tract tend to be rich in sialoglycans (Figure 1.4a and 1.4b) (Lewis and Lewis, 2012). Many of the O-glycans present on gut mucins include both α2,3- and α2,6-linked sialic acid at the terminus, which provides a main source of diversity (Figure 1.4c) (Tailford *et al.,* 2015b). Interestingly, human saliva contains predominantly α2,6-linked sialic acid which is due to the most abundant salivary mucin (MUC5B) containing almost exclusively α2,6-linked sialic acid (Limsuwat e*t al.,* 2016). It is thought that theses salivary α2,6-linked sialic acids are well positioned as they can act as decoy receptors by binding to human influenza viruses that would bind to similar α2,6-linked sialic acids receptors in the tissues of the human upper airways (Limsuwat e*t al.,* 2016; Kumlin *et al.,* 2008).

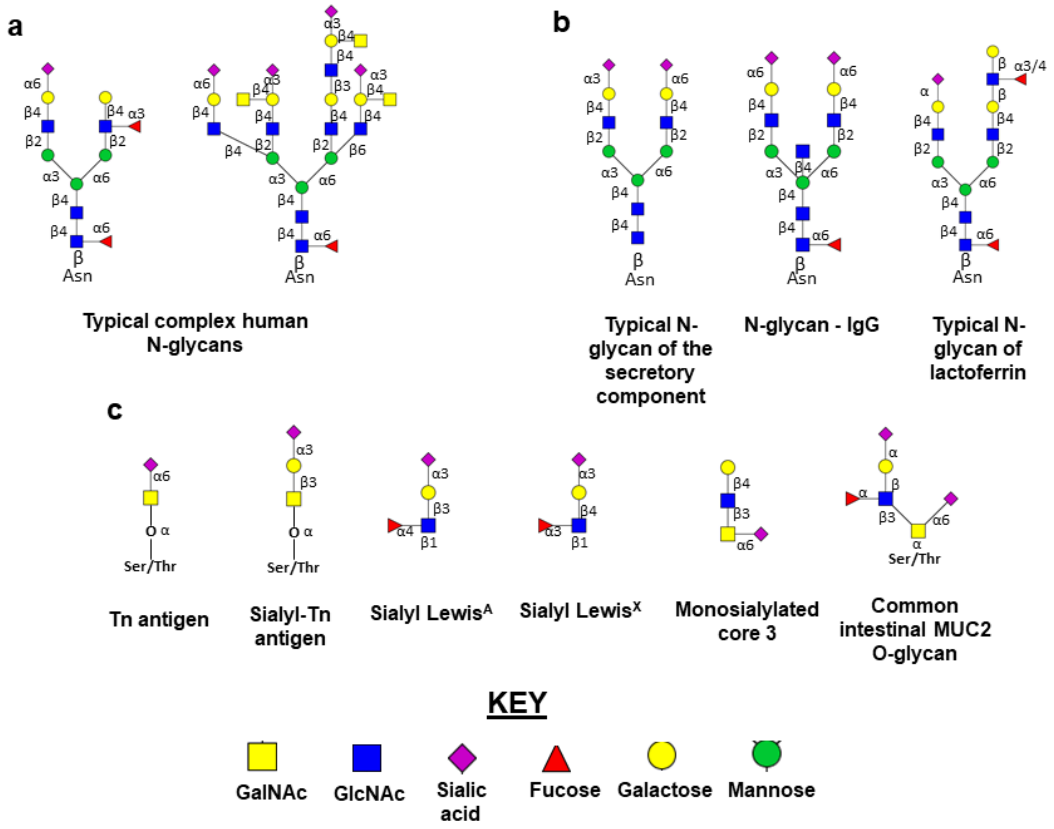


Figure 1.4: Types of human sialylated glycan structures. (a) Typical complex human N-glycans. **(b)** Human mucosal sialoglycans. **(c)** Gut mucin O-glycans. This figure was produced using GlycoWorkbench (Ceroni *et al*., 2008).

1.2.2 Sialic acid release - sialidases

Sialic acid is often released from glycoconjugates at different points during the life cycle of the molecule by sialidases (or neuraminidases), which cleave the α-glycosidic link (Corfield, 1992). Four human neuraminidases have been characterised and are referred to as neuraminidase 1, 2, 3 and 4 (NEU1, 2, 3 and 4) (Monti *et al*., 2002). NEU1 is a lysosomal neuraminidase and removes sialic acid from glycoproteins and glycolipids, which can be returned to the Golgi body to undergo re-sialylation (Monti *et al*., 2002; Varki and Schauer, 2009), however more recently NEU1 has also been found on plasma membranes where it can be involved in cell signalling for immune responses. For example, NEU1 is induced on human T-cells upon T-cell receptor stimulation, which in turn induces production of cytokines such as interleukin (IL)-2 and IL-13 which suggests an immunoregulatory role for the neuraminidase (Stamatos *et al*., 2005; Miyagi and Yamaguchi, 2012). In addition, NEU1 hydrolyses sialic acid from α2,3-linked β-galactosyl residues of Toll-like receptor 4 (TLR4), which is considered

essential for receptor activation and signalling through the activation of NFkB in macrophage and dendritic cells (Amith *et al*., 2010).

NEU2 is a cytosolic neuraminidase and has been suggested to function in myoblast differentiation and could also have a physiological role in hydrolysing sialic acids from misfolded glycoproteins that are released into the cytoplasm (Monti *et al*., 1999; Miyagi and Yamaguchi, 2012). Similarly to NEU1, NEU3 has also been identified as a plasma membrane associated neuraminidase and is thought to be involved in cell surface events such as immune responses and additionally modulate gangliosides in the lipid bilayer (Miyagi and Yamaguchi, 2012; Monti *et al*., 2000). Studies on the *NEU4* gene have suggested that the enzyme is associated with lysosomes, the mitochondria and the ER and is thought to be involved in neuronal cell apoptosis (Miyagi and Yamaguchi, 2012).

Neuraminidases are also widely distributed in microorganisms such as bacteria, viruses and protozoa (Varki and Schauer, 2009). Two proteins are present on the influenza virus envelope and are referred to as hemagglutinin (HA) and neuraminidase (NA), both of which recognise sialic acid and are involved in infection. HA initiates infection by binding to sialic acid receptors on cell surfaces (Russell *et al*., 2006), while NAs are required at the time of budding of newly replicated viral particles from the host and enable this by cleaving the sialic acid from glycoproteins on host cells. Influenza NA also hydrolyses sialic acid from mucins in the respiratory tract to facilitate viral particle movement to target cells (McAuley *et al*., 2019). 9 sub-types of NA have been described for influenza A (N1-9), whereas there is only 1 sub-type described for influenza B. The 9 sub-types of NA common to influenza A are further divided into two groups based on the presence or absence of a small cavity next to the active site (caused by movement in 3D space of residues (147-152)). N1, N4, N5 and N8 are suggested to form group one and N2, N3, N6, N7 and N9 form group two (Russell *et al*., 2006).

Bacterial NAs are referred to as sialidases and have been found to reside in the respiratory tract, the gastrointestinal tract, the reproductive tract and the oral cavity (Lewis and Lewis, 2012). *Streptococcus pneumoniae* produces three sialidases (NanA, NanB and NanC), which desialylate host cell surfaces for bacterial adhesion. NanA contributes to mucosal colonisation, as it is involved in biofilm formation and contributes to infection by promoting pro-inflammatory responses (Parker *et al*., 2009; Chang *et al*., 2012). Human milk oligosaccharides (HMOs) are a rich source of sialylated glycans and promote the colonisation of the intestine with commensal microbes such as *Bacteroides thetaiotaomicron* and *Bifidobacterium longum* subsp. *Infantis*. Both bacteria secrete glycoside hydrolases, including sialidases, which are likely to participate in the host metabolism by facilitating milk and future food digestion and perhaps limit pathogen colonisation by competing for microbial adhesins (Marcobal *et al*., 2011; Sela *et al*., 2011). The sialidase secreted by the gut pathogen *Vibrio cholerae,* VCNA, plays a

large role in the pathogenesis of cholera by cleaving sialic acid from gangliosides and unmasking the receptor which is targeted by the cholera toxin (Moustafa *et al*., 2004). This targeting of glycans is not restricted to gut microbiology but is now beginning to be explored in the oral cavity (Corfield, 1992). Indeed, the oral pathogen, *Tannerella forsythia* is one of the major pathogens associated with periodontitis and can utilise a sialidase (NanH) that is either associated with the outer-membrane or the secreted fractions (Honma *et al*., 2011). A study by Honma and colleagues found that NanH is involved in bacterial colonisation by cleaving sialic acid to expose epitopes on epithelial cells (Honma *et al*., 2011). Furthermore, Roy and colleagues showed that NanH is involved in biofilm initial adhesion and growth and that glycoprotein-associated sialic acid is likely to be a nutrient source for *T. forsythia* during biofilm growth (Roy *et al*., 2011).

In humans, a second O-acetyl group can be found at the C-9 position forming Neu5,9Ac_2 (diacetylated sialic acid), which the majority of sialidases struggle to cleave efficiently (Corfield, 1992). To access this sialic acid, *T. forsythia* produces a sialate-O-acetylesterase (NanS) which removes the O-acetyl group and increases the overall sialic acid release in the presence of the cognate sialidase (NanH), (Phansopa *et al*., 2015), in a process thought to give *T. forsythia* a competitive advantage *in vivo*. Gut *Bacteroides* such as *B. thetaiotaomicron* and *Bacteroides fragilis* also produce sialate-O-acetylesterases, which have strong homology to NanS of *T. forsythia* and suggests that it is also important to harvest diacetylated sialic acid in mucin rich regions like the gut (Briliute *et al*., 2019; Phansopa *et al*., 2015; Corfield *et al*., 1992).

1.2.3 Utilisation of sialic acid by pathogenic and commensal bacteria

The gut has a complex community of non-pathogenic microorganisms (microflora) which play an important role in the functional activity and stability of the intestine (Thursby and Juge, 2017). Some of these organisms have complex glycan degrading enzymes, including sialidases that can utilise the sialic acid for nutritional purposes, such as *Ruminococcus gnavus*. The *R. gnavus* sialidase (RgNanH) has an additional nutritional advantage as it is able to scavenge sialic acid in the form of 2,7-anhydro-Neu5Ac, which may not be accessible to the other microflora in the gut community (Tailford *et al*., 2015).

B. thetaiotaomicron is an important member of the gut microflora and has a polysaccharide utilisation locus (PUL), which encodes an extensive network of 19 enzymes, including a sialidase, which enable the breakdown of complex N-glycans, some of which could be utilised by the organism as nutrients (Briliute *et al*., 2019). This process is initiated by an extracellular facing membrane bound endo-glucosamidase which has been found to cleave bi-antennary and some tri-antennary N-glycans from α_1AGp, immunoglobulin (Ig) A and IgG from human serum and IgA from human colostrum, which are

then translocated into the periplasm via an outer membrane receptor complex where the N-glycan faces additional cleavage from periplasmic enzymes (Briliute *et al*., 2019).

Whilst *B. thetaiotaomicron* encodes a sialidase that hydrolyses sialic acid, it is incapable of consuming sialic acid, as it lacks a complete Nan operon (Marcobal *et al*., 2011; Brigham *et al*., 2009), which causes an increase in free sialic acid in the gut. This free sialic acid can cross-feed other mutualist bacteria in the gut community that may not be able to hydrolyse sialic acid themselves. However, when the normal microflora is depleted, for example during the use of antibiotics, the free sialic acid can be utilised by antibiotic resistant pathogens such as *Clostridium difficile* and some strains of *Salmonella typhimurium* that do not produce sialidases themselves but do have genes for sialic acid transport and metabolism, leading to their expansion in gut (Ng *et al*., 2013). McDonald and colleagues have highlighted that the ability of gut bacteria to use sialic acid as a carbon source is common, as *V. cholerae* also has genes for sialic acid uptake and catabolism, which could enable it to outcompete the non-catabolizers in the gut (McDonald *et al*., 2016).

In addition to sialic acid catabolism, some pathogenic bacteria can coat their outer surfaces with sialic acid, allowing them to mimic host cells and therefore resist the host's immune response. *E. coli* K1 and *C. jejuni* can produce sialic acid *de novo*, whereas other bacteria acquire sialic acid from the environment using secreted sialidases (Severi *et al*., 2007). Once obtained, sialic acid is converted into its activated form, CMP-Neu5Ac, which is in turn added to a bacterial cell surface target such as LPS by a sialyltransferase. LPS is usually recognised by the C3 component of the complement pathway but LPS sialylation inhibits this pathway (Severi *et al*., 2007).

1.2.4 Sialic acid uptake systems in bacteria

There is great diversity in transportation mechanisms of sialic acid from extracellular to intracellular environments amongst bacteria. Two uptake systems have been recognised for the transport of sialic acid across the inner membrane; Major Facilitator permeases, which are referred to as NanT and SiaPQM Tripartite ATP-independent periplasmic (TRAP) transporters (Thomas, 2016). Two outer membrane sialic-acid-specific transporters have also been identified, which are NanC from *E. coli* K-12 and NanO and NanU of *T. forsythia* and *B. fragilis* (Roy *et al*., 2010).

T. forsythia, along with other *Bacteroides* species, encodes a sialic acid utilisation and transport locus (TF0030-TF0038), which consists of approximately 16-kb of its genome (Stafford *et al.,* 2012). This locus encodes enzymes that are involved in the catabolism of sialic acid (*TF0030* and *TF0031*) and proteins that scavenge and transport sialic acid. The proteins included in transport include NanH, a secreted sialidase that scavenges sialic acid in the oral cavity (Stafford *et al.,* 2012; Roy *et al.,* 2011).

Another protein is NanT, an inner membrane permease, which is coupled to a TonB-dependent outer membrane transporter. As mentioned above, the outer membrane transport proteins, *TF0033-TF0034,* are referred to as *nanO* and *nanU* respectively (Roy *et al.,* 2010). NanU binds free or released sialic acid and interacts with NanO, which in turn passes the sialic acid into the periplasm, relying on energy from the TonB-ExbB-ExbD complex. Sialic acid is then transported to the cytoplasm by NanT before utilisation by the metabolism pathway (Figure 1.5) (Roy *et al*., 2010).

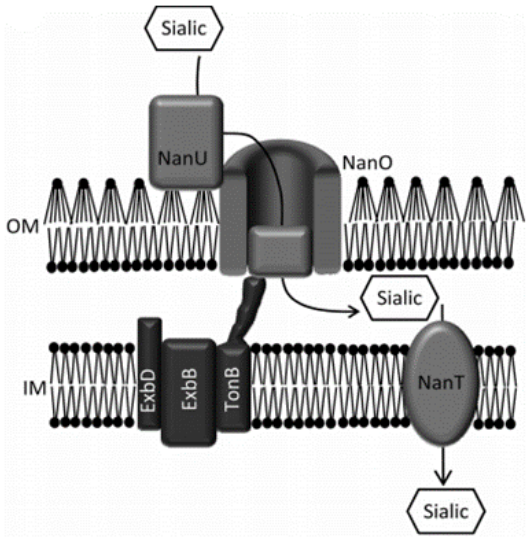


Figure 1.5: The sialic acid uptake pathway in *T. forsythia*. Sialic acid passes through the outer membrane (OM) via the NanOU transport proteins, which is energised by TonB-ExbB-ExbD. Once in the periplasm, sialic acid then passes the inner membrane (IM) into the cytoplasm via NanT. (Permission to use from G.Stafford, School of Clinical Dentistry, University of Sheffield).

1.2.5 Sialic acid - implications in disease

Neu5Gc is a common form of sialic acid in mammalian cells, however it is not synthesised in humans. Neu5Gc differs to Neu5Ac through the addition of a hydroxyl group at the 5-N-acetyl, which occurs in the cytoplasm of the cell (Varki and Schauer, 2009). The lack of Neu5Gc in humans is caused by a mutation that inactivates CMP-Neu5Ac hydroxylase (the enzyme responsible for Neu5Gc biosynthesis). Humans have no alternative synthesis pathways for Neu5Gc production, although it can enter the body by eating red meat (Tangvoranuntakul *et al*., 2003). Neu5Gc is referred to as an xeno-autoantigen, as when it is incorporated onto human glycoproteins, it can result in an immune response. Neu5Gc can access tissue cells such as carcinomas and when anti-Neu5Gc antibodies recognise tumour cells with this antigen, it can result in inflammation and tumour progression (Samraj *et al.,* 2015). Consuming red meat can also increase susceptibility to gastrointestinal disease, as a toxin

secreted by Shiga toxigenic *E. coli* has a strong binding preference for glycans terminating in Neu5Gc. This specificity of has been shown in a crystal model of the toxin-Neu5Gc complex (Byres *et al*., 2008).

1.3 Bacterial sialidases: structural and biochemical features and mechanism of action

1.3.1 Glycoside hydrolases

The Carbohydrate-active enzymes (CAZy) database was developed to classify the enzymes involved in the synthesis and catabolism on complex carbohydrates and for the display and analysis of genomic, structural and biochemical information on CAZymes (Lombard *et al*., 2014). The database is a sequenced based classification of CAZymes that supports the functional, structural and mechanistic properties of the enzymes (Davies and Sinnott, 2008). The database consists of families of glycoside hydrolases (GH), glycosyltransferases, polysaccharide lyases, carbohydrate esterases and non-catalytic carbohydrate-binding modules (Davies and Sinnott, 2008).

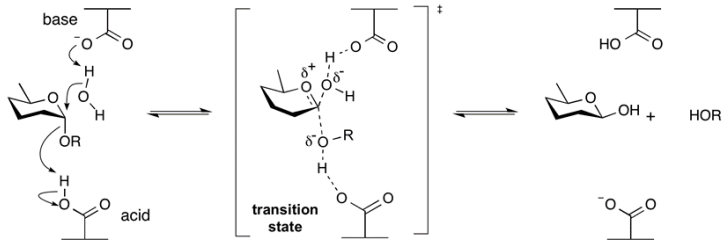
The GH families are of interest in this project, as sialidases are classified into these families (see section 1.3.2). GHs catalyse the hydrolysis of the glycosidic linkages of glycosides and they have been classified into more than 100 families via a sequence-based classification which is available on the CAZy database discussed previously (Henrissat, 1991). Each individual GH family contains proteins that are related by sequence and therefore fold, which allows predictions to be made regarding the catalytic machinery and molecular mechanism, which appear to be conserved throughout GH families (Henrissat *et al*., 1995). In addition, a nomenclature for the enzymatic sub-sites of carbohydrate processing GHs has been proposed. This rule states that sub-sites are numbered with increasingly negative numbers away from the hydrolysis position towards the non-reducing terminus (-1, -2, etc) and with increasingly positive numbers towards the reducing terminus (+1, +2, etc) (Davies *et al*., 1997).

The stereochemical outcome of the hydrolysis mechanism (retaining or inverting) is also considered in GH classification, which tends to be conserved in GH families (McCarter and Withers, 1994). GH hydrolysis can occur via two mechanisms that gives rise to an overall retention or inversion of the anomeric configuration of the glycoside. A one step, single-displacement mechanism with an oxocarbenium ion-like transition state is responsible for the hydrolysis of a glycoside with an inversion of the anomeric configuration (Figure 1.6a). This reaction usually involves two amino acid side chains (usually glutamates or aspartates) as a general acid and a general base (McCarter and Withers, 1994)). The retaining mechanism involves hydrolysis in which the net retention of configuration occurs via a two step, double-displacement mechanism involving a covalent intermediate, with each step passing through an oxocarbenium ion-like transition state (Figure 1.6b). This reaction requires acid/base and

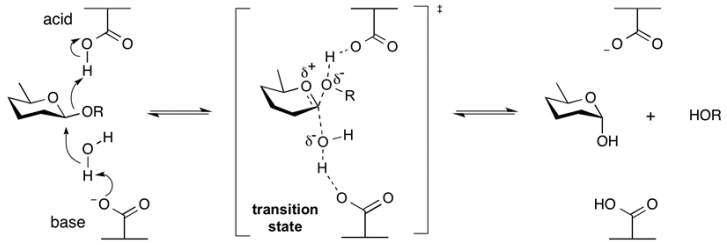
nucleophilic assistance, which is provided by two amino acid side chains, which tend to be aspartate and glutamate (described in more detail in section 4.1.3) (McCarter and Withers, 1994).

a

Inverting mechanism for an α-glycosidase:

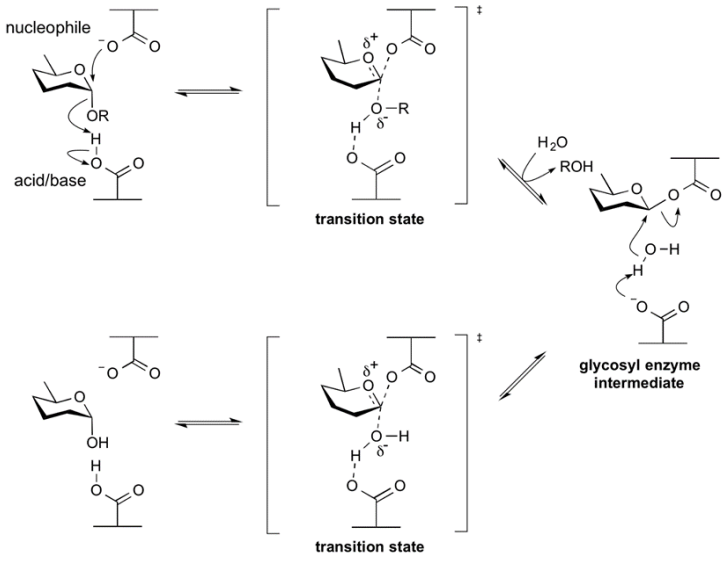


Inverting mechanism for a β-glycosidase:



b

Retaining mechanism for an α-glycosidase:



Retaining mechanism for a β-glycosidase:

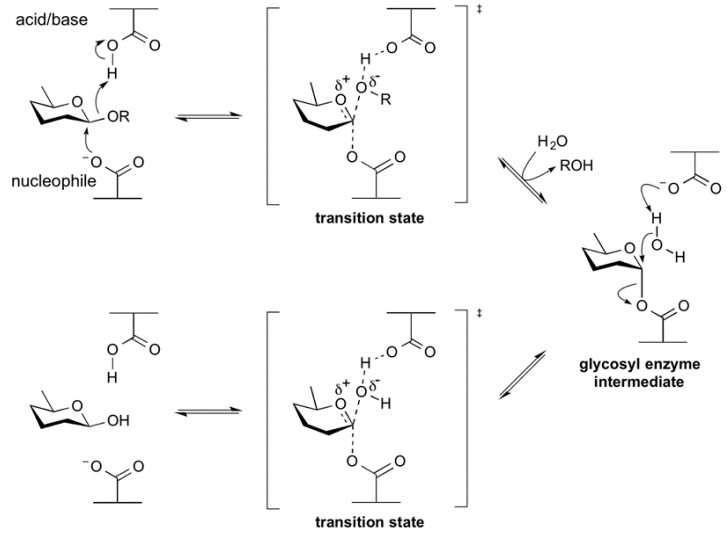


Figure 1.6: The two major mechanisms of glycosidic bond hydrolysis. (a) Inverting glycoside hydrolases and **(b)** retaining glycoside hydrolases. This figure was sourced from CAZypedia (Withers and Williams "Glycoside Hydrolases" in *CAZypedia,* available at URL https://www.cazypedia.org/index.php/Glycoside_hydrolases, accessed 17th September 2019).

1.3.2 Sialidase classification

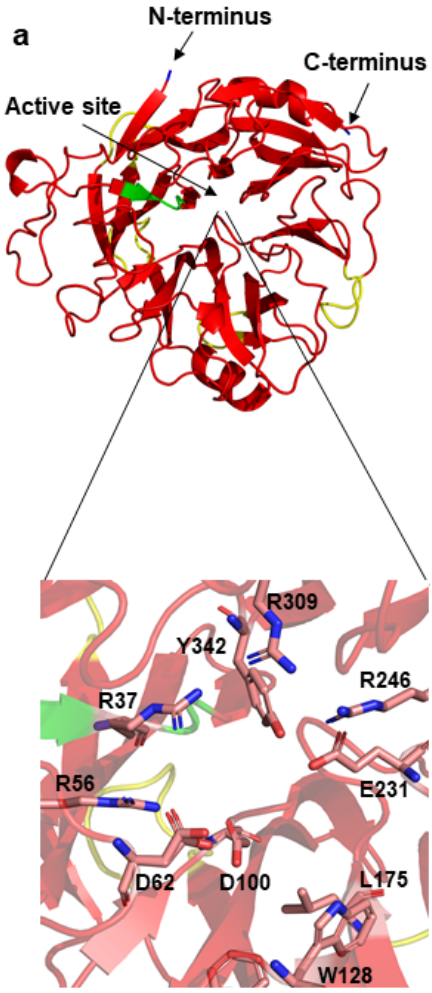
Sialidases (also known as neuraminidases) hydrolyse sialic acid from sialo-glycoconjugates and are widely distributed amongst bacterial, viral, mammalian and fungal species (Varki and Schauer, 2009). There are two main classes of sialidases; those that hydrolyse terminal sialic acids (exo-sialidases) and those that target endo poly-sialic acid. This project focuses on the exo-sialidases which are classified into CAZy subfamilies (such as glycoside hydrolase (GH) 33, GH34 and GH83), based upon amino acid sequence similarities, however they all take the same protein fold (section 1.3.3) (Lee, 2013; Chan *et al.,* 2012). The GH34 family refers to influenza viral sialidases, whilst the GH83 family represents other viral sialidases such as the paramyxovirus HA-NA (Chan *et al*., 2012). The majority of bacterial and eukaryotic sialidases are classified as the GH33 subgroup of the CAZy database classification (Lee, 2013).

Exo-sialidases are further divided into hydrolytic sialidases, trans-sialidases and intramolecular tran-sialidases (IT-sialidase) (Juge *et al*., 2016). Hydrolytic sialidases hydrolyse the glycosidic bond and release free sialic acid from α2,3-, α2,6- and α2,8-linked sialic acid, showing they have a wide substrate specificity (Juge *et al*., 2016). Trans-sialidases hydrolyse the glycosidic linkage and subsequently transfer the sialic acid to alternate glycoconjugates. Tran-sialidases have a preference for α2,3-linked sialic acid and the first to be discovered was TcTS of *Trypanosoma cruzi* (Schenkman *et al*., 1991). Similarly, the IT-sialidases discovered to date have a strict specificity for α2,3-linked sialic acid and the cleavage product is 2,7-anhydro-Neu5Ac instead of sialic acid. Currently only NanB of *S. pneumoniae*, NanL of *Macrobdella decora* and RgNanH of *R. gnavus* have been found to be IT-sialidases (Gut *et al*., 2008; Luo *et al*., 1998; Tailford *et al*., 2015).

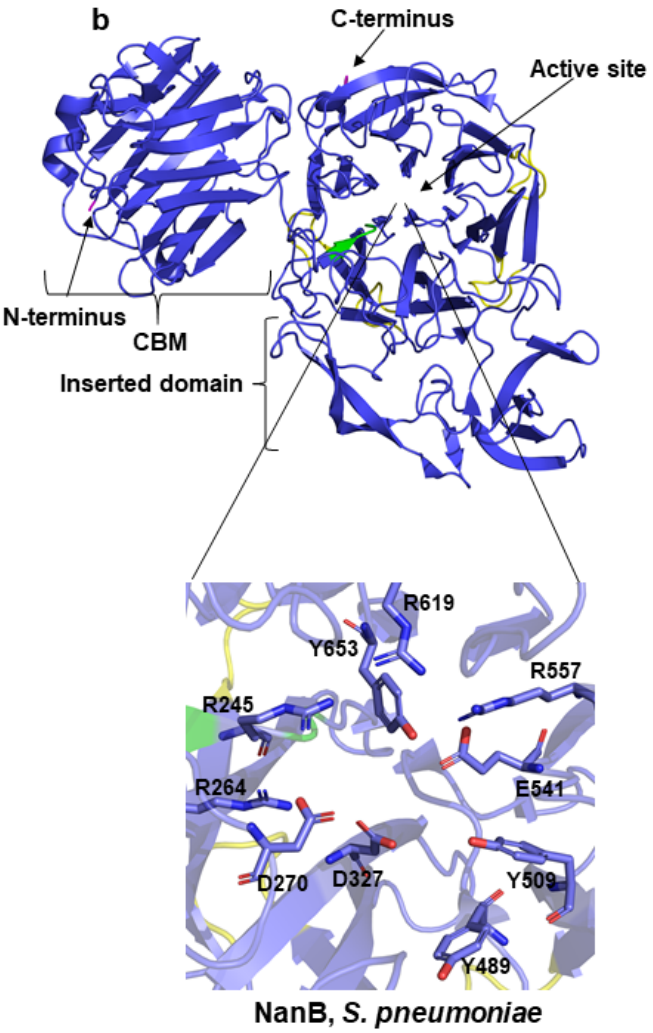
1.3.3 Structural features and key residues of bacterial sialidases

Sialidases that are expressed by bacteria are generally secreted or anchored to the cell membrane (Lewis and Lewis, 2012), and the molecular weight can vary between 40 and 150 kDa, depending on the presence of additional domains (Kim *et al*., 2011). The protein sequence homologies are less than 30 % between sialidases from different species, however the topology of the structure is conserved, and common motifs are shared (Kim *et al*., 2011).

Crystal structures of bacterial sialidases that have been deposited in the Protein Data Bank (PDB) (rcsb.org) (Table 4.1, Chapter 4) show that the catalytic domain folds into a 6-bladed beta (β)-propeller, with each blade being composed of a 4 stranded anti-parallel β sheet (Figure 1.7a and 1.7b) (Kim *et al*., 2011; Park *et al*., 2013; Xu *et al*., 2008; Crennell *et al*., 1994; Tailford *et al*., 2015). Some bacterial sialidases have carbohydrate-binding motifs (CBM) that act as lectin domains and can be attached at the N- or C-termini (Figure 1.7b) (Xu *et al*., 2008; Kim *et al*., 2011). The active site in bacterial sialidases is found in the central cleft of the propeller and contains a conserved arginine triad which interacts with the negatively charged carboxyl group at the C1 position of sialic acid (Figure 1.7a and 1.7b) (Buschiazzo and Alzari, 2008). Other conserved catalytic site residues are those that are involved in catalysis and include an aspartate (Asp) that acts as an acid/base catalyst and a tyrosine/glutamate (Tyr/Glu) pair whose role as a catalytic nucleophile has long been the matter of debate (section 4.1.3) (Buschiazzo and Alzari, 2008). In addition, bacterial sialidases have four to five copies of an aspartate box (Asp-box) motif in the catalytic domain with the consensus, Ser/Thr-x-Asp-x-Gly-x-Thr-Trp/Phe, where 'x' can represent any amino acid. The function of this motif is unknown, although it is thought that it could help to dictate the protein fold, as the repeats are conserved at topologically identical positions in the β-sheet folding (Roggentin *et al*., 1989; Kim *et al*., 2011). Finally, a Y/FRIP motif (tyrosine/phenylalanine-arginine-isoleucine/leucine-proline) is also conserved amongst the bacterial sialidase catalytic domain and is located upstream of the first Asp-box and has been determined to donate its arginine to the arginine triad (Kim *et al*., 2011). All sialidases have a hydrophobic pocket in their active sites to accommodate the N-acetyl group of sialic acid, although these residues are not usually conserved (Xu *et al*., 2008; Crennell *et al*., 1993).



STSA, *S. typhimurium*



NanB, *S. pneumoniae*

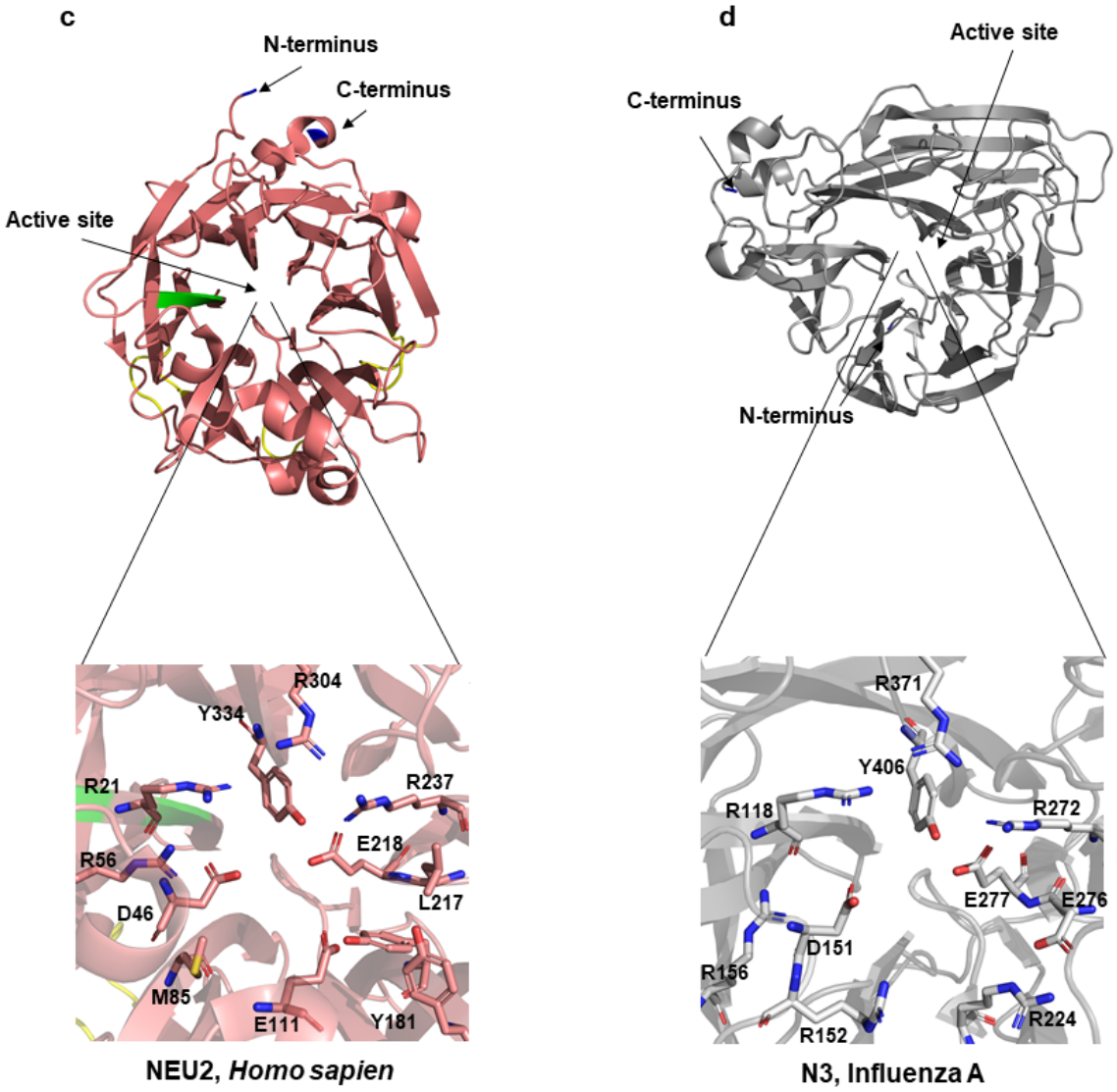


Figure 1.7: Crystal structures of bacterial, human and influenza sialidases. The canonical 6-bladed β-propeller is conserved amongst sialidases of different genera. Y/FRIP motifs are shown in green and Asp-box motifs are shown in yellow **(a-c)**. The active site is located in the central cleft of the propellers, which has been magnified in each figure to highlight key conserved residues. **(a)** STSA (PDB: 3SIL) is solely composed of a catalytic domain (propeller). R37, R309 and R246 form the arginine triad and D62, E231 and Y342 are the residues responsible for catalysis. **(b)** NanB (PDB: 2VW0) consists of an N-terminal CBM, a catalytic domain and a domain of unknown function inserted into the third domain. R245, R619 and R557 form the arginine triad and D270, E541 and Y653 are the residues involved in catalysis. **(c)** NEU2 (PDB: 1VCU) consists only of a catalytic domain. R21, R304 and R237 form the arginine triad and D46, E218 and Y334 are the residues involved in catalysis. **(d)** N3 (PDB: 4HZX) is composed solely of a catalytic domain. R118, R371 and R272 form the arginine triad and D151, E277 and Y406 are the residues responsible for catalysis. These figures were created using PyMOL.

The crystal structures of all 9 of the NA sub-types of influenza A have been solved (Sun *et al*., 2014) along with the NA of influenza B, which has conserved secondary and tertiary features with the NAs of influenza A (Bossart-Whitaker *et al*., 1993). The structure of NA is comprised of a head (with the catalytic activity) that is linked to a transmembrane domain by a stem, forming a homotetramer (approximately 240 kDa) in the viral envelope (Varghese and Colman, 1991; McAuley et al., 2019). In common with bacterial sialidases, influenza NAs fold into 6-bladed β-propellers (Figure 1.7d). The active site is found in the central cleft of the propeller, where highly conserved residues are found such as the arginine triad, the conserved Asp, the Tyr/Glu nucleophilic pair and a hydrophobic pocket (Figure 1.7d) (Bossart-Whitaker *et al*., 1993; Schwerdtfeger and Melzig, 2010). Influenza NAs tend to contain at least 8 disulphide bridges which bind Ca^{2+} and enhance activity. Disulphide bridges are limited in bacterial sialidases, with the sialidase from *S. typhimurium* containing only one that does not bind to Ca^{2+} (Crennell *et al*., 1993). Asp-box motifs, F/YRIP motifs and CBMs have not yet been highlighted in influenza NAs.

Currently, NEU2 is the only human sialidase for which the crystal structure has been determined. Once again, the canonical 6-bladed β-propeller fold is conserved in NEU2 and the active site is located in the central cleft of the propeller, which contains the conserved residues described above for both bacterial and influenza sialidases (Figure 1.7c) (Chavas *et al*., 2005). Similarly to bacterial sialidases, NEU2 has a conserved YRIP motif and 3 Asp-box motifs of which the function remains unknown (Chavas *et al*., 2005).

1.3.4 Carbohydrate-binding motifs

Lectin-like domains (section 1.1.6) have been found to be attached to the N or C-termini of several of the bacterial sialidases deposited in the PDB and are referred to as carbohydrate-binding motifs (CBM) (See Table 4.1, Chapter 4) (Kim *et al.,* 2011). The majority of CBMs that have been deposited in the PDB are have been classified into the CBM40 family of the CAZy database. CBM40 domains are approximately 200 residues and fold into β-sandwiches composed of two sheets, each consisting of 5-7 β-strands (Figure 1.7b) and they function in binding sialic acid (Boraston *et al*., 2004; Moustafa *et al*., 2004).

It was previously thought that the large sialidase (83 kDa) secreted by *V. cholerae* only had enzymatic activity, until the crystal model highlighted the presence of two identical CBM domains; one at the N-terminus and the other inserted between the second and third blades of the propeller, taking the form of β-sandwiches (Crennell *et al*., 1994). It has been observed that *V. cholerae* sialidase remains attached to cells treated with the sialidase and that the sialidase is able to degrade mucin layers, enabling the organism to reach the intestinal microvilli. These observations suggest that the CBM

adheres to epithelial cells, whilst the catalytic domain performs its catalytic function, so perhaps the CBM could be targeted in the future by new inhibitors (Crennell *et al*., 1994). A study by Thobhani and colleagues demonstrated that sialidases with CBM's (from *V. cholerae* and *C. perfringens*) had an increased catalytic activity than those in the absence of a CBM (STSA from *S. typhimurium*) towards polysaccharide substrates (Thobhani *et al*., 2003). Other sialidases such as NanB of *S. pneumoniae* and RgNanH of *R. gnavus* also have N-terminal domains, which are CBM40 classified (Xu *et al*., 2008; Tailford *et al*., 2015).

NanH of *T. forsythia* has an N-terminal CBM domain, the structure of which is yet to be deposited in the PDB. Frey and colleagues have suggested that this domain represents the founding member of a novel CBM family, as it does not share any homology with other documented lectin domains or the CBM40 family (Frey *et al.,* 2018). Alignments of the NanH N-terminal domain with those of other *Bacteroides* species revealed significant homology and indicated that this novel domain is conserved between these groups of bacteria (Frey *et al*., 2018). In isolation from the catalytic domain, the NanH-CBM is able to bind both sialylated and non-sialylated glycoconjugates, but not monosaccharides of Neu5Ac. Frey and colleagues therefore propose that this domain has a role in docking and orientation of an oligosaccharide into the active site (Frey *et al*., 2018). The commensal sialidase BTSA of *B. thetaiotaomicron* also has a unique N-terminal β-barrel CBM, which is not CBM40 classified. This CBM is considered to bind and position the polysaccharide substrate directly into the active site in the catalytic domain (Park *et al*., 2013).

1.3.5 The substrate specificity of sialidases

Whilst some sialidases have dual specificity (the ability to hydrolyse a variety of glycosidic linkages), some sialidases are linkage specific despite the structural similarities they may share with others. It is known that there are sialidases that have higher activities towards artificial substrates such as 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid (MU-NANA), than natural substrates like sialyllactose (Kim *et al.,* 2011). *Arthrobacter sialophilus* sialidase hydrolyses α2,3-, α2,6- or α2,8-linked sialic acids, however its activity towards sialyllactose is less than 10 % of that towards MU-NANA (Kessler *et al*., 1982). Currently, there is no knowledge of any sialidases that are specific for α2,6-linked sialic acid in the literature. The sialidases of *S. typhimurium* (LT2), *Clostriudium septicum, Clostridium sordellii,* and *Clostridium tertium* have higher activity towards α2,3-linked sialic acid than α2,6-linked (Zenz *et al*., 1993; Roggentin *et al*., 1987; Grobe *et al*., 1998), whereas the sialidase from *Corynebacterium diptheriae* has a slight preference for α2,6-linked sialic acid, although the data in this paper are not conclusive (Kim *et al*., 2010). NanL of *M. decora,* NanB of *S. pnuemoniae* and RgNanH of *R. gnavus* are sialidases that have been found to be specific for α2,3-linked sialic acid (Luo *et al*., 1998;

Gut *et al*., 2008; Tailford *et al*., 2015). The structural features responsible for this specificity are discussed in section 5.1.2. In terms of the human NAs, NEU1 and NEU2 both have higher activity towards oligosaccharide substrates with α2,3-linked sialic acid (Miyagi and Tsuiki, 1984; Miyagi and Tsuiki, 1985) and NEU3 also has a higher activity towards α2,3-linked sialic acid but with gangliosides (Hata *et al*., 1998). The specificity of influenza A NA depends on the isolate examined. For example, the specificity of human virus N2 NA has changed over time since its introduction from the avian virus (Baum and Paulson, 1991). The earliest human N2 virus isolate in 1957 was specific for α2,3-linked sialic but by 1972 it was found to have similar specificities for both α2,3- and α2,6-linked sialic acid. α2,6-linked sialic acid is the preferred receptor determinant of some of the human influenza virus HA receptors, suggesting there was a drift in specificity towards the HA receptor determinant to facilitate the release of progeny viruses (Baum and Paulson, 1991).

1.3.6 Sialidase mechanism of action

Two catalytic mechanisms have been proposed for sialidases; one is based on the influenza neuraminidase and the second based on the *T. cruzi* sialidase (TcTS) (see section 4.1.3 and Figure 4.2 for more detail) (Taylor and von Itzstein, 1994; Amaya *et al.,* 2004). The TcTS method involves a 'ping-pong' mechanism which uses the conserved 'Tyr/Glu' pair as a nucleophile and seems to be more widely accepted for bacterial sialidases. This mechanism is initially the same for hydrolytic, trans- and IT-sialidases, as they share the same catalytic residues discussed in section 1.3.3 (Amaya *et al*., 2004; Park *et al*., 2013; Juge *et al*., 2016). A proton is donated from the conserved Asp, which initiates the formation of a glycosyl enzyme intermediate. The conserved Tyr acts as a nucleophile, after activation by an adjacent conserved Glu (base catalyst). The Asp then acts as a base catalyst during the nucleophilic attack and the Neu5Ac product is released by hydrolytic sialidases yet transferred to another glycoconjugate by trans-sialidases (Lee, 2013; Amaya *et al*., 2004). A conserved Thr residue amongst IT-sialidases sterically hinders the glycerol group of sialic acid and forces it into an axial position. Therefore, during IT-sialidase reactions, the O7-hydroxy group of the bound sialic acid glycerol group is able to attack the positively charged C2 atom of the positively charged oxocarbenium intermediate. This results in the release of 2,7-anhydro-Neu5Ac as opposed to Neu5Ac, which is true for RgNanH of *R. gnavus* and NanB of *S. pneumoniae* (Juge *et al*., 2016; Tailford *et al*., 2015; Gut *et al*., 2008).

1.3.7 Sialidase inhibition

Sialic acid cleavage is a crucial part of the influenza virus replication and blocking the action of influenza NA with inhibitors is an effective way to treat influenza infection (McKimm-Breschkin, 2013). Viral proliferation is prevented by the NA inhibitors oseltamivir and zanamivir, which are licensed

worldwide, under the names Tamiflu and Relenza respectively. These inhibitors are structural analogues of sialic acid and were produced based upon the transition state analog, N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid (DANA), which was found to be a weak inhibitor of NA and sits in the active site where sialic acid would be expected to bind (Figure 1.8 and Figure 1.9) (McKimm-Breschkin, 2013). Influenza NAs have IC_{50} values with oseltamivir and zanamivir which are in the nanomolar range, with ranges between 0.3-0.6 nM and 0.7-3.1 nM with oseltamivir and zanamivir respectively for influenza A subtypes H1N1 and H3N2 (Gubareva *et al*., 2001). However, mutations in the NAs are enabling influenza to gain an increased resistance to the most heavily used drug, oseltamivir (Li *et al*., 2013). Two more recently developed NA inhibitors are peramivir and laninamivir (Figure 1.8), which have both been approved in Japan, with peramivir also being approved by the US FDA. Peramivir is based upon DANA but has a cyclopentane ring and features of both zanamivir and oseltamivir, whereas laninamivir is based upon zanamivir with a 7-OCH_3 substitution (McKimm-Breschkin, 2013).

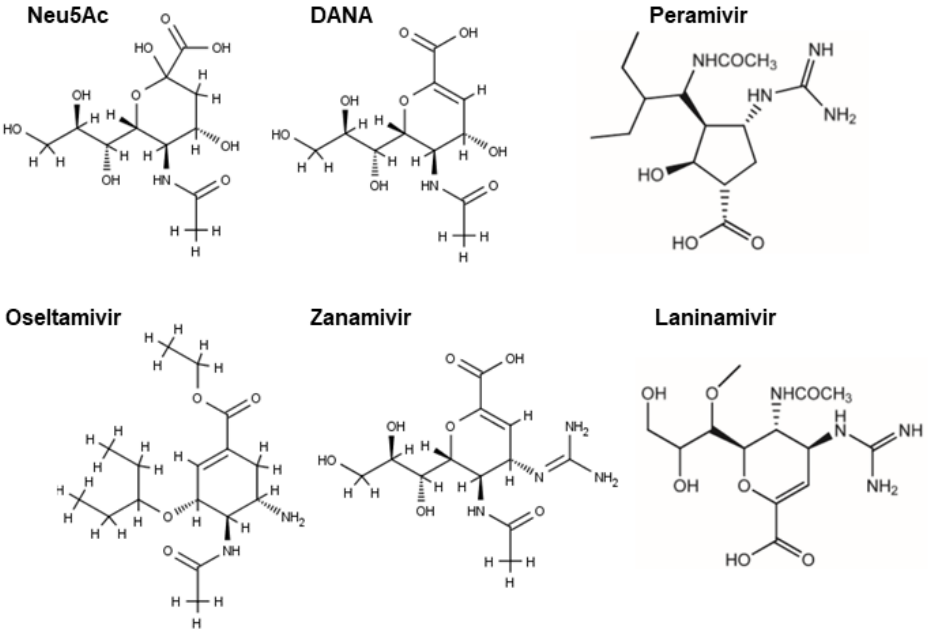


Figure 1.8: Sialidase inhibitors. Chemical structures of the Neu5Ac structural analogues (Neu5Ac, DANA, oseltamivir and zanamivir were provided by C. Phansopa, Stafford research group, whereas peramivir and laninamivir were adapted from (McKimm-Breschkin, 2013).

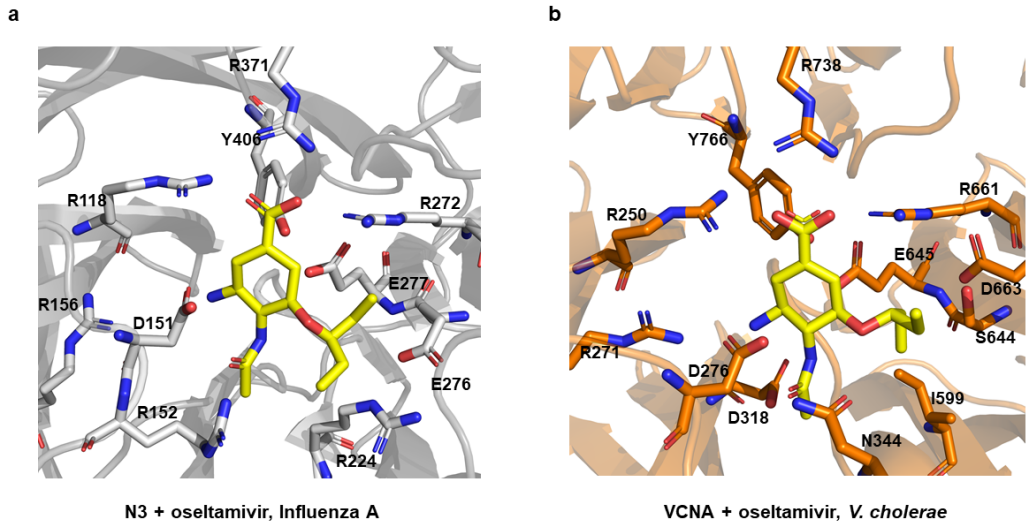


Figure 1.9: Oseltamivir in complex with viral and bacterial sialidases. (a) Oseltamivir (yellow sticks) in complex with N3 of influenza A (grey) (PDB: 4HZX), orientated through the interactions between the carboxyl group of oseltamivir and the arginine triad (R118, R371 and R272). **(b)** Oseltamivir (yellow sticks) in complex with VCNA of *V. cholerae* (orange) (PDB: 6EKS), orientated by the interactions between the carboxyl group of oseltamivir and the arginine triad (R250, R738 and R661). This figure was made using PyMOL.

Some studies have involved investigating the effect of NA inhibitors on endogenous enzymes that are involved in sialic acid metabolism. A study by Hata and colleagues discovered that oseltamivir did not affect the activity of any of the four human NAs, even when concentrations as high as 1 mM were used. Zanamivir was able to inhibit NEU2 and NEU3 in the micromolar range with NEU2 having an IC_{50} of 16.4 µM with the MU-NANA substrate, however this is still a contrast to the lower nanomolar levels at which it can inhibit influenza NA (Hata *et al*., 2008). Oseltamivir and zanamivir also inhibit bacterial sialidases, but less efficiently than influenza NAs. The IC_{50} values for VCNA from *V. cholerae* with oseltamivir and zanamivir were 144 µM and 52 µM respectively, meaning that off-target effects in human NEU enzymes are possible (Figure 1.9) (Quosdorf *et al*., 2017). Targeting bacterial sialidases with a more efficient inhibitor could prove beneficial in halting activity and therefore bacterial pathogenesis of some of the sialidases discussed in section 1.2.2. Recently, difluorosialic acids have been designed as mechanism based sialidase inhibitors and have proved effective against sialidases such as *C. perfringens*. 9-Azido-9-deoxy-2,3-difluorosialic acid was the most effective inhibitor with an IC_{50} value of 24 nM for the *C. perfringens* sialidase (Li *et al*., 2019).

1.4 Glycan profiling in the biopharmaceutical industry

1.4.1 Biopharmaceuticals are produced recombinantly

Biopharmaceuticals or biologics are proteinaceous drugs that are extracted from or semi-synthesised from biological sources with many being produced by recombinant methods in living cells, such as Chinese hamster ovary (CHO) cells, rather than by chemical processes. As of 2018, there were 316 biopharmaceuticals with an active license in the United States and/or the European Union (Walsh, 2018). Products approved over the past four and a half years include 68 monoclonal antibodies (mAbs), 23 hormones, 16 clotting factors, 9 enzymes (Walsh, 2018). Examples of well-known recombinant biopharmaceuticals are immunoglobulin G (IgG) (Jefferis, 2005), erythropoietin (EPO) (Sasaski *et al.,* 1987) and follicle-stimulating hormone (FSH) (Prevost, 1998).

Recombinant biopharmaceuticals possess and often require post-translational modifications such as N- and O-glycosylation, which add great complexity to therapeutic proteins and can in turn affect the pharmokinetics and immunogenicity of the drug (Goh and Ng, 2018). Mammalian expression systems (CHO cells) are preferred for commercial production because their glycosylation modification machinery is closer to that in humans than other systems such as yeast (*Pichia pastoris*) or insect cells and can therefore produce glycans that are compatible with humans (Kim *et al*., 2012). However, glycosylation is sensitive to changes in the production environment, such as scale-ups and method adaptions. It is important to maintain consistency in glycosylation patterns for safety and efficacy reasons, which still remains a challenge for the biopharmaceutical industry (Fernandes, 2004). Glycan structures are designated a critical quality attribute (CQA) and the glycan profile of the final product must be analysed to assure batch consistency.

In particular, sialic acid analysis is important when analysing glycoprotein glycosylation, as it can influence drug safety and efficacy (Byrne *et al.,* 2007). Sialic acid is found at the terminus of glycans and is important in the stability of glycoproteins and is involved in several biological interactions, as well as being a label of ‘self’ in humans. The presence of sialic acid on biopharmaceuticals can be beneficial to the circulating half-life and therefore activity of a drug (Varki, 2008). For example, sialic acid ‘hides’ the penultimate galactose residue, which would otherwise be recognised by C-type lectin asialoglycoprotein receptors in hepatocytes, leading to the removal of the glycoprotein from circulation (Ashwell and Harford, 1982). Another sialic acid derivative, Neu5Gc (section 1.2.5), can be added to glycoproteins that have been expressed in non-human cell lines. This non-human sialic acid can cause an immunological response, leading to inflammation and/or neutralisation of the drug, as it is considered a xeno-autoantigen in humans. CHO cells are commonly used for biopharmaceutical expression as they produce only low levels of Neu5Gc (Bertozzi *et al*., 2009; Hokke *et al*., 1990).

1.4.2 Recombinant antibody therapeutics

One of the most common biopharmaceuticals are recombinant antibodies or hybrids thereof which are a group of disease modifying drugs with many disease targets, including cancers and rheumatoid arthritis. By the end of 2017, the global market had a total of 57 monoclonal antibodies (mABs) and 11 biosimilars in clinical use (Grilo and Mantalaris, 2019). The efficacy of mABs depends upon how well they form a complex with the target and the activation of downstream effector mechanisms to remove the target. Currently, only mABs from one of the IgG subclasses are used as therapeutics (Jefferis, 2005). IgG is composed of two identical heavy chains and two identical light chains which are in covalent and noncovalent association to form three protein moieties (Figure 1.10). The upper two moieties of this structure (Fab region) are structurally identical and both express a specific antigen-binding site. The lower portion of IgG is a homodimer (Fc region), which contain binding sites for ligands that are involved in activating clearance (Jefferis, 2005).

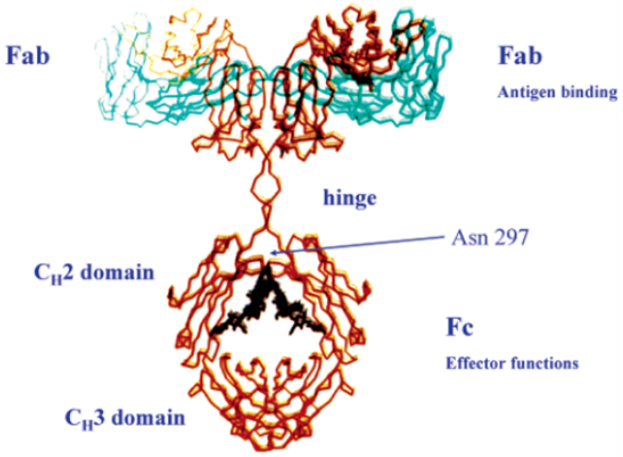


Figure 1.10: A model of human IgG, with the N-glycosylation site highlighted in the Fc region. The light chains are shown in blue, whilst the heavy chains are shown in orange. The dark mass below where Asn297 is labelled represents the glycosylation. This figure was adapted from (Jefferis, 2005).

An N-glycosylation site is located at Asn297 in in the Fc portion and the oligosaccharide attached in polyclonal IgG is of the complex bi-antennary type, with a heptasaccharide core. Heterogeneity is increased as variable sugar residues can be added to the outer arms of the glycan chain, allowing up to 32 different oligosaccharides and the structure of the glycan plays an important role in effector functions such as antibody-dependent cellular cytotoxity (ADCC) (Jefferis, 2005). CHO cells can produce recombinant IgG antibodies with glycoforms that are identical to those in human IgG. However, as discussed previously (section 1.4.1), under non-optimal conditions abnormally glycosylated products may be produced that are immunogenic or lack potency (Jefferis, 2005).

Biopharmaceutical companies work on the development of mAbs with improved effector functions, for example, obinutuzumab has been found to have a better efficacy than rituximab in clinical trials for chronic lymphocytic leukaemia treatment (Owen and Stewart, 2015).

1.4.3 Recombinant erythropoietin

EPO is an endogenous cytokine hormone that is secreted by the kidneys as a response to a decrease in oxygen levels in tissues, which stimulates the production of red blood cells (and therefore oxygen carrying capacity) and is now produced recombinantly as a treatment for anaemia (Sasaki *et al*., 1987). Recombinant human EPO (rHuEPO) is a successful biopharmaceutical that can be produced in CHO cells where the glycosylation pattern is indistinguishable from human EPO, except for minor differences in sialylation (Sasaki *et al*., 1987). The polypeptide represents 60 % of human endogenous EPO's molecular weight, with the remaining 40 % consisting of N- and O-glycans. Three N-glycosylation sites are present at Asn24, Asn38 and Asn83 and one O-glycosylation site is found at Ser126 (Lai *et al*., 1986).

It is of common knowledge that rHuEPO has been used illegally by sportsmen as a performance enhancing drug. Several rHuEPO detection methods exist, including the measurement of hematologic parameters, glycan analysis and gene-based detection methods (Citaran *et al*., 2015). Glycan analysis of N-glycans involves obtaining glycan profiles for the sample expected to contain rHuEPO and comparing it to that of endogenous EPO. Techniques such as high-performance liquid chromatography (HPLC) and mass spectrometry (MS) can be used to distinguish between endogenous EPO from rHuEPO (Citaran *et al*., 2015).

1.4.4 Glycan characterisation – release, labelling and profiling

Several analytical techniques are used to analyse the glycosylation patterns of biopharmaceuticals. Firstly, the attached glycans must be released from the protein of interest, followed by the derivatisation and analysis of the released glycan. It is important to understand the composition of the units that form the glycan and the linkages between these units and HPLC and MS are techniques that can be used in combination to characterise glycans (Kozak *et al*., 2015).

PNGase F is the most common enzyme used to release N-glycans from the protein they are attached to. This enzyme cleaves between GlcNAc and the asparagine in the protein chain, resulting in the asparagine being converted to an aspartate residue (Stanley *et al*., 2017). O-glycans are released from proteins using chemical cleavage, as glycan cleaving enzymes tend to be specific for N-glycans, although chemical cleavage is applicable to N-glycans. β-elimination is one method of chemical cleavage in which a glycoprotein is exposed to an alkaline environment, which releases the glycan

(Roth *et al*., 2012). Additional β-elimination reactions can be performed to further degrade the glycan into monosaccharides. However, β-elimination has its disadvantages as the reduced end aldehyde is eliminated during the reaction, meaning glycan labelling (more details below) prior to HPLC is not possible (Roth *et al*., 2012). A second method of chemical cleavage is hydrazinolysis, which involves the addition of anhydrous hydrazine to initiate hydrolysis. This method is applicable to both N- and O-glycans and altering the reaction temperature can allow the release of these two different glycan forms separately (Roth *et al*., 2012). Recently, a more rapid and efficient N-glycan deglycosylation technique has been developed, which has been found to have higher activity for the removal of N-glycans from IgG, ribonuclease B and fetuin than in a conventional in-solution enzyme digestion. This involves immobilising PNGase F onto the surface of magnetic microparticles, which can then be reused by separating the magnetic particles from the digest using a magnet (Bodnar *et al*., 2016).

After release, glycans are labelled with a fluorescent probe at the reducing terminus to enable detection by HPLC. This conjugation occurs through a reductive animation reaction with 2-aminobenzamide $(2AB)^2$ or 2-anthranilic acid (2-AA), both of which have been shown to be efficient labels (Bigge *et al*., 1995). More recently, procainamide [4-amino-*N*-(w-diethylaminoethyl) benzamide] has been found to be a more efficient label than 2AB for N-glycans released from IgG. Kozak and colleagues determined that after analysis by ultra (U)HPLC-fluorescence (FLR) coupled to electrospray ionisation-mass spectrometry (ESI-MS), the fluorescence intensity for procainamide labelled IgG was higher than that of 2AB labelled IgG and the ESI efficiency for procainamide was up to 30 times that of 2AB (Kozak *et al*., 2015).

HPLC is the most widely used method for glycan separation and tends to be performed in HILIC mode (hydrophilic interaction liquid chromatography), with a fluorescent detector to detect the intensity of the labelled glycans (Reusch *et al*., 2015a). Generally, dextran ladder mobility (GU values) and mass data are obtained for each glycan peak on the chromatogram, the structures of which can be determined when compared to similar data from known N-glycans in reference databases (Zhao *et al*., 2018). HILIC-(U)HPLC-FLR can be followed ESI-MS, allowing the production and detection of ions which are separated according to their mass-to-charge (*m/*z) ratios, which provides qualitative (structural) and quantitative (molecular mass) information about the glycans inputted (Han and Costello, 2013).

ESI-MS is a soft ionisation technique and it is not strong enough to fragment the glycans. This allows the intact glycan molecule to be observed as a precursor in a mass spectrum (the plot of the relative abundance of the ions as a function of their *m/z* ratio). Tandem mass spectrometry (MS/MS) is a more sensitive analytical technique that is performed in addition to ESI-MS within the same instrument. MS/MS involves the fragmentation of precursor ions of a specific mass-to-charge ratio, which enables

the determination of the chemical structure of the glycan from the fragments detected (Han and Costello, 2013; Ho *et al*., 2003; Reusch *et al*., 2015b). Capillary electrophoresis (CE) with fluorescence detection and high pH anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) have also been shown to have high accuracy in glycan separation (Reusch *et al*., 2015a). These methods can be coupled to (ES)-MS or Matrix-Assisted Laser Desorption/Ionisation (MALDI)-MS to determine the masses of the glycans present in the profiles produced (Reusch *et al*., 2015a).

Variations in the pattern of glycosylation in human cells are associated with inflammatory diseases, neurodegenerative diseases and cancer, so it is also important to study glycosylation patterns to determine disease biomarkers (Kam and Poon, 2008). An example of where glycan profiling could be beneficial is in the detection of biomarkers for maturity onset diabetes of the young (MODY). It is common for MODY patients to be misdiagnosed and therefore not treated effectively, as current genetic tests are difficult to interpret (Thanabalasingham and Owen, 2011). Hepatocyte nuclear factor 1-α is (HNF1A) is a key regulator of fucosylation (Lauc *et al*., 2010) and variants in this protein can cause the most common form of MODY (Shields *et al*., 2010). Patients with defects in the HNF1A gene have reduced plasma N-glycan outer arm fucosylation and therefore plasma N-glycans could be a promising HNF1A-MODY diagnostic marker (Thanabalasingham *et al*., 2013).

1.4.5 The role of exoglycosidases in glycan profiling

The profiling methods described in section 1.4.4 have their limitations, as similar glycan species cannot be discriminated by only their mobility during chromatography or mass. This problem tends to arise for complex samples such as plasma and it can be avoided by subjecting the labelled glycan to sequential exoglycosidase digests, to remove terminal monosaccharides from the non-reducing end of a glycan. The digests are analysed by HPLC methods and the removal of a monosaccharide is detected by a peak shift when compared to an undigested standard (Mulloy *et al*., 2009). Exoglycosidase digests are fast and cost-effective methods of oligosaccharide sequencing and the use of positionally specific exoglycosidases allows the identification of glycosidic linkages, as well as the sugar residue. Several exoglycosidases are sold commercially and include α-sialidases, β-galactosidases, α-mannosidases and α-fucosidases, some of which have versions that are specific to particular glycosidic linkages, with others that are dual specific (Guthrie and Magnelli, 2016).

1.5 The basic principles of X-ray crystallography

1.5.1 An overview

Macromolecular crystallography has progressed since the first protein structure of myoglobin was published in 1958 (Kendrew *et al*., 1958). Since then, a repository for crystal structures, the PDB, has been developed and is shared amongst the scientific community (Berman *et al*., 2000). Crystallography is highly accurate and can reveal structural details at an atomic resolution level and therefore provide an insight into protein function and also structure guided drug discovery (Rupp, 2009). In brief, the determination of a protein structure by X-ray crystallography involves producing high-quality crystals from purified protein and measuring the intensities of X-ray beams that are diffracted by the crystals. The diffraction patterns are then analysed by computer programs to form an image of the crystal content (Rhodes, 1999).

1.5.2 Protein crystallisation

Protein crystals are required for structure determination by X-ray crystallography because diffraction from a single molecule would be too weak. Larger protein crystals diffract a larger proportion of the X-ray beam which increases the signal providing structural information about the protein to a higher resolution. Under certain conditions which favour interactions between individual protein molecules, the proteins will aggregate in a spatially and rotationally symmetric manner, which is referred to as a crystal lattice. The simplest component of the lattice is called the unit cell, which is representative of the whole crystal, as the crystal lattice is composed of repeating unit cells that are translationally repeated in 3 directions to form a 3-dimensional (3D) crystal (Figure 1.11) (Berman *et al*., 2000). However, the smallest portion of a crystal structure is the asymmetric unit, which is the unique structural element within a unit cell that is acted upon by symmetry operators to generate the complete unit cell (Figure 1.11). Symmetry operations that are permitted in crystals of biological molecules are rotation, translation and screw operations but since proteins are inherently asymmetric, due to the chirality of amino acids, mirror planes are not found in unit cells. It is the contents of the asymmetric unit which are deposited in the PDB (rcsb.org - Berman *et al*., 2000) however, it must be emphasised that the asymmetric unit may not represent a whole biologically functional assembly and that techniques such as size exclusion chromatography must be applied to determine the multimeric state of a protein (Berman *et al*., 2000).

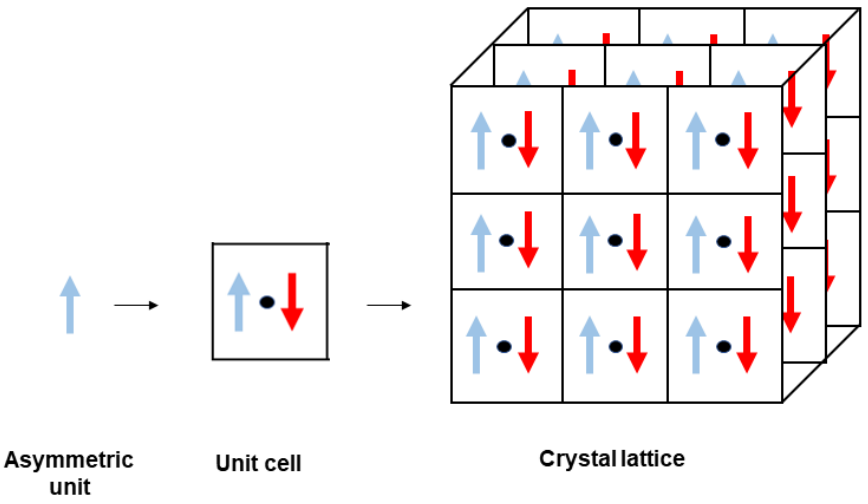


Figure 1.11: The asymmetric unit, unit cell and crystal lattice. A two-fold crystallographic symmetry axis (black oval) has been applied to the asymmetric unit (blue arrow) to produce a second copy (red arrow), both of which represent the unit cell. The unit cell is repeated in three directions to form a 3D crystal. This figure was derived from a figure at rcsb.org (Berman *et al.*, 2000).

In order to produce crystals, the protein solution (in the mg ml^{-1} range) is combined with an aqueous buffer containing a precipitant such as polyethylene glycol (PEG). The initial conditions are altered to induce the protein from solution and whether or not this happens depends upon the properties of the solution such as the pH, temperature and protein concentration. The most common method used for crystal formation is vapour diffusion by hanging or sitting-drops (Rupp, 2009) (discussed in more detail in section 4.2.2). To determine a suitable crystallisation condition for the protein of interest, high through-put methods have been developed to enable researchers to screen a variety of potential crystallisation conditions in a short space of time (Lin, 2018). For example, commercial kits containing preassembled crystallisation conditions are readily available, along with liquid handling robots to lay down the screen. Not only can these robots save time, they can also handle volumes in the nanolitre scale, which is not accessible reproducibly by hand and therefore reduces the volume of purified protein required for the experiment and removes human error (Lin, 2018).

Once a suitable crystal has formed, it can be harvested from the mother liquor using nylon loops. However, protein crystals can be extremely fragile as the intermolecular interactions that hold the protein molecules together tend to be weak interactions such as hydrogen bonding, van der Waals interactions and hydrophobic interactions, so great care needs to be taken during handling. As well as this, there are large channels between the protein molecules that are filled with solvent, which is essential for retaining the crystal structure, as drying destroys crystals (Rhodes, 1999). When looped,

crystals need to remain surrounded by mother liquor to prevent dehydration before being flash cooled in liquid nitrogen (77 K) to provide protection from radiation damage. Prior to storage, crystals are usually passed through the original crystallisation solution that has been supplemented with a cryoprotectant, as the formation of ice is likely to destroy the crystal (Pflugrath, 2015).

1.5.3 X-ray diffraction

Once a suitable protein crystal has been produced, it is exposed to X-ray radiation, usually at a synchrotron light source, where the X-ray beam is extremely intense compared to that generated by electrons colliding with a copper target in a standard X-ray generator in a routine laboratory setting (Berg *et al*., 2011). The crystal is subjected to a narrow beam of X-rays, the majority of which pass through the crystal, whilst some are scattered by their interaction with the electron cloud around the atoms in the crystal. The resulting diffraction pattern is detected by an electronic detector and it consists of a regular array of spots on the detector resulting from diffracted beams that are referred to as reflections. An electron density map of the molecular structure can then be reconstructed by Fourier transform techniques (Rupp, 2009). A Fourier reconstruction from reciprocal diffraction space back to direct molecular space requires the amplitude of the diffracted X-rays and relative phase angles of each reflection. Whilst the structure factor amplitudes are readily accessible (they are proportional to the square root of the measured reflection intensities), the phases are not. The phases are necessary for reconstruction of the electron density of the molecule through Fourier transforms and this lack of phases is referred to as 'the phase problem' (Rupp 2009).

1.5.4 Phase determination

The two main methods used for phase determination are molecular replacement and experimental phasing. Approximately 60 % of the protein structures that have been deposited in the PDB have had the phase problem solved by molecular replacement (Scapin, 2013), which is a technique that exploits initial phases from a structurally similar model. Molecular replacement involves systematic rotation and translation of the known structure model, until a best fit between the calculated diffraction data from the known model and the observed data from the unknown structure is obtained (Rossmann, 1990).

Unfortunately, there is not always a suitable known structure with which a molecular replacement experiment can be performed. Under these circumstances, experimental phasing can be used, with one method being referred to as isomorphous replacement, which involves the attachment of heavy atoms (high atomic number) to the protein in the crystal. The crystal is soaked in a solution containing the salt of a heavy atom, for example mercury, platinum and silver, which bind to one or a few specific

places in the protein without altering the conformation or crystal packing, meaning the crystal should be 'isomorphous' to a native crystal (Rhodes, 1999). The heavy atom crystals should cause a measurable difference in the intensities of some of the reflections in the diffraction pattern when compared to the native crystal. The difference data can be used to locate the heavy atom that is responsible for the difference in intensity, which provides a reference point from which the protein phases can be calculated (Taylor, 2003).

Once the phases have been determined, the Fourier transform can be applied to the amplitudes and phases of every reflection in the diffraction pattern to produce a 3D electron density map (Rupp, 2009). A model of the protein structure is then built into the electron density using computer graphics programs (Emsley *et al.,* 2010).

1.6 Aims and objectives

This project has the ultimate aim to produce a sialidase or sialidases that have specific activity against either the α2,3 or α2,6 sialic acid linkage with a view to their potential use in biopharmaceutical glycan-analysis pipelines. In particular, plasma N-glycans contain both α2,3- and α2,6-linked sialic acid and α2,3-linked sialic acid, with and without antennary fucose, is implicated in many diseases. The removal of α2,6-linked sialic acid would enable the study of the α2,3 glycoprofile by removing a lot of noise in a chromatogram and therefore enhance the diagnostic potential of plasma (Personal communication – Daniel Spencer, Ludger Ltd, Oxford UK). To achieve this, the overall aim of this work was initially set out as:

1) **Fully understand the biochemical properties of existing sialidases.**

In the case of this project, our laboratory team has gathered significant preliminary data on the NanH sialidase from *Tannerella forsythia* and determined it to be a stable, broad specificity enzyme. However, information is limited regarding the specific glycan preferences of NanH (i.e. in the context of larger glycan structures) or inhibitor sensitivity profiles. In addition, the optimal storage conditions of NanH were yet be determined to ensure the enzyme remains highly active after long periods of time.

2) **Determine the structural features that are related to the substrate specificity of sialidases and the whether the specificity can be engineered.**

If one is to target specific residues for mutagenesis that might be key to enzymatic activity or substrate binding, we need to understand how substrates interact with these sialidases. One way to achieve this is by solving the crystal models of NanH, both in the apo-forms but also preferably in complex with substrates to reveal residues key to substrate interactions that might then be targeted.

Chapter 2

Materials and Methods

2.1 Bacterial strains and culture, plasmids and primers

Escherichia coli strains were used for cloning and protein expression, see Table 2.1. The *E. coli* was cultured on Luria-Bertani (LB) agar (Fisher Scientific) or in 2xYT broth (Fisher Scientific). All cultures were supplemented with 50 µg mL^{-1} ampicillin (Fisher Scientific), as the plasmids used contained genes for ampicillin resistance. LB-agar cultures were grown at 37 °C, with the 2xYT broth cultures also requiring agitation at 200-220 rpm.

Strain	Genotype	Supplier
E. coli DH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 endA1 thi-1 hsdR17*	New England Biolabs (NEB)
E. coli BL21 Origami B (DE3)	F- *ompT hsd*S_B(r_B- m_B-) *gal dcm lacY1 ahpC* (DE3) *gor522*:: Tn*10 trxB* (Kan^R, Tet^R)	Novagen (MerckMillipore)
E. coli XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*^{q} *ZΔM15* Tn*10* (Tet^r)]	Agilent Technologies

Table 2.1: *E. coli* strains. Details of the *E. coli* DH5α, *E. coli* Origami B (DE3) and *E. coli* XL1-Blue strains used in this project.

---

pET21a was the plasmid used for cloning and protein expression, the features of which are summarised in Table 2.2.

Plasmid	Description	Phenotype	Description
pET21a	A suitable cloning and protein expression vector with a His-tag for ease of purification by affinity chromatography.	Ampicillin resistant.	G.Stafford strain collection.
NanH-pET21a	Synthesised by Eurofins, codon optimised for *E. coli* based on the *T. forsythia* 92.A2 genome.	Ampicillin resistant.	G.Stafford strain collection. Credit – Dr Andrew Frey.

Table 2.2: Cloning and expression plasmids.

Several pairs of primers (Sigma-Aldrich) were designed to incorporate point mutations into NanH-pET21a by site-directed mutagenesis. Table 2.3 lists the primers used throughout the project, which

includes T7 forward and reverse primers, which were required for deoxyribonucleic acid (DNA) sequencing (GATC Services, Eurofins Genomincs).

Primer	Sequence
T7 Forward	TAATACGACTCACTATAGGG
T7 Reverse	GCTAGTTATTGCTCAGCGG
D237A F	TATAACAATAGCGTGGCGCTGCAAGAACATATTG
D237A R	CAATATGTTCTTGCAGCGCCACGCTATTGTTATA
SDM1 F	GCACTGCCGGAAAGCCATATTTGTATGGCAAG
SDM1 R	CTTGCCATACAAATATGGCTTTCCGGCAGTGC
SDM2 F	CCGAATACCACAGAAGGTTATACCCGTCATCATATTACC
SDM2 R	GGTAATATGATGACGGGTATAACCTTCTGTGGTATTCGG
SDM3 F	CCGAATACCACAGAATATCGTCATCATATTACC
SDM3 R	GGTAATATGATGACGATATTCTGTGGTATTCGG
SDM4 F	CGAATACCACAGAAGGTTATCGTCATCATATTACC
SDM4 R	GGTAATATGATGACGATAACCTTCTGTGGTATTCG
SDM6 F	GATGTGCGTTATAACAATTATGTGGATCTGCAAGAACAT
SDM6 R	ATGTTCTTGCAGATCCACATAATTGTTATAACGCACATC
SDM7 F	GATGTGCGTTATAACAATTATAGCGTGGATCTGCAAGAA
SDM7 R	TTCTTGCAGATCCACGCTATAATTGTTATAACGCACATC
SDM8 F	GGTAGCGCAAGCTTTTATATTCCGGGTCTG
SDM8 R	CAGACCCGGAATATAAAAGCTTGCGCTACC
SDM9 F	CTGCCGGAAAGCTATTGTATGGCAAGC
SDM9 R	GCTTGCCATACAATAGCTTTCCGGCAG
SDM10 F	GAATATGCGTGATTGGCGTGGTGGTAGC
SDM10 R	GCTACCACCACGCCAATCACGCATATTC
SDM11 F	GTTATAACAATAGCGTGGATTGGCAAGAACATATTGAT
SDM11 R	ATCAATATGTTCTTGCCAATCCACGCTATTGTTATAAC
SDM12 F	GGTAATGCACGTTATTGGACCAATAGCATG
SDM12 R	CATGCTATTGGTCCAATAACGTGCATTACC
SDM13 F	GCGTTATAACAATAGCGTGTATCTGCAAGAACAT
SDM13 R	ATGTTCTTGCAGATACACGCTATTGTTATAACGC
SDM14 F	TATAACAATAGCGTGGAACTGCAAGAACATATTG
SDM14 R	CAATATGTTCTTGCAGTTCCACGCTATTGTTATA
SDM15 F	TATAACAATAGCGTGAACCTGCAAGAACATATTG
SDM15 R	CAATATGTTCTTGCAGGTTCACGCTATTGTTATA

Primer	Sequence
SDM16 F	TATAACAATAGCGTGCAGCTGCAAGAACATATTG
SDM16 R	CAATATGTTCTTGCAGCTGCACGCTATTGTTATA
SDM17 F	TATAACAATAGCGTGAGCCTGCAAGAACATATTG
SDM17 R	CAATATGTTCTTGCAGGCTCACGCTATTGTTATA
SDM18 F	GATGGTTGGGGTGCGAGCTGTCTGAGC
SDM18 R	GCTCAGACAGCTCGCACCCCAACCATC
SDM19 F	GATGGTTGGGGTCAGAGCTGTCTGAGC
SDM19 R	GCTCAGACAGCTCTGACCCCAACCATC
SDM20 F	GATGGTTGGGGTAGCAGCTGTCTGAGC
SDM20 R	GCTCAGACAGCTGCTACCCCAACCATC
SDM21 F	GATGGTTGGGGTACCAGCTGTCTGAGC
SDM21 R	GCTCAGACAGCTGGTACCCCAACCATC
SDM22 F	GATGGTTGGGGTTTTAGCTGTCTGAGC
SDM22 R	GCTCAGACAGCTAAAACCCCAACCATC
SDM23 F	CGTTATAACAATAGCCAGGATCTGCAAGAACAT
SDM23 R	ATGTTCTTGCAGATCCTGGCTATTGTTATAACG
SDM24 F	CGTTATAACAATAGCTATGATCTGCAAGAACAT
SDM24 R	ATGTTCTTGCAGATCATAGCTATTGTTATAACG
SDM25 F	GGTATGGGTAATGCAGCGGCCTGGACCAATAGC
SDM25 R	GCTATTGGTCCAGGCCGCTGCATTACCCATACC

Table 2.3: Primers used throughout the project. The black underline highlights the codon of newly substituted residues, whereas the red underline shows the codon of residues to be inserted.

2.2 DNA extraction – plasmid DNA

E. coli which had been transformed with plasmid DNA was cultured on LB-agar supplemented with 50 µg mL^{-1} ampicillin and incubated at 37 °C, overnight (O/N). An *E. coli* colony was cultured in 2xYT broth as previously described in section 2.1 before undergoing plasmid extraction using an ISOLATE II Plasmid Miniprep kit (Bioline), according to the manufacturer's protocol.

2.3 Site-directed mutagenesis

Wild-type (WT) NanH plasmid DNA was used as the template for the site-directed mutagenesis (SDM) reactions carried out in the project, unless a mutant already produced by SDM was to act as a

template. SDM reactions were set up as seen in Table 2.4, using reagents from a QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) and primers designed to incorporate specific mutations (Sigma-Aldrich) (see Table 2.3), and subjected to the thermocycling conditions shown in Table 2.5. The reactions were then placed on ice for 2 min before the addition on 1 µL DpnI (stock at 10 U/µl) and incubated at 37 °C for 1 h to digest any template plasmid DNA that remained after the cycling.

During optimisation of the SDM protocol provided by the manufacturer, it was discovered that it was necessary to concentrate the mutant plasmid DNA to produce enough DNA for transformation into *E. coli* DH5α. The DpnI treated mutant plasmid DNA was precipitated and concentrated by adding 1/10 of the total volume 3 M sodium acetate pH 5.2 and 2 total volumes worth 100 % ethanol. The mutant plasmid DNA mix was incubated at -20 °C for 1 h before centrifugation at 11,000 x g, for 5 min at 4 °C. The supernatant was removed, and the DNA pellet was resuspended in 20 µL of 70 % ethanol. The resuspension and centrifugation step were repeated, and the supernatant removed once more. The DNA pellet was left for 30 min to air dry. The DNA pellet was resuspended in 4 µL nuclease free H_2O and the 4 µL of resuspension was immediately transformed into *E. coli* XL1-Blue supercompetent cells as in (section 2.7.1).

Component	50 µL reaction
10 x reaction buffer	1 x (5 µL)
Plasmid DNA template	5-50 ng
Oligonucleotide primer (Forward)	125 ng
Oligonucleotide primer (Reverse)	125 ng
dNTP mix	1 µL
Nuclease free H_2O	Up to 50 µL
Then add 2.5 U *PfuUltra* HF DNA Polymerase	

Table 2.4: Reagents, volumes and amounts used in SDM reactions.

Cycles	Temperature (°C)	Time
1	95	30 s
18	95	30 s
18	55	1 min
18	68	1 min/kb plasmid length

Table 2.5: Thermocycling used in SDM according to the manufacturer's protocol.

2.4 DreamTaq colony PCR

Colonies of transformed *E. coli* XL1-Blue (section 2.7.1) were patched onto LB-agar containing 50 µg mL^{-1} ampicillin and incubated O/N at 37 °C. DreamTaq colony PCR (Thermo Fisher) was used to screen the patched colonies to determine whether plasmid DNA was present in *E. coli* after transformation. Reactions were set up as described in Table 2.6 and to each reaction a small swab of a patched colony was added. Positive controls for expression vector pET21a involved using a swab of a colony of a construct of that vector where the size of the DNA insert is known. The negative controls were always simply one reaction aliquot without *E. coli*. Primers used depended upon the vector the insert was cloned into and in this case T7 primers were used for pET21a Table 2.3. The reactions were then subjected to the PCR cycle as shown in Table 2.7 before analysis by agarose gel electrophoresis (section 2.5).

Component	20 µL reaction
2 x DreamTaq Green	1 x (10 µL)
Forward primer (vector dependent)	0.5 µM
Reverse primer (vector dependent)	0.5 µM
Nuclease free H_2O	Up to 20 µL

Table 2.6: DreamTaq colony PCR components.

Temperature (°C)	Time
95 (Initial denaturation)	5 min
95 (Denaturation)	30 s
50 (Annealing)	30 s
72 (Extension)	1 min per kb
72 (Final extension)	7 min

Table 2.7: Colony PCR cycle used in this project, adapted from the protocol provided by ThermoFisher.

2.5 Agarose gel electrophoresis

DNA analysis was performed using 1 % ($^{w}/_{v}$) agarose gels. Gels were made by adding agarose powder (Sigma-Aldrich) to 1 X TAE buffer (TAE 50 X stock: Tris Base (242 g), Acetic acid (57.1 mL), 0.5 M EDTA pH 8.0 (100 mL) and top up to 1 L with dH_2O), along with 5 µg mL^{-1} ethidium bromide. The mixture was placed in a microwave until all the agarose had dissolved before being poured into a prepared DNA gel mould. Agarose gels were run in either BioRad minisub-cell GT systems or BioRad sub-cell GT systems (depending on the size of the gel) in 1 X TAE buffer. The DNA was visualised on a UV transilluminator-GBOX (Syngene) with Genesys (Syngene) software.

2.6 DNA sequencing

Once colony PCR confirmed presence of plasmid DNA in *E. coli*, plasmids were extracted as in section 2.2 and sent for Sanger sequencing by GATC Services (Eurofins Genomics) in the forward and reverse direction as required. Samples contained 5 µL of plasmid DNA and 5 µL of the appropriate T7 primer (Table 2.3).

2.7 Bacterial transformation

2.7.1 Transformation of chemically competent *E. coli* XL1-Blue supercompetent cells

E. coli XL1-Blue supercompetent cells (Agilent Technologies) were used for transformation of NanH variant plasmid DNA (section 2.3) according to the manufacturer's protocol. 50 µL aliquots of XL1-Blue cells were thawed on ice for 10 min and transferred to 13 mL round bottomed tubes. 4 µL of concentrated plasmid DNA (section 2.3) was added to the aliquot of cells and incubated on ice for 30 min before being subjected to a 42 °C heat shock for 45 s. The cells were returned to ice for 2 min before the addition of 0.5 mL preheated SOC media (42 °C) and incubation at 37 °C, 225 rpm for 1 h. 250 µL of the transformed *E. coli* cells were plated onto two separate LB-agar plates supplemented with 50 µg mL^{-1} ampicillin and incubated O/N at 37 °C.

2.7.2 Preparation of chemically competent *E. coli* BL21 Origami

E. coli Origami was cultured in 0.5 L 2xYT media until OD_{600} 0.6 was reached. Cells were pelleted by centrifugation at 8671 x g for 20 min before being resuspended in 50 mL 0.1 M ice cold $CaCl_2$ and incubated on ice for 1 h. Cells were pelleted at 1976 x g for 15 min before being resuspended in 10 mL ice cold $CaCl_2$. Cells were pelleted using the same method as previous, but this time resuspended in 5 mL 0.1 M ice cold $CaCl_2$, 10 % glycerol and again pelleted as before. Cells were resuspended in 2 mL 0.1 M ice cold $CaCl_2$, 10 % glycerol, separated into 75 µL aliquots and either immediately transformed with plasmid DNA or stored at -80 °C.

2.7.3 Transformation of chemically competent *E. coli* BL21 Origami

200 µL aliquots of Origami cells were thawed on ice for 10 min before the addition of 5 µL (1 pg-100 ng) plasmid DNA. The cells were incubated on ice for 30 min, followed by a 90 second heat shock at 42 °C, before returning to ice for a further 2 min. 1 mL of SOC media was added to the cells, which were then incubated for 1 h, 220 rpm at 37 °C before being plated on LB-agar. 50 µL of transformed *E. coli* and 50 µL of recovered *E. coli* cell suspension were plated onto LB-agar, which was supplemented with 50 µg mL^{-1} ampicillin and incubated O/N at 37 °C.

2.8 Protein production and purification

2.8.1 Small-scale protein overexpression trials

E. coli O/N cultures, as described in section 2.1, were used to seed 100 mL 2xYT broth supplemented with 50 µg mL^{-1} ampicillin. The cultures were incubated at 37 °C, 220 rpm until mid-log phase was reached (OD_{600} 0.6). Once approximately OD_{600} 0.6 was reached, the cultures were induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) (Calbiochem). The cultures were incubated O/N at 25 °C and 200 rpm. 1 mL samples of cells were collected prior to IPTG induction, 4 h after induction and 20 h after induction. The OD_{600} was recorded for these cells before they were spun down for 2 min at 13,000 x g. Cell pellets were resuspended in 2X sodium dodecyl sulphate (SDS) loading buffer (Table 2.9) before analysis by SDS-polyacrylamide gel electrophoresis (PAGE) (section 2.9).

2.8.2 Small-scale purification

A small-scale purification method was designed to clarify protein expression by removing cell homogenate and impurities seen by SDS-PAGE after overexpression. 20 h post induction the 100 mL cell cultures (section 2.8.1) were pelleted at 6370 x g and resuspended in 2 mL 50 mM sodium phosphate, 0.5 M NaCl pH 7.4, 20 mM imidazole (resuspension buffer). The cells were subjected to 3 x 3 s of sonication at 16 microns amplitude, with 30 s between each pulse where the samples remained on ice. The cell lysates were spun at 6000 x g for 2 min to remove the cell debris. The supernatant was spun at 13000 x g for 2 min to remove the insoluble fraction, leaving the cell free extract (supernatant) ready for purification.

200 µL NiNTA nickel resin (Qiagen) was loaded onto Micro Bio-Spin chromatography columns (BioRad) and allowed to settle for 10 min before centrifugation at 1000 x g for 30 s. NiNTA resin (Qiagen) was equilibrated by loading 1 mL of 50 mM sodium phosphate, 0.5 M NaCl pH 7.4, 20 mM imidazole (equilibration buffer) followed by centrifugation at 1000 x g for 30 s. The cell free extract (approximately 0.5-1 mL) was loaded onto the resin and incubated at room temperature for 10 min before centrifugation at 1000 x g for 30 s. The resin was washed in 1 mL of 50 mM sodium phosphate,

0.5 M NaCl pH 7.4, 50 mM imidazole (wash buffer) and spun at 1000 x g for 30 s. The wash step was repeated before 150 µL of 50 mM sodium phosphate, 0.5 M NaCl, pH 7.4 250 mM imidazole (elution buffer) was loaded onto the resin and spun at 1000 x g for 30 s to elute the pure protein. Samples were taken from each step above and mixed with 2X SDS loading buffer (Table 2.9) for analysis by SDS-PAGE (section 2.9). To remove imidazole and decrease the salt concentration, the pure protein was subjected to buffer exchange. The pure protein (150 µL) was loaded onto Amicon Ultra – 0.5 mL Centrifugal Filter Units with 10 kDa molecular weight (MW) cut off (Merck Millipore). The volume was topped up to 0.5 mL with 50 mM sodium phosphate, 200 mM NaCl pH 7.4 and centrifuged at 13,000 x g for approximately 7 min. The flow through (approximately 0.45 mL) was discarded and the filter unit was topped up to 0.5 mL using 50 mM sodium phosphate, 200 mM NaCl pH 7.4 and centrifuged at 13,000 x g for approximately 7 min. The addition of low salt phosphate buffer and subsequent centrifugation were repeated 8 times.

2.8.3 Large-scale protein expression

Large-scale protein expressions were carried out when a high yield of protein was required. O/N *E. coli* cultures (see section 2.1) were used to seed 0.5 L 2xYT broth in a 1:50 cell to broth ratio. Cultures were grown at 37 ^{o}C, 200 rpm until OD_{600} 0.6 (mid log phase) was reached. The cells were induced with 1 mM IPTG before being left to incubate at 25 ^{o}C, 200 rpm O/N. After approximately 20 h of incubation, the *E. coli* was pelleted at 8671 x g for 20 min. The cell pellet was resuspended in 20 mL resuspension buffer (section 2.8.2), with the total volume reaching approximately 25 mL, and either immediately used for purification (section 2.8.4) or stored at -20 ^{o}C until required.

2.8.4 Large-scale protein purification

A protease inhibitor tablet (cOmplete ULTRA Tablets, EDTA-free EASYpack, Roche) and 10 µg mL^{-1} DNase (Sigma Aldrich) were added to the resuspended *E. coli* cells produced as in section 2.8.3. Cells were lysed by passing through a French pressure cell three times at 1000 psi, followed by centrifugation of the lysate at 15000 x g for 20 min. The supernatant was clarified by centrifugation at 15000 x g for a further 40 min, resulting in the separation of cell free extract from insoluble material. The cell free extract was purified on 5 mL Ni^{2+} columns (HisTrap FF, GE Healthcare Life Sciences) using a peristaltic pump, with flow rates no higher than 5 mL min^{-1} to perform the protein loading, the column washes and the protein elution steps of the purification. Initially the column was equilibrated with 5 column volumes of equilibration buffer (section 2.8.2) before approximately 25 mL cell free extract sample was loaded. The column was washed with 10 column volumes wash buffer and the protein then eluted in 20x 2 mL fractions using 50 mM sodium phosphate, 0.5 M NaCl, pH 7.4, 300 mM imidazole (elution buffer). An AKTA Prime plus purification system (GE Healthcare) was

occasionally used for the elution step. His-tagged proteins were eluted on a 0-0.5 M imidazole gradient at a flow rate of 1 mL min^{-1} over 15 1.5 mL fractions. The protein content and purity of the fractions was determined by SDS-PAGE (section 2.9). Fractions with over 95 % purity were pooled and dialysed against 50 mM sodium phosphate, 200 mM NaCl, pH 7.4 in membrane with a 14 kDa MW cut off (Medicell Membranes Ltd) to remove imidazole and reduce the salt concentration.

2.9 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins by their electrophoretic mobility based on their molecular weight. The size of the protein to be analysed determines the acrylamide content in the gel. Thermo Fisher (2012) suggested that for proteins with a MW between 20 and 150 kDa, a 12 % gel is appropriate. Table 2.8 shows the reagents used in the resolve and stack regions of a 12 % gel. Once set, the gel was placed in a BioRad SDS-PAGE tank (mini PROTEAN Tetra System) and the tank was filled with 1X SDS running buffer (made from 10X stock, Table 2.9). 2X SDS loading buffer was added to the samples (Table 2.9), which were then boiled at 95 ^{o}C for 5 min before loading them onto the gel, along with a pre-stained MW marker (EZ-run protein ladder – Thermo Fisher). Electrophoresis was carried out at 200 V for approximately 1 h before staining the gel with InstantBlue Coomassie Protein stain (Expedeon).

Reagent	Resolve (12 %)	Stack
Lower Tris	2.5 mL	-
Upper Tris	-	2.1 mL
Acrylamide	3 mL	4.725 mL
dH_2O (deionised water)	4.3 mL	0.975 mL
TEMED	5 µL	17 µL
APS	350 µL	100 µL

Table 2.8: Reagents used to make 12 % SDS gels. TEMED–Tetramethylethylenediamine, APS–Ammonium persulphate. Instructions on how to make up lower and upper Tris are found in Table 2.9.

Buffer	Reagents
Upper Tris	6.1 g Tris Base (Thermo Fisher), 0.4 g SDS (Thermo Fisher). Dissolved in dH_2O and adjusted to pH 6.8 in a total volume of 100 mL.
Lower Tris	18.2 g Tris Base, 0.4 g SDS Dissolved in dH_2O and adjusted to pH 8.8 in a total volume of 100 mL.
10x SDS-PAGE running buffer	30 g Tris Base, 144 g glycine (Thermo Fisher), 10 g SDS was made up to a volume of 1000 mL with dH_2O.
2x SDS loading buffer	40 mL 100 mM Tris-HCl (Thermo Fisher), 10 mL 20 % glycerol (Thermo Fisher), 1 g SDS, 0.1 g Bromophenol Blue (Thermo Fisher), 200 mM DTT (Thermo Fisher) (50 mL total volume).

Table 2.9: Buffers used in SDS-PAGE. DTT = Dithiothreitol.

2.10 Protein concentration determination – bicinchoninic assay

Bicinchoninic assays (BCA assays) were used to determine protein concentrations using the Pierce BCA assay kit (Thermo Fisher) according to the manufacturer's protocol. Bovine Serum Albumin (BSA) was used to generate a standard curve of known protein concentrations. BSA standards were diluted in the same buffer as the protein of interest.

2.11 Biochemical characterisation of sialidases

2.11.1 Size-exclusion chromatography

After affinity purification on a Ni^{2+} column, dialysed WT NanH was concentrated using a Viva Spin (GE Healthcare Sciences) to approximately 3 mg mL^{-1}. 2 mL of concentrated NanH was loaded onto a 16x600 HiLoad Superdex column equilibrated in 0.5 M NaCl, 50 mM Tris, pH 8.0. The gel filtration involved the collection of 2 mL fractions at a flow rate of 1.5 mL min^{-1}. The fractions that formed the main peak on the chromatogram were combined and analysed by SDS-PAGE. The apparent MW of the peak was determined by calculating the Kav and applying this to a standard curve of proteins of known MW's.

2.11.2 Native polyacrylamide gel electrophoresis

3-12 % NativePAGE Bis Tris Gels (Novex) were used to study the composition and structure of native NanH. The gel was placed into a XCell SureLock Mini-Cell tank and the wells filled with 1X NativePAGE Blue Cathode Buffer (Novex), before loading a native page marker (Novex) and protein samples which were prepared in NativePAGE sample buffer (Novex). 200 mL 1X NativePAGE Dark Blue Cathode Buffer

and 550 mL Anode Buffer (Novex) were loaded into the upper cathode chamber and lower anode chamber of the tank respectively, before electrophoresis was carried out at 150 V.

2.11.3 Size exclusion chromatography-multiangle laser light scattering (SEC-MALLS)

Native NanH was purified as described in section 2.8.4 before analysis by size-exclusion chromatography-multiangle laser light scattering (SEC-MALLS). NanH (0.5 and 3.0 mg mL^{-1}) was analysed on a Superdex S200 10/300 column (GE Healthcare) in 50 mM sodium phosphate, 200 mM NaCl pH 7.4. This was carried out at approximately 20 ^{o}C and conducted on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu HPLC system (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler). Data were analysed using the Astra V software (Wyatt Technologies). The experiment was performed by the Molecular Interactions Lab, University of York, UK.

2.11.4 Reaction kinetics of sialidases with 4-Methylumbelliferyl-N-acetyl-α-D-neuraminic acid

4-Methylumbelliferyl-N-acetyl-a-D-neuraminic acid (MU-NANA) (Carbosynth) was used to determine the activity of purified WT NanH and NanH variants. Increasing concentrations of MU-NANA were exposed to 2.5-100 nM of sialidase in 50 mM sodium phosphate, 200 mM NaCl, pH 7.4. The reactions were incubated for up to 3 min at room temperature. At 1 min intervals, 50 µL of each reaction was removed and quenched by the addition of 75 µL 100 mM sodium carbonate, pH 10.5. Fluorescence of the 4-Methylumbelliferyl (4-MU) released at each time point was measured via excitation at 350 nm and emission at 450 nm, using a Tecan Infinite M200. Unless specified otherwise, 3 technical replicates were taken to calculate the reaction velocity for each substrate concentration. The results were found to be replicable in biological repeats (not shown). Application of a 4-MU standard curve allowed quantification of the 4-MU release rate. Rate of reaction was determined by first finding the initial rate of reaction, which occurred between 1 and 3 min for NanH and mutants. Experiments varied as to whether the 1 and 2 min or 2 and 3-min time point was used to determine initial rate. Initial rate, also referred to as velocity (V_0, 4-MU release µmol min^{-1} mg^{-1}), was plotted against MU-NANA concentration (µM) and the K_M and V_{max} were determined in GraphPad PRISM 7.03 software by fitting the following nonlinear equation using the least-squares method:

$$Y = V_{max} \times \frac{X}{K_M + X}$$

X refers to the substrate concentration, Y is the enzyme velocity, V_{max} is the maximal enzyme velocity and K_M is the concentration at half maximal velocity when the enzyme becomes saturated. The k_{cat}

values were determined in GraphPad PRISM 7.03 software by fitting the following nonlinear equation using the least-squares method:

$$Y = E_t \times k_{cat} \times \frac{X}{K_M + X}$$

X refers to the substrate concentration, Y the enzyme velocity, E_t is the concentration of enzyme active sites, k_{cat} is the turnover number and K_M is the concentration at half maximal velocity when the enzyme becomes saturated.

2.11.5 Sialidase inhibition assays

2.5 nM NanH was incubated with 0.1 mM MU-NANA at room temperature in 50 mM sodium phosphate, 200 mM NaCl, pH 7.4, in the presence of increasing concentrations of inhibitor. Inhibitors studied in this project were N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid (DANA) (Carbosynth) and oseltamivir acid (Carbosynth). The reactions were incubated for up to 3 min at room temperature. 50 µL reactions were removed at 1 min intervals and quenched by the addition of 75 µL 100 mM sodium carbonate, pH 10.5. Fluorescence of 4-MU release at each time point was measured via excitation at 350 nm and emission at 450 nm, using a Tecan Infinite M200. 3 technical replicates were taken to calculate the reaction velocity for each inhibitor concentration. Sialidase inhibition was expressed as the percentage change in fluorescence with a given concentration of inhibitor compared to fluorescence in the absence of inhibitor, as described by the formula below, where Top is 100 % (activity), Bottom is the percent of activity at maximum inhibition and the IC_{50} is the inhibitor concentration which reduces activity by 50 %.

$$Y = \frac{Bottom + (Top - Bottom)}{(1 + \left(\frac{X}{IC_{50}}\right))}$$

This equation was fitted to the data using GraphPad PRISM 7.03 [inhibitor] vs. response (three parameters) nonlinear fit (by least-squares method).

2.11.6 Quantification of sialic acid release from glycans (thiobarbituric acid assay)

Sialidase assays were carried out with the substrates 3 and 6-sialyllactose (3- and 6-SL) (Carbosynth), which have the α2,3 and α2,6 linkages, respectively, and also 3-Sialyl-Lewis A and 3-Sialyl-Lewis X (SLe^A or SLe^X) (Carbosynth), which both contain α2,3-linked sialic acid. WT NanH and NanH variants were incubated with these substrates and sialic acid release was quantified. Sialic acid is not readily quantifiable by fluorescence like 4-MU from MU-NANA (section 2.11.4). The thiobarbituric acid assay

(TBA assay) by Aminoff (Aminoff, 1961), allows quantification of free sialic acid by labelling the sialic acid with a thiol group which enables absorbance measurements to be taken.

50 µL reactions were set up to include 50-100 nM sialidase and increasing concentrations of 3- and 6-SL in 50 mM sodium phosphate, 200 mM NaCl, pH 7.4. Reactions were incubated at 37 °C and halted at 1, 3 and 5 min after initiation through the addition of 25 µL 20 mM sodium periodate (Sigma-Aldrich) in 60 mM H_2SO_4, which oxidised the sialic acid. The reactions were left to oxidise for 30 min at 37 °C. Oxidation was then halted by the addition of 20 µL 2 % sodium meta arsenite (Sigma-Aldrich) in 0.5 M HCl. The oxidised sialic acid was labelled with a thiol group through the addition of 100 µL 100 mM thiobarbituric acid pH 9.0 (Sigma-Aldrich), and incubated at 95 °C for 7.5 min. 100 µL of each reaction was then transferred to a clear bottomed 96-well plate and the absorbance at 549 nm was measured using a TECAN Infinite M200. Unless specified otherwise, 3 technical replicates were taken to calculate the reaction velocity for each substrate concentration.

The initial rate was determined at all concentrations tested and Neu5Ac release (V_0, Neu5Ac release µmol min^{-1} mg^{-1}) was plotted against ligand concentration (mM) and the K_m and V_{max} were determined in GraphPad PRISM 7.03 software by fitting the following nonlinear equation using the least-squares method:

$$Y = V_{max} \times \frac{X}{K_M + X}$$

X refers to the substrate concentration, Y is the enzyme velocity, V_{max} is the maximal enzyme velocity and K_M is the concentration at half maximal velocity when the enzyme becomes saturated. The k_{cat} values were determined in GraphPad PRISM 7.03 software by fitting the following nonlinear equation using the least-squares method:

$$Y = E_t \times k_{cat} \times \frac{X}{K_M + X}$$

X refers to the substrate concentration, Y the enzyme velocity, E_t is the concentration of enzyme active sites, k_{cat} is the turnover number and K_M is the concentration at half maximal velocity when the enzyme becomes saturated.

2.11.7 Protein-ligand binding assay

To determine the ability of NanH mutants to bind to 3- and 6-SL, a fluorescence based tryptophan quenching assay was used. The method relies upon tryptophan residues present in the protein structure that fluoresce between 300-350 nm on excitation at 280 nm. Ligand binding results in a

conformational change that either exposes or shields the tryptophan from solvent, resulting in increased or decreased excitation, respectively.

5 µL of 3- or 6-SL was pipetted into a UV clear 96-well plate (Greiner Bio-One) at 40x concentration. Protein was diluted to 0.1 µM in 50 mM sodium phosphate, 200 mM NaCl pH 7.4 and 195 µL of this was added to the ligand in the 96-well plate. The plate was then incubated at 25 °C for 5 min in a Tecan infinite M200 plate reader, before the fluorescence was measured using an excitation wavelength of 280 nm and the emission was scanned between 300-380 nm. The change in protein fluorescence (of 3 technical replicates) due to an increase in concentration of ligand was plotted as a percentage change relative to a protein only baseline condition. The B_{max} and K_d of the ligands were determined using the following equation:

$$Y = B_{max} \times \frac{X}{(K_d + X)}$$

X refers to the final concentration of ligand; Y is the change in protein fluorescence as a percentage compared to the protein only condition and B_{max} is the maximum number of binding sites. This equation was fitted to the data using GraphPad PRISM 7.03 one site specific binding nonlinear fit (by least-squares method).

2.11.8 NanH storage optimisation

NanH was stored at 1 µM in 20 µL aliquots at both 4 and -20 °C, with varying concentration of glycerol (0, 5, 10, 15, 20 %) and either 200 mM or 500 mM NaCl, as shown in Table 2.10. An aliquot of NanH from each condition was thawed after 1, 2, 3, 4, 5, 6 and 9 months of storage (28 day intervals) and diluted to 2.5 nM before performing a MU-NANA activity assay (section 2.11.4) to compare to the activity at day 0 for this specific protein batch before storage.

4 °C		-20 °C	
200 mM NaCl	500 mM NaCl	200 mM NaCl	500 mM NaCl
0 % glycerol	0 % glycerol	0 % glycerol	0 % glycerol
5 % glycerol	5 % glycerol	5 % glycerol	5 % glycerol
10 % glycerol	10 % glycerol	10 % glycerol	10 % glycerol
15 % glycerol	15 % glycerol	15 % glycerol	15 % glycerol
20 % glycerol	20 % glycerol	20 % glycerol	20 % glycerol

Table 2.10: Conditions NanH was stored at for up to 9 months.

In addition, 20 µL aliquots of 1 µM NanH in 50 mM sodium phosphate, 200 mM NaCl, pH 7.4 were stored at 37 and 42 ^{o}C for both 24 and 48 h to determine how a temperature increase would affect the stability of the enzyme. At the end of the allocated time periods, NanH was diluted to 2.5 nM before performing a MU-NANA activity assay (section 2.11.4) and comparing the data to the activity of the enzyme stored at 4 ^{o}C, that was assayed at the same time.

2.12 Analysis of sialidase activity with complex glycans

The experiments performed using the methods in this subsection were performed at Ludger Ltd, Oxford.

2.12.1 PNGaseF in solution digest and clean up

N-glycans were removed from plasma (human serum) in a PNGaseF digest using reagents and consumables from Ludger Ltd kits. dH_2O was added to 100 µg of glycoprotein to make a total volume of 17.5 µL. 5 µL 5X reaction buffer and 1.25 µL denaturation solution were added to the glycoprotein and the solution was vortexed and briefly centrifuged, followed by an incubation at 100 ^{o}C (± 5^{o}C) for 5 to 10 min. N-glycans were released by the addition of 1.25 µL Triton X-100 and 1 µL PNGaseF to the glycoproteins, followed by a 3 h incubation at 37 ^{o}C. Released N-glycans were converted to aldolases through the addition of 5 µL of 5 % formic acid with an incubation period at room temperature for 30 to 40 min.

Protein and acid were immediately removed from the N-glycans using a protein binding 96-well plate. 100 µL of methanol was applied to wet the membranes in the wells. A vacuum between -0.05 and -0.1 bar was applied until all the liquid had passed through the well. 300 µL of dH_2O was used to wash the membrane and a vacuum was applied using the same method as above. The N-glycans were pipetted into the wells of the protein binding plate. Each N-glycan sample vial was washed with 100 µL dH_2O and added to the plate. The plate was centrifuged at 600 rcf for 3 min to allow the sample to pass through the membrane. A further 100 µL dH_2O was added to the protein binding plate and the centrifugation was repeated. The eluted glycans were dried in a vacuum centrifuge for 6 h.

2.12.2 Procainamide labelling of glycans

N-glycans digested in section 2.12.1, along with several other glycan standards (3'-Sialyl-Lewis-X, FA2G2S2, and mannose mix) were labelled with procainamide through the use of the Ludger procainamide labelling kit (LT-KPROC-VP24 kit). 150 µL of dimethyl sulfoxide (DMSO)-acetic acid (30 % acetic acid in DMSO) was added to a vial of procainamide dye and pipette mixed until dissolved. The total volume of dye mixture was added to a vial of 2-picoline borane (reductant) and mixed until dissolved, to make the final labelling reagent. Glycans were diluted to 2 µg in a total volume of 10 µL

with dH_2O. 10 µL labelling reagent was added to each glycan sample and incubated at 65 °C (± 2 °C) for 1 h and then cooled to room temperature.

2.12.3 O-glycan clean up

Labelled O-glycans (3'-Sialyl-Lewis-X) were cleaned using LudgerClean S cartridges to remove excess labelling mix. LudgerClean S cartridge membranes were primed with 1 mL H_2O, 5 mL 30 % acetic acid and 1 mL acetonitrile. The labelled O-glycans were spotted onto the membranes, whilst still damp with acetonitrile. The sample was left to adsorb onto the membrane for 20 min and the sample vial and pipette tip used to load the sample was rinsed with 100 µL acetonitrile and applied to the membrane. The cartridges were washed with 1 mL acetonitrile followed by 2 x 5 mL acetonitrile and the glycans were eluted with 2 x 0.5 mL water. Eluted glycans were dried in a vacuum centrifuge for 9 h and dissolved in 200 µL dH_2O.

2.12.4 N-glycan clean up

N-glycans (plasma, FA2G2S2 and ManMix) were cleaned using a LC-PROC-96 plate to remove excess labelling mix. The membranes of an LC-PROC-96 plate were primed with 200 µL 70 % ethanol, 200 µL water and 200 µL acetonitrile. Between the addition of each liquid, a vacuum between -0.05 and -0.3 bar was applied until all the liquid had passed through the membranes. 230 µL acetonitrile was added to the procainamide labelled N-glycans before transferring to the primed membranes. The samples were drained by gravity for 10 min before a minimum vacuum was applied to allow the samples to pass through the membranes. The membranes were washed 3X by the addition of 200 µL acetonitrile to each well, which were then drained as above. The following elution step was performed twice, to allow the glycans to elute slowly. 100 µL dH_2O was applied to the membranes and minimum vacuum was briefly applied to elute the glycans before allowing the remainder of the glycans to elute by gravity for 10 min.

2.12.5 Glycan sialidase digestions

Glycan sialidase digests were performed in a microtiter plate followed by the purification of glycans as in section 2.12.1. Approximately 0.07-0.2 µg purified procainamide labelled glycans were incubated with 10 nM NanH sialidase in a final volume of 10 µl in a 250 mM sodium phosphate, pH 6.0 buffer for 18 h at 37 °C. Preparation of a protein binding plate involved wetting the membranes with 100 µL methanol and washing the membranes with 300 µL dH_2O as in section 2.12.1. Digested samples were pipetted onto the membranes and the sample tube was rinsed with 100 µL dH_2O which was also pipetted onto the membrane. The plate was placed on a shaker at room temperature (RT) for 45 min. Glycans were eluted by centrifugation of the plate at 800 rcf for 3 min. 100 µL dH_2O was pipetted onto

each membrane and the plate was centrifuged as above. The eluted glycans (approximately 200 µL per sample) were transferred to a microtiter plate and dried in a vacuum centrifuge for 8.5 h before being reconstituted with dH_2O to the desired volume for high-performance liquid chromatography (HPLC).

2.12.6 Accelerated degradation study

180 µL aliquots of NanH at a concentration of 4.4 µM were incubated for both 24 and 48 h at -20 °C, RT, 37 °C and 55 °C. A sialidase digest was performed with NanH samples being diluted to 10 nM in a total volume of 10 µL with 3'-Sialyl-Lewis-X (as in section 2.12.5). Aliquots of NanH stored at -20 °C were subjected to either 2 or 5 cycles of freeze-thawing over 48 h before use in a sialidase digest. NanH aliquots were left to thaw for 1-2 h, however the time allowed for freezing the aliquots varied between 1 and 18 h. After digestion the samples were purified to remove protein as in section 2.12.5.

2.12.7 Analysis of sialidase digests

Sialidase digests were analysed using hydrophilic interaction (HILIC) mode ultrahigh-performance liquid chromatography (UHPLC) (Dionex Ultimate 3000 UHPLC) coupled to electrospray ionisation - mass spectrometry (ESI-MS) (Bruker Amazon Speed ETD). Undigested forms of the glycans were also analysed by UHPLC ESI-MS.

2.13 Structural characterisation of sialidases

2.13.1 Crystallisation screens

Protein was dialysed into 10 mM Tris, 100 mM NaCl, pH7.5 using a membrane with a 14 kDa MW cut off (Spectrum Labs). Protein was concentrated to at least 20 mg mL^{-1} using a Viva Spin concentrator with a 10 kDa MW cut off (GE Healthcare Sciences). A Matrix Hydra II PlusOne robot (Thermo Scientific) was used to perform the initial screening for protein crystallisation. For each protein, three 96 crystallisation condition screens referred to as the PACT, JCSG and ProPlex suites (Molecular Dimensions) (Appendix 2), were laid down as sitting drop trials on 96-well 2-drop plates (Molecular Dimensions). Each drop contained 200 nL native protein + 200 nL reservoir solution comprised of one of the conditions from a suite. The protein crystallisation trial screens were stored at 17 °C. This Hydra robot was updated to a Mosquito (TTP Labtech) during this project. This new robot enabled the same screens to be laid down at a faster pace than with the Hydra. When using the Mosquito, each sitting drop contained 100 nL protein + 100 nL reservoir solution in this project.

2.13.2 Crystal optimisations

Hits from the initial robot screens were optimised using the hanging drop technique, with siliconized cover slips, on a larger scale to produce a higher concentration of larger crystals. Optimisations were performed in 24 well plates with a reservoir volume of 500 µl and a 2 µL: 2 µL ratio of protein to reservoir solution. The protein concentrations used were identical to the concentration in the initial screening (section 2.13.1). A Formulator Robot (Formulatrix) became available to design customised optimisation screens in 96-well plates. Once the reservoir solutions had been dispensed into the 96-well plate by the Formulator, the Mosquito was used to dispense the protein of interest in the same method discussed in section 2.13.1.

2.13.3 Ligand incorporation – soaking

Soaking involved producing crystals of the apo protein and soaking them in a compound of interest. The apo crystal was removed from the crystallisation drop and passed through a droplet containing the compound for 10-60 s. The soaking droplet contained 5 mM of the compound of interest within the initial crystallisation growth condition. Occasionally, soaking the crystals resulted in the destruction of the crystal. This led to a different approach of only passing the crystal briefly (less than 20 s) through the appropriate cryo-protectant that also contained 5 mM of the selected compound.

2.13.4 Cryoprotection

Protein crystals are susceptible to damage from X-ray radiation, so diffraction data are collected from crystals of interest at typically 100 K. Crystals were often stored at 77 K prior to data collection. To prevent ice crystals forming, the crystals were placed in a cryoprotectant solution. The X-ray data collected in this project were from optimisations of crystals grown in C4 from the JCSG screen (Appendix 2). The cryoprotectant for this condition contained the components of the crystallisation reservoir solution with a 2 % increase in the polyethylene glycol concentration and a supplement of 15 % ethylene glycol.

2.13.5 Crystal mounting

Crystals were looped from the droplets using loops with diameters that were appropriate to the size of the crystal. Loops were attached to metal pins, which in turn were attached to magnetic bases that were easily transferred to a storage puck. The loops were passed beneath the crystal and lifted from the droplet with extreme care.

2.13.6 Data collection

Crystals produced in this project were subjected to X-ray radiation at the Diamond Synchrotron Light Source near Oxford (UK) on I02, I03, I04 and I24 beamlines.

2.13.7 Data processing

X-ray diffraction images were indexed and integrated to produce a complete dataset using the xia2 XDS software (Winter, 2010) at the Diamond Light Source. The XScale program within xia2 was used to merge the symmetry related reflections to generate a set of unique diffraction intensity data and to assess the data quality. Calculated values for the space group, mosaicity and unit cell dimensions of the crystal were obtained.

2.13.8 Structure determination and refinement

The CCP4 suite of programs (Winn *et al*., 2011) was used to manipulate the data, which enabled calculation of the Matthews coefficient (V_m) (Kantardjieff and Rupp, 2003) to estimate the number of protein molecules in the asymmetric unit.

Molecular replacement was used to solve the crystal structure of native NanH using the PHASER program from the CCP4 suite (McCoy *et al*., 2007; Winn *et al*., 2011). The phase problem for the NanH dataset was solved by using BTSA sialidase from *Bacteroides thetaiotaomicron* as a model in the molecular replacement (PDB: 4BBW). BTSA has a high protein sequence identity (approximately 65 %) to NanH, which allowed the generation of phases for NanH. Further NanH or NanH variant crystal structures were solved by inputting the new data set into Refmac5 refinement (CCP4 suite) (Vagin *et al*., 2004; Winn *et al*., 2011) against the coordinates of native NanH, which had previously been refined several times. The models underwent several rounds of model building in Coot (CCP4 suite) (Emsley *et al*., 2010) and refinement in Refmac5. The reliability of the structures was validated using the options within Coot and by use of the MOLPROBITY program (Chen *et al*., 2010). All subsequent structure manipulation such as structure superimposition was carried out using the programs from within the CCP4 suite.

2.13.9 Structure visualisation and analysis

The crystal model figures in Chapter 4 and 5 were produced using The PyMOL Molecular Graphics System, Version 2.2.3 Schrödinger, LLC (www.pymol.org). The electrostatic potential (section 4.2.3.2) was calculated using the 'APBS Electrostatics' plugin (Baker *et al*., 2001; Jurrus *et al*., 2018) in PyMOL 2.2.3 using the default settings. The interfaces between NanH subunits (section 4.2.3.3) were analysed using PDBePISA at EMBL-EBI, which can be accessed via this address:

https://www.ebi.ac.uk/pdbe/pisa/ (Krissinel and Henrick, 2007). The 'wizard mutagenesis' function in PyMOL 2.2.3 was used throughout Chapter 5 to predict the best rotamer for the mutation being made.

Chapter 3

Biochemical characterisation of NanH from *Tannerella forsythia*

3.1 Introduction

3.1.1 The sialidases of *Tannerella forsythia*

The oral microbiota contains several enzymes that are secreted by its bacterial residents. One of which are the sialidases, which hydrolyse the terminal N-acetylneuraminic acid (sialic acid, Neu5Ac) from glycoproteins (Stafford *et al.,* 2012). *Tannerella forsythia* is an anaerobic, Gram-negative pathogen that resides in the oral cavity and is one of the keystone pathogens responsible for periodontal disease (a destructive gum disease that results in tooth loss). Several studies have demonstrated sialidase activity in *T. forsythia* isolates, formerly known as *Bacteroides forsythus* (Moncla *et al*., 1990; Braham and Moncla, 1992).

Later studies have revealed the molecular basis of this activity, with the first study identifying the *siaHI* gene in a clone library screen as having sialidase activity. However, the *siaHI* gene sequence does not contain classical sialidase features such as Asp-box and Y/FRIP motifs that are common to GH33 CAZymes (carbohydrate active enzymes) (Ishikura *et al*., 2003). In this initial study, the SiaHI protein purified from *E. coli* showed sialidase activity against MU-NANA (Ishikura *et al*., 2003). However, subsequent studies using lysates expressing *siaHI* (Thompson *et al*., 2009) have not been able to replicate this finding, reporting no activity against MU-NANA. A *siaHI* mutant of *T. forsythia* still retained sialidase activity in vitro (Honma *et al*., 2011) which suggests siaHI is not responsible for sialidase activity. In addition, researchers at SUNY-Buffalo have failed to reproduce any activity from *siaHI* in a His-tagged format (Sharma, personal communication). It is therefore clear that SiaHI was not responsible for the sialidase activity present in *T. forsythia* and that there must have been an alternative source of activity in the original Ishikura paper (Ishikura *et al*., 2003).

3.1.2 NanH of *Tannerella forsythia*

Subsequent studies into the characterisation of a second sialidase in *T. forsythia* and the determination of the role of this enzyme in epithelial cell attachment and biofilm formation showed that the product of the *nanH* gene is the principle sialidase in *T. forsythia* (Thompson *et al*., 2009; Honma *et al*., 2011; Roy *et al*., 2010). *T. forsythia* NanH is an orthologue of *B. fragilis* NanH, with 65 % sequence identity. NanH contains four perfect Asp-boxes (SxDxGxTW), along with one imperfect Asp-box (TRDLGKSW) and the conserved FRIP region (residues 211-214) at the N-terminus of the catalytic

domain (residues 196-552), which exhibits glycoside hydrolase activity (GH33) (Figure 3.1). A 20-amino-acid signal sequence is present at the N-terminus of NanH (residues 14-33), which suggested that the enzyme was likely to be secreted or surface localised (Figure 3.1) (Thompson *et al*., 2009). Furthermore, proteome work by Friedrich and colleagues showed that NanH could be found in outer membrane vesicles, confirming that NanH is indeed a secretory protein (Friedrich *et al*., 2015).

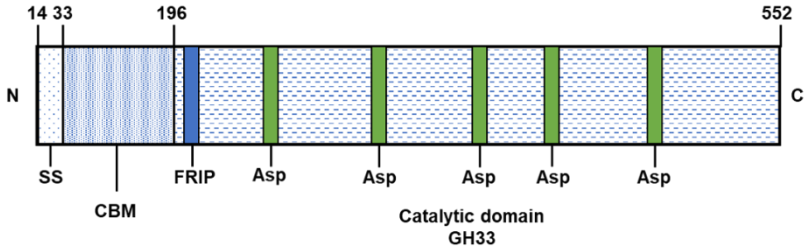


Figure 3.1: Linear schematic of NanH from *T. forsythia.* This schematic shows the N-terminal signal sequence (SS) followed by an N-terminal carbohydrate binding module (CBM) and a catalytic domain that contains the conserved FRIP and repeating Asp-box motifs. The N- and C-termini are represented by the letters N and C respectively. The original annotation of NanH included 13 amino acids at the N-terminus (TF0035, oralgen.org), however these residues showed no significant sequence similarity to *B. fragilis* NanH and an alternative start codon at position 14 was proposed to be to be the true N-terminus (Thompson *et al*., 2009).

3.1.3 The carbohydrate-binding module of NanH

As mentioned above, the GH33-sialidase domain of NanH is preceded by a 160 amino acid N-terminal domain (Figure 3.1). Sialidases such as NanA of *S. pneumoniae* and VCNA of *V. cholerae* are also preceded by N-terminal domains that have been determined to be carbohydrate binding motifs (CAZy CBM40) (Moustafa *et al*., 2004; Yang *et al*., 2015). However, the N-terminal domain of NanH is distinct from CBM40 domains and only shares similarities with closely related sialidases from *Bacteroides* spp. such as *B. thetaiotaomicron* (section 4.2.3.1) (Frey *et al.,* 2018). We recently reported that this N-terminal domain is capable of binding various sialoglycans and non-sialylated derivatives but has no catalytic activity when isolated from the catalytic domain (Frey *et al*., 2018). A catalytically inactive version of NanH also retained the ability to bind sialic acid-containing ligands showing that the binding activity of the N-terminal domain is enhanced when associated with the catalytic domain. This domain is therefore likely to belong to a novel subfamily of carbohydrate-binding module that does not yet have a designated CAzy classification (Frey *et al*., 2018).

3.1.4 The catalytic activities of NanH

Purified recombinant NanH (from expression in *E. coli)* has catalytic activity against the fluorogenic sialidase substrate MU-NANA between pH 4.0 and pH 8.0 (Frey *et al*., 2018). The specificity of NanH in whole cell lysate was tested on a range of substrates and it was found to have activity against both α2,3- and α2,6-linked sialic (Thompson *et al*., 2009). The same lysates containing NanH were said to have no sialidase activity against complex substrates such as fetuin, whereas Roy and colleagues demonstrated that purified, recombinant NanH has activity against fetuin (Roy *et al*., 2011). This suggested that it is important to perform sialidase assays using purified NanH, as other proteins in the *E. coli* cell lysate may be reducing the activity of NanH.

3.1.5 Chapter aims

As stated in the introduction, the overall aim of this thesis was to gain increased understanding of microbial sialidases with a view to engineering their capabilities for glycoanalytical pipelines – specifically asking whether the production of an α2,6- linkage specific enzyme was possible. Our initial hypothesis was that the *T. forsythia* enzyme NanH was a candidate for this purpose, given its biological properties as a highly active and potentially stable secreted enzyme whose natural environment is the oral cavity. Therefore, this chapter describes the first stage of that process, by assessing in detail biochemical properties of the NanH sialidase from *T. forsythia*. A second aim, given our intention to develop this enzyme for use in bioanalytical pathways with our industrial collaborator, Ludger Ltd, was to assess the storage and stability properties of NanH.

3.2 Results

3.2.1 Justification for using NanH as a model sialidase

In the early stages of this project, there was the intention to search for sialidases in the literature that may already have specificity towards either α2,3- or α2,6-linked sialic acid, alongside the studies on *T. forsythia* NanH. This search highlighted SiaBb2 from *B. bifidum* (Kiyohara *et al*., 2011) and NanH from *C. diptheriae* (Kim *et al*., 2010) as having higher preferences for α2,3- and α2,6-linked sialic acid respectively. Unfortunately, the NanH gene from *C. diptheriae* (codon optimised for *E. coli*) was unable to be cloned into a pET-21a plasmid for unknown reasons. SiaBb2 was successfully cloned into pET-21a however complications began at the expression and purification steps. A 0.5 L culture of *E. coli* yielded an extremely low concentration of protein, which was therefore concentrated to enable biochemical analysis to be performed. However, protein concentration resulted in a low volume of purified protein, which limited the amount of analysis that could be carried out. Preliminary analysis of SiaBb2 in a high-throughput version of the TBA assay (section 5.2.2) suggested that this enzyme was entirely specific for α2,3-linked sialic acid but unfortunately in the time-frame of this project it was not possible to optimise a protein expression and purification strategy to yield enough protein for analysis with complex glycans. If a higher yield of SiaBb2 could be produced, it would be interesting to also solve the crystal structure to determine whether there are any structural features that could be responsible for the higher specificity towards α2,3-linked sialic acid, that differ to the aromatic stacks seen in trans- and IT-sialidases.

Recombinant His-tagged NanH was readily available in the laboratory of G. Stafford and showed potential for specificity engineering. For these reasons, NanH was chosen as the model sialidase to be used in this body of work.

3.2.2 Expression and purification of NanH

The first step in the process of biochemical and structural understanding of NanH was to produce high yields of pure protein. NanH was previously cloned as a synthetic gene that was codon-optimised for *E. coli* into the pET-21a expression vector to incorporate a C-terminal hexa-Histidine tag onto the protein for ease of purification (Frey *et al*., 2018). Expression conditions for this enzyme were established previously in the laboratory using *E. coli* BL21 Origami and purification was carried out using Nickel-affinity chromatography (section 2.8.3 and 2.8.4). The ultimate aim was to produce enough protein for crystallisation trials (Chapter 4). Therefore, in order to gain the high yields of protein needed for crystallisation, the purification method was adapted to use 5 mL Ni^{2+} Fast Flow columns (GE Healthcare Life Sciences).

A NanH purification (Figure 3.2a) was analysed by SDS-PAGE and the elution fractions that corresponded to the single peak in in the elution profile (10-17) presented bands that were the correct size for a monomer of NanH (57.4 kDa) (Figure 3.2b). There was some level of impurity seen in these fractions (perhaps only visible because the gel was overloaded), however the purification was successful as NanH had over 95 % purity. Elution fractions containing protein were then subjected to dialysis in 14 kDa cut-off tubing (Spectrum Labs), to not only reduce the concentration of imidazole (which has previously been shown to interfere with sialidase activity, G. Stafford personal communication) but to also lower the NaCl concentration. The concentration of NanH obtained from 0.5 L of culture in this case was approximately 10 mL at 4 mg mL^{-1}.

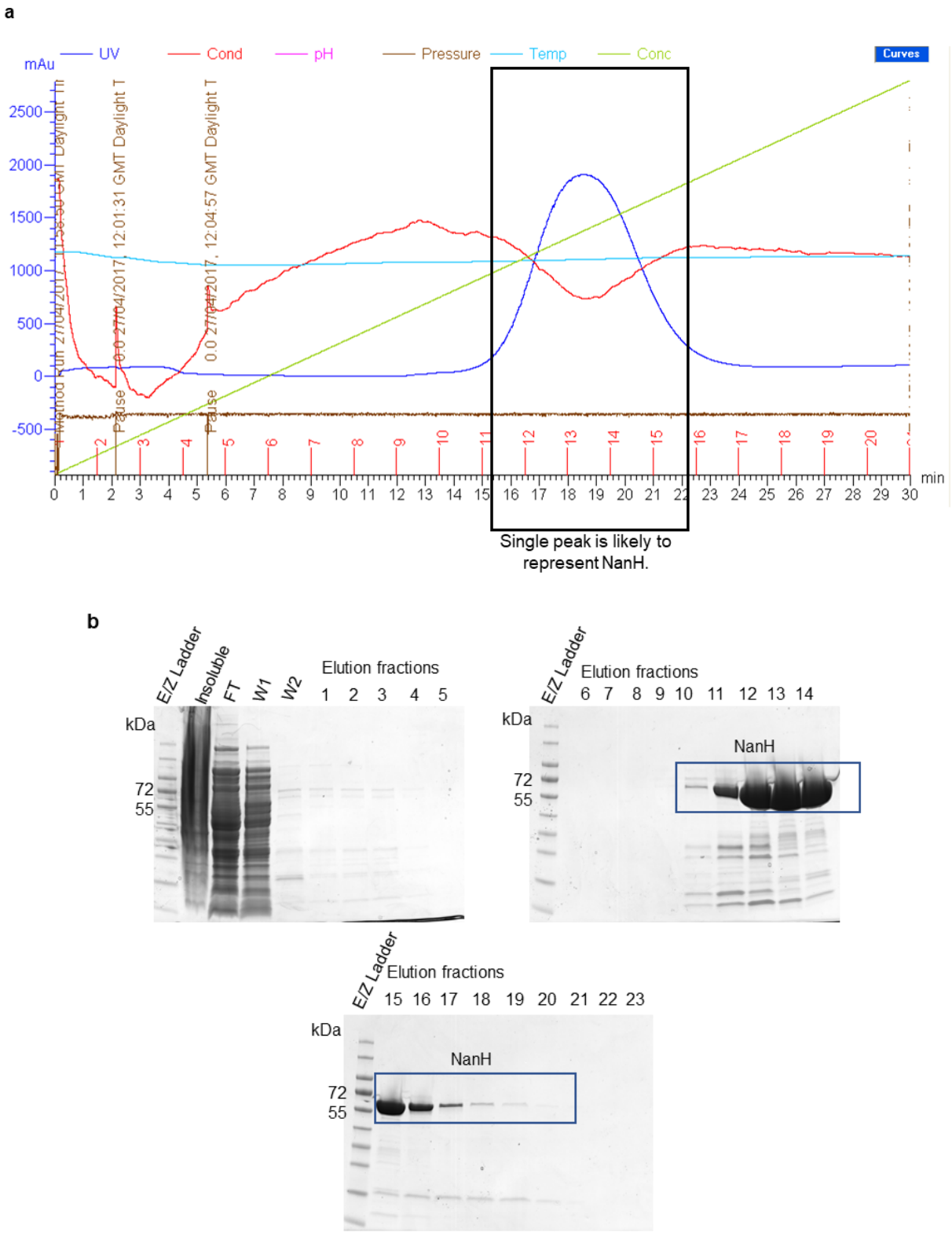


Figure 3.2: The purification profile of NanH. (a) The elution profile of NanH with the peak that is likely to represent NanH highlighted. **(b)** SDS-PAGE of the elution fractions from the purification in part (a) (FT = flow through, W1 = wash 1 and W2 = wash 2).

After purification and dialysis into 50 mM sodium phosphate, 200 mM NaCl, pH 7.4 buffer, it was important to check that NanH had enzymatic activity. A quick method for doing this was to incubate NanH in a 96-well plate with the synthetic sialidase substrate, MU-NANA, for 10 minutes at room temperature. The sialidase cleaves the sialic acid from 4-methylumberlliferone (4-MU), which fluoresces when measured under ultraviolet (UV) light. For all batches of NanH purified, it was confirmed that there was enzymatic activity, with an example shown in Figure 3.3.

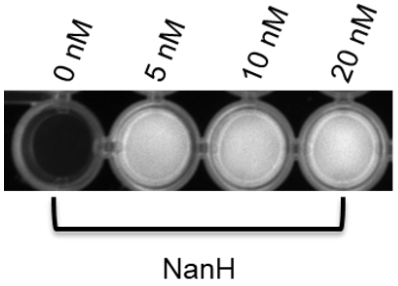


Figure 3.3: Visualisation of 4-MU release when MU-NANA is cleaved by NanH. NanH activity is represented by the fluorescence under UV light at increasing concentrations of enzyme (0, 5, 10 and 20 nM – determined by the bicinchoninic acid assay (BCA)).

3.2.3 Establishing the multimeric form of NanH

In order to try and establish the multimeric state of the NanH sialidase, a variety of techniques were attempted. Firstly, size-exclusion chromatography (SEC) was employed, which is a method used to separate molecules in solution by their size. After affinity purification and dialysis (section 3.2.1), 2 mL of NanH at 3 mg mL^{-1} was loaded onto a Superdex200 gel filtration column (favours proteins in the molecular weight range of 10,000 to 600,000 Da) and eluted at a flow rate of 1.5 mL min^{-1}. The elution profile contained peaks at 55.35 mL and 65.39 mL (Figure 3.4). The apparent molecular weight of the peak at 55.35 mL is 240 kDa (by comparison to a native protein standard), which corresponds to a tetrameric complex. The approximate molecular weight for the peak at 65.39 mL is 105 kDa, which corresponds to a predicted dimeric complex (2 x 57.4). The elution profile clearly shows that most of the protein lies in the fractions that correspond to a dimeric complex, suggesting that NanH is predominantly a dimer in solution (Figure 3.4).

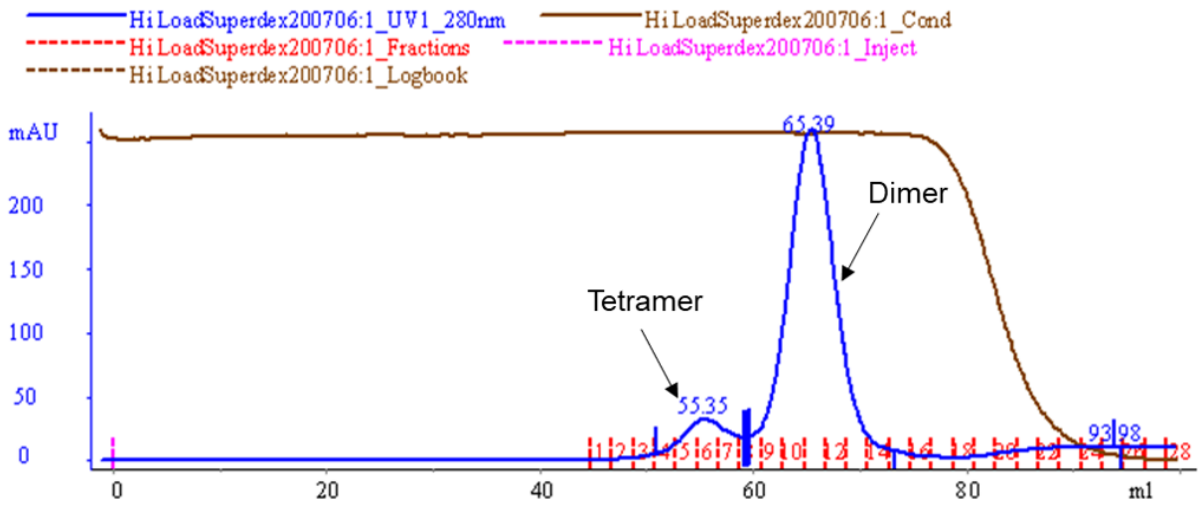


Figure 3.4: NanH elution profile after gel-filtration on a Superdex200 gel filtration column. NanH protein (3 mg mL^{-1}) was loaded onto a Superdex200 gel filtration column and eluted at a flow rate of 1.5 mL min^{-1} on an AKTA Purifier 10 FPLC System. The blue trace represents the absorbance at 280 nm, the brown trace represents the conductivity and the red numbers represent the 2 mL fractions that were collected after a void volume of 42 mL.

To confirm that the peaks from the SEC column contained a polypeptide of the right size to be a monomer of NanH, SDS-PAGE was performed on the fractions that corresponded to the peaks at 55.35 mL and 65.39 mL in (Figure 3.4). The size of the bands on the gel (Figure 3.5) confirmed that the protein in both the peaks at 55.35 mL and 65.39 mL represented polypeptides of the right size to be a monomer of NanH (57.4 kDa).

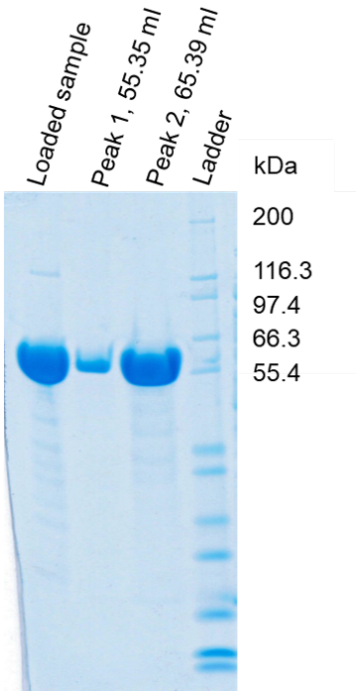


Figure 3.5: NU-PAGE analysis of SEC fractions that correspond to the two peaks on the elution profile. (Lane1) Original loaded sample, (2) Peak 1 – fraction 6, 55.35 ml, (3) Peak 2 – fraction 10-12, 65.39 ml, (4) Mark12 protein ladder.

In order to further probe the multimeric state and validate the SEC data, NativePAGE electrophoresis was used to investigate the multimeric form of NanH in more detail. NativePAGE determines the molecular weight of a protein in its native form and was therefore a further test used to confirm whether NanH is predominantly a dimer in solution. NanH (pooled SEC fractions 10-12) was ran on a NativePAGE gel at different concentrations. Surprisingly, the most predominant band was present at approximately 242 kDa, indicating that in the conditions of NativePAGE, NanH forms a tetramer (Figure 3.6).

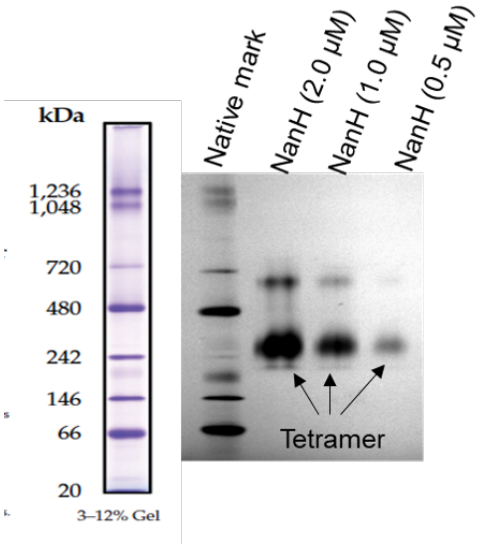


Figure 3.6: NanH NativePAGE. (Lane1) Native mark, (2) 2.0 µM NanH, (3) 1.0 µM NanH, (4) 0.5 µM NanH.

A sample of freshly purified NanH was analysed by a Size-Exclusion Chromatography Multi-Angle Laser Light Scattering (SEC-MALLS) service at the University of York (section 2.11.3), in a final attempt to understand the multimeric form of NanH in solution. The result confirmed what was see in the SEC above, as NanH was again found predominantly as a dimer in solution with a small amount of protein being in the tetrameric form.

3.2.4 Reaction kinetics of NanH using synthetic substrate 4-Methylumbelliferyl N-acetyl-α-D-neuraminic acid

As mentioned in section 3.2.2, MU-NANA is a synthetic sialidase substrate that is widely used to determine sialidase activity. Reaction kinetics for NanH were obtained by quantifying the 4-MU release from increasing concentrations of MU-NANA and applying the raw data to a standard curve of known concentrations of 4-MU. The initial rate of reaction was determined for all substrate concentrations of interest and plotted as an x,y pair with the appropriate concentration of substrate (MU-NANA) to obtain a Michaelis-Menten curve. The initial rate of reaction is the amount of product made per unit time at the beginning of the reaction, when the concentration of product is increasing linearly. From the Michaelis-Menten curve, the maximum rate of reaction (V_{max}) and the concentration of substrate for which a rate of reaction equal to half V_{max} (K_M) can be determined. A lower K_M is often consistent with higher binding affinity as K_M can sometimes be treated as an apparent dissociation rate constant (Fersht, 1985).

The reaction kinetics of NanH with MU-NANA were determined in neutral pH buffers (50 mM sodium phosphate, pH 7.4 with or without the presence of 200 mM sodium chloride (NaCl) (Figure 3.7, Table 3.1). This condition was selected as it was indicated that assays were best performed at a neutral pH at Ludger Ltd for the convenience of allowing multiple enzymes to be processed in a single reaction (personal communication – Daniel Spencer). Conveniently, pH optima studies have already confirmed that NanH is still highly active at neutral pH (Frey *et al*., 2018). The results showed that NanH has a higher activity in the presence of 200 mM NaCl as the V_{max} was 13.8 µmol min^{-1} mg^{-1} NanH compared to 9.0 µmol min^{-1} mg^{-1} NanH in the absence of NaCl. NanH also had a slightly lower K_M with MU-NANA in the absence of 200 mM NaCl, as the K_M was 3.5 µM compared to 6.2 µM. The data was plotted as the 4-MU release per concentration of NanH against the concentration of MU-NANA, enabling the k_{cat} to be calculated, which is the rate constant for the transition of the enzyme substrate complex into the enzyme product complex ($time^{-1}$). We define the catalytic efficiency as, k_{cat}/K_M (sometimes referred to as the specificity constant) which is a common measure of the overall efficiency of an enzyme and determines the specificity for competing substrates (Fersht, 1985). The catalytic efficiency values for NanH were similar in the presence and absence of 200 mM NaCl at 129.2 min^{-1} $µM^{-1}$ and 147.2 min^{-1} $µM^{-1}$ respectively. Since the k_{cat}/K_M values were similar, the rest of the assays in this project were carried out at pH 7.4, 200 mM NaCl, which is a good representation of physiological conditions in the human body.

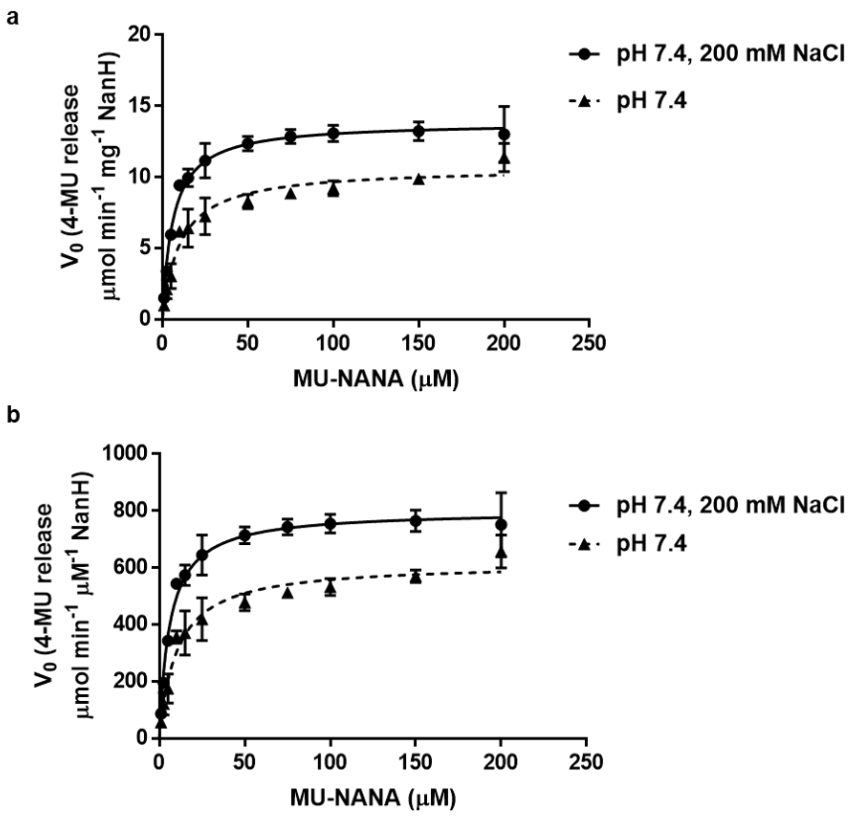


Figure 3.7: The reaction kinetics of NanH and MU-NANA. (a) Increasing concentrations of MU-NANA were exposed to 2.5 nM NanH. The reaction was stopped at 1-, 2- and 3-minute intervals and the rate of reaction (4-MU release) was determined by applying the raw data to a standard curve of known 4-MU concentrations. The rate was plotted against [MU-NANA] in a Michaelis-Menten curve from which the V_{max} and K_M could be calculated (Table 3.1). **(b)** The 4-MU release (per µM NanH) was plotted against [MU-NANA] enabling the k_{cat} to be calculated and in turn the catalytic efficiency (Table 3.1). In both cases the data represents the mean of three repeats, with the error bars representing the standard deviation.

	pH 7.4, 200 mM NaCl	pH 7.4
V_{max} (µmol min^{-1} mg^{-1})	13.8 ± 0.2 (13.4 to 14.3)	9.0 ± 0.22 (8.5 to 9.4)
K_M ([MU-NANA], µM)	6.2 ± 0.5 (5.2 to 7.3)	3.5 ± 0.5 (2.7 to 4.6)
k_{cat} (min^{-1})	799.6 ± 13.7 (772.7 to 827.3)	518.2 ± 12.9 (493.2 to 544.1)
k_{cat}/K_M (min^{-1} µM^{-1})	129.2	147.2

Table 3.1: Reaction kinetics values obtained from Michaelis-Menten curves of NanH digest with MU-NANA. The V_{max}, K_M, k_{cat} and k_{cat}/K_M values are presented (± standard error) for the reaction kinetics of NanH with MU-NANA using neutral pH but varying the NaCl concentration. The values in parentheses display the 95 % confidence intervals.

Since NanH is a predicted multimer the MU-NANA kinetics data for the reaction ran at pH 7.4 + 200 mM NaCl was assessed for cooperativity. A Hill coefficient was calculated using the 'Allosteric sigmoidal enzyme kinetics' function on GraphPad Prism 7.03. The Hill coefficient was found to be 1.2, which is in the range that represents non-cooperative enzymes.

3.2.5 Inhibition of NanH using common sialidase inhibitors

Structure-function data is crucial to this project so that an understanding of the key catalytic residues can be gained and therefore which residues could be mutated to alter the specificity of NanH. It is important to gain a model of NanH in complex with an inhibitor (Neu5Ac analogue) to understand where sialic acid binds and provide insights into the enzyme mechanism. Previous studies on the inhibition of NanH proved that zanamivir (also known as Relenza, which is used worldwide to treat influenza) was capable of inhibiting NanH (Frey, 2016). Therefore, it was decided to investigate NanH inhibition with alternative sialidase inhibitors. The inhibitors chosen were DANA and oseltamivir acid (also known as Tamiflu, again used to treat influenza), which are also structural mimics of Neu5Ac (Figure 1.8, section 1.3.7).

Inhibition assays using DANA and oseltamivir acid have been performed on purified NanH to determine their inhibition abilities compared to zanamivir. NanH activity in the presence of increasing concentrations of DANA or oseltamivir was observed by incorporating the inhibitor into the MU-NANA assay described in section 3.2.4 and quantifying the 4-MU release. The sialidase activity in the presence of inhibitor was expressed as the percentage change in fluorescence at a given concentration of inhibitor, compared to that in the absence of inhibitor. The inhibitory constant (IC_{50}) is a value that

represents the efficacy of an inhibitor and is defined as the concentration of an inhibitor that is needed to inhibit enzyme activity by half. The IC_{50} of zanamivir with NanH was determined to be 6309 µM by A. Frey (Frey, 2016). This study shows both DANA and oseltamivir acid to be more efficient than zanamivir with IC_{50} values of 11.0 µM and 371.6 µM respectively and therefore being approximately 500 and 16 times more effective respectively (Figure 3.8, Table 3.2).

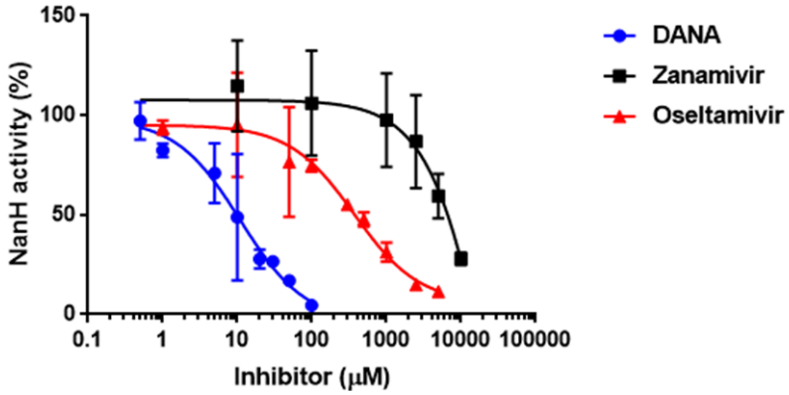


Figure 3.8: The effect of DANA and oseltamivir on NanH activity. The [inhibitor] was plotted (on a logarithmic, base 10, scale) against the percentage of sialidase activity relative to a condition that contains no inhibitor. The data represents the mean of three repeats and the error bars represent standard deviation.

Inhibitor	Inhibitory constant (IC50) - NanH
N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid (DANA)	11.0 ± 3.5 µM (5.6 to 21.7)
Oseltamivir acid	371.6 ± 114.1 µM (185.1 to 732.5)
Zanamivir	6309.0 ± 8.3 µM

Table 3.2: IC_{50} values obtained for DANA and oseltamivir acid inhibition of NanH. The IC_{50} values were calculated from the curves in Figure 3.8 (± standard error). The values in parentheses display the 95 % confidence intervals.

3.2.6 Determining NanH specificity for glycosidic linkages

The primary aim of this project is to engineer the specificity of NanH to produce mutants that are specific to either the α2,3- or α2,6-linked sialic acid. To do this, it was required to understand the specificity of NanH wild-type enzyme. NanH was incubated with the sialylated substrates α2,3- and α2,6-sialyllactose (3-/6-SL), which contain an α2,3- and α2,6-linked sialic acid respectively (Figure 3.9).

3-/6-SL are readily available from Carbosynth, easy to use in the digestions and represent a common glycan ending (galactose followed by sialic acid) that is seen in nature.

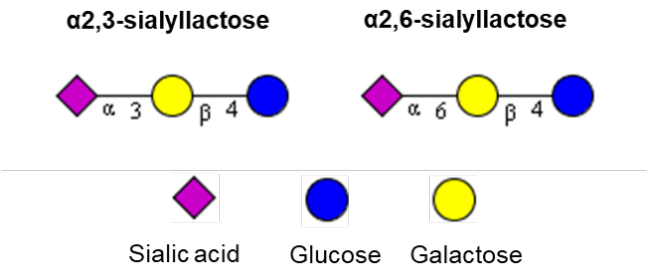


Figure 3.9: Composition of 3- and 6-SL. 3-/6-SL are trisaccharides composed of sialic acid, galactose and glucose with sialic acid at one of the termini.

Unlike the MU-NANA kinetics discussed in section 3.2.4, the products after a NanH digest with 3-/6-SL are not easily quantifiable by fluorometric methods like 4-MU. Free sialic acid released from the digest was quantified using an adaption of the thiobarbituric acid (TBA) assay (Aminoff, 1961). After a NanH digest of 3-/6-SL, thiobarbituric acid is added to the digest and labels free sialic acid with a thiol group. This introduces a colour change which varies depending on the concentration of sialic acid present, allowing quantification by spectrophotometry (section 2.11.6). A standard curve of known concentrations of sialic acid was run alongside the TBA assay to enable the determination of the concentration free sialic acid that was produced.

From the Michaelis-Menten curve, the V_{max} value suggests that NanH had a similar maximum rate of reaction against 3- and 6-SL with rates of 46.4 and 43.4 µmol min^{-1} mg^{-1} NanH respectively (Figure 3.10, Table 3.3). However, when considering the K_M (ligand-binding affinity), it can be seen that NanH has a higher affinity for 3-SL over 6-SL, which have K_M values of 0.3 and 0.6 mM respectively. The data was also plotted as Neu5Ac release per concentration of NanH against the concentration of 3-/6-SL (µmol min^{-1} $µM^{-1}$), allowing the k_{cat} to be calculated. The k_{cat}/K_M for NanH against 3-SL was more than twice that of NanH against 6-SL, with values of 8654.8 and 3920.2 min^{-1} mM^{-1} respectively, both of which are much higher than with the synthetic substrate MU-NANA. This suggested that although NanH is dual specific, it had a preference for the α2,3-linked sialic acid.

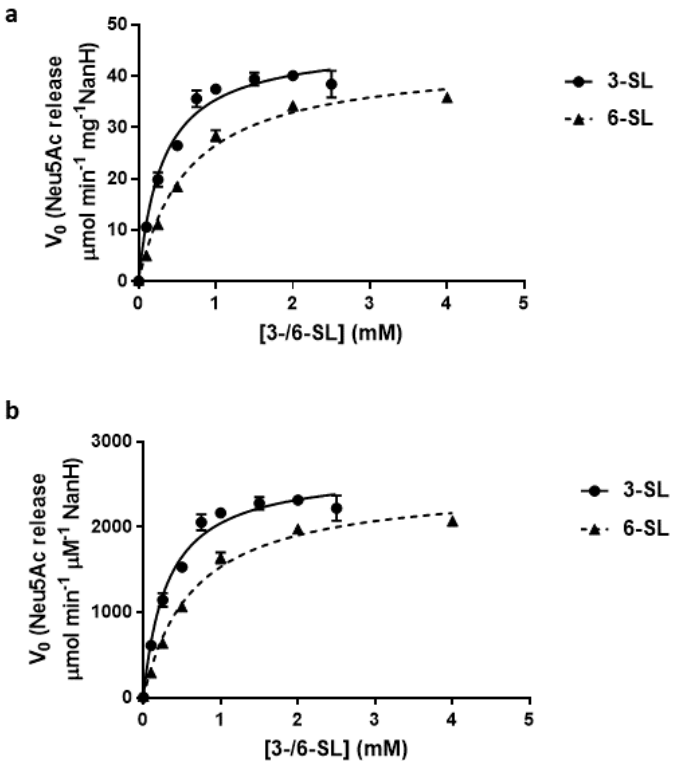


Figure 3.10: The reaction kinetics of 3-/6-SL and NanH. (a) Increasing concentrations of 3-/6-SL were exposed to 50 nM NanH. The reaction was halted at 1-, 3- and 5-minute intervals before performing the TBA assay so that free sialic acid could be quantified. The free sialic acid was quantified by applying the data to a standard curve of known concentrations of sialic acid that had also been passed through the TBA assay. The rate was plotted against [3-/6-SL] in a Michaelis-Menten curve from which the V_{max} and K_M could be calculated (Table 3.3). **(b)** The Neu5Ac release (per µM NanH) was plotted against [3-/6-SL] enabling the k_{cat} to be calculated and in turn the catalytic efficiency (k_{cat}/K_M) (Table 3.3). In both cases the data represents the mean of three repeats, with the error bars representing the standard deviation.

	3-SL	6-SL
V_{max} (µmol min^{-1} mg^{-1})	46.4 ± 1.2 (44.1 to 49.0)	43.4 ± 1.1 (41.3 to 45.7)
K_M ([3-/6-SL], mM)	0.3 ± 0.0 (0.3-0.4)	0.6 ± 0.1 (0.5-0.7)
k_{cat} (min^{-1})	2683.0 ± 69.7 (2550.0 to 2829.0)	2509.0 ± 63.5 (2385.0 to 2642.0)
k_{cat}/K_M (min^{-1} mM^{-1})	8654.8	3920.3

Table 3.3: Reaction kinetics values obtained from Michaelis-Menten curves of NanH and 3-/6-SL digests. The V_{max}, K_M, k_{cat} and k_{cat}/K_M values are presented (± standard error) for the reaction kinetics of 3-/6-SL and NanH. The values in parentheses display the 95 % confidence intervals.

The TBA assay was also performed using the same method as above but using 3-Sialyl-Lewis A and 3-Sialyl-Lewis X ($SLe^{A/X}$) (Table 3.5) as the substrates for sialic acid digestion. These glycans are present in secretions and cell surfaces in the oral cavity, human cells and human plasma and are sialic acid antigens. Both of these ligands contain α2,3-linked sialic acid but differ due to the linkages between their other constituent sugars. NanH displayed a similar K_M and V_{max} with SLe^A and SLe^X (K_M = 1.3 ± 0.2 and 1.1 ± 0.2 mM, V_{max} = 36.1 ± 2.6 and 31.2 ± 2.1 µmol min^{-1} mg^{-1} NanH for SLe^A and SLe^X, respectively) (Figure 3.11, Table 3.4). The catalytic efficiency of NanH with the two substrates was also similar (k_{cat}/K_M = 1645.2 and 1675.7 min^{-1} mM^{-1} for SLe^A and SLe^X respectively). It was noticed that the catalytic activity of NanH with $SLe^{A/X}$ was four-fold less than that of 3-SL when all three of these ligands contain α2,3-linked sialic acid, showing the importance of the underlying sugar residues in sialidase activity.

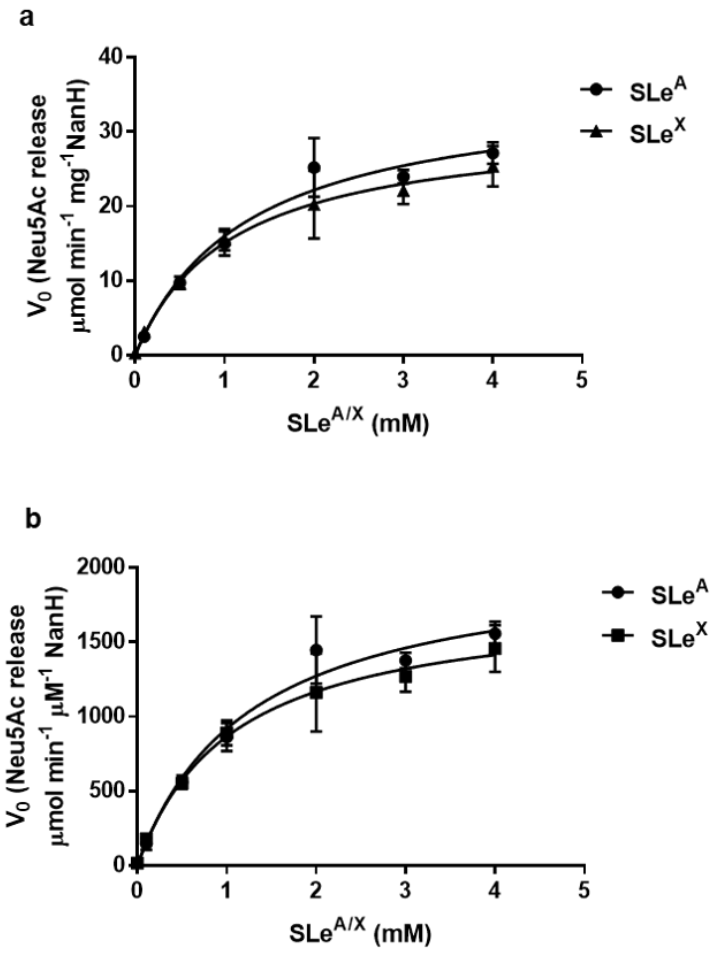


Figure 3.11: The reaction kinetics of NanH and SLe$^{A/X}$. (a) Increasing concentrations of SLe$^{A/X}$ was exposed to 50 nM NanH. The reaction was halted at 1-, 3- and 5-minute intervals before performing the TBA assay to quantify the concentration of free sialic acid. The sialic acid product was quantified by applying the data to a standard curve of known concentrations of sialic acid that had also been passed through the TBA assay. The rate was plotted against [SLe$^{A/X}$] in a Michaelis-Menten curve from which the V_{max} and K_M could be calculated (Table 3.4). **(b)** The Neu5Ac release (per µM NanH) was plotted against [SLe$^{A/X}$] enabling the K_{cat} to be calculated and in turn the catalytic efficiency (k_{cat}/K_M) (Table 3.4). In both cases the data represents the mean of three repeats, with the error bars representing the standard deviation.

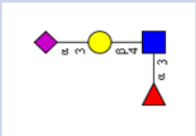
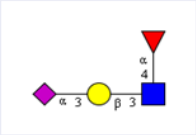
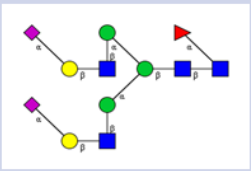
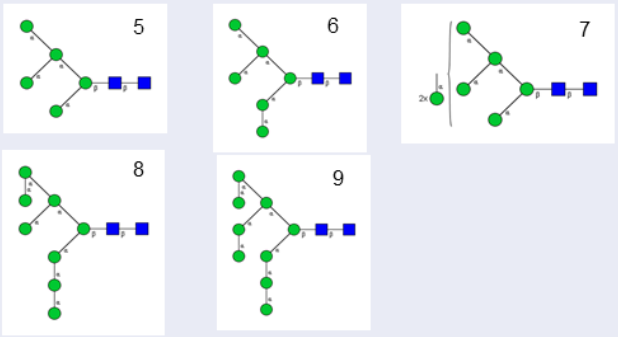
	SLe^A	SLe^x
V_{max} (µmol min^{-1} mg^{-1})	36.1 ± 2.6 (31.6 to 42.4)	31.2 ± 2.1 (27.3 to 36.6)
K_M ([$SLe^{A/X}$], mM)	1.3 ± 0.2 (0.9 to 1.9)	1.1 ± 0.2 (0.7 to 1.6)
k_{cat} (min^{-1})	2073.0 ± 148.5 (1812.0 to 2431.0)	1793.0 ± 122.1 (1568.0 to 2102.0)
k_{cat}/K_M ($min^{-1}mM^{-1}$)	1645.2	1675.7

Table 3.4: Reaction kinetics values obtained from Michaelis-Menten curves of NanH and $SLe^{A/x}$ digests. The V_{max}, K_M, k_{cat} and k_{cat}/K_M values are presented (± standard error) for the reaction kinetics of $SLe^{A/X}$ and NanH. The values in parentheses display the 95 % confidence intervals.

Once again, there was no reason to believe there was cooperativity of NanH with any of the substrates in these specificity assays. The Hill coefficients were calculated to be 1.2 and 1.3 for 3 and 6-SL respectively and 1.3 and 0.9 for SLe^A and SLe^X respectively.

3.2.7 Assessment of NanH activity on small and complex glycans relevant to glycoanalytical pipelines

Purified NanH was taken to Ludger Ltd (Oxford, UK) to investigate its activity on small glycan standards and more complex glycans. This analysis was performed by Marianne Satur, alongside Paulina Urbanowicz. Three different categories of NanH analysis were performed; NanH activity towards α2,3- and α2,6-linked sialic acid, whether NanH had any alternative activity towards non-sialylated glycans and studies to determine the stability of NanH after exposure to different stress conditions (elevated temperature and freeze-thawing). Table 3.5 lists the structures of the standards referred to throughout this section. The more complex glycan structures such as human plasma N-glycans are described in the appropriate section.

Standard name	Description	Structure
3-Sialyl-Lewis X	NeuAcα2-3Galβ1-4[Fucα1-3]GlcNAc	
3-Sialyl-Lewis A	NeuAcα2-3Galβ1-3[Fucα1 4]GlcNAc	
CAB-A2F-01	F(6)A2G(4)2S(6)2 (FA2G2S2)	
MAN Mix	Man5, Man6, Man7, Man8 and Man9	

Residue	Structure
Neu5Ac	
Neu5Gc	
N-acetylglucosamine (GlcNAc)	
N-acetylgalactosamine (GalNAc)	
Mannose	
Glucose	
Galactose	
Fucose	

Table 3.5: The standards used in enzyme digests at Ludger Ltd followed by a key showing the Consortium of Functional Glycomics.

3.2.7.1 NanH activity towards α2,3- and α2,6-linked sialic acid

NanH activity on α2,3- and α2,6-linked sialic acid was tested on simple substrates to confirm the data produced in the TBA assay (section 3.2.6). Ludger Ltd had access to SLe^X that contains α2,3-linked sialic acid and had produced a FA2G2S2 standard that contained α2,6-linked sialic acid (Table 3.5). Each of these glycans were labelled, at the reducing terminus, with a fluorescent dye called procainamide. Approximately 0.07-0.2 µg of glycans were digested with 10 nM NanH to release sialic acid and the residual molecules after cleavage were purified and analysed by hydrophilic interaction (HILIC) mode ultrahigh-performance liquid chromatography (UHPLC) (Dionex Ultimate 3000 UHPLC) coupled to an electrospray ionisation - mass spectrometer (ESI-MS) (Bruker Amazon Speed ETD). NanH digests were performed for 1 h and an O/N period (approximately 18 h), alongside an identical digestion in which NanH was replaced with Ludger Ltd alpha3,6,8,9 (E0001) sialidase (positive control) before analysis by UHPLC ESI-MS. Undigested forms of the glycans were also analysed by UHPLC ESI-MS to allow easy comparison of whether sialic acid hydrolysis occurs upon addition of a sialidase.

NanH is highly active towards α2,3-linked sialic acid on SLe^X, as sialic acid digestion goes to completion after the O/N period (Figure 3.12). It was noticed that the NanH digest did not go to completion after 1 h like in the E0001 digest, which is caused by the fact that the concentration of E0001 was unknown and the experiment was not quantitative, which is the same for the rest of the digest data in this section.

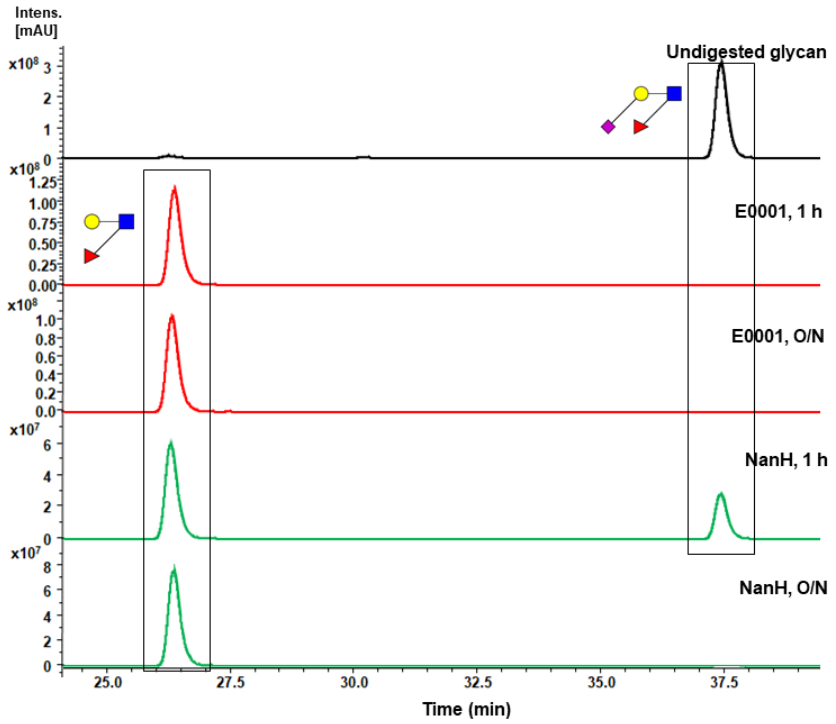


Figure 3.12: HILIC-UHPLC fluorescence (LC-FLR) trace of procainamide labelled sialidase digestions with SLe^X. Undigested (sialylated) SLe^X (black trace). E0001 1h and O/N digests (red traces). NanH 1h and O/N digests (green traces).

The FA2G2S2 glycan standard (Table 3.5) was digested separately by NanH and E0001 using the same method as that used for SLe^X and analysed using UHPLC ESI-MS. Undigested FA2G2S2 was also analysed as a negative control to demonstrate that no sialidase activity existed unless enzyme was added. It was noticed that the undigested FA2G2S2 trace contains impurities, including a NeuGc containing glycan impurity. There is also a non-fucosylated variant (A2G2S2) and a monosialylated form (FA2G2S1) of FA2G2S2 present in the undigested sample, which is presumed to be the nature of the batch produced (Figure 3.13).

The E0001 FA2G2S2 digest went to completion within 1 h, with sialic acid being shown to have been lost from both the fucosylated (FA2G2) and non-fucosylated variants (A2G2) as well as the NeuGc containing glycan impurity (Figure 3.13). After 1 h, the NanH FA2G2S2 digest resulted in 3 major fluorescence peaks representing FA2G2S2, FA2G2S1 and FA2G2 showing that this reaction had not gone to completion. After the O/N incubation period the NanH FA2G2S2 digestion had gone to completion, as the only major peak seen on the trace represented FA2G2 and therefore highlighted the removal of both sialic acid residues. Once again, the incomplete digestion after 1 h by NanH could

possibly be overcome by increasing the concentration of NanH in the digest but overall, it can be concluded that NanH is able to hydrolyse α2,6-linked sialic acids.

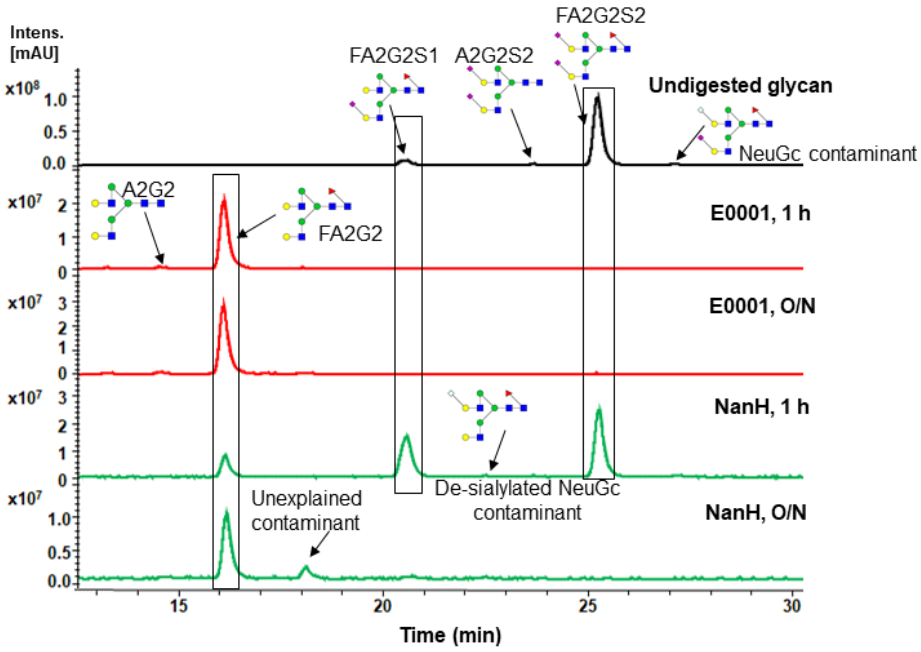


Figure 3.13: LC-FLR trace of procainamide labelled sialidase digestions with FA2G2S2. Undigested (sialylated) FA2G2S2 (black trace). E0001 1h and O/N digests (red traces). NanH 1h and O/N digests (green traces).

Understanding NanH activity with complex glycans allows analysis of how well NanH hydrolyses sialic acid from bi-, tri- and tetra-antennary glycans. This is more relevant to the biopharmaceuticals or proteins that Ludger Ltd analyse as they are likely to have complex glycosylation patterns compared to the small glycans previously used in this project. Human plasma N-glycans (contains both α2,3- and α2,6-linked sialic acid) were digested separately by NanH and E0001 using the same method as that used for SLe^X and analysed using UHPLC ESI-MS. Undigested plasma N-glycans were also analysed as a negative control. Complete digestion of sialic acid from plasma N-glycans occurred when incubated with NanH and E0001 O/N, as no sialylated glycans were observed in either fluorescence trace, as determined by UHPLC ESI-MS analysis (Figure 3.14). Once again, this highlights that NanH is a dual specific sialidase, as it successfully hydrolyses both α2,3- and α2,6-linked sialic acid from bi-, tri- and tetra-antennary glycans, as well as the small standards mentioned previously in this section.

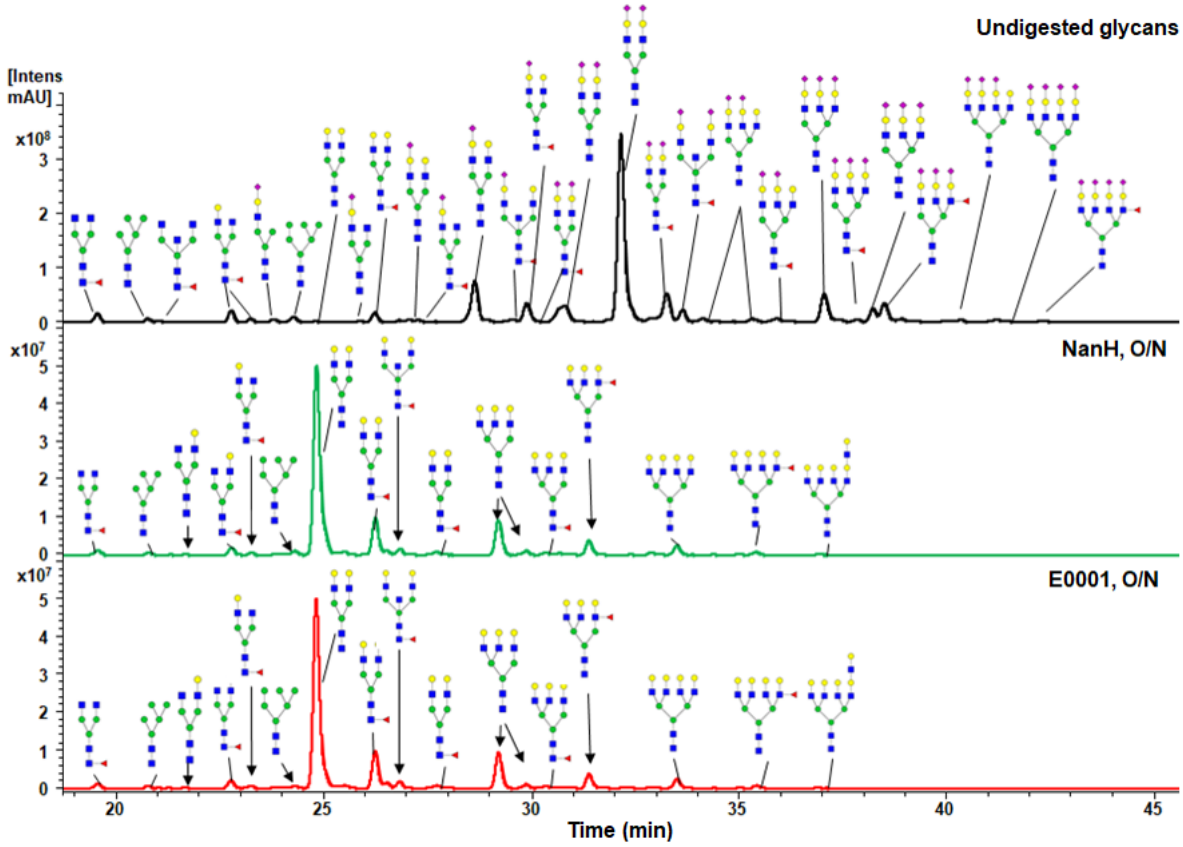


Figure 3.14: LC-FLR of procainamide labelled sialidase digestions with human plasma N-glycans. Undigested (sialylated) plasma N-glycans (black trace). E0001 O/N (red trace). NanH O/N (green trace).

3.2.7.2 Establishing non-target activity towards non-sialylated glycans

NanH activity towards non-sialylated glycans had not yet been analysed. 'MAN Mix' is a glycan mix produced by Ludger Ltd that contains 5 different non-sialylated glycans that vary due to the number of mannose residues that they contain (Table 3.5). NanH activity towards the various glycans in the MAN Mix was assessed. MAN Mix was exposed to 10 nM NanH for an O/N period of approximately 18 h, which was followed by purification and analysis by UHPLC ESI-MS. An undigested sample of MAN Mix was also analysed by UHPLC ESI-MS as a negative control. NanH did not show any alternative hydrolytic properties specific to mannose residues after an O/N digestion (Figure 3.15).

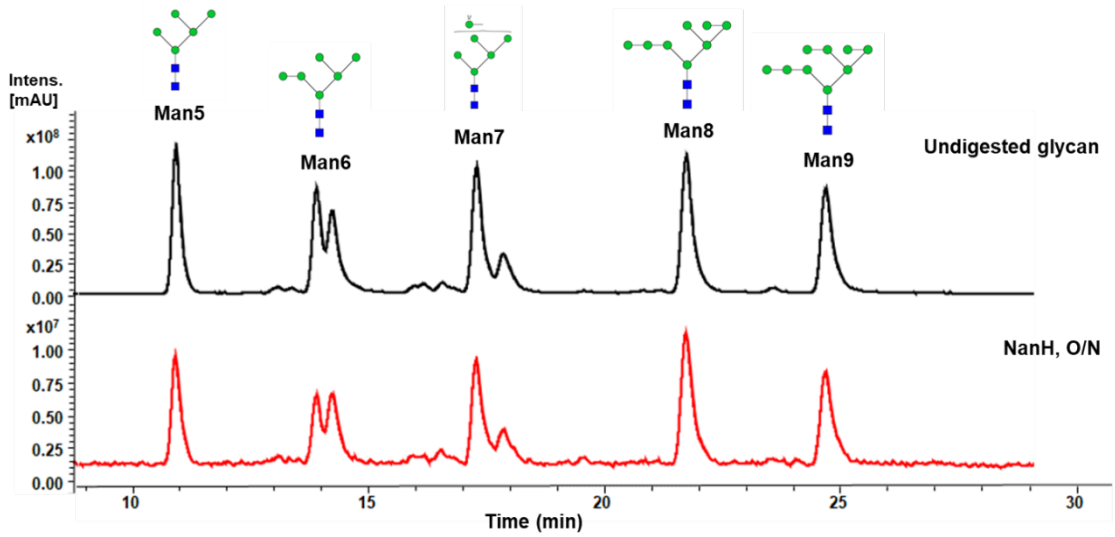


Figure 3.15: LC-FLR of procainamide labelled NanH digest with MAN Mix. Undigested MAN Mix (black trace). MAN Mix NanH digest (red trace).

3.2.7.3 NanH stability studies

It is important to assess NanH activity after being subjected to different stress conditions if the enzyme is to be used in Ludger Ltd products in the future. If an engineered NanH variant were to be made a publicly available product the ability to transport this enzyme without the need for maintenance of a constant temperature would be a valuable asset, reducing the cost of transport. The marketability of the product would also rely on an enzyme that was stable and active for a relatively long period of time. Accelerated thermal stability studies are a common method of determining product shelf life. At the same time, various stress conditions, such as elevated temperature and freeze-thawing, are good indications of the overall stability of an enzyme, for example in transport or handling. NanH was subjected to a variety of conditions relating to temperature before assessing the enzymatic activity in NanH [10 nM] digests with SLe^X and analysis by UHPLC ESI-MS (as in section 3.2.7.1). The conditions included 2 and 5 cycles of freeze-thawing (FT) of NanH aliquots and incubation of NanH aliquots for 24 and 48 h at room temperature, 37 °C and 55°C. As always, an undigested sample of SLe^X was analysed as a negative control, but this time a digest was also set up to contain 'untreated NanH'. Untreated NanH refers to NanH that has been stored at its regular temperature of 4 °C, in which it was believed to remain stable for up to 4 months (personal communication, A. Frey – Stafford research group).

NanH did not lose activity after 2x FT cycles, when the SLe^X digest was compared to the NanH untreated SLe^X digest trace, suggesting that NanH remained stable. However, after 5x FT cycles the

NanH and SLe^X digest went to completion, which was not consistent with the 2x FT data or untreated data, where some undigested SLe^X remained. Overall, it can be said that NanH will retain almost 100 % of its activity even after 5x FT cycles (Figure 3.16a). When NanH was stored at RT for 24 and 48 h, the digests with SLe^X almost went to completion, which was similar to the untreated NanH digest with SLe^X. This suggested that NanH can work just as well when stored at room temperature for up to 48 h as when it is stored at 4 °C (Figure 3.16b). When NanH was incubated at 37 °C, the digest with SLe^X went to completion after 48 h but remained partially incomplete after 24 h, which was consistent with the untreated enzyme (Figure 3.16c), which again showed that NanH is a stable enzyme. Unfortunately, after both 24 and 48 h at 55 °C, NanH lost all activity as it was unable to digest SLe^X at all (Figure 3.16d).

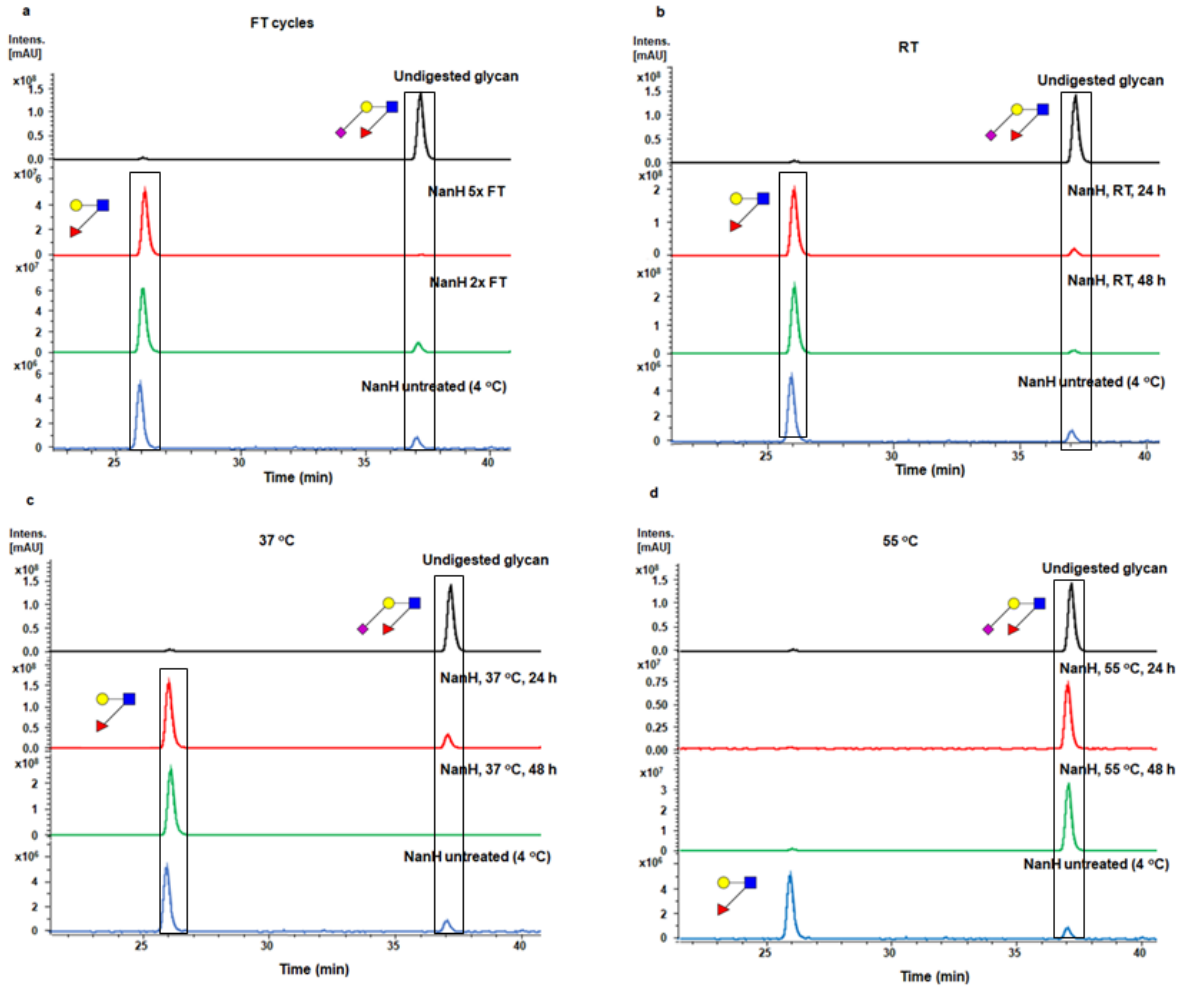


Figure 3.16: LC-FLR trace of procainamide labelled NanH digest with SLe^X when NanH was subjected to different conditions. (a) NanH digest of SLe^X after NanH was subjected to 2 or 5 FT cycles. Undigested SLe^X (black trace), 5x FT cycles (red trace), 2x FT cycles (green trace), untreated NanH (blue trace). **(b), (c)** and **(d)** represent NanH digests with SLe^X after NanH was subjected to RT, 37 °C and 55

°C for 24 and 48 h respectively. The black and blue traces represent undigested SLe^X and the digest using untreated NanH respectively. The red traces represent NanH digests with SLe^X for 24 h at RT, 37 °C and 55 °C in (b), (c) and (d) respectively. The green traces represent NanH digests with SLe^X for 48 h at RT, 37 °C and 55 °C in (b), (c) and (d) respectively.

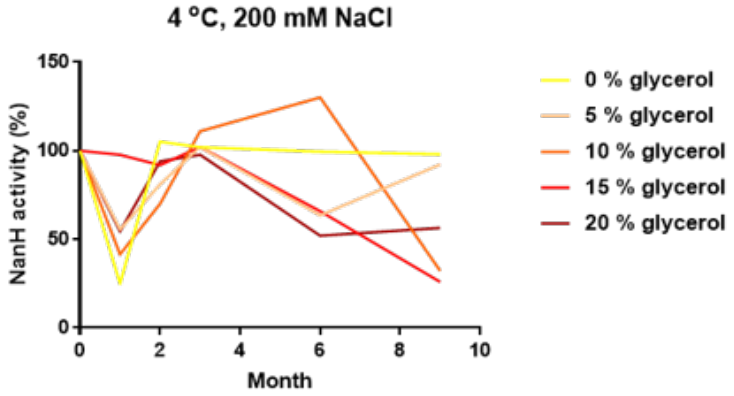
3.2.8 Long term storage studies

Prior to this project, anecdotal evidence suggested that NanH was stable when stored at 4 °C for up to approximately 4 months but lost activity when stored at -20 °C in 50 mM sodium phosphate, 200 mM NaCl, pH 7.4. Since this project involved the use of NanH and related mutants by the industrial partner Ludger Ltd, it was important to find a storage condition or set of conditions that would allow NanH to retain a large percentage (over 50 %) of its activity for the best part of a year. In this project, glycerol was used due to its cryoprotectant properties and ready availability.

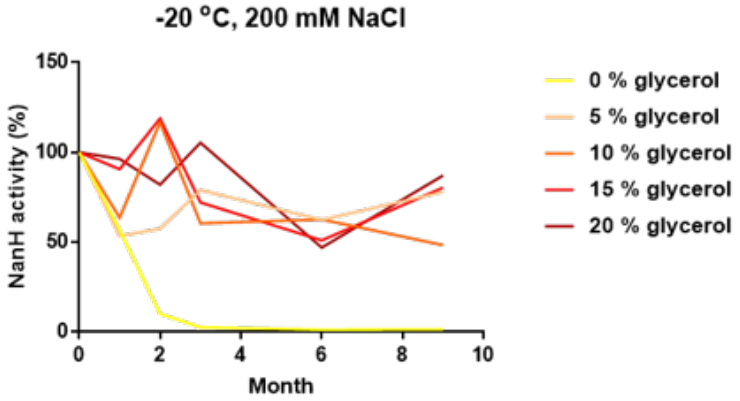
Pure NanH was dialysed to 50 mM sodium phosphate, 200 mM NaCl, pH 7.4 and divided into 20 µl aliquots at a final concentration of 1 µM including 0, 5, 10, 15 and 20 % glycerol. These aliquots were then stored at both 4 and -20 °C for up to 9 months and their activities were assessed using 2.5 nM enzyme in the MU-NANA assay (section 3.2.4) at 1, 2, 3, 6 and 9 months. The MU-NANA assay was chosen due to the speed of the assay, which enabled screening of the effect of each condition on NanH separately and calculation of the K_M and V_{max}. Identical storage was also performed at an increased salt concentration of 0.5 M, at both 4 and -20 °C, including 0, 5, 10, 15 and 20 % glycerol.

For each condition, the k_{cat}/K_M was determined and plotted as percentage activity when compared to the k_{cat}/K_M of NanH at 0 month (100 %) (Figure 3.17). The data clearly shows that when stored at either salt concentration at -20 °C, glycerol is advisable, as after 2 months, all enzyme activity has been lost when 0 % glycerol is used. All conditions appear to show a gradual decrease in activity as the months pass, however it can be confidently said that in most of these conditions, storage in the presence of 10, 15 and 20 % glycerol allow retention of approximately 50 % of the activity that is seen at 0 months. Surprisingly even the enzyme stored with 0 % glycerol in the 4 °C conditions still retains more than 50 % activity after 9 months, which suggests freeze-thawing from -20 °C with 0 % glycerol is not advisable for this enzyme.

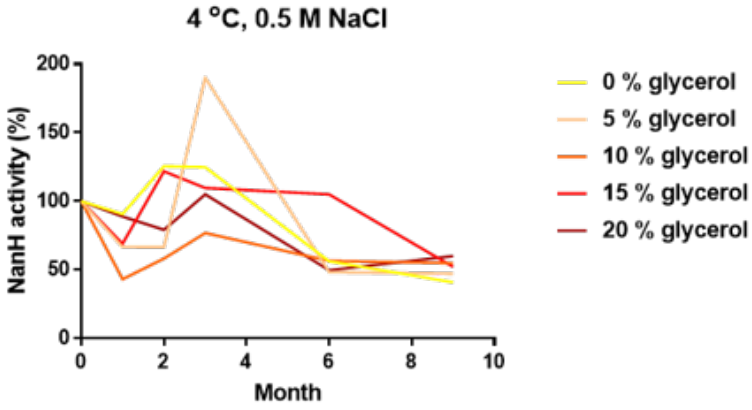
a


4 °C, 200 mM NaCl
NanH activity (%)
150
100
50
0
0
2
4
6
8
10
Month
0 % glycerol
5 % glycerol
10 % glycerol
15 % glycerol
20 % glycerol

b


-20 °C, 200 mM NaCl
NanH activity (%)
150
100
50
0
0
2
4
6
8
10
Month
0 % glycerol
5 % glycerol
10 % glycerol
15 % glycerol
20 % glycerol

c


4 °C, 0.5 M NaCl
NanH activity (%)
200
150
100
50
0
0
2
4
6
8
10
Month
0 % glycerol
5 % glycerol
10 % glycerol
15 % glycerol
20 % glycerol

d

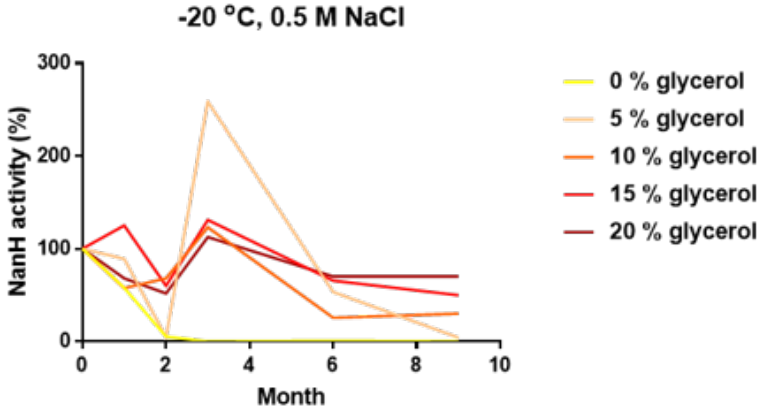

-20 °C, 0.5 M NaCl
NanH activity (%)
300
200
100
0
0
2
4
6
8
10
Month
0 % glycerol
5 % glycerol
10 % glycerol
15 % glycerol
20 % glycerol

Figure 3.17: Storage assay data shown as NanH % activity. MU-NANA reaction kinetics was performed for each individual storage condition at 1, 2, 3, 6- and 9-months post purification. The k_{cat}/K_M was calculated and plotted as a percentage when the 0-month value was classed as 100 %. **(a)** represents the aliquots stored at 4 °C, 200 mM NaCl **(b)** -20 °C, 200 mM NaCl, **(c)** 4 °C, 0.5M NaCl and **(d)** -20 °C, 0.5 M NaCl.

3.2.9 NanH accelerated degradation studies

It was also important to determine the activity of NanH after subjecting the protein to increased temperatures. If Ludger Ltd are interested in using NanH in the sialidase kits that they sell, they need to know that NanH will remain active after delivery to customers. Aliquots of 20 µl NanH at 1 µM were heated to 37 and 42 °C for 24 and 48 h. The enzyme was then diluted to 2.5 nM and ran through the MU-NANA assay (section 3.2.4) to enable the production of Michaelis-Menten curves to determine the activity compared to NanH that remained at 4 °C for 1 week (Figure 3.18, Table 3.6).

It is clear from the data that the protein degrades rapidly at temperatures above 37 °C. After 24 h at 42 °C, NanH loses approximately 92 % of its activity when comparing the k_{cat}/K_M to that at 4 °C, which then showed 100 % activity loss after 48 h at 42 °C. Heating NanH to 37 °C did not have a substantial effect on enzyme activity as after 48 h, there was a slightly increased V_{max} at 6.9 µmol min^{-1} mg^{-1} NanH compared to 6.0 µmol min^{-1} mg^{-1} NanH when at 4 °C for 1 week. However, when comparing the k_{cat}/K_M, NanH was slightly more efficient at 4 °C than 37 °C with values of 75.7 and 68.8 min^{-1} μM^{-1} respectively. Overall, it can be concluded that NanH is stable up to 37 °C for 48 h.

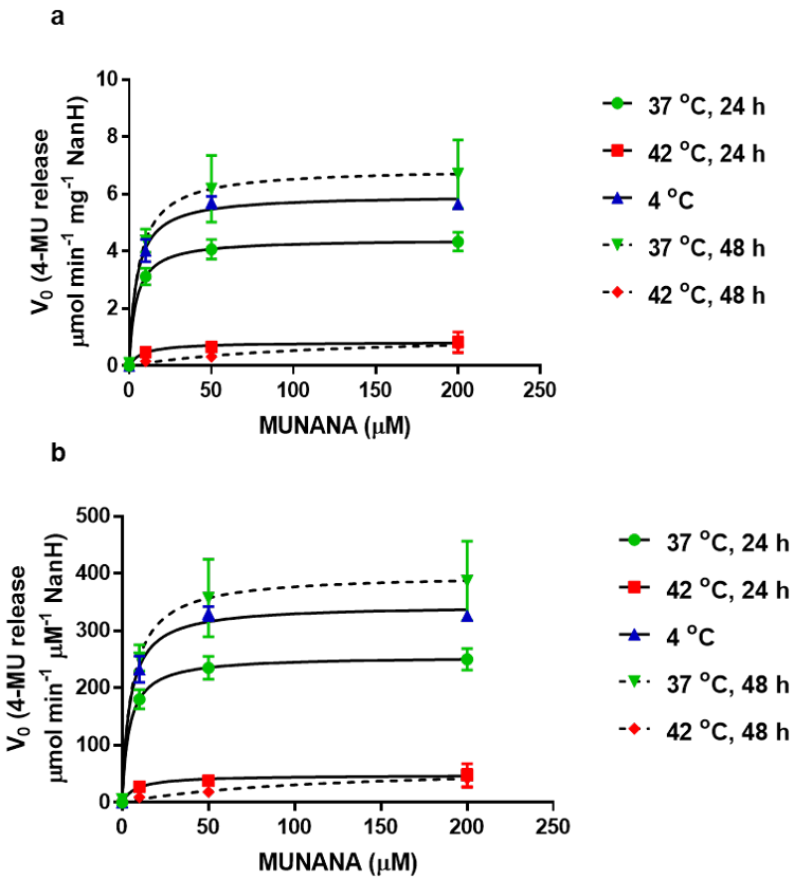


Figure 3.18: Michaelis-Menten curves representing the effect of temperature and time on NanH. (a) 2.5 nM NanH, that had been incubated at different temperatures and for 24 or 48 h, was assayed with increasing concentrations of MU-NANA from 0 to 200 µM. The reaction was stopped at 1-, 2- and 3-minute intervals and the rate of reaction (4-MU release) was determined by applying the raw data to a standard curve of known 4-MU concentrations. The rate was plotted against [MU-NANA] in a Michaelis-Menten curve from which the V_{max} and K_M could be calculated (Table 3.6). **(b)** The 4-MU release (per µM NanH) was plotted against [MU-NANA] enabling the k_{cat} to be calculated and in turn the catalytic efficiency (k_{cat}/K_M) (Table 3.6). In both cases the data represents the mean of three repeats, with the error bars representing the standard deviation.

	37 °C, 24 h	37 °C, 48 h	42 °C, 24 h	42 °C, 48 h	4 °C
V_{max} (µmol min^{-1} mg^{-1})	4.4 ± 0.1 (4.3 to 4.6)	6.9 ± 0.3 (6.4 to 7.5)	0.8 ± 0.1 (0.7 to 1.0)	1.1 ± 0.2 (0.8 to 2.3)	6.0 ± 0.1 (5.7 to 6.3)
K_M ([MU-NANA], µM)	4.2 ± 0.5 (3.2 to 5.4)	5.8 ± 1.3 (3.4 to 8.8)	8.1 ± 3.2 (2.8 to 17.7)	110.1 ± 49.2 (44.7 to 395.1)	4.6 ± 0.7 (3.1 to 6.3)
k_{cat} (min^{-1})	255.3 ± 4.8 (245.7 to 265.2)	398.6 ± 15.4 (368.1 to 430.6)	47.8 ± 3.8 (40.4 to 56.3)	64.8 ± 13.0 (45.4 to 131.4)	344.6 ± 8.3 (326.5 to 363.2)
k_{cat}/K_M (min^{-1} µM^{-1})	60.9	68.8	5.9	0.6	75.7

Table 3.6: Reaction kinetics values obtained from the Michaelis-Menten curves of the NanH digests of MU-NANA, when NanH had previously been subjected to high temperatures for differing time periods. The V_{max}, K_M, k_{cat} and k_{cat}/K_M values are presented along with the associated error (standard deviation) for the reaction kinetics of MU-NANA and NanH. The values in parentheses display the 95 % confidence intervals.

3.3 Discussion

3.3.1 Production of pure protein

NanH expression using 0.5 L *E. coli* BL21 Origami cells, followed by purification using a 5ml Ni^{2+} Fast Flow column, produced protein yields (10 mL at 4 mg mL^{-1}) that were high enough to use for future structural studies and the biochemical assays discussed. This purification method was easily repeated throughout the project, producing yields between 10-15 mL at 1.0-4.0 mg mL^{-1} from expression using 0.5 L of *E. coli* cells. Purified NanH (approximately 57.4 kDa) always had above 95 % purity when analysed by SDS-PAGE, which is an important trait to have for protein crystallisation prior to structural studies. Any impurities seen in the SDS-PAGE could have created defects in the crystal lattice and prevented the crystallisation of NanH.

3.3.2 The multimeric form of NanH

Analysis of the multimeric form of NanH initially suggested that the protein is predominantly dimeric in solution through SEC analysis. Native-PAGE data suggested the predominant form is tetrameric, which was rather unexpected since it was the dimeric fractions from SEC analysed on NativePAGE. A sample of freshly purified NanH was then analysed by SEC-MALLS, which confirmed the SEC result. Therefore, it is likely that the *in situ* form of NanH is dimeric.

The Protein Data Bank (PDB) was screened for bacterial sialidases (CAZy - GH33) and structures for 10 were found. 8 of these sialidases were described as biological monomers from their PDB entry which includes the sialidases from *Micromonospora viridifaciens* (PDB: 1EUU), *S. typhimurium* (PDB: 3SIL), *V. cholerae* (PDB: 1KIT), *C. perfringens* (PDB: 2VK5), *R. gnavus* (PDB: 4X47) and *S. pneumoniae* (PDB: 2W20, 2VW1, 4YW4). *Parabacteroides distasonis* sialidase (PDB: 4FJ6) is presented in the PDB as a biological dimer and has high sequence similarity (73 %) to NanH. Unfortunately, there are not yet any publications that discuss the crystal model or biochemical characterisation of this enzyme so there is not yet any evidence for cooperativity between subunits. *B. thetaiotaomicron* sialidase, BTSA (PDB: 4BBW), also has high sequence similarity (65 %) to NanH, although it has been described as a biological monomer in the PDB. However, when the structure of BTSA was analysed by PDBePISA at EMBL-EBI it was suggested that BTSA could perhaps be a biological dimer. Since no cooperativity was observed in biochemical assays, the reason NanH takes a dimeric form is currently unknown. Solving the crystal structure of NanH (Chapter 4) has provided further insight into the multimeric form, as it has identified how the two protein chains could perhaps be interacting and therefore provided ideas upon why NanH could be a biological dimer.

3.3.3 Reaction kinetics against synthetic substrate 4-Methylumbelliferyl N-acetyl-α-D-neuraminic acid

NanH showed sialidase activity with MU-NANA with a V_{max} of 13.8 µmol min^{-1} mg^{-1} NanH, a K_M of 6.2 µM, a k_{cat} of 799.6 min^{-1} and k_{cat}/K_M of 129.2 min^{-1} $µM^{-1}$ when in 50 mM sodium phosphate, 200 mM NaCl, pH 7.4 (the condition used for biochemical assays throughout the project). It is often difficult to compare the activity of different enzymes in nature as they perform vastly different functions with rates differing by orders of magnitude. However, since many sialidase studies involve kinetics studies using MU-NANA, some brief comparisons can be made. Purified NanH from *S. typhimurium* had the highest k_{cat} and k_{cat}/K_M when measured using MU-NANA (2700 s^{-1} and 10.8 s^{-1} $µM^{-1}$) (Table 3.7) (Hoyer *et al*., 1991). The study presented in this thesis has shown that *T. forsythia* NanH has a higher binding affinity (lower K_M) when compared with that of NanH of *S. typhimurium*, RgNanH of *R. gnavus* and BTSA of *B. thetaiotaomicron* respectively (Table 3.7) (Tailford *et al*., 2015; Park *et al*., 2013). One obvious difference between the sialidases in Table 3.7 is that *T. forsythia* NanH is a highly likely to have a dimeric structure whereas the others are suggested to be monomers. Perhaps the difference in multimeric form is what leads to the increased k_{cat}/K_M seen for *T. forsythia* NanH compared to RgNanH and BTSA. However, the Michaelis-Menten kinetics studies (sections 3.2.4 and 3.2.6) suggested no evidence of cooperativity, which suggests that dimeric structure offers no advantage. Perhaps the 'novel' CBM domain of *T. forsythia* NanH is providing additional or improved binding and therefore a higher k_{cat}/K_M derived from the lower K_M (Frey *et al*., 2018).

Organism	Enzyme	K_M (µM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} $µM^{-1}$)	Reference
S. typhimurium	NanH	250.0	2700.0	10.8	Hoyer et al., 1991
R. gnavus	RgNanH	590.0	103.5	0.2	Tailford et al., 2015
B. thetaiotaomicron	BTSA	160.0	146.7	0.9	Park et al., 2013
T. forsythia	NanH	6.2	13.3	2.1	This study

Table 3.7: Kinetics values for sialidase activity measured with MU-NANA.

3.3.4 NanH inhibition using common sialidase inhibitors

Zanamivir and oseltamivir are approved therapeutics against neuraminidases of influenza A viruses. The inhibition data collected in this study showed that DANA was the most potent inhibitor of NanH, followed by oseltamivir and then zanamivir. The IC_{50} values obtained showed that in general, these inhibitors do not inhibit NanH as efficiently (micromolar and millimolar range, Table 3.2) as they do

their original viral targets, such as neuraminidases from the Influenza A virus (nanomolar range) (Gubareva *et al*., 2001). Similar situations are seen in inhibition studies with DANA from other bacterial sialidases, with BTSA from *B. thetaiotaomicron,* NanH from *S. pneumoniae*, NanB from *S. pneumoniae* and NanH from *S. typhimurium* being found to have K_i values of 35 µM, 2µM, 0.3 mM and 0.38 mM respectively (Park *et al*., 2013; Xu *et al*., 2008). VCNA from *V. cholerae* was inhibited in the micromolar range by both oseltamivir and zanamivir with IC_{50} values of 144 µM and 52 µM (Quosdorf *et al*., 2017). Effective inhibitors against bacterial sialidases are still being pursued, which in some studies still involves designing several derivatives of DANA (Slack *et al*., 2018).

3.3.5 NanH specificity against glycosidic linkages

The experiments in section 3.2.6 proved that NanH is a dual specific sialidase that hydrolyses both the α2,3- and α2,6-glycosidic linkages. Purified NanH from *T. forsythia* was discovered to hydrolyse both α2,3- and α2,6-linked sialic acid with k_{cat}/K_M values of 8654.8 and 3920.3 min^{-1} mM^{-1} respectively, which suggested a preference for α2,3-linked sialic acid. This finding agreed with that of a study by Thompson and colleagues, who claimed that NanH present in *E. coli* cell lysate was dual specific, with a preference for α2,3-linked sialic acid (Thompson *et al*., 2009). The aim of this project was to produce a sialidase specific for either α2,3- or α2,6-linked sialic acid, which means it has been necessary to make modifications upon NanH in order to change the specificity (Chapter 5). There are no known bacterial sialidases in the literature that are specific to α2,6-linked sialic acid, with the majority being dual specific and only a few that are specific to α2,3-linked sialic acid (Table 3.8).

It was noticed that the catalytic activity of NanH for both $SLe^{A/X}$ were four-fold less than that of 3-SL when all three of these ligands contain α,2,3-linked sialic acid (Figure 3.19), showing the importance of the underlying sugar residues in sialidase activity. 3-SL is composed of a terminal sialic acid, followed by a galactose and glucose, whereas sialic acid in SLe^X is followed by galactose, GlcNAc and fucose and in SLe^A by a galactose, GlcNAc and fucose.

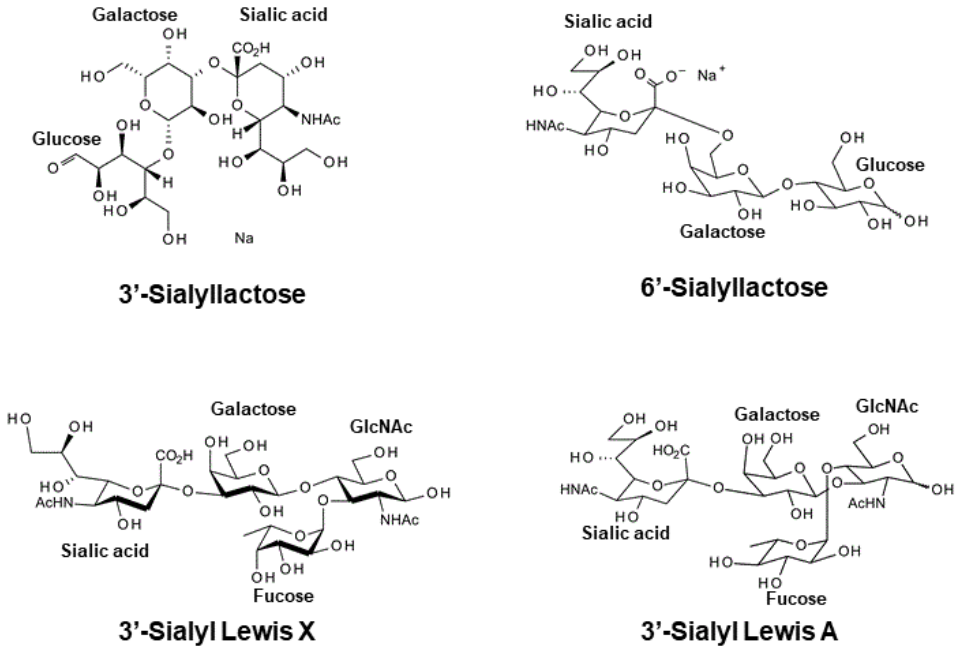


Figure 3.19: The chemical structures of substrates used in the assays described in section 3.2.6.

There are no known bacterial sialidases in the literature that are specific to α2,6-linked sialic acid, with the majority being dual specific and only a few that are specific to α2,3-linked sialic acid (Table 3.8). The α2,3 specific sialidases listed in Table 3.8, differ from NanH of *T. forsythia,* which is described as a hydrolytic sialidase because it simply cleaves and releases sialic acid. The α2,3 specific sialidases are either trans-sialidases, which transfer the cleaved sialic acid to a glycoconjugate or IT-sialidases, which produce 2,7-anhydro-Neu5Ac rather than sialic acid (Chou *et al*., 1996; Gut *et al*., 2008; Tailford *et al*., 2015). Perhaps NanH of *T. forsythia* is too evolutionary different to be modified to an α2,3 specific sialidase, as the hydrolytic, trans and IT-sialidases have evolved to have different functions. NanH of *S. typhimurium* is a hydrolytic sialidase and has a preference for α2,3-linked sialic (Hoyer *et al*., 1991). If the structure of NanH from *T. forsythia* was known and compared to that of NanH from *S. typhimurium*, modifications could be made to *T. forsythia* NanH in attempt to alter the specificity towards α2,3-linked sialic.

Organism	Enzyme	Linkages targeted	Preference	Reference
Streptococcus pneumoniae (respiratory tract)	NanA	α2,3; α2,6	No noticeable preference	Xu *et al*., 2008; Gut *et al*., 2008
	NanB	α2,3	α2,3	Xu *et al*., 2008; Gut *et al*., 2008
	NanC	α2,3	α2,3	Xu *et al*., 2011; Owen *et al*., 2015
Vibrio Cholerae (GI tract)	VCNA	α2,3; α2,6	Unknown	Eneva *et al*., 2015
Clostridium perfringens (GI tract)	NanH	α2,3; α2,6	α2,3	Li and McClane, 2014
	NanI	α2,3; α2,6	α2,3	Li and McClane, 2014
	NanJ	α2,3; α2,6	α2,6	Li and McClane, 2014
Salmonella typhimurium (GI tract)	NanH	α2,3; α2,6	α2,3	Hoyer *et al*., 1991
Corynebacterium diphtheriae (GI tract)	NanH	α2,3; α2,6	α2,6	Kim *et al*., 2010
Bacteroides thetaiotaomicron (GI tract)	BTSA	α2,3; α2,6	α2,6	Park *et al*., 2013
Ruminococcus gnavus (GI tract)	RgNanH	α2,3	α2,3	Tailford *et al*., 2015
Pasturella multocida (GI tract)	NanH	α2,3; α2,6	α2,3	Mizan *et al*., 2000
	NanB	α2,3; α2,6	α2,6	Mizan *et al*., 2000
Tannerella forsythia (oral cavity)	NanH	α2,3; α2,6	α2,3	This study Thompson *et al*., 2009
Streptococcus oralis (oral cavity)	Unknown	α2,3; α2,6	α2,3	Byers *et al*., 2000
Trypanasoma cruzi (protozoa, blood)	TcTS	α2,3	α2,3	Vandekerckhove *et al*., 1992
Macrobdella decora (leech)	NanL	α2,3	α2,3	Chou *et al*., 1996

Table 3.8: Sialic acid-linkage targets and preferences of sialidases from a range of organisms.

3.3.6 NanH activity on small and complex glycans relevant to glycoanalytical pipelines

One of the experiments performed at Ludger Ltd was to investigate the specificity of NanH using a different substrate to those used in Sheffield, as discussed in section 2.3.7.1. The data showed NanH had activity against both α2,3- and α2,6-linked sialic acid after digests with the small glycans, $3SLe^X$ and FA2G2S2. NanH was unable to hydrolyse mannose residues from MAN mix glycans or fucose residues (data not shown) from an FA2G2 standard, which has so far suggested that NanH does not have activity towards alternative glycans. NanH was able to hydrolyse sialic acid from bi- and tri-antennary plasma-N glycans which suggested it is capable of targeting complex glycans. Overall NanH performed well against the current dual specific sialidase used at Ludger Ltd (E0001) in the conditions they were tested under. Therefore, NanH has potential to be used in Ludger Ltd assays and could be even be more effective than E0001 if concentrations higher than 10 nM are used (as stated before, the concentration of E0001 was unknown and could have been much higher than that of NanH).

3.3.7 Stability and accelerated degradation studies

The stability of a commercial enzyme is important because this determines the way in which it can be transported. For example, a study showed that EcoRI and HindIII restriction enzymes were stable and retained full activity for up to 8 weeks and 6 months respectively, at 37 ^{o}C in the manufacturers buffer (Clark et al., 2000). This suggested that these enzymes have potential to be transported by cheaper methods such as conventional mail (Clark et al., 2000). It was found that NanH remained stable after 48 h at both 37 ^{o}C and RT in the accelerated degradation study (section 3.2.8) and the stability study (section 3.2.6.3) and therefore ideal for use by Ludger Ltd. This assay of stability of recombinant NanH could relate to NanH being a stable secreted virulence and scavenging enzyme in the mouth. Perhaps the studies in section 3.2.9 and 3.2.7.3 needed a longer incubation period, however it is likely NanH could be transported by conventional mail, at least over short distances in the UK and Europe. In future, NanH could also be stored at -20 ^{o}C with higher concentrations of glycerol in the solution, as manufactured restriction enzymes usually contain at least 50 % glycerol, to provide protection from freeze-thawing, such as those produced by New England Biolabs (neb.com).

3.4 Conclusion

Overall NanH would be a more than acceptable enzyme to use as a commercial product. The methods used in NanH expression and purification allowed a high yield of pure and soluble protein to be produced. Storage studies have shown that NanH can be stored cheaply in glycerol at both 4 ^{o}C and -20 ^{o}C for up to 9 months and retain 50 % or more of the original activity. Evidence in this chapter has proved that NanH is a highly active, dual specific sialidase, which can act upon bi- and tri-antennary glycans, as well as small glycans. The high yield and purity of NanH have made it a good candidate for crystallisation. In chapter 4, a crystal structure of NanH was solved, which has enabled the understanding of the catalytic site and mechanism and therefore highlighted residues that could be modified to alter the specificity and hopefully produce a sialidase the fits the needs of Ludger Ltd.

Chapter 4

Solving the crystal structure of NanH and NanH variants

4.1 Introduction

4.1.1 Bacterial sialidases – key features

Crystal models for 10 bacterial sialidases (GH33) have been deposited in the PDB either in the apo form (no compound bound) or in complex with inhibitors (Table 4.1). The models include sialidases from the three exo-sialidase families: hydrolytic sialidases, trans-sialidases and IT-sialidases. The catalytic domain of bacterial sialidases share the same architecture and fold into 6-bladed β-propellers, some of which have additional domains at the N- or C-terminus (Kim *et al*., 2011; Crennell *et al*., 1994; Xu *et al*., 2008).

Organism	Name	CAZy	Mechanism	CBM present	PDB entry
Clostridium perfringens	NanI	GH33	Exo-sialidase: Hydrolytic	✓	2VK5, 5TSP
Micromonospora viridifaciens	NedA	GH33	Exo-sialidase: Hydrolytic	X Extra domains but does not seem to be SA specific CBM	1EUU, 1EUR, 1EUS, 1EUT
Streptococcus pneumoniae	NanA	GH33	Exo-sialidase: Hydrolytic	✓	2W20, 2VVZ
	NanB	GH33	Anhydrosialidase: IT-sialidase	✓	2VW0, 2VW1, 2JKB
	NanC	GH33	Exo-sialidase: Hydrolytic	✓	5F9T
Vibrio cholerae	VCNA	GH33	Exo-sialidase: Hydrolytic	✓	1KIT
Salmonella typhimurium	STSA	GH33	Exo-sialidase: Hydrolytic	X	3SIL, 1DIL, 1DIM, 2SIL, 2SIM
Parabacteroides distanosis	N/A	GH33	Exo-sialidase: Hydrolytic	✓	4FJ6
Ruminococcus gnavus	RgNanH	GH33	Anhydrosialidase: IT-sialidase	✓	4X47, 4X49, 4X4A, 4X6K
Bacteroides thetaiotaomicron	BTSA	GH33	Exo-sialidase: Hydrolytic	✓	4BBW

Table 4.1: Crystal structures of bacterial sialidases that have been deposited in the PDB. The catalytic mechanism and whether a CBM domains are present are also listed. SA stands for sialic acid.

Strictly conserved residues cluster at the centre of the propeller and form the active site. These residues include an arginine triad, which binds to the carboxyl group of sialic acid, a Tyr/Glu

nucleophilic pair and an aspartate that acts as an acid/base catalyst (Buschiazzo and Alzari, 2008; Crennell *et al*., 1994). It is in this region where terminal sialic acids bind and the glycosidic linkage is hydrolysed, releasing sialic acid from the glycan. Other conserved domains include Asp-box motifs (S/T-x-D-x-G-T-W/F, where 'x' can represent any amino acid), which are repeated 4-5 times throughout the sialidase and found on the outskirts of the propeller blades, remote from the active site. The function of Asp-box motifs is unknown, but they are predicted to stabilise the protein fold (Kim *et al.,* 2011). A conserved Y/FRIP is also conserved amongst bacterial sialidases and again, the function is unknown, but the arginine is one of the triad members (Kim *et al*., 2011).

4.1.2 Bacterial sialidases – carbohydrate-binding motifs

GH33 bacterial sialidases can be preceded, flanked or followed in the protein structure by additional domains. Some of these domains have lectin-like functions and help mediate adherence to the glycoconjugate of interest and therefore enhance the catalytic activity (Boraston *et al*., 2004). These domains are referred to as CBM's and are composed of approximately 200 residues. The sialidases of *V. cholerae* (VCNA), *S. pneumoniae* (NanA and NanB) and *C. perfringens* (NanJ) contain lectin-like domains (CBM40 classified) (Crennell *et al*., 1994; Moustafa *et al*., 2004; Boraston *et al*., 2007; Yang *et al*., 2015). Models have been solved for each of the CBM40s of VCNA, NanA and NanJ in complex with 3-/6-SL and the binding pockets appear to coordinate the sialic acid portion of the substrate (Moustafa *et al*., 2004; Boraston *et al*., 2007; Yang *et al*., 2015). CBM40 domains fold into β-sandwiches composed of two sheets, each between 5 and 7 antiparallel β-strands (Owen *et al*., 2017; Yang *et al*., 2015).

4.1.3 Bacterial sialidases – proposed mechanism

Two mechanisms have been proposed for the catalysis of the glycosidic linkage by sialidases: one based on the mechanism of the influenza neuraminidase (Taylor and von Itzstein, 1994; von Itzstein, 2007) and the second based on *Trypanosoma cruzi* sialidase (TcTS) (Amaya *et al*., 2004). In both cases, sialic acid is orientated in the active site by its carboxyl group, which can form a hydrogen bonding network with the guanidinium groups of the arginine triad. The carboxylate group of sialic acid transforms from an axial to a pseudoequatorial position due to charge-charge interactions with the arginine triad and steric constraints with the nearby conserved tyrosine (Taylor and von Itzstein, 1994). The extensive hydrogen bond network binding Neu5Ac to the sialidase compensates for the free energy lost because of these distortions. In solution, the sialic acid pyranose ring adopts a $^{2}C_{5}$ chair conformation. The sialic acid binding process to sialidases leads to distortion of the pyranose ring to a distorted boat conformation (Figure 4.1) (Taylor and von Itzstein, 1994).

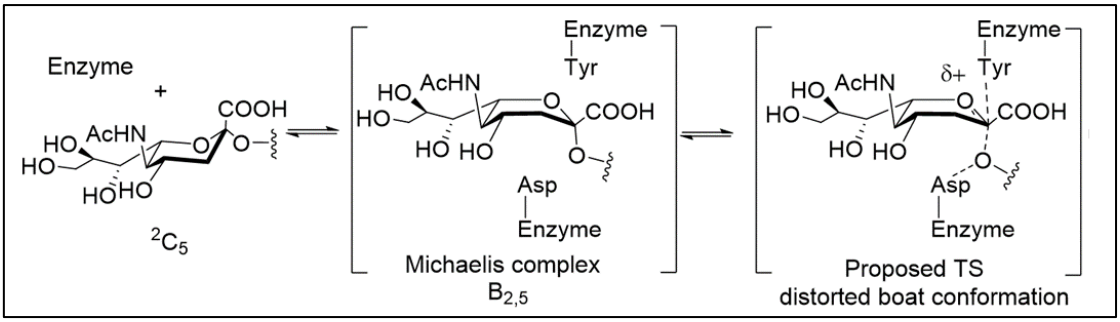


Figure 4.1: Sialic acid ring distortion during binding to sialidase. This figure is based upon influenza A neuraminidase (Colombo *et al*., 2018).

The original proposed mechanism based on the influenza neuraminidase states that the conserved aspartate sidechain (or perhaps the aspartate via an intervening water molecule) acts as a general acid catalyst, facilitating cleavage of the glycosidic bond between sialic acid and the sub-terminal aglycon via the donation of a proton (Taylor and von Itzstein, 1994). The resulting oxocarbenium reaction intermediate is stabilised by a conserved deprotonated tyrosine (deprotonated through a nucleophilic attack involving the conserved glutamate from the Tyr/Glu pair). A water molecule then acts as a catalytic nucleophile, attacking the positive charge on the C2 of the oxocarbenium, forming sialic acid. This attack is facilitated by base catalysis from the conserved aspartate, which regains its proton (Chong *et al*., 1992; Burmeister *et al*., 1993; Kati *et al*., 2001; Newstead *et al*., 2008).

An alternative mechanism (Figure 4.2) was proposed for the *Trypanosoma cruzi* sialidase (Amaya *et al*., 2004). This involves trapping the oxocarbenium ion to additionally form a glycosyl-covalent intermediate. In this case, the conserved tyrosine acts as the catalytic nucleophile with the idea that the use of a negatively charged carboxylate as a nucleophile would be disfavoured, as sialic acid itself is negatively charged and also that no carboxylate was suitably placed (Buschiazzo and Alzari, 2008). Distortion of the sialic acid pyranose ring upon substrate binding places the leaving group (ROH) in a pseudo-axial position (Amaya *et al.,* 2004). This optimises proton donation from the conserved aspartate (perhaps activated by a water molecule), which is in hydrogen bonding distance of the glycosidic oxygen and therefore acts as the acid catalyst in the initial glycosyl-enzyme formation. The conserved tyrosine then acts as an atypical catalytic nucleophile, after activation by the adjacent conserved glutamate (base catalyst) (Amaya *et al*., 2004; Lee, 2013). During the nucleophilic attack, the conserved aspartate acts as a general base catalyst and sialic acid is released (Amaya *et al*., 2004; Lee, 2013). The conserved tyrosine has now also been considered as the key catalytic residue that may act as a nucleophile in influenza A and B neuraminidases (Vavricka *et al*., 2013) and this mechanism is widely accepted for bacterial sialidases (Juge *et al*., 2016).

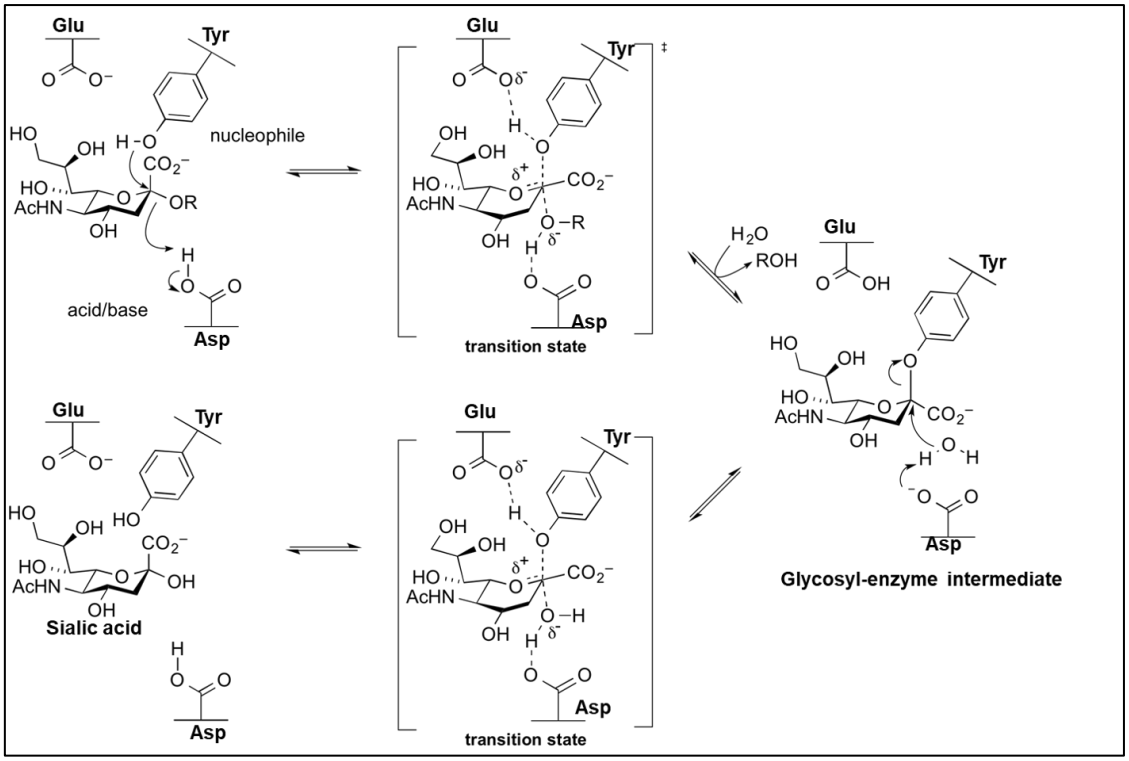


Figure 4.2: The enzymatic mechanism of glycosidase hydrolases based upon the mechanism of *T. cruzi* sialidase. This figure was sourced from CAZypedia, with permission to use granted by the curator of the page (Withers and Williams "Glycoside Hydrolases" in *CAZypedia,* available at URL https://www.cazypedia.org/index.php/Glycoside_hydrolases#Alternative_nucleophiles, accessed 30th April 2019). Small text additions were made to include the amino acid names, 'sialic acid' and 'glycosyl-enzyme intermediate'.

4.1.4 Bacterial sialidases in complex with inhibitors and ligands

Sialidase inhibitors, such as zanamivir, oseltamivir and DANA are structural mimics of sialic acid (Figure 1.8, section 1.3.7) (McKimm-Breschkin, 2013). Some of the sialidases listed in Table 4.1 also have had models solved in complex with these inhibitors. The carboxyl groups of these inhibitors interact with the conserved arginine triad and the N-acetyl tails sit in a conserved hydrophobic pocket, however the residues that form this pocket are not generally conserved (Xu *et al*., 2008). Unfortunately, there are no crystal models in the PDB of bacterial sialidases in complex with ligands such as 3- and 6-SL, although there are models with ligands bound to the CBM (Moustafa *et al*., 2004).

A study that determined the crystal structure of an influenza A NA has also been successful in solving the structure in complex with ligands (Zhu *et al*., 2012). For this to be possible, the researchers inactivated the NA by mutating the conserved catalytic aspartate to a glycine (D151G). This enabled

models of the D151G mutant to be solved in complex with 3'-sialyllactosamine (3'-SLN) and 6'-sialyllactosamine (6'-SLN), which contain α2,3- and α2,6-linked sialic acid respectively (PDB: 4GZW and 4GZX) (Zhu *et al*., 2012). The sialic acid portion of the ligands were bound in similar conformations to the inhibitors discussed above. Both the α2,3- and the α2,6-glycosidic linkages appeared to be sitting in the same position within the active site (Figure 4.3) (Zhu *et al*., 2012).

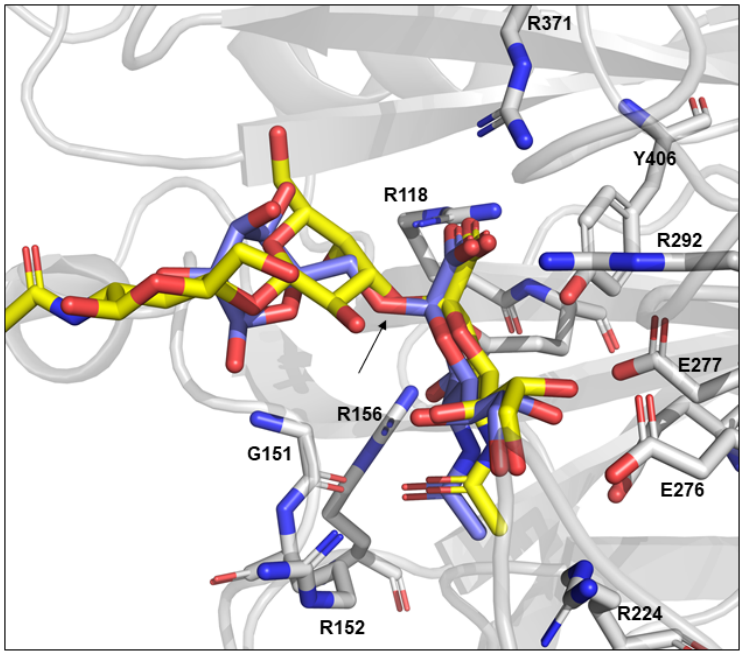


Figure 4.3: Influenza A NA (D151G variant) in complex with 3-SLN and 6-SLN. The neuraminidase protein backbone is in cartoon representation with key catalytic residues shown as grey sticks. 3-SLN is shown as yellow sticks and 6-SLN is shown as blue sticks with the glycosidic links highlighted by a black arrow. The protein backbone is only shown for the model in complex with 3-SLN (PDB: 4GZW), as little difference was seen between this and the backbone in the 6-SLN model (PDB: 4GZX).

4.1.5 Chapter aims

As stated in this introduction, the aim of this project was to engineer a sialidase to be specific to either α2,3- and α2,6-linked sialic acid. An understanding of a crystal structure of a sialidase in complex with substrates that contain either α2,3- and α2,6-linked sialic acid would greatly enable this. The data in Chapter 3 showed that *T. forsythia* NanH (a dual specific sialidase) would be a good crystallisation candidate since high yields of pure protein can be produced, which increases the chances of successful crystallisation. This chapter describes the attempts to solve the crystal model of wild-type (WT) NanH in complex with common sialidase inhibitors. The production of an inactive NanH variant is also presented, the purpose of which was to study it in complex with ligands, such as 3- and 6-SL, which

contain α2,3- and α2,6-linked sialic acid respectively. It was hoped that these models would enable the understanding of where these inhibitors and substrates interact with the active site and therefore which residues could be of importance in the specificity of the enzyme.

4.2 Results

4.2.1 The production of a NanH inactive variant

In order to crystallise a ligand within the active site of NanH it was crucial to produce a NanH inactive variant, otherwise a build-up of product (sialic acid) could be present in the crystallisation drop until reaction equilibrium is formed. SDM was employed to generate the single point mutant, NanH-D237A. D237 is likely to be the conserved catalytic aspartate that is heavily involved in sialidase activity. The aspartate is also seen in the catalytic site of influenza A virus neuraminidase and Zhu and colleagues showed that stripping this residue back to a glycine abolishes sialidase catalytic activity (Zhu et al., 2012). After confirmation of mutagenesis by sequencing, NanH-D237A-pET21a was expressed when incubated at 25 °C, 200 rpm for 18 h in *E. coli* BL21 Origami cells and had a level of over 95 % purity after French press and Ni-NTA chromatography (Figure 4.4) (sections 2.8). A 0.1 L culture of *E. coli* yielded approximately 0.2 mL of protein at 2.4 mg mL^{-1}, which was dialysed to 50 mM sodium phosphate, 200 mM NaCl pH7.4 to reduce salt and imidazole concentrations that were present in the purification buffer.

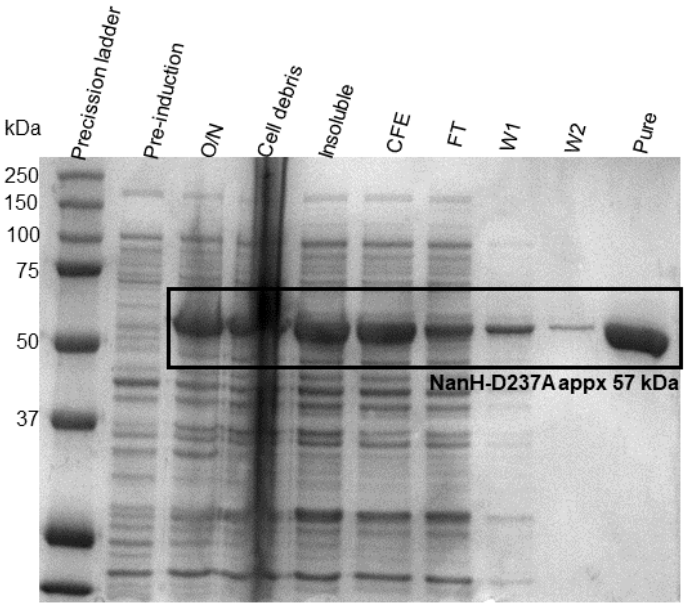


Figure 4.4: SDS-PAGE of NanH-D237A overexpression and small-scale purification. *E. coli* cells containing NanH-D237A-pET-21a were induced after reaching OD600 0.6**.** Cell samples were collected prior to induction (**pre-induction**) and 18 h after induction (**O/N**). The *E. coli* were lysed, and **cell debris** and **insoluble protein** were collected after two rounds of centrifugation. The cell free extract **(CFE)** was loaded to nickel resin for purification. The flow through **(FT)** and wash fractions **(W1, W2)** were collected before eluting the **pure protein** (section 2.8).

NanH-D237A was incubated in a 96-well plate with 0.1 mM of the synthetic sialidase substrate, MU-NANA, for 10 min at RT to quickly screen for activity, as in section 3.2.2 (Figure 4.5). No fluorescence was detected at concentrations of 10 and 100 nM NanH-D237A, in contrast with wild-type (WT) NanH which showed detectable levels of fluorescence at 5 nM concentrations of enzyme and 0.1 mM of MU-NANA (Figure 3.3). This confirmed that the D237A substitution had successfully inactivated NanH. NanH-D237A was assayed with increasing concentration of MU-NANA, as in section 3.2.4. No activity was detected, even after increasing the concentration of D237A from 2.5 nM to 100 nM. A small-scale TBA assay was also performed on 500 µM 3- and 6-SL to check NanH-D237A was inactive against differing substrates. Even when a concentration of 100 nM NanH-D237A was used in an O/N (approximately 18 h) digest, sialic acid release was not detected. This confirmed D237A was inactive.

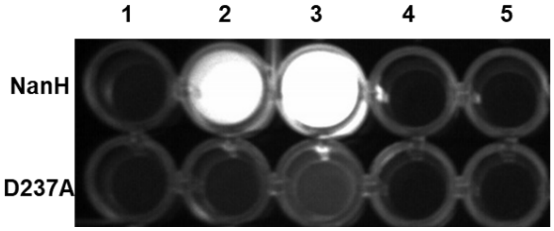


Figure 4.5: Visualisation of 4-MU release when MU-NANA is cleaved by NanH-D237A. NanH-D237A inactivity is represented by the lack of fluorescence under ultra-violet (UV) light. **(1)** 0.1 mM MU-NANA, **(2)** 10 nM enzyme + 0.1 mM MU-NANA, **(3)** 100 nM enzyme + 0.1 mM MU-NANA **(4)** 10 nM enzyme, **(5)** 100 nM enzyme. Enzyme concentrations were determined by the BCA assay.

Before attempting to solve the crystal model of NanH-D237A complexed with substrates, it was necessary to determine whether this variant was still able to bind substrates. This query was solved using a fluorescence-based tryptophan quenching assay, which relied on the presence of tryptophan residues in the NanH-D237A. Tryptophan residues become either exposed or enclosed during a conformational change in the protein upon ligand binding, which can be detected by an increase or decrease in fluorescence respectively. NanH-D237A was shown to bind both 3- and 6-SL with micromolar affinities of 304.8 ± 91.5 µM and 133.5 ± 48.3 µM respectively (Figure 4.6 Table 4.2).

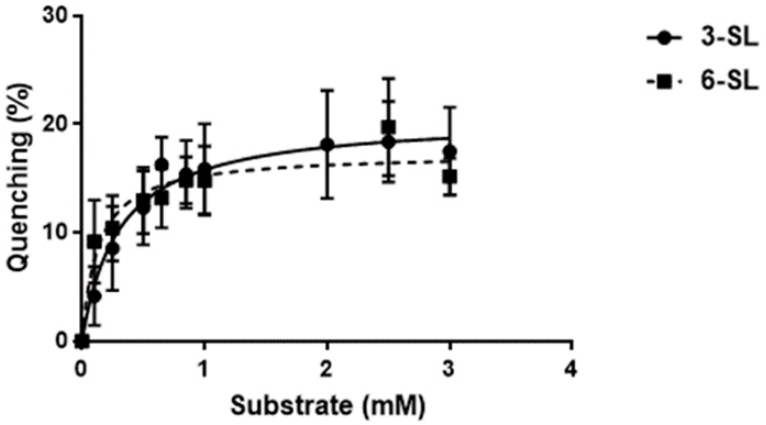


Figure 4.6: NanH-D237A ligand binding assay. 0.1 µM NanH-D237A was exposed to increasing concentrations of 3-/6-SL and the change in protein fluorescence (%) compared to a no ligand dataset was used to obtain the binding affinity (K_d) of 3-/6-SL with NanH-D237A. The data presented represents the mean of 3 repeats, with the error bars representing standard deviation.

Substrate	Binding affinity (K_d) (µM)
3-SL	304.8 ± 91.5 (164.2 to 532.1)
6-SL	133.5 ± 48.3 (50.3 to 285.6)

Table 4.2: The binding affinities of NanH-D237A with 3- and 6-SL. The K_d values are presented (± standard error). The values in parentheses represent the 95 % confidence intervals.

4.2.2 Crystallisation of NanH and the NanH inactive variant

The structures of NanH and the NanH-D237A variant were determined using X-ray crystallography. The crystallisation and model building for all proteins discusses in this chapter was carried out by Marianne Satur in the Department of Molecular Biology and Biotechnology, University of Sheffield. Purified NanH and NanH-D237A were dialysed into a low concentration salt buffer to prevent disruption of hydrogen bonding networks, and in turn the crystal lattice, during crystallisation. Generally, a protein concentration of 10 mg mL^{-1} or higher is a requirement for successful protein crystallisation. In order to produce protein crystals, NanH and NanH-D237A were trialled under the PACT, JCSG and Proplex screens (Appendix 2) (Molecular Dimensions), using a Matrix Hydra II PlusOne Robot (Thermo Fisher) or a Mosquito Robot (TTP Labtech), in 96 well sitting drop plates (Figure 4.7)

(section 2.13.1). Each screen contained 96 different conditions, which differ according to the precipitant present and its concentration, the buffer, the concentration and pH of the buffer and whether any salts or additives are present.

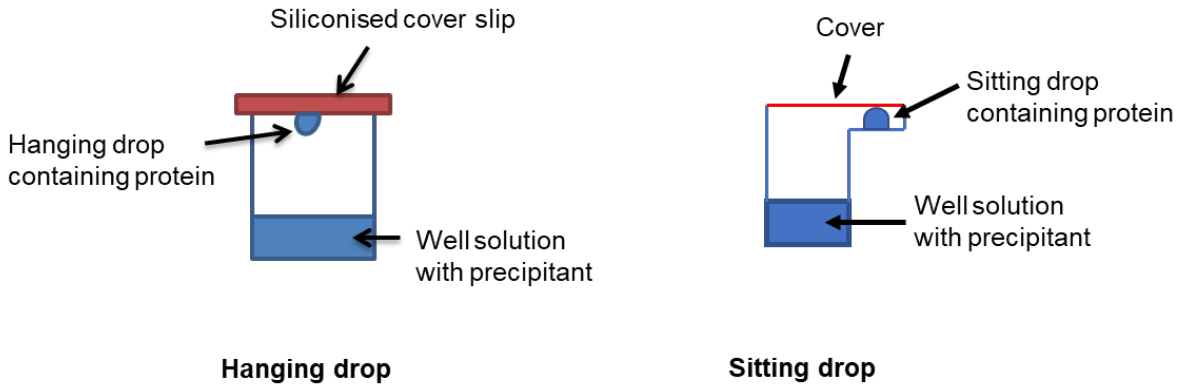


Figure 4.7: Hanging and sitting drop vapour diffusion set up.

NanH robot screens were performed using the apo enzyme (no compound bound) at a concentration of 40 mg mL^{-1}. Several crystal hits were produced within 24 hours in condition C4 of the JCSG screen (0.1 M HEPEs pH 7.0, 10 % $^{w}/_{v}$ PEG 6000) (Figure 4.8a). The JCSG C4 condition was optimised across a 24 well plate using the hanging drop vapour diffusion technique (Figure 4.7) to produce larger crystals. The optimisation involved varying the concentration of PEG 6000 (4-17 % $^{w}/_{v}$) and varying the pH of HEPES (pH 6.0-7.5). A droplet containing protein and well solution in a 1:1 ratio was placed in vapour equilibrium above the well solution that contained precipitant. Initially, the droplet had a lower concentration of precipitant than that of the well solution, resulting in a higher vapour pressure. This concentration gradient resulted in water leaving the protein droplet, which reduced in volume and increased the protein concentration. An equilibrium was reached when the reagent concentration in the droplet was the same as that in the well solution.

The optimisation led to formation of several NanH crystals, with the best selection (according to appearance and size) being produced in 0.1 M HEPES pH 6.0, 7 % $^{w}/_{v}$ PEG 6000 (Figure 4.8b). Prior to the collection of X-ray diffraction data, NanH crystals were passed for approximately 5 seconds through a cryoprotectant that was composed of the crystallisation solution with an addition of 15 % ethylene glycol.

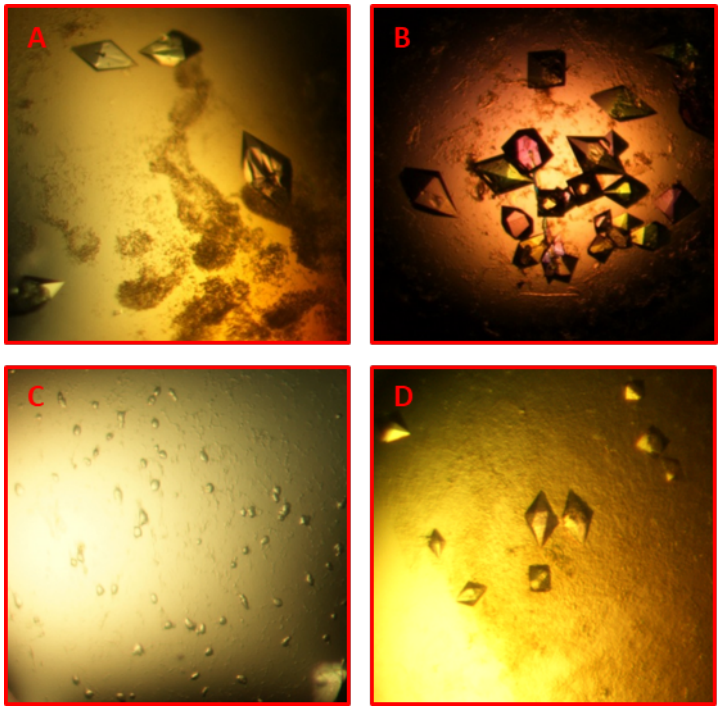


Figure 4.8: NanH-apo and NanH-D237A-apo crystals. (a) NanH-apo crystals grown in condition C4 of the JCSG robot screen. **(b)** NanH-apo crystals grown from the optimisation of condition C4 of the JCSG screen. **(c)** NanH-D237A-apo crystals grown in condition C4 of the JCSG robot screen. **(d)** NanH-D237A-apo crystals grown from the optimisation of condition C4 of the JCSG screen using the Formulator robot.

Additional NanH-apo crystal were grown using fresh protein at a later date, but this time at a reduced protein concentration of 20 mg mL^{-1} and also using 10 mM HEPES as opposed to 0.1 M (the reasons for which are discussed in section 4.2.3), which produced the same conformation of crystal, as in Figure 4.8b (not shown). These crystals were soaked in a cryoprotectant (crystallisation solution that contained 15 % ethylene glycol), which contained 5 mM oseltamivir acid, for between 5-30 sec and then stored appropriately (section 2.13.4) before X-ray data was collected. Oseltamivir was added to the cryoprotectant because the crystals degraded when initially soaked for longer periods in a crystallisation condition containing oseltamivir prior to cryo-protection.

Section 4.2.1 described the production of the inactive NanH variant, NanH-D237A. NanH-D237A-apo was trialled for crystallisation using the JCSG, PACT and Proplex screens in sitting drops using the Mosquito robot. Hits appeared within 48 h, in C4 of the JCSG screen (0.1 M HEPEs pH 7.0, 10 % w/v PEG 6000) (Figure 4.8c). At this point in the project, there was access to a Formulator Robot (Formulatrix), which was used to optimise C4 of the JCSG screen. The optimisation was performed using a larger variety of conditions, as a 96-well plate was used compared to a 24-well plate and

smaller protein droplets were used. The Formulator was beneficial as it calculated and dispensed the crystallisation reagents into the plate, saving time and allowing the coverage of more variations of JCSG C4. The optimisation involved adjusting the concentration of PEG 6000 (5, 6, 7, 8… 15, 16 %) and the pH values (pH 6.7, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, 7.4). Once again, a reduced concentration of 10 mM HEPES was used for the optimisation (section 4.2.3). Crystal hits appeared after 48 h, some of which were analysed in the NanH-D237A-apo form, with others being soaked in a cryoprotectant containing either 5 mM 3-SL or 5 mM 6-SL and stored appropriately before X-ray data collection. An example of crystals sent for data collection are shown in Figure 4.8d.

4.2.3 The crystal structure of NanH-apo and NanH-HEPES

The determination of the crystal structure of NanH revealed a model showing the overall molecular architecture of the enzyme, how the domains fold and the details of the binding site, which also allowed comparison with other sialidases. In the first crystal structure of NanH to be solved, the active site was occupied by a molecule of HEPES (Figure 4.9). This structure was determined, with a resolution of 2.0 Å, using data collection on the I03 beamline at Diamond Synchrotron (Oxford, UK), on a crystal from the condition shown in Figure 4.8b. The data collection and refinement statistics for NanH-HEPES are shown in Table 4.3. The phase problem for this dataset was solved by molecular replacement using the BTSA sialidase from *B. thetaiotaomicron* as a model in the molecular replacement (PDB: 4BBW). This allowed the generation of phases for NanH, as BTSA had a high protein sequence identity (approximately 65 %) to NanH (section 2.13.8). A clear solution was indicated by the quality of the map produced, which was supported by the RFZ, TFZ and LLG statistic values, which were 5.7, 38 and 1307 respectively.

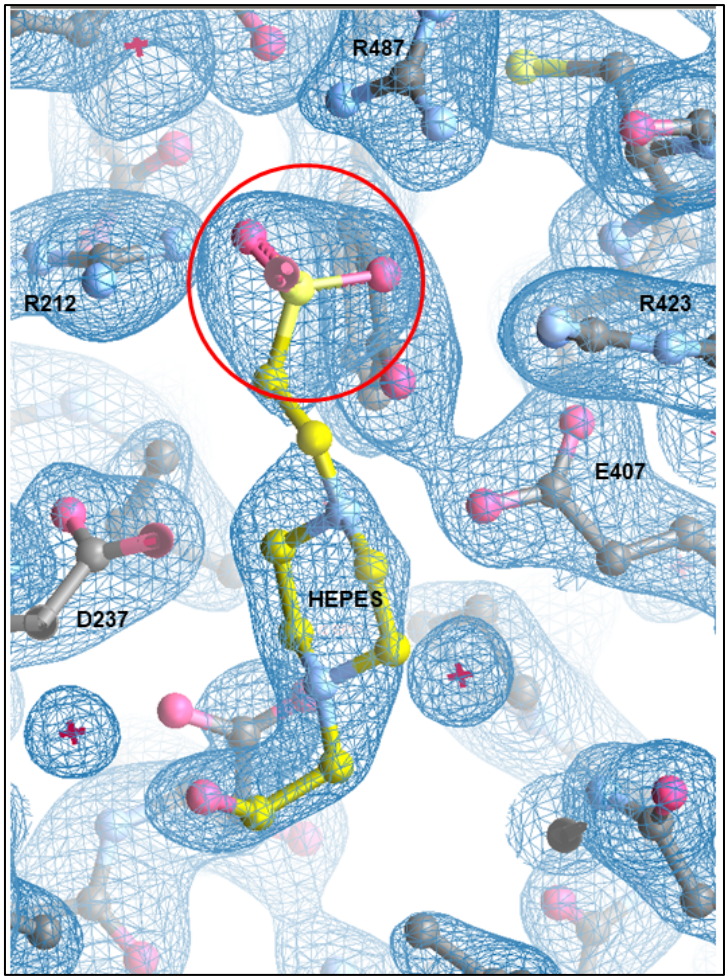


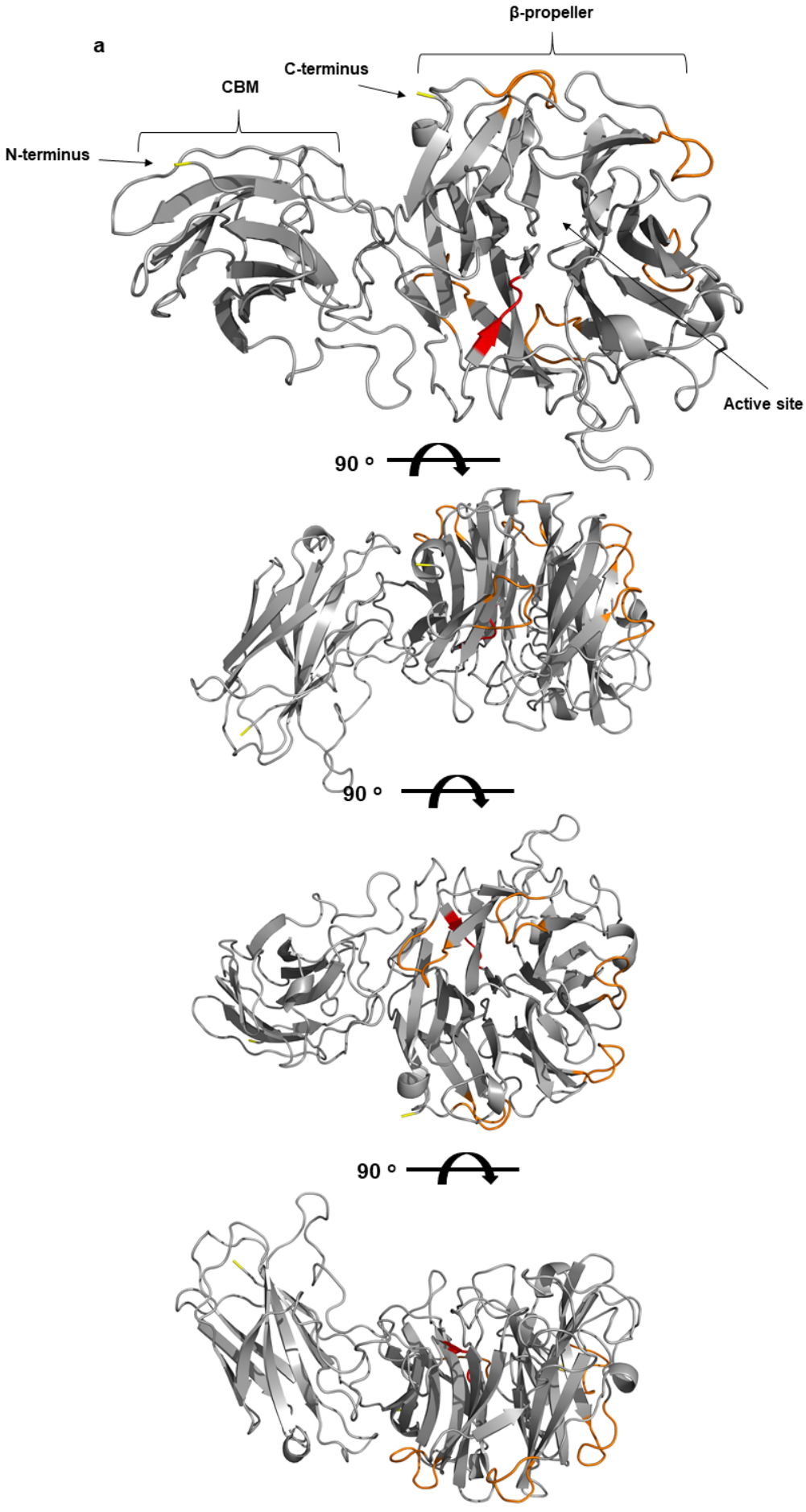
Figure 4.9: A molecule of HEPES buffer in complex with the active site of NanH. The negatively charged sulphate group (circled in red) of HEPES (yellow) was found to interact with the arginine triad of NanH. This figure was produced using Coot from the CCP4 software and the map used was the $2F_O$-F_C double difference at a root-mean-square deviation (RMSD) of 0.98 Å.

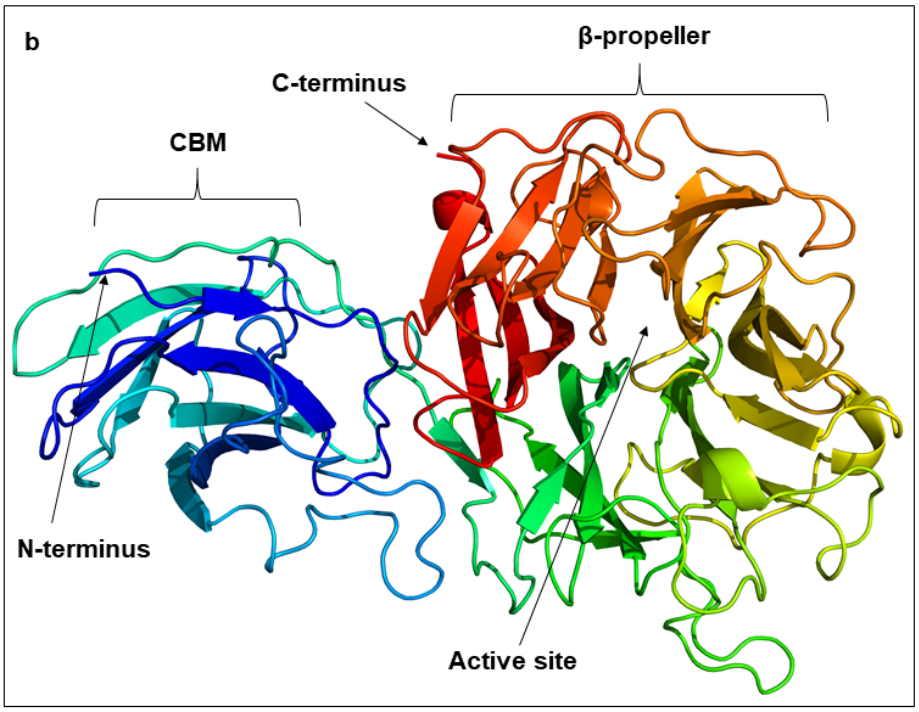
Parameters	NanH-apo	NanH-HEPES	NanH-Oseltamivir	NanH-D237-apo	NanH-D237-3-SL	NanH-D237-6-SL
Space group	P 41 21 2	P 41 21 2	P 41 21 2	P 41 21 2	P 41 21 2	P 41 21 2
Unit cell parameters:						
a (Å)	79.9	79.7	79.5	80.1	80.1	79.9
b (Å)	79.9	79.7	79.5	80.1	80.1	79.9
c (Å)	349.3	349.5	348.9	349.9	349.8	349.6
Resolution range (Å)	1.66 (1.70 – 1.66)	1.97 (2.02 - 1.97)	1.92 (1.97 – 1.92)	2.11 (2.16 - 2.11)	2.06 (2.11 – 2.06)	1.90 (1.95 – 1.90)
Number in asymmetric unit (AU)	2	2	2	2	2	2
Rmerge (all I^+ & I^-)	0.17 (0.89)	0.06 (0.90)	0.15 (1.02)	0.13 (2.27)	0.16 (2.91)	0.14 (2.70)
Rpim (all I^+ & I^-)	0.05 (0.34)	0.02 (0.49)	0.04 (0.30)	0.04 (0.66)	0.03 (0.58)	0.03 (0.59)
Mean ((I) / sd (I))	6.9 (0.9)	21.8 (1.4)	10.6 (1.9)	11.4 (1.0)	14.8 (1.5)	14.4 (1.3)
Completeness (%)	100.0 (100.0)	99.4 (97.4)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Multiplicity	12.1 (7.9)	10.5 (4.0)	12.6 (12.2)	12.8 (12.8)	25.6 (26.1)	25.3 (21.8)
R-factor (%)	0.23	0.21	0.31	0.25	0.23	0.26
R_{free} (%)	0.27	0.27	0.36	0.30	0.29	0.31
Number of refined atoms: protein/water	762/8066	485/7990	320/7963	134/7964	331/8064	423/8021
Ramachandran plot:						
In preferred regions (%)	94.7	93.6	92.7	92.9	93.5	93.0
In allowed regions (%)	3.9	5.2	5.9	5.8	5.1	5.7
Outliers (%)	1.4	1.2	1.5	1.3	1.4	1.3

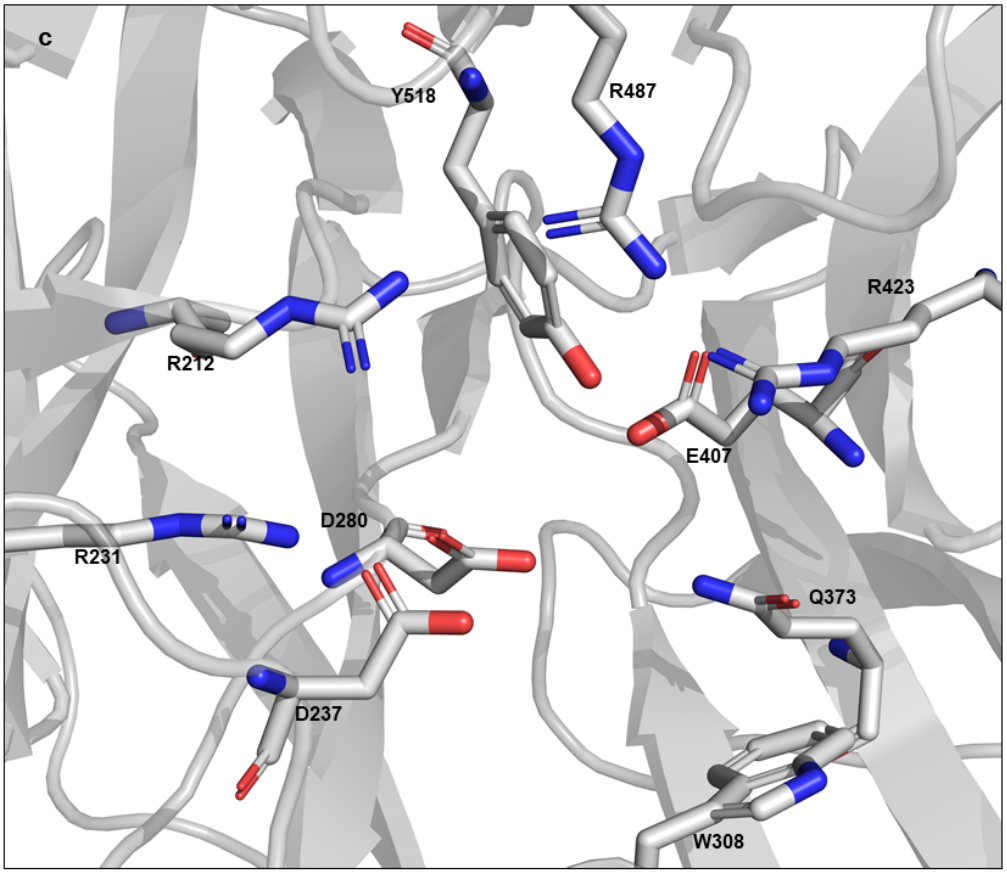
Table 4.3: Data collection and refinement statistics for NanH and NanH variants. Numbers in parentheses are statistics for the outer shell.

The NanH-HEPES crystal model showed that there were two copies of NanH in the asymmetric unit, which will be referred to as subunit A and subunit B (Figure 4.10d and e). The model consisted of 519 residues (34-552) in each subunit. When subunit A was superimposed onto subunit B, using the program 'SUPERPOSE' from the CCP4 suite, the root mean square (RMS) on the xyz displacement for the Cα's for each of the 519 residues was 0.311 Å. This was strong evidence that the two subunits are structurally identical. The catalytic domain (residues 197-552) folds into the conserved 6-bladed β-propeller that is common amongst bacterial sialidases, with each blade being composed of 3-4 anti-parallel β-strands (Figure 4.10a and 4.10b). The outer (fourth) strand of each sheet connects to the inner strand in the adjacent sheet through long looping regions. The N-terminal domain (residues 34-196), which was previously determined to be a carbohydrate binding module (CBM) (Frey *et al*., 2018), is composed of 8 antiparallel β-strands (Figure 4.10a). The topology for both the CBM and catalytic domain can be viewed in Figure 4.11.

The location of the active site in NanH was highlighted by the conserved catalytic residues at the centre of the propeller (Figure 4.10c), which agreed with the active site location in other bacterial sialidases. R212, R423 and R487 form the arginine triad, which is known to bind to the negatively charged carboxyl group of sialic acid (Newstead *et al*., 2008). D237 was discovered to be the conserved aspartate that is believed to be involved in catalytic mechanism, alongside Y518 and E407, which form the Tyr/Glu nucleophilic pair. The remaining highlighted residues were predicted to have some involvement in forming a hydrogen bonding network or van der Waals contacts with sialic acid. A total of 5 conserved Asp-boxes were highlighted and located at the edge of individual blades in the propeller, which suggests that they are involved in maintaining the protein fold. An open cleft region is present between the two NanH subunits, along with smaller clefts in both N-terminal domains (Figure 4.10d and 4.10e).


a
β-propeller
C-terminus
CBM
N-terminus
Active site
90 °
90 °
90 °


b
β-propeller
C-terminus
CBM
N-terminus
Active site


c
Y518
R487
R423
R212
E407
D280
R231
Q373
D237
W308

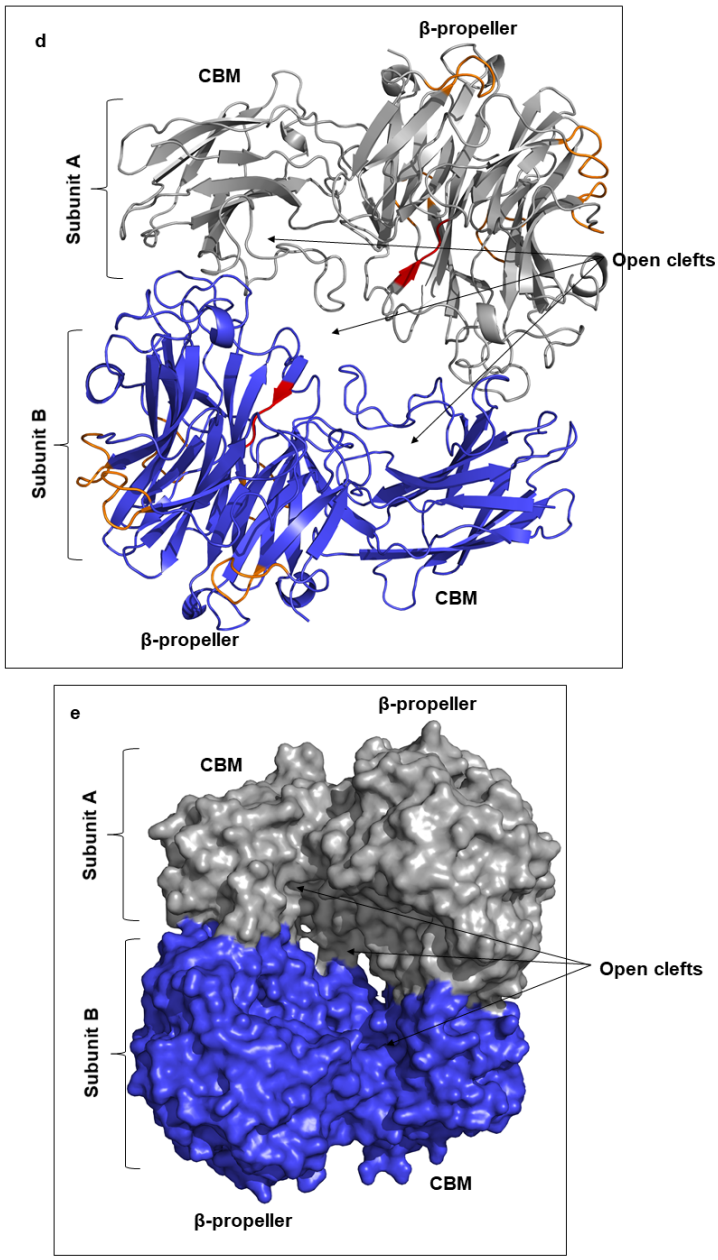


Figure 4.10: The crystal model of NanH-apo. (a) Subunit A (monomer) of NanH-apo (grey), highlighting a 6-bladed β-propeller, with the FRIP domain (red), the Asp-boxes highlighted (orange), and an N-terminal CBM in 4 orthogonal views when consecutively rotated by 90°. The N and C-termini residues a highlighted in yellow and labelled. **(b)** Subunit A (monomer) of NanH-apo, highlighting a 6-bladed β-propeller in rainbow representation to highlight the individual blades. **(c)** The active site of NanH-apo with the conserved catalytic residues highlighted as sticks. **(d)** The dimeric form of NanH-apo. **(e)** A surface representation of the dimeric form of NanH-apo with subunit A (grey) and subunit B (blue). There figures were made using PyMOL.

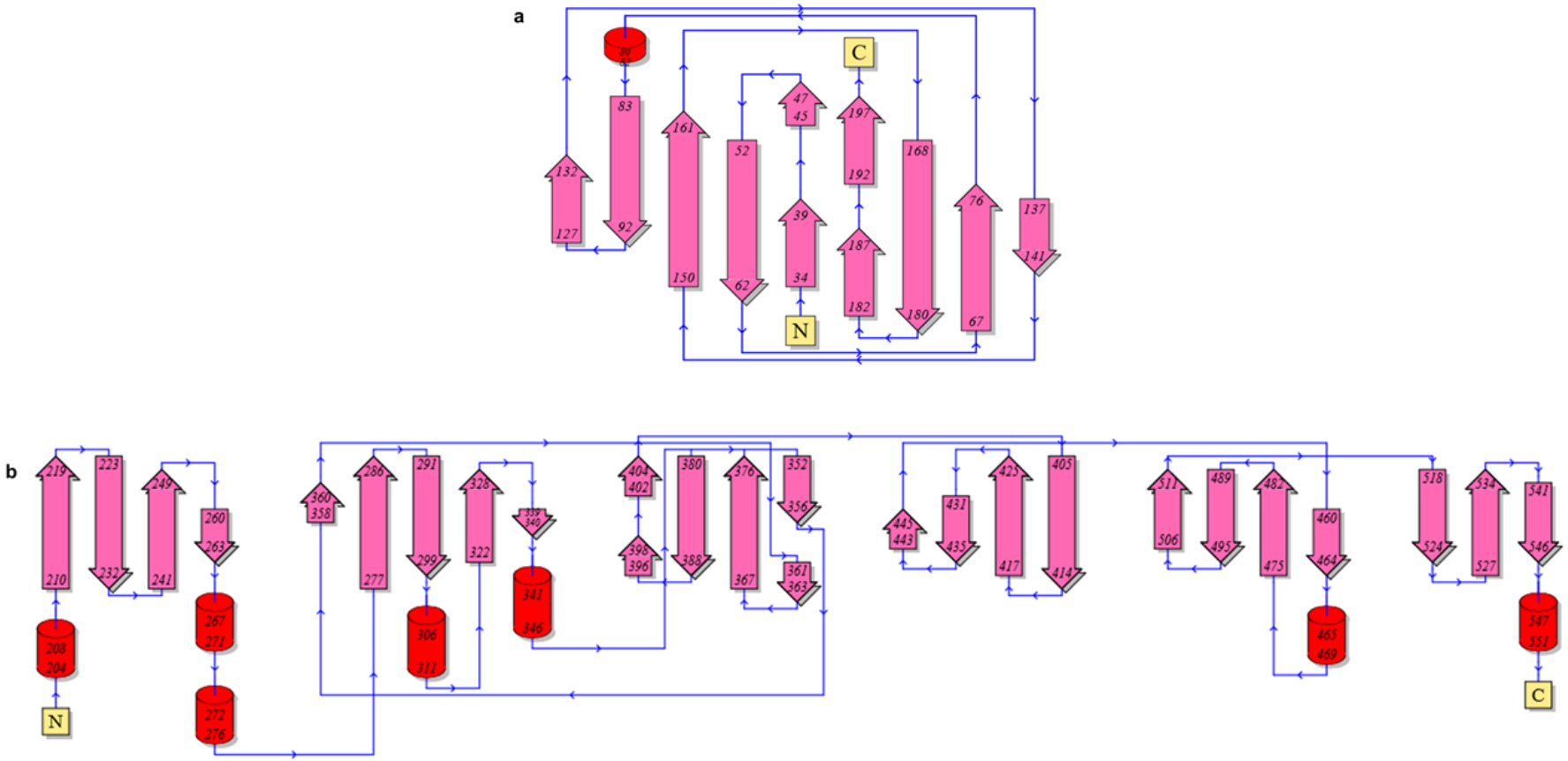


Figure 4.11: Topology diagram to represent the secondary structural elements of NanH. a) NanH-CBM domain. **b)** NanH-catalytic domain. The β-strands are represented as pink arrows and the α-helices are represented by red cylinders. This figure was generated on PDB sum at EMBL-EBI .

As previously noted (Figure 4.9), extra electron density was discovered in the active site of NanH and was determined to be a molecule of HEPES which originated from the buffer used in the crystallisation condition (Figure 4.12a). HEPES contains a negatively charged sulphate ion, which is able to interact via a hydrogen bonding network with the positively charged guanidinium groups in the arginine triad (R212, R423 and R487) (Figure 4.12b). The hydroxyethyl group at the base of the piperazine ring only appears to form van der Waals interaction with D280 and W308 in the NanH protein backbone.

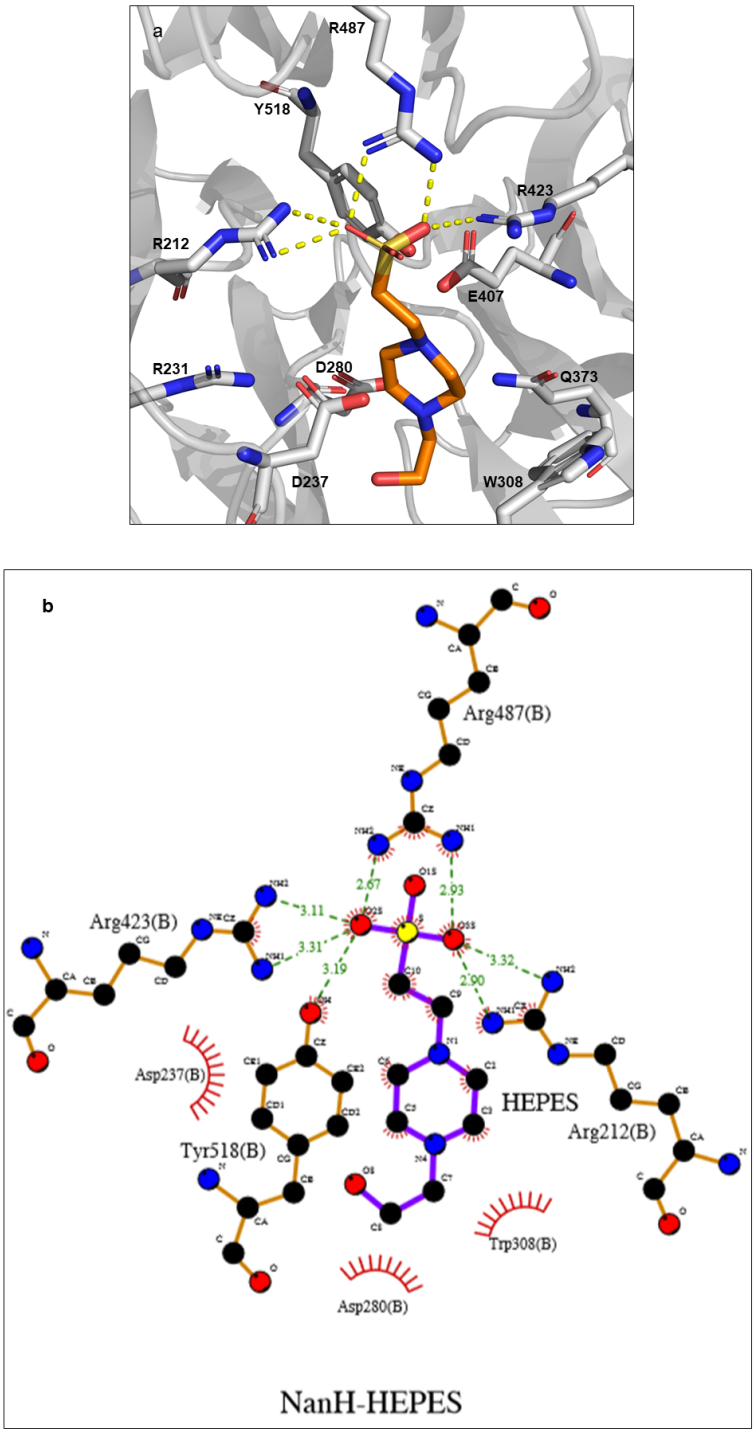


Figure 4.12: NanH in complex with HEPES. (a) A molecule of HEPES (orange) was found to interact with the active site of NanH. This figure was made using PyMOL. **(b)** The interactions between NanH and HEPES, with hydrogen bonding (green dashes) and van der Waals interactions (red-dashed semi-circles) shown. HEPES is shown in purple. Subunit B of NanH was used to produce (b), as this subunit had the best electron density for building the HEPES molecule. This figure was produced using LigPlot+ at EMBL-EBI.

In this project it was necessary to bind inhibitors and ligands into the active site of NanH in attempt to understand how the enzyme discriminates between substrates and whether this can be modified. Formation of a NanH-HEPES complex during crystallisation was problematic as there was the possibility that the binding of HEPES could outcompete inhibitors and ligands for the active site. Many other buffers used in crystallisation also contain negatively charged groups and had potential to interact with the arginine triad. Therefore, it was tested whether NanH-apo crystals would grow when the concentration of HEPES was reduced by a factor of 10 (100 mM HEPES to 10 mM HEPES). Luckily this did not affect the growth of crystals and X-ray data were collected on the i04 beamline at the Diamond Synchrotron Light Source. The crystals for NanH-apo were isomorphous with those of NanH-HEPES, therefore the phase problem for NanH-apo was solved using the phases from the NanH-HEPES dataset. The data collection and current refinement statistics for NanH-apo are shown in Table 4.3. The reduction in the concentration of HEPES allowed the structure of NanH-apo to be solved, as only solvent (H_2O) was present in the active site. There were no outstanding differences between the protein backbone in the NanH-apo model compared to that in NanH-HEPES. When subunit A of NanH-apo was superimposed onto subunit A of NanH-HEPES, the RMS on the xyz displacement for the Cα's for each of the 519 residues was 0.165 Å. When the same analysis was performed for subunit B, the RMS was 0.173 Å, meaning Figure 4.10 is also a valid representation of NanH-apo, as there were no differences between the structures.

Even though there are high structural similarities amongst bacterial sialidases, there are features that are present in some sialidases but not others. Interesting features found in the active site region of NanH included a non-proline cis-peptide between residue L238 and Q239 (Figure 4.13). This non-proline cis-peptide is conserved in the sialidase from *P. distasonis* but not in BTSA of *B. thetaiotaomicron*. R306 was also of interest due to its side chain taking an uncommon rotamer (Figure 4.13) and is only conserved in sialidases with high sequence identity to NanH, such as BTSA and the sialidase from *P. distasonis*.

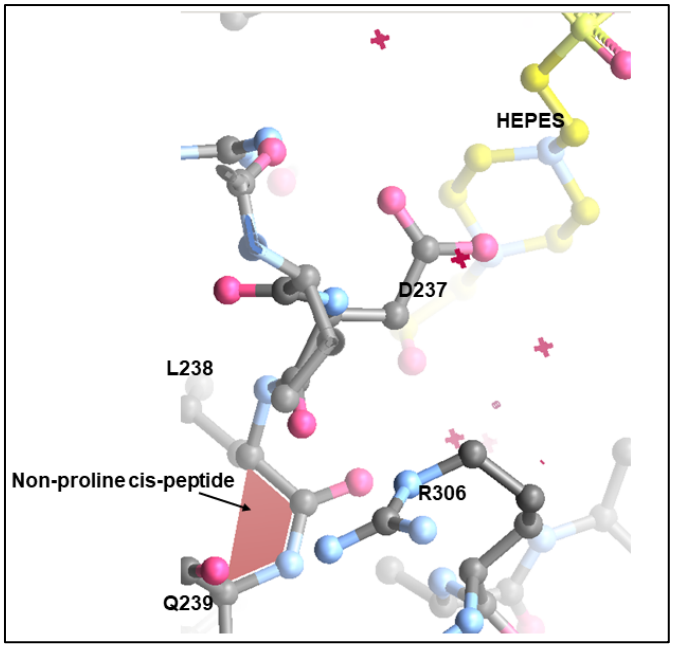


Figure 4.13: The non-proline cis-peptide and R306 are found in proximity of the active site. The non-proline cis-peptide is highlighted (red) between L238 and Q239, with R306 found slightly to the right.

4.2.3.1 The carbohydrate-binding module

Previous studies on NanH have determined the presence of a novel NanH N-terminal domain. This domain was distinct from the catalytic domain and could bind sialoglycans in isolation from the catalytic domain (Frey *et al*., 2018). The domain has now been highlighted in the crystal structure (Figure 4.10a) and will be referred to as the carbohydrate-binding module (CBM). BLAST and Pfam database searches on NanH-CBM (residues 34-196) did not reveal any homology with CBM domains that have been documented in the CAZy database, such as the CBM40 family. Crystal models of CBM40 domains of sialidases from *S. pneumoniae* and *R. gnavus* are shown in Figure 4.14 and were noticeably distinct from the NanH-CBM as they are larger due to their higher content of β-strands. However, it could be argued that the NanH-CBM is similar, as it also has a β-sandwich style fold. These CBM40s did not have any homology to NanH-CBM when BLAST protein sequence alignment was used. In this study, the Phyre2 server was also used to analyse the NanH-CBM in isolation (residues 34-196) in attempt to find homologous proteins. NanH-CBM had 42 % and 51 % sequence identity to N-terminal domains from the sialidases of *B. thetaiotaomicron* and *P. distasonis* respectively, which was expected since WT NanH has high sequence identity to these sialidases (section 3.3.2). Crystal structures are available for these sialidases and the structure of their CBM domains were similar to that of NanH (Figure 4.14), suggesting that these sialidases are part of a novel CBM family along with NanH.

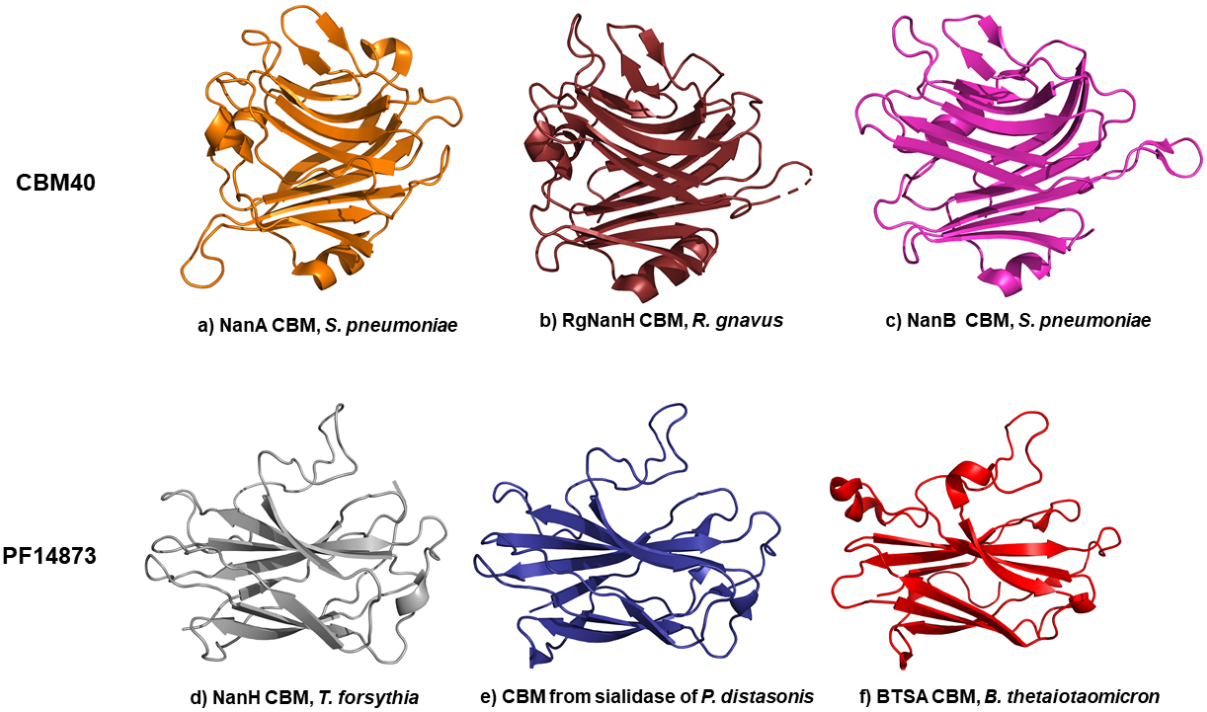


Figure 4.14: CBM domains from the CBM40 and PF14873 families. a) The CBM of NanA from *S. pneumoniae* (PDB: 4ZXK), **b)** The CBM of RgNanH from *R. gnavus* (PDB: 6ER2), **c)** The CBM of NanB from *S. pneumoniae* (PDB: 2VW0), **d)** The CBM of NanH from *T. forsythia* (this study), **e)** The CBM of a sialidase from *P. distasonis* (PDB: 4FJ6), **f)** The CBM of BTSA from *B. thetaiotaomicron* (PDB: 4BBW). Images produced using PyMOL.

A Pfam search using the NanH-CBM sequence (residues 34-196) was performed to determine which protein family NanH-CBM belongs to. NanH-CBM fell under the PF14873 family and the definition of this is a 'domain usually found at the N-terminus of the bacterial neuraminidase'. A Pfam alignment was performed with 39 members of the PF14873 family, with a selection of these shown in the alignment Figure 4.15. There was extremely high sequence conservation amongst PF14873 proteins when consensus>50 (greater than 50 % conservation) was used (not shown), perhaps because the majority are of *Bacteroides* origin (*T. forsythia* was previously referred to as *B. forsythus*). The last three sequences in the alignment (Figure 4.15) did not show as high sequence conservation to NanH-CBM at consensus>50 as the other sequences. This lack of conservation could be because that the PF14873 protein from *Bacteroides sp.* is a domain from a GH18 enzyme as opposed to GH33 like the others. The PF14873 proteins from *Sphingobacterium sp.* and *Pedobacter heparinus* were thought to be sialidase related and potentially don't have as high sequence similarity because they are not *Bacteroides*. Therefore, to pick out the regions of highest conservation, consensus>80 was used to remove the conservation that seemed to represent the *Bacteroides* species only. Conserved regions

in the protein sequence alignment were highlighted onto the NanH crystal model Figure 4.16. Interestingly, 6 consecutive amino acids close to the N-terminal read FRIRIP, which is a similar sequence to the FRIP motif in the catalytic domain. FRIRIP was located at the deepest point of the cleft in the CBM.

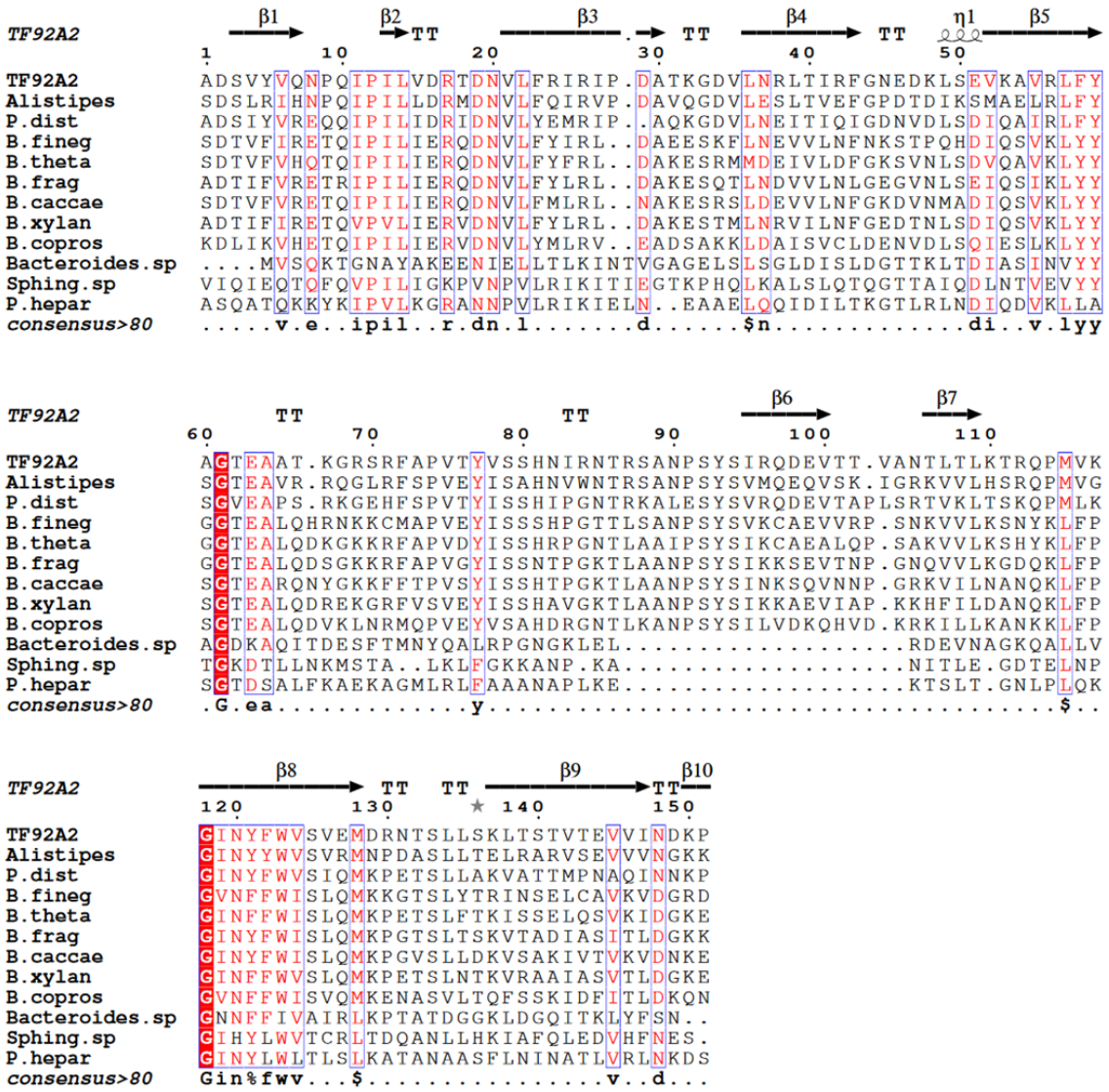


Figure 4.15: Alignment of NanH-CBM with other PF14873 family members. The alignment and image production were performed using MultAlin and ESPript. Strictly conserved residues are shown in white with a red background. The alignment consists of CBMs from the sialidases from *T. forsythia* (TF) strain 92A2, Alistipes sp., NCBI Taxonomy ID: 1262695, *P. distasonis*, *B. finegoldii, B. thetaiotaomicron,* one of the *B. fragilis* sialidases (strain ATCC 25285, nanH1), *B. caccae, B. xylanisolvens, B. coprosuis, Bacteroides. sp., Sphingobacterium sp.* and *P. heparinus.* NCBI accession numbers were WP_046826229.1, CDD16645.1, WP_011967072.1, CDC53424.1, WP_008766031.1, CAH07505.1, CCZ75115.1, CBK66099.1, WP_006744190.1, CCY51394.1, WP_013668360.1 and WP_015808031.1 respectively.

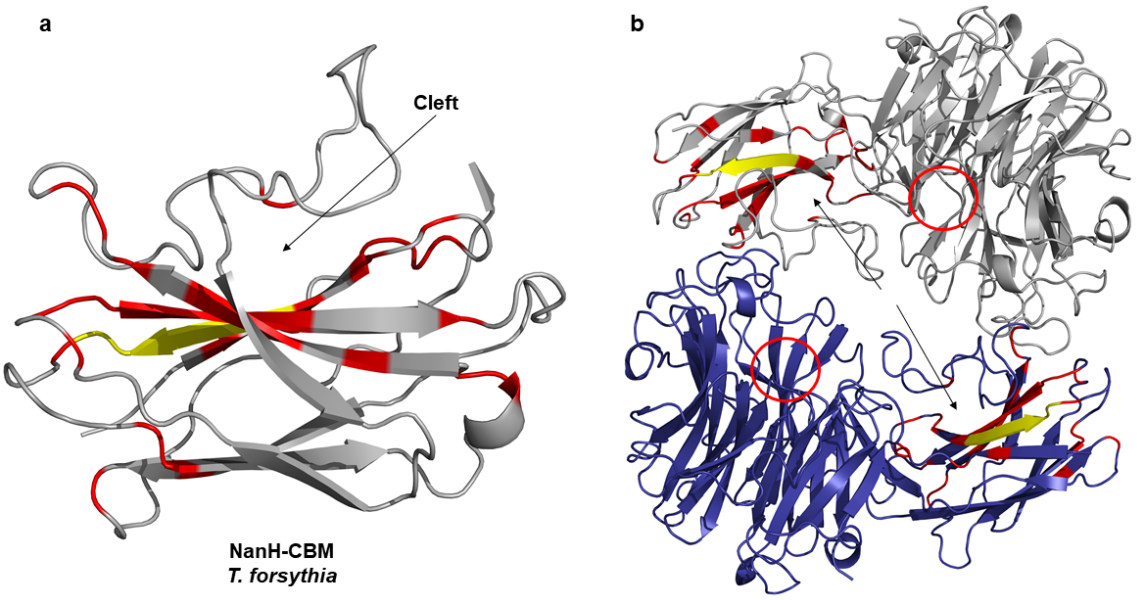


Figure 4.16: Conserved residues of the PF14873 family highlighted on the CBM of NanH from *T. forsythia*. The conserved regions from the consensus>80 alignment in Figure 4.15 are shown in red on the NanH crystal model. The yellow region represents the FRIRIP sequence. **(a)** A close-up image of the conserved residues in NanH-CBM from subunit A. **(b)** The CBM conserved residues are highlighted (red) in both subunit A (grey) and subunit B (blue) to show their location relative to the dimeric enzyme. The CBM clefts are highlighted using the black arrows and the active site region is circled. These figures were created using PyMOL.

4.2.3.2 The electrostatic surface potential of NanH

When the electrostatic surface potential of NanH was calculated using the APBS Electrostatics plugin in PyMOL, it was noticed that the active site face of the enzyme is mostly positive, apart from a section of the active site which is negatively charged (Figure 4.17a). The residues responsible for this negativity are D237 and E407 (Figure 4.17c). The region that surrounds the active site has a positive surface potential which is caused by the arginine triad (Figure 4.17c). The electrostatic surface potential throughout the opposite face was negatively charged (Figure 4.17b). A second highly negatively charged region can be seen in the CBM region of NanH (Figure 4.17a, circled in red). The residues that correspond to this region were highlighted on a cartoon representation of the CBM (Figure4.17d) and are located at the entrance of the cleft that was described in section 4.2.3.1.

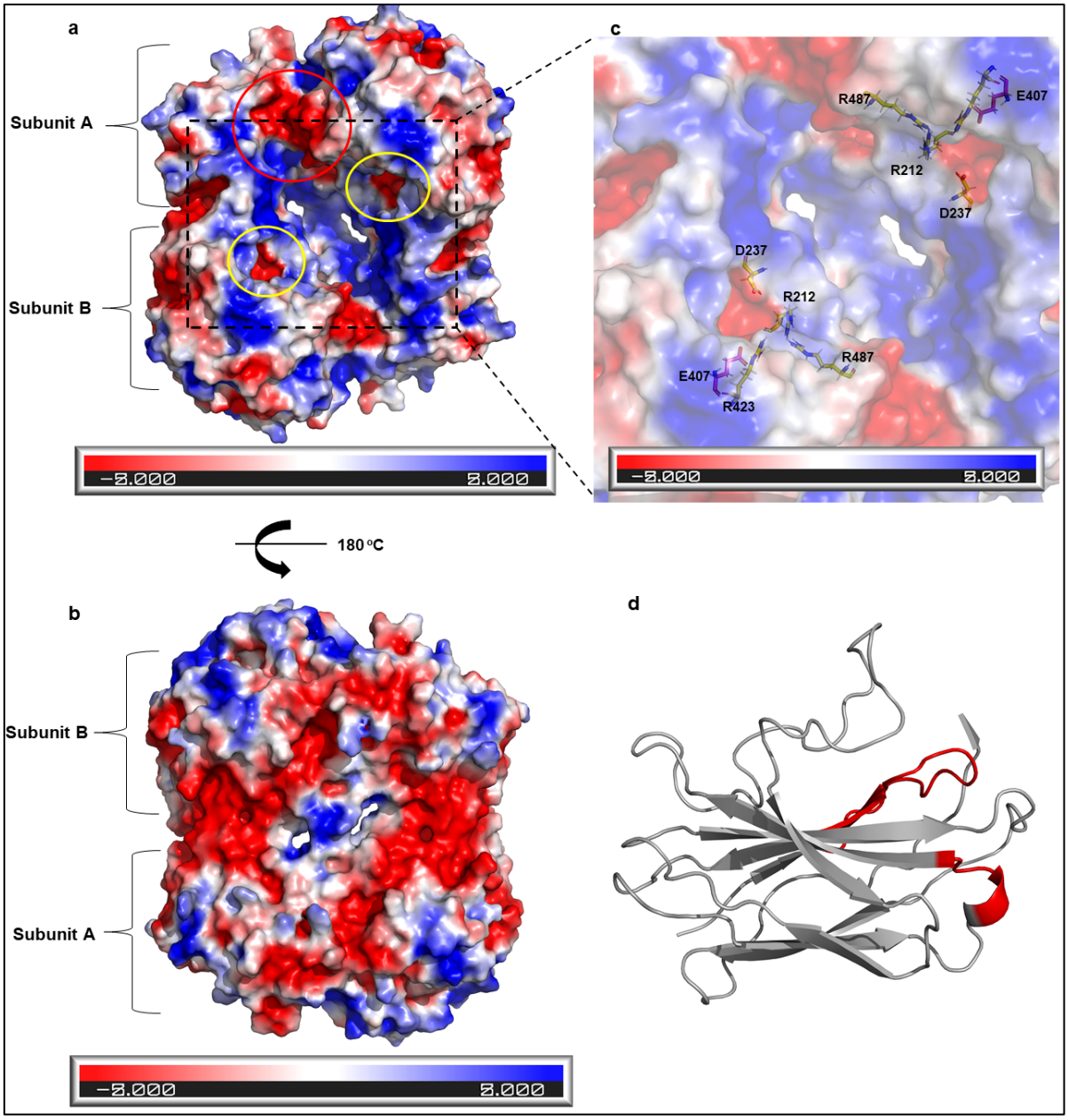


Figure 4.17: The electrostatic potential of NanH. Calculated by the APBS tool using the default settings (PyMOL), the solvent excluded surface of NanH has been coloured based on the electrostatic potential from -5 to +5 kT/e and the APBS_ramp02 was set to Range +/- 2. Red represents negative charge and blue represents positive charge. **(a)** The active site face of NanH, with the catalytic region circled in yellow. **(b)** The opposite face of NanH. **(c)** An enlarged view of the area surrounded by a black, dashed box in Figure 4.17a. **(d)** NanH-CBM with the negatively charged region (region surrounded by red circle in Figure 4.17a) shown in red.

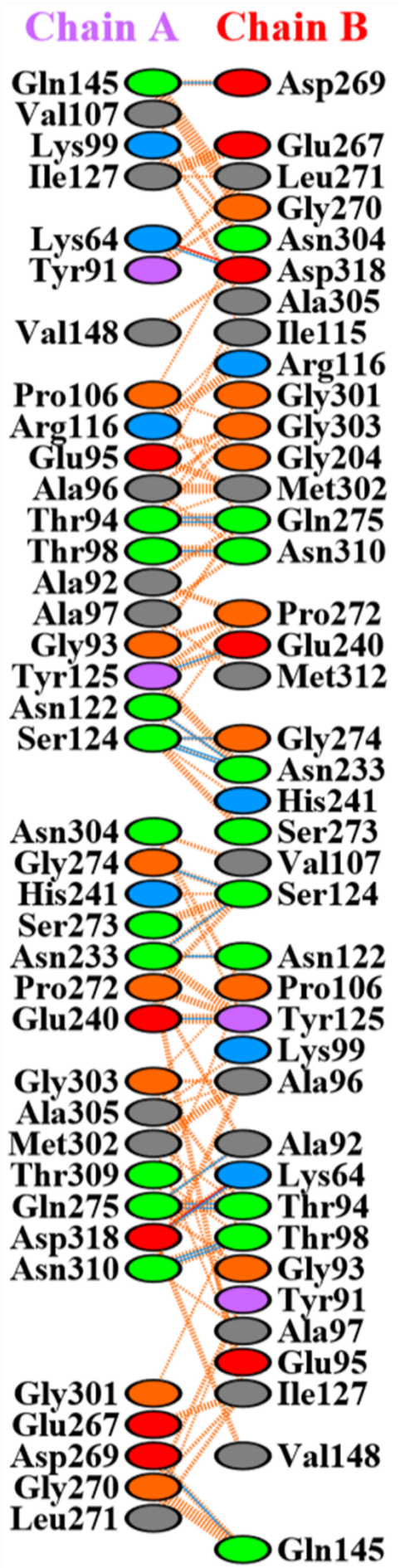
4.2.3.3 The interfaces of NanH

The interface environment between subunit A and B was analysed using PDBePISA at EMBL-EBI. The total solvent accessible surface area is 19292 $Å^2$ and 19385 $Å^2$ for subunit A and subunit B respectively. When assembled, the solvent accessible surface area is 35530.4 $Å^2$. The interface area between subunit A (47 residues) and subunit B (47 residues) is 1575 $Å^2$, which allowed the percentage buried surface area to be calculated as 8.9 %. Even though the buried surface area is rather small, PDBePISA predicted that NanH is likely to be a stable biological dimer as the ΔG^{diss} was calculated to be 10.1 kcal/mol. This data is backed up with the SEC data in section 3.2.3, where NanH was found to be a dimer in solution. It is also likely that two salt bridges form; the first between K64 (subunit A) and D318 (subunit B) and the second between D318 (subunit A) and K64 (subunit B). It was also suggested that there are 30 hydrogen bonds and 234 non-bonded contacts which are represented in Figure 4.18a and b.

PDBePISA also contains a function that searches for proteins with similar interfaces in the PDB by structural similarity, which was applied to the interface of NanH. The results suggested that BTSA of *B. thetaiotaomicron* and the sialidase from *P. distanosis* have interfaces similar to that of NanH. BTSA (65 % sequence identity to NanH) has been deposited in the PDB with one molecule (molecule A) in the asymmetric unit and its biological assembly was also suggested to be monomeric (PDB: 4BBW). However, PDBePISA analysis (in this project) suggested that there might be a real interface in BTSA with an area of 1548 $Å^2$ and that it is likely that there are two identical copies of subunit A, which could interact and behave as a stable dimer in solution. The results suggested that the BTSA interface is formed of 46 residues from each molecule and when these were compared to the interface residues of NanH, there was approximately 51 % sequence identity, however there was no conservation of either salt bridge.

The sialidase of *P. distanosis* (PDB: 4FJ6) (72 % sequence identity to NanH) has 4 copies of the protein (homo-4-mer) in the asymmetric unit (molecules A, B, C and D). However according to PDBePISA there might only be stable interfaces between A and B, and C and D, which suggests the biological assembly could be a stable dimer in solution and that the homo-4-mer is likely to be a crystal artefact. The suggested interface area is 1440 $Å^2$ and is formed by 44 residues from each molecule, which have approximately 60 % sequence identity to the interface residues in NanH. Only one salt bridge was conserved and was found between K50 in molecule A and D306 in molecule B, which correspond to K64 in subunit A of NanH and D318 in subunit B of NanH.

a


Chain A
Chain B
Gln145
Val107
Lys99
Ile127
Lys64
Tyr91
Val148
Pro106
Arg116
Glu95
Ala96
Thr94
Thr98
Ala92
Ala97
Gly93
Tyr125
Asn122
Ser124
Asn304
Gly274
His241
Ser273
Asn233
Pro272
Glu240
Gly303
Ala305
Met302
Thr309
Gln275
Asp318
Asn310
Gly301
Glu267
Asp269
Gly270
Leu271
Asp269
Glu267
Leu271
Gly270
Asn304
Asp318
Ala305
Ile115
Arg116
Gly301
Gly303
Gly204
Met302
Gln275
Asn310
Pro272
Glu240
Met312
Gly274
Asn233
His241
Ser273
Val107
Ser124
Asn122
Pro106
Tyr125
Lys99
Ala96
Ala92
Lys64
Thr94
Thr98
Gly93
Tyr91
Ala97
Glu95
Ile127
Val148
Gln145

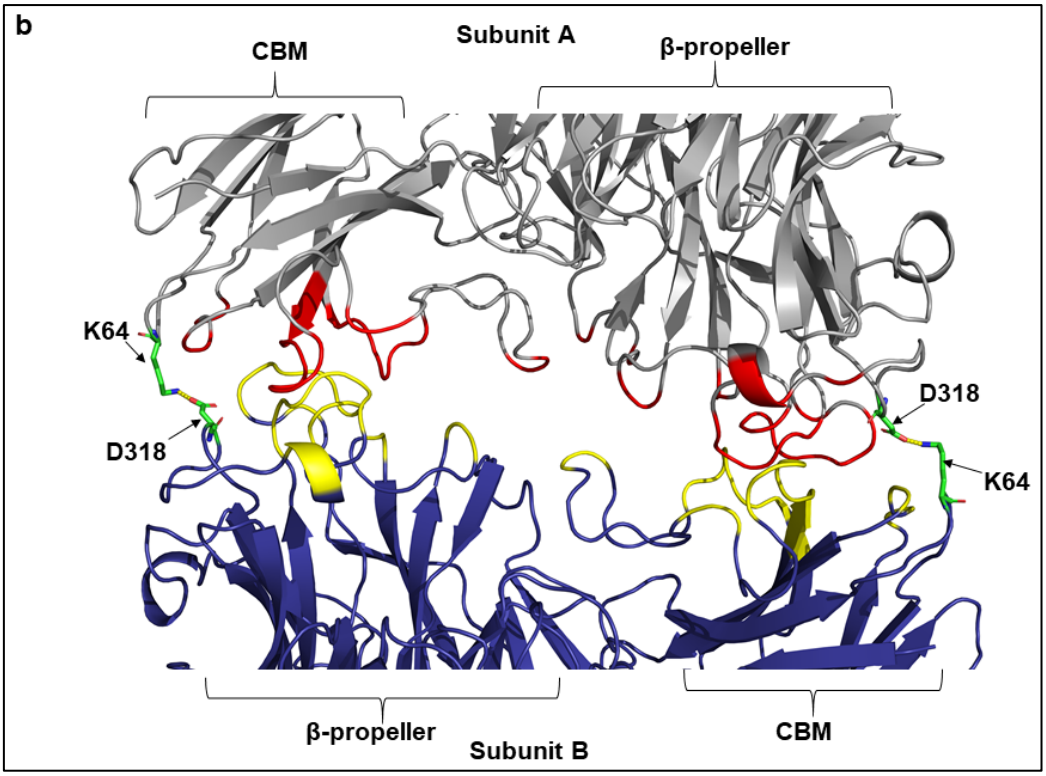


Figure 4.18: The protein-protein interface between subunit A and subunit B of NanH. (a) A schematic diagram of the interactions between subunit A and subunit B. Salt bridges are represented by red lines, non-bonded contacts by orange dashed lines and hydrogen bonds by blue lines. Positively charged residues are shown in blue, negatively charged in red, neutral in green, aliphatics in grey, aromatics in purple, proline and glycine in orange and cysteine in yellow. The figure was created using PDB sum from EMBL-EBI. **(b)** The protein-protein interface residues determined by PDBePISA are highlighted on the dimeric crystal model of NanH. The interface residues from subunit A (grey) are highlighted in red and those from subunit B (blue) are highlighted in yellow. The residues involved in salt bridge formation are highlighted as green sticks and labelled.

4.2.4 The crystal structure of NanH-oseltamivir

Oseltamivir is an acetamido cyclohexene that is a structural homolog of sialic acid, initially designed as an inhibitor of influenza's NA. Here, the aim was to gain a crystal structure of NanH in complex with oseltamivir acid to determine how sialic acid binds to the active site. NanH-apo crystals that were grown at the reduced concentration of HEPES, section 4.2.2 were soaked with a cryoprotectant solution that contained 5 mM oseltamivir (a significantly higher concentration than the IC_{50}, section 3.2.5) and an X-ray diffraction pattern was collected using Beamline I03, Diamond Synchrotron Beamline, Oxford, UK. The structure of NanH-oseltamivir was solved at 1.9 Å and since the crystals

were isomorphous with those of NanH-HEPES, the phase problem for NanH-oseltamivir was solved using the phases from NanH-HEPES. The current refinement statistics for NanH-oseltamivir are shown in Table 4.3.

The NanH-oseltamivir dataset presented two copies of NanH in the asymmetric unit, which is in agreement with the NanH-apo and NanH-HEPES datasets. Oseltamivir is orientated in the active site by its negatively charged carboxyl group, which is held equatorial to the cyclohexene ring through hydrogen bonding interactions with the guanidiniums of the arginine triad (Figure 4.19a, b and d). R212 forms one hydrogen bond at 2.82 Å, whilst R487 and R231 both form two hydrogen bonds at lengths of 2.97 Å and 2.78 Å, and 3.32 Å and 3.08 Å respectively. The cyclohexene ring is sitting in a distorted boat conformation caused by the carboxyl group taking a pseudoequatorial position upon the binding the arginine triad, in agreement with Taylor and von Itzstein (Taylor and von Itzstein, 1994). The C4-amino group of oseltamivir (position C9, Figure 4.19d) interacts with the carboxylates of D237 and D280 via hydrogen bonding at lengths of 2.45 Å and 2.9 Å respectively. The N-acetyl tail of oseltamivir is sitting in a hydrophobic pocket and a hydrogen bond of 2.87 Å is able to form between the amide of the N-acetyl and the carboxylate of D280. The oxygen of the N-acetyl tail is able to form a hydrogen bond with a nearby water molecule (Figure 4.19b). Analysis of the NanH active site when in complex with oseltamivir showed that all but one residue had identical conformations to NanH-apo. The sidechain of W308 appears to have been displaced by the pentyl ether group of oseltamivir, as each carbon in the aromatic rings of the tryptophan had moved distances between 0.8-2.0 Å when compared to NanH-apo (Figure 4.19c). Overall, there were no crucial differences in the protein backbone compared to NanH-apo. When subunit A of NanH-apo was superimposed onto subunit A of NanH-oseltamivir, the RMS on the xyz displacement for the Cα's for each of the 519 residues was 0.176 Å. When the same analysis was performed for subunit B, the RMS was 0.189. In addition, a non-proline cis-peptide is present between residues L238 and G239, and R306 still assumes an uncommon rotamer in the NanH-oseltamivir model, as previously shown in, Figure 4.13.

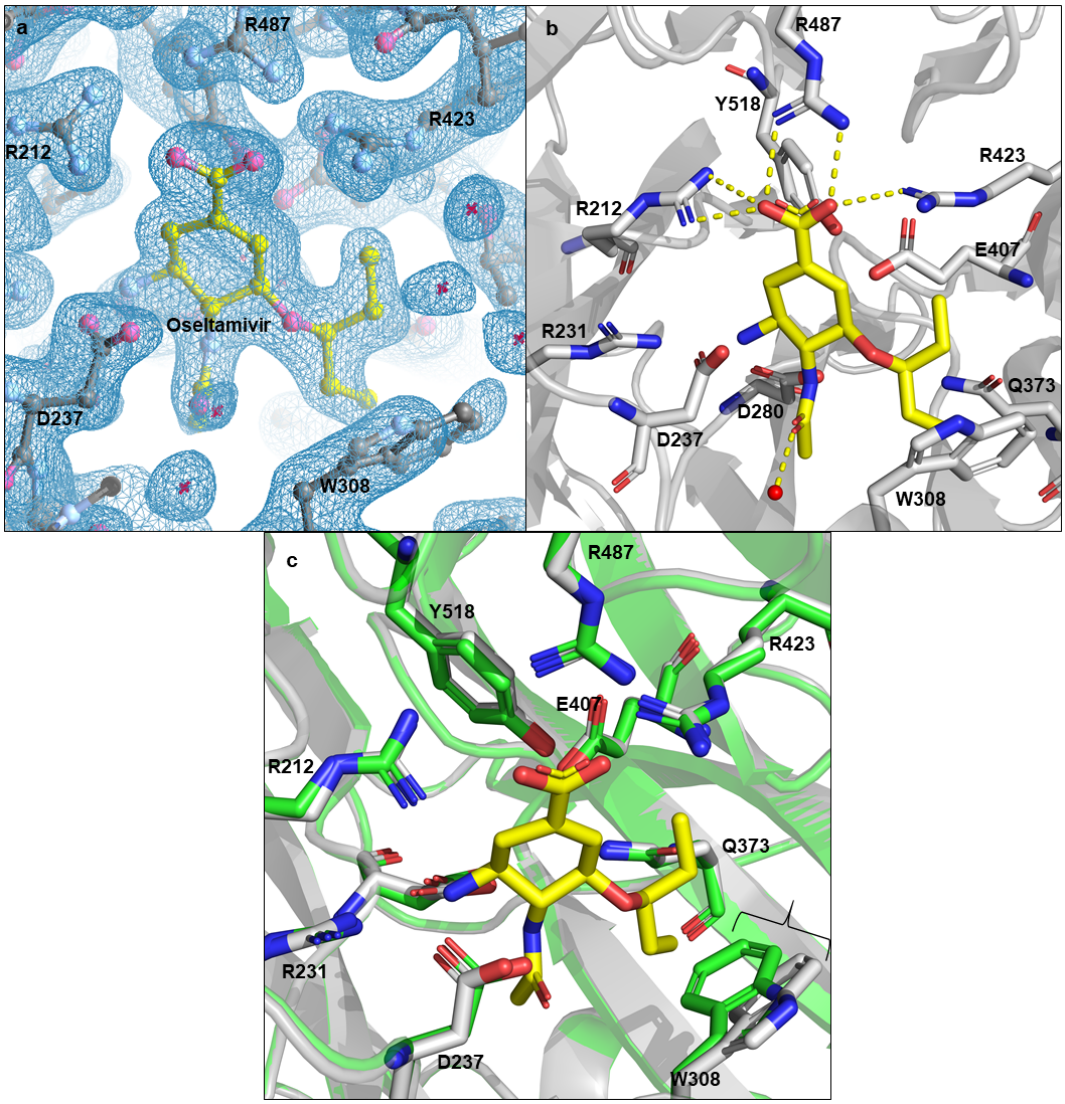

a
R487
R423
R212
Oseltamivir
D237
W308
b
R487
Y518
R423
R212
E407
R231
Q373
D280
D237
W308
c
R487
Y518
R423
E407
R212
Q373
R231
D237
W308
d
Arg212(A)
Asp237(A)
Arg487(A)
Tyr518(A)
Trp308(A)
Asp280(A)
Oseltamivir
Ile213(A)
Arg423(A)
Glu407(A)
Thr406(A)
Gln373(A)
NanH-oseltamivir

Figure 4.19: NanH active site in complex with the inhibitor oseltamivir. (a) Oseltamivir (yellow) was modelled correctly into the electron density (blue) of the F_0-F_C double difference map at an RMSD 1.04 Å. This figure was produced using Coot from the CCP4 software. **(b)** NanH sialidase in cartoon representation (grey), with the catalytic residues (grey sticks) and oseltamivir (yellow sticks). A single water molecule is highlighted as a red sphere and the yellow dashed lines represent the hydrogen bonding network. This figure was created on PyMOL. **(c)** Overlay of NanH-apo (green cartoon) with NanH-oseltamivir (grey cartoon) illustrating the movement of W308 upon the binding of oseltamivir (yellow) to the NanH active site. This figure was created using PyMOL. **(d)** The interactions between NanH and oseltamivir with hydrogen bonding (green dashes) and van der Waals interactions (red-dashed semi-circles) highlighted. Oseltamivir is highlighted in purple. Subunit A of NanH was used to produce (d), as it had the best electron density for building the oseltamivir compound. This figure was produced using LigPlot+ at EMBL-EBI.

4.2.5 The crystal structure of NanH-D237A in complex with 3'-sialyllactose and 6'-sialyllactose

In order to further understand the positioning of each glycosidic link within the active site and considering that NanH is a dual-specific sialidase, targeting α2,3- and α2,6-linked sialic acid, it was necessary to solve two more crystal structures of NanH: one in which α2,3-linked sialic acid is in complex with the active site and another in complex with α2,6-linked sialic acid. Furthermore, this would permit the determination of residues to be mutated in an attempt to alter the specificity of NanH for either α2,3- or α2,6-linked sialic acid (Chapter 5). However, it would be impossible to use active NanH for this purpose, as any bound sialic acid molecules would be rapidly cleaved and therefore likely to not be present in the asymmetric unit. Therefore, it was necessary to use an inactive NanH variant in an attempt to trap substrates in the active site.

NanH-D237A was shown to be inactive but able to retain binding ability to 3- and 6-SL (section 4.2.1). NanH-D237A crystallised under JCSG C4 when HEPES buffer was reduced to a concentration of 10 mM (section 4.2.2). NanH-D237A crystals were soaked in a cryo-protectant containing 5 mM 3- or 6-SL and an X-ray diffraction data were collected on Beamline I04 (Diamond Synchrotron, Oxford). NanH-D237A-3-SL and NanH-D237A-6-SL crystals were isomorphous to NanH-HEPES crystals, allowing the phase problem to be solved using the phases from the NanH-HEPES dataset, with the refinement statistics shown in Table 4.3.

Complexes for NanH-D237A-3-SL and NanH-D237A-6-SL were solved at 2.06 Å and 1.90 Å respectively (Figure 4.20a and 4.20b). There were two copies of NanH-D237A in the asymmetric unit, which was

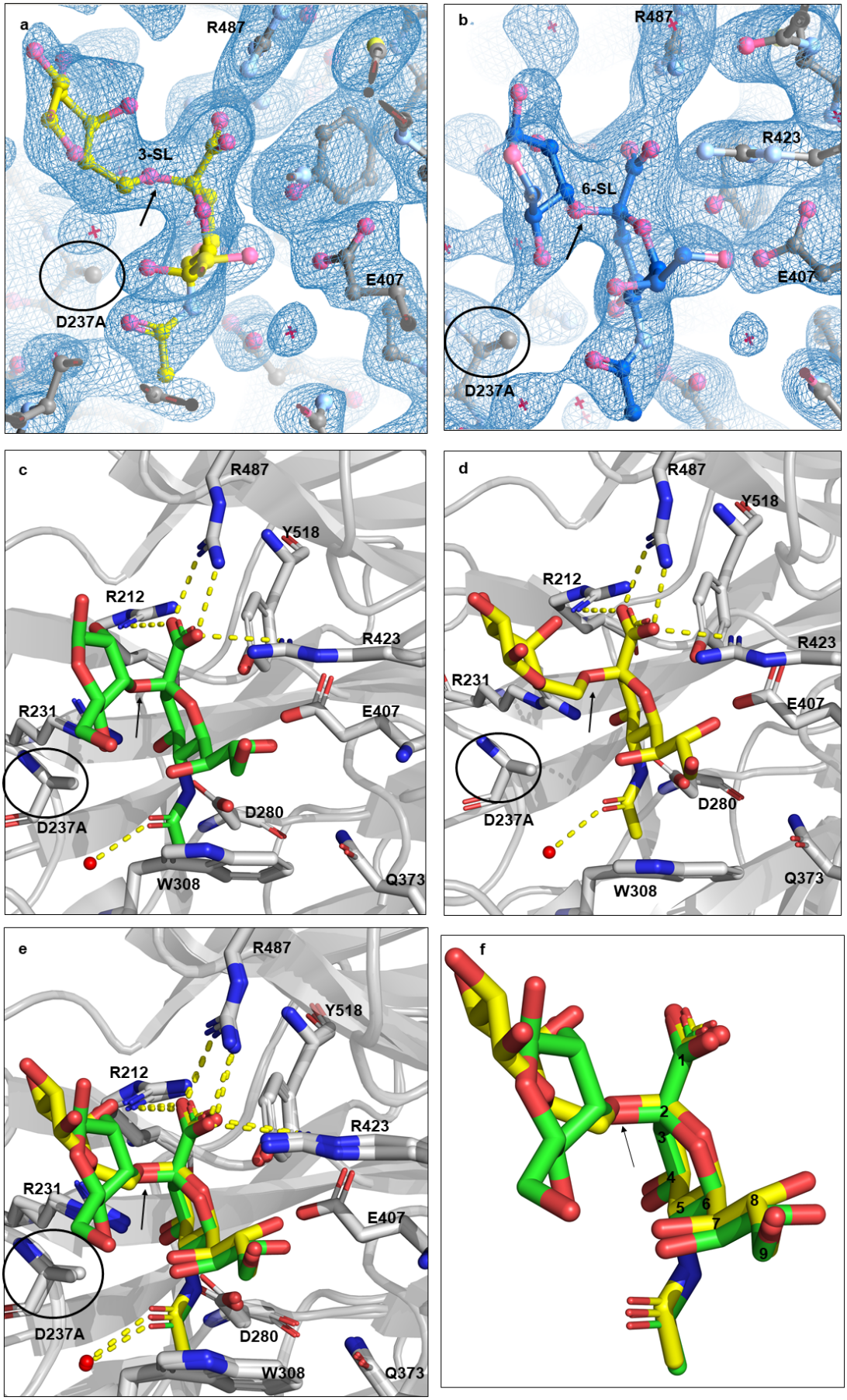
the seen in all 3 datasets (NanH-D237A-apo, NanH-D237A-3-SL and NanH-D237A-6SL). The overall structures of these three models were identical with no noticeable differences in the protein backbones, dimer interfaces or within the active sites. No noticeable structural differences were found when comparing the inactive variant to NanH-HEPES model, which was backed up by low RMS values when superimposing the models (Table 4.4). The only minor difference seen was the lack of electron density at residue 237, which highlighted the substitution of the larger aspartate side-chain to a smaller alanine (Figure 4.20a and b).

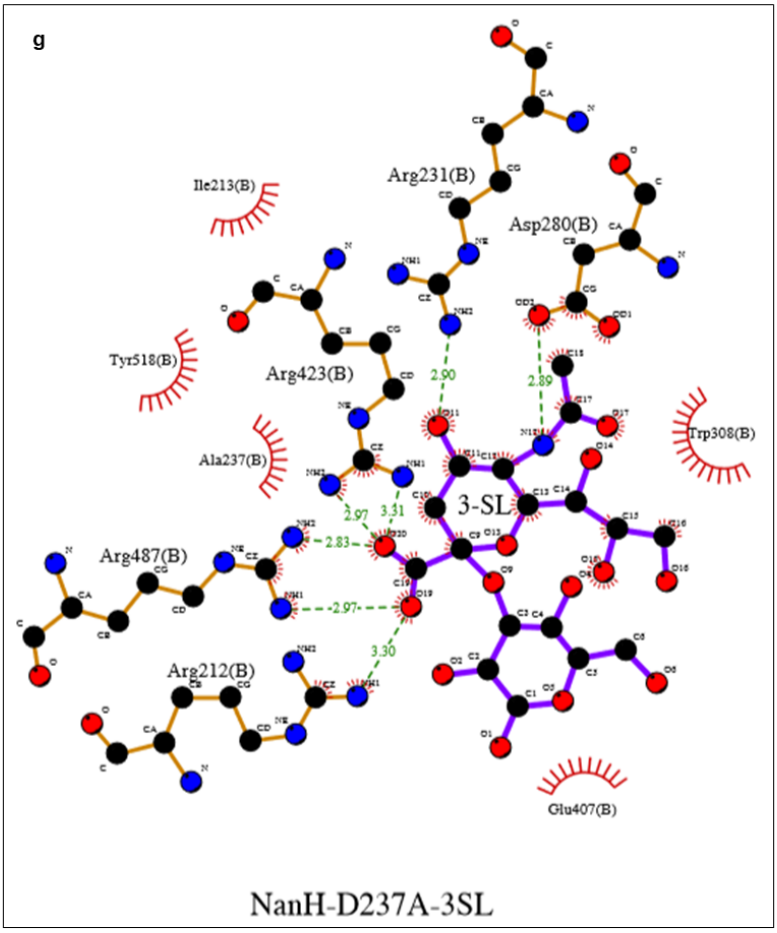
Fixed	Moving	RMS (Å)
NanH-HEPEs CA NanH-HEPES CB	D237A-apo CA D237A-apo CB	0.15 0.17
NanH-HEPEs CA NanH-HEPES CB	D237A-3SL CA D237A-3SL CB	0.15 0.15
NanH-HEPEs CA NanH-HEPES CB	D237A-6SL CA D237A-6SL CB	0.14 0.15

Table 4.4: The RMS on the xyz displacement of the Cα's for each of the 519 residues when NanH-D237A subunits were superimposed onto the equivalent subunit of NanH-HEPES. SUPERPOSE program from the CCP4 software suite was used to superimpose subunits A and B from the NanH-D237A models onto the equivalent subunits in the NanH-HEPES model.

The terminal sialic acid of 3-SL is positioned in the active site of NanH-D237A in a similar way to that seen for oseltamivir in WT NanH (Figure 4.20c). A hydrogen bonding network was seen between the carboxyl group of sialic acid and the guanidinium groups of the arginine triad. R423 and R487 each form two hydrogen bonds to the sialic acid carboxyl group with bond distances of 3.31 Å and 2.97 Å, and 2.83 Å and 2.97 Å respectively (Figure 4.20g). R212 forms one hydrogen bond to the carboxyl with a bond distance of 3.30 Å. The D237A substitution resulted in the 4-hydroxyl of sialic acid forming a hydrogen bond of 2.90 Å to the guanidinium group of R231. A hydrogen bond of 2.89 Å is formed between D280 and the amide of the N-acetyl tail of the sialic acid portion of 3-SL, a similar organisation to that seen in oseltamivir (Figure 4.20g). The pyranose ring of sialic acid adopts a distorted boat conformation with a pseudoequatorial conformation that is adopted by the carboxyl bound to the arginine triad. The underlying galactose of 3-SL is positioned so that it is perpendicular to the sialic acid and directed out of the active site. A hydrogen bonding network cannot be seen between galactose at subsite +1 and the protein backbone, which is why it is not surprising that electron density for the third sugar of 3-SL (glucose) at subsite +2 is not ordered. A similar situation was seen by Zhu and co-workers (Zhu *et al*., 2012) when studying an influenza neuraminidase. A non-proline cis-peptide is present again between residues L238 and Q239, R306 adopts the same uncommon rotamer

as in all the previous NanH models and there was no displacement of W308 compared to that seen in the NanH-oseltamivir model.


a
R487
3-SL
E407
D237A
b
R487
R423
6-SL
E407
D237A
c
R487
Y518
R212
R423
R231
E407
D237A
D280
W308
Q373
d
R487
Y518
R212
R423
R231
E407
D237A
D280
W308
Q373
e
R487
Y518
R212
R423
R231
E407
D237A
D280
W308
Q373
f
1
2
3
4
5
6
7
8
9


g
Ile213(B)
Arg231(B)
Asp280(B)
Tyr518(B)
Arg423(B)
2.90
2.89
Trp308(B)
Ala237(B)
3-SL
2.97
3.31
Arg487(B)
2.83
2.97
3.30
Arg212(B)
Glu407(B)
NanH-D237A-3SL

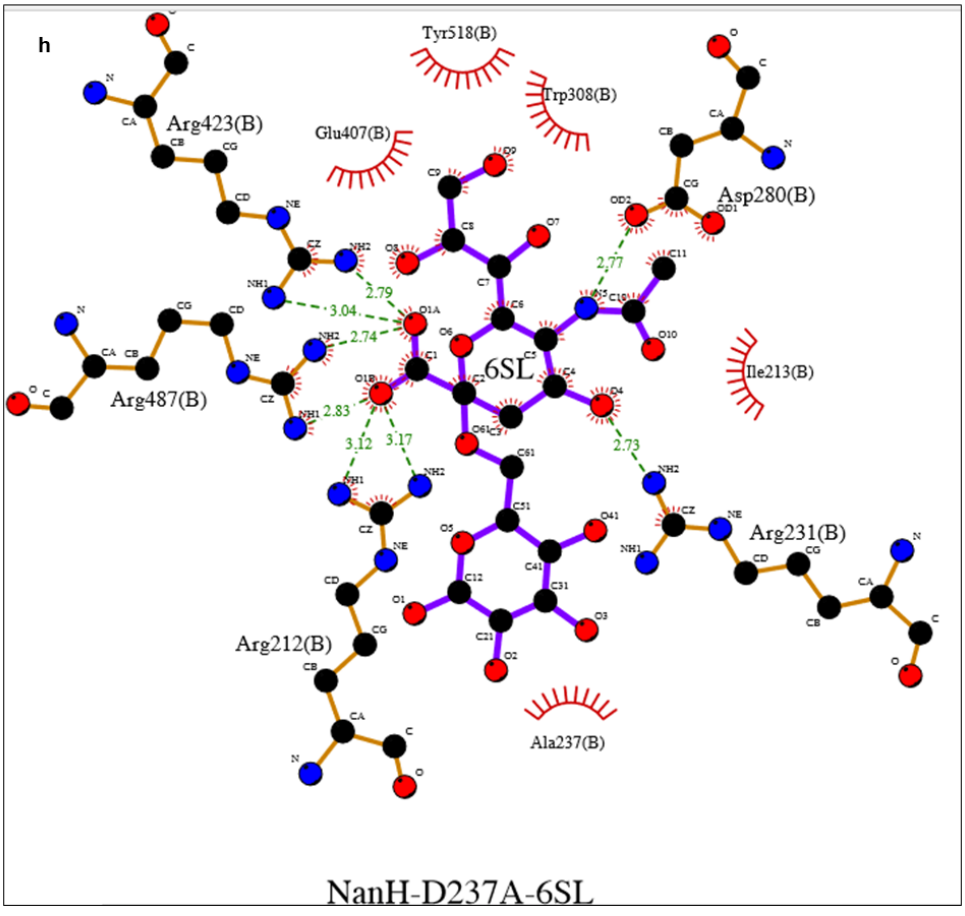

h
Tyr518(B)
Trp308(B)
Arg423(B)
Glu407(B)
Asp280(B)
2.77
2.79
3.04
2.74
6SL
Ile213(B)
Arg487(B)
2.83
3.12
3.17
2.73
Arg231(B)
Arg212(B)
Ala237(B)
NanH-D237A-6SL

Figure 4.20: NanH-D237A bound in complex with 3- and 6-SL. (a) 3-SL (yellow) was modelled into the electron density (blue) of the $2F_0$-F_C double difference map at an RMSD of 1.02 Å. **(b)** 6-SL (dark blue) was modelled into the electron density (blue) of the $2F_0$-F_C double difference map at an RMSD of 0.99 Å. Both (a) and (b) were produced in the Coot program from CCP4 software suite. **(c)** NanH-D237A (grey cartoon) in complex with 3-SL (green sticks). **(d)** NanH-D237A in complex with 6-SL (yellow sticks). **(e)** NanH-D237A in complex with 3-SL (green sticks) overlaid with NanH-D237A in complex with 6-SL (yellow sticks). **(f)** The overlay of 3-SL (green sticks) and 6-SL (yellow sticks) isolated from the protein backbone to emphasise the similar position of the glycosidic linkage. (c), (d), (e) and (f) were created in PyMOL and the protein backbone of NanH-D237A is shown in grey in cartoon representation with the key catalytic residues as sticks. **(g/h)** The interactions between NanH-D237A and 3-SL and NanH-D237A and 6-SL respectively, with hydrogen bonding (green dashes) and van der Waals interactions (red-dashed semi-circles) highlighted. Ligands are shown in purple. Subunit B of NanH-D237A was used to produce (g/h), as this subunit had the best electron density to model the ligands into. This figure was produced using LigPlot+ at EMBL-EBI.

6-SL sits similarly to 3-SL in the active site of NanH-D237A (Figure 4.20d, e and f). The terminal sialic acid is in a hydrogen bonding network with the guanidinium groups of the arginine triad and the carboxyl group takes up a pseudoequatorial position to the pyranose ring. The pyranose ring adopts a distorted boat conformation. R212, R487 and R423 are all positioned so that they can each form two hydrogen bonds with the carboxyl group at 3.17 Å and 3.12 Å, 2.83 Å and 2.74 Å and 3.04 Å and 2.79 Å respectively (Figure 4.20h). The lack of D237 resulted in hydrogen bond formation (2.87 Å) between the 4-hydroxyl of sialic acid and the guanidinium group of R231 (Figure 4.20h). When overlaid with the NanH-D237A-3-SL model, the sialic acid portion of 6-SL overlaid almost perfectly with that of the 3-SL. There were no substantial differences (≥ 0.1 Å) between the hydrogen bond lengths connecting the 3-SL and 6-SL terminal sialic acids to the conserved catalytic residues (Figure 4.20g and h). The structural similarity is shown in Figure 4.20f which highlights that the α2,3 and α2,6 glycosidic linkages were in almost identical conformations. The underlying galactose at subsite +1 did not appear to make any contacts with the protein backbone, so may exist in various orientations and the third sugar (glucose) at subsite +2 was not seen.

4.3 Discussion

4.3.1 The production of NanH-D237A

In order to crystallise NanH with ligands that contain α2,3- or α2,6-linked sialic acid, an inactive NanH variant was produced in the hope that ligands would be seen in complex in a crystal structure. The conserved catalytic aspartate, D237, was substituted for an alanine. NanH-D237A was expressed and purified under identical conditions to WT NanH and the activity was successfully abolished when assayed against MU-NANA and 3- and 6-SL. A ligand-binding experiment showed that NanH-D237A remained capable of binding both 3- and 6-SL. Similar sialidase de-activating experiments have been performed for an influenza A NA and TcTS sialidase from *Trypansoma cruzi* (protozoan). Substitutions D151G and D59A were performed on the influenza NA and TcTs respectively, both of which involved the conserved catalytic aspartate and successfully inactivated the enzymes (Zhu *et al*., 2012; Amaya *et al*., 2004).

4.3.2 The NanH protein fold and multimeric form

The crystal structure of *T. forsythia* NanH was solved using a molecular replacement model from the structure of BTSA from *B. thetaiotaomicron*, with which the *T. forsythia* enzyme shares 65 % sequence identity. The structure of the NanH catalytic domain showed the canonical 6-bladed β-propeller fold that is common to bacterial sialidases (Kim *et al*., 2011). Like other sialidases, the active site is located at the centre of the propeller and hosts the conserved residues such as the arginine triad (R212, R423 and R487) and the catalytic residues D237, Y518 and E407 (Xu *et al*., 2008; Park *et al*., 2013; Tailford *et al*. 2015). The structure of the N-terminal domain was also seen for the first time and consisted of 8 of β-strands.

Analysis of the interfaces between subunit A and B of NanH using PDBePISA suggested that NanH is likely to be a stable biological dimer, agreeing with the SEC data in section 3.2.3 and what were believed to be real interfaces in the asymmetric unit. However, since the NativePAGE data suggested that NanH is an even larger multimer, NanH was also analysed by SEC-MALLS, which found that NanH is predominantly a dimer in solution. When studying Figures 4.10d and 4.10e, a large cleft can be seen between the two subunits in NanH, which suggests that NanH may be able to bind and cleave sialic acid from large, complex glycans such as mucins, agreeing with data from Roy and colleagues (Roy *et al*., 2011).

Sialidase ATCC 8503 (PDB: 4FJ6) of *P. distasonis* is described as a homo-4-mer in the asymmetric unit (72 % protein sequence identity to NanH), although there have yet to be any publications discussing the determination of the multimeric form using other techniques such as SEC. Results from analysis

using PDBePISA have suggested that this sialidase is likely to be a stable dimer in solution and has a similar interface area to the interface between the molecules in a NanH dimer (1575 $Å^2$ and 1440 $Å^2$ respectively). BTSA (65 % sequence identity to NanH) (PDB: 4BBW) has been reported as monomeric in both its biological assembly and asymmetric unit in the PDB. However, the publication linked to this model does not appear to have carried out further analysis on this finding (Park *et al*., 2013). Interestingly, BTSA was a hit when PDBePISA was used to search for proteins with similar interfaces to NanH. The result suggested that BTSA could in fact be a stable biological dimer with an interface surface area of 1548 $Å^2$. NanA of *S. pneumoniae* has two molecules in the asymmetric unit, however SEC data suggested that NanA is monomeric (Xu *et al*., 2008). All of the other bacterial sialidases that have structures published in the PDB have been described as biological monomers presumably through analysis using PDBePISA, with little indication as to whether other experiments suggest otherwise (Crennell *et al*., 1993; Crennell *et al*., 1994; Tailford *et al*., 2015; Newstead *et al*., 2008; Xu *et al*., 2008). Data in Chapter 3 showed that there is no cooperativity between the two subunits in NanH and since the dimer does not appear to confer any biochemical advantage for NanH one might suggest that the dimeric structure increases the stability of the protein in an aqueous environment since it can be a secreted protein (Thompson *et al*., 2009). However, monomeric NanA can also be secreted (Gratz *et al*., 2017) which suggests that the dimeric structure of NanH may not be necessary for extracellular stability leaving the question of why NanH is a dimer unanswered. Perhaps the dimeric structure assists NanH in accessing sialic acid on complex glycans with tri- and tetra-antennary termini (section 3.2.7.1), which could be tested by interfering with the interface using SDM.

Non-proline cis-peptides are extremely uncommon but when other bacterial sialidases models from the PDB were analysed, non-proline cis-peptides were found in the same position in NanB of *S. pnuemoniae* (PDB: 2VW1) and the sialidase of *P. distasonis* (PDB: 4FJ6). A proline cis-peptide was discovered in the same location in NanI of *C. perfringens* (PDB: 2VK5). The function of non-proline cis-peptides in these sialidases is unknown, but since they are positioned next to the catalytic aspartate, perhaps they help orientate it towards the active site. R306 of NanH was only conserved in BTSA and the sialidase of *P. distasnosis,* suggesting there is not a specific function that is responsible for the uncommon conformation.

The electrostatic potential of NanH (Figure 4.17) is typical of secreted bacterial sialidases. The conserved catalytic residues (D237, E407 and D280) were highlighted as a negatively charged pocket, which is bordered by the positively charged arginine triad. A second negatively charged region was highlighted in the CBM cleft which was shown in Figures 4.10d, 4.10e and 4.16a. The function of this negatively charged region is unknown but perhaps it could represent the CBM binding pocket, although the specific sugar that would bind is unknown. Most of the active site face has a positive

charge (Figure 4.17a), which should encourage the binding of anionic sialoglycoconjugates. The opposing face (180^{o} rotation, Figure 4.17b) is negatively charged which could be responsible for the prevention of non-specific binding. Similar electrostatic potentials for sialidase catalytic domains have been confirmed for NanI of *C. perfringens*, RgNanH of *R. gnavus* and NanB of *S. pneumoniae*, which suggests that it is a conserved feature in bacterial sialidases (Newstead *et al*., 2008; Tailford *et al*., 2015; Xu *et al*., 2008).

4.3.3 The carbohydrate-binding domain

The NanH crystal structure also revealed the structure of the additional N-terminal domain (residues 34-196) which had previously been characterised by other researchers (Frey *et al*., 2018). This work showed that the domain is indeed a CBM, as it is able to bind substrates such as 3- and 6-SL as well as more complex substrates such as fetuin. The NanH CBM domain lacked sequence and structural similarities to the CBM40 domains that have been identified in sialidases such as NanA, NanB and NanC (*S. pneumoniae*), NanI (*C. perfingens*) and RgNanH (*R. gnavus*) (Yang *et al*., 2015; Xu *et al*., 2008; Owen *et al*., 2015; Ribeiro *et al*., 2016; Tailford *et al*., 2015) and it is therefore believed that the NanH CBM is part of a novel family.

When analysed on Pfam, the NanH-CBM was categorised into the PF14873 family. PF14873 was described as protein domains that are located at the N-termini of sialidases and included the N-terminal domains of *B. thetaiotaomicron* and *P. distasnosis*, which was expected due to the high sequence similarity of NanH to these proteins. The sequence alignment performed in section 4.2.3.1 highlighted the highly conserved regions which were in turn highlighted on the model. The conserved residues seemed to cluster in a CBM cleft that was highlighted in Figure 4.16a, suggesting that this could be a binding pocket. The fact that the CBM of subunit A faces the active site face of the catalytic domain of subunit B (and vice versa) suggested that the CBM can bind the substrate and orientate the sialic acid into the active site in the opposing subunit (Figure 4.16b). However, this model does not fit with those sialidases that have a CBM but are monomeric. Since NanH can also behave as a surface bound enzyme, perhaps the function of the CBM is to assist in glycan foraging or cellular attachment to in vivo human cell surfaces or pellicles for biofilm formation.

4.3.4 The NanH-HEPES model

The first crystal model solved for NanH was discovered to be in complex with a molecule of HEPES, which originated from the buffer used in the crystallisation condition. The negatively charged sulphate group of HEPES is able to interact with the positively charged guanidinium groups of the arginine triad, which is where the negatively charged carboxyl group of sialic acid is predicted to bind. This is not an

uncommon finding, as HEPES was also discovered in the active site in an influenza A neuraminidase (TZ205 neuraminidase). Again, HEPES originated from the crystallisation buffer and interacted with the arginine triad by its sulphate group (Zhu *et al*., 2012). NanB of *S. pneumoniae* was found to have a molecule of CHES bound to the active site. CHES also has a sulphate that interacts with the arginine triad of NanB and originated from the crystallisation condition (Xu *et al*., 2008). The inhibition of NanH by HEPES was briefly assessed in assays, as in section 3.2.5, however there was no reduction in NanH activity when concentrations as high as 100 mM HEPES were used.

4.3.5 The NanH-oseltamivir model

RgNanH, VCNA, NanA and NanC have all been deposited in the PDB in complex with oseltamivir (PDB: 4X49, 6EKS, 2YA6 and 4YW5 respectively). As in *T. forsythia* NanH, these enzymes orientate oseltamivir in the active site through the binding of the carboxyl group to the arginine triad and the conserved aspartates (D237 and D280) are in hydrogen binding distance of the C4-amide of oseltamivir. The data in section 3.2.5 showed oseltamivir to have one of lower IC_{50} values with NanH and that it was therefore an efficient inhibitor of NanH when compared to zanamivir. Perhaps this is reason why a crystal model of NanH in complex with either zanamivir was unable to be solved in the time scale of this project.

4.3.6 The NanH-D237A models

The *T. forsythia* inactivated mutant (NanH-D237A) crystal model was solved with 3- and 6-SL in complex with the active site. This is the first time that ligands have been seen in complex with the active site of a bacterial sialidase, as 3- and 6-SL have only been seen bound to CBM40 domains. As stated in section 4.2.1, NanH-D237A was based upon a study by Zhu and colleagues, which showed the ligands 3- and 6-SLN (containing α2,3- and α2,6-linked sialic acid respectively) in complex with the active site in crystal models of an inactivated NA from influenza A virus (Zhu *et al*., 2012). In this thesis it has been shown that there is no difference between the positioning of the glycosidic linkage of α2,3- and α2,6-linked sialic acid ligands in the active site (Figure 4.20e) and the same situation was seen in the influenza A NA models (Figure 4.3). This was in agreement with evidence in the PDB that sialic acid is orientated the same way in the active sites of different sialidases (through inhibitor and ligand bound models discussed in this chapter) and reflects the fact that α2,3- and α2,6-linkages both stem from the second carbon of sialic acid.

Since both NanH and the influenza A NA are dual specific and showed the linkages to be in the same position, there were no obvious residues that stood out in either crystal model as having potential to control the specificity if mutated. However, it was noticed that the linkages were near to the catalytic

aspartate. This suggests that perhaps engineering residues that surround the aspartate (but are not crucial to activity) could be a starting point when trying to alter the specificity of NanH.

TcTS of *T. cruzi* (trans-sialidase) has been confirmed to have specificity for α2,3-linked sialic acid. This sialidase was inactivated via the substitution of its catalytic aspartate (D59A) and a crystal model of the mutant was solved where the active site was in complex with 3-SL (Amaya *et al*., 2004). The ligand was sitting in a similar position to 3-SL in NanH and when overlaid, it could be seen that the glycosidic linkage was in the same location as in NanH (Figure 4.21). In wild-type TcTS, aromatic residues W312 and Y119 are thought to be involved in binding the lactose portion of 3-SL through hydrophobic stacking interactions (Figure 4.21) (Amaya *et al*., 2004). It has been suggested that the specificity of TcTS for α2,3-linked sialic acid arises from the aromatic residue W312. When W312 was mutated to alanine, the W312A variant was able to hydrolyse both α2,3- and α2,6-linked sialic acid. This same mutation also abolished transferase activity (Paris *et al*., 2001). W312 is part of a loop extension that is not present in other microbial sialidases such as *T. forsythia* NanH (Paris *et al*., 2001). This suggests that this loop and residue could be engineered into NanH to alter the specificity towards α2,3-linked sialic acid.

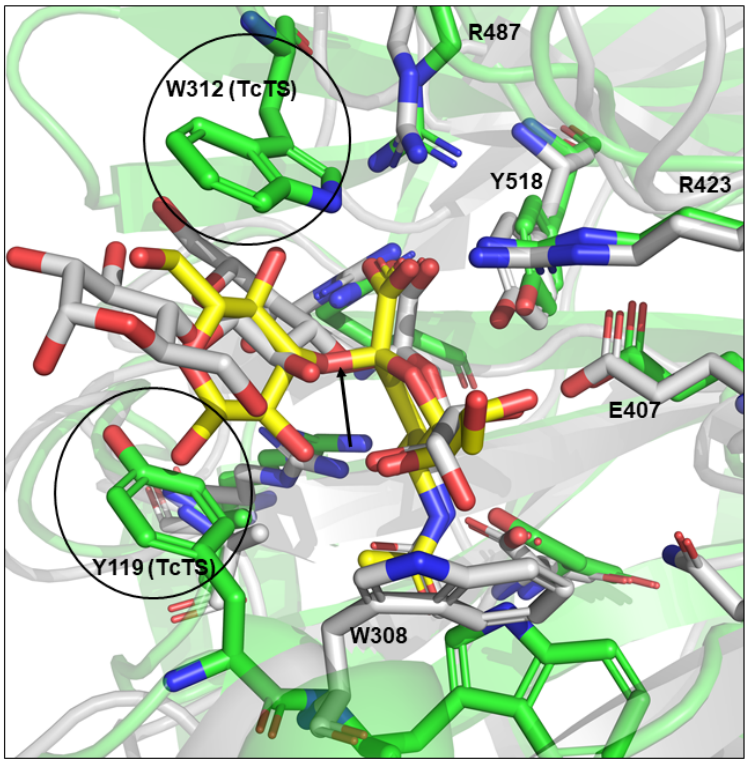


Figure 4.21: NanH-D237A in complex with 3-SL overlaid with TcTS in complex with 3-SL. The D237A-NanH protein backbone is highlighted in grey, with the 3-SL highlighted as yellow sticks. The TcTS protein backbone is shown in green with 3-SL highlighted as grey sticks (PDB: 1S0I). W312 and Y119

that are only present in TcTS have been circled and the arrow indicates the location of the glycosidic linkages. All other residue numbering corresponds to that of NanH.

4.4 Conclusion

This chapter has presented several key findings regarding the crystal structure of NanH. First of all, the overall fold of the protein has been determined as a 6-bladed β-propeller with a novel N-terminal CBM domain. The crystal structure revealed two copies of NanH in the asymmetric unit, with what appeared to be a real interface. Further analysis of the interface between the two subunits using PDBePISA confirmed that NanH is likely to be a stable biological dimer. The active site, which is conserved amongst sialidases, has been heavily analysed in an inhibitor bound complex (NanH-oseltamivir), along with two substrate bound complexes (NanH-D237A-3SL and NanH-D237A-6SL). Now the location and orientation of 2,3- and α2,6-linked sialic acid in the active site is known, residues that may interfere with enzyme specificity can be determined and mutations can be made in attempt to alter the specificity (Chapter 5).

Chapter 5

Engineering the specificity of NanH

5.1 Introduction

5.1.1 Bacterial sialidase specificities

Bacterial sialidases hydrolyse terminal sialic acids that are linked by α2,3, α2,6, or α2,8 glycosidic linkages to a diverse range of substrates (Figure 1.3, section 1.2.1) (Kim *et al*., 2011). Some bacterial sialidases show preferences for a specific glycosidic linkage and have different affinities towards sialoglycoconjugate substrates despite their shared structural architecture. Sialidases produced by *C. septicum*, *C. sordellii* and *C. tertium* are dual specific (can cleave either α2,3- and α2,6-linked sialic acid) but have a higher hydrolysis rate (V_{max}) towards α2,3-linked sialic acid (Zenz *et al*., 1993; Roggentin *et al*., 1987; Grobe *et al*., 1998). In contrast, the sialidase produced by *C. diptheriae* could be considered to have a slight preference for α2,6-linked sialic acid, however the relative sialidase activity was only 10 % higher for α2,6-linkages over α2,3-linkages (Kim *et al*., 2010). Currently, there are no known sialidases that have been determined to be entirely specific to α2,6-linked sialic acid and therefore no structural features that are responsible for this specificity have been determined.

S. typhimurium LT2 sialidase is a hydrolytic sialidase which produces the Neu5Ac as the reaction product and was the first bacterial sialidase proposed to have a preference for α2,3-linked sialic acid. Kinetics data showed the enzyme to exhibit a 260-fold kinetic preference (V_{max}) for α2,3-linked sialic acid over α2,6-linked sialic acid and that it is able to hydrolyse sialic acid from glycoproteins and gangliosides, but not mucins (Hoyer *et al*., 1991). In addition, SiaBb2 from B*ifidobacterium bifidum* is said to have a strong preference for α2,3-linked sialic acid, however a crystal structure has not yet been solved and structural features responsible for this specificity have not yet been determined (Kiyohara *et al*., 2011). Another well-studied example is the NanL enzyme of *M. decora* (North American Leech) that has a strict specificity for α2,3-linked sialic acid, but is actually classed as an IT-sialidase (section 1.3.2) with a reaction product of 2,7-anhydro-Neu5Ac rather than Neu5Ac (Luo *et al*., 1998). NanB of *S. pneumoniae* and RgNanH of *R. gnavus* also have strict specificities for α2,3-linked sialic acid and again are IT-sialidases (Gut *et al*., 2008; Tailford *et al*., 2015).

5.1.2 Relating sialidase structure to specificity

There are few studies in which structure-function and specificity have been considered, however some studies allude to specificity dependent structural features, one example is NanL from *M. decora* which has two additional bulky residues (W734 and Y643) above the arginine triad, which are said to be responsible for the strict α2,3 specificity due to the stacking between the two sidechains (Figure 5.1a) (Luo *et al*., 1998). W734 has a bulky aromatic sidechain which fills a vacant space that is observed in some bacterial sialidase structures and introduces predicted steric hindrance above the 2-carboxylate of the sialic acid in the substrate. It is thought that the position of W734 encourages 2-deoxy-2, 3-didehydro-D-N-acetylneuraminic acid (Neu5Ac2en), an unsaturated sialic acid, to bind deeper into the active site of NanL compared to other sialidases (Luo *et al*., 1998). There is space directly above the sialyl group that can hold the underlying sugar to sialic acid, regardless of the type of glycosidic linkage. However, based upon molecular modelling, the number of orientations of the terminal sialic acid are reduced and therefore substrates that contain anything but α2,3-linked sialic acid will clash with the W734 sidechain (Luo *et al*., 1998).

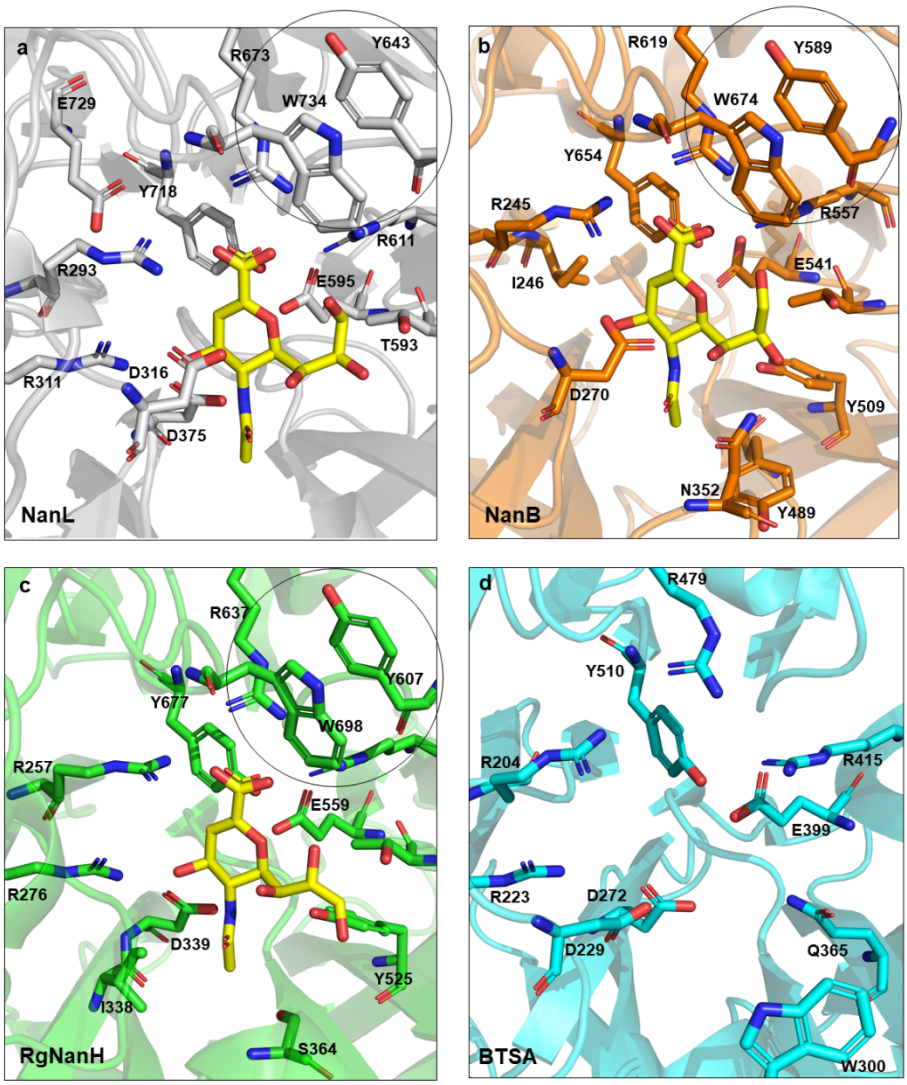


Figure 5.1: The stacking of the aromatic sidechains from tryptophan and tyrosine residues are the structural features proposed to be responsible for the α2,3 specificity of IT-sialidases. (a) NanL from *M. decora* in complex with DANA (yellow) (PDB: 1SLI). **(b)** NanB from *S. pneumoniae* in complex with DANA (yellow) (PDB: 2VW1). **(c)** RgNanH from *R. gnavus* in complex with DANA (yellow) (PDB: 4X47). **(d)** BTSA from *B. thetaiotaomicron* (wide substrate specificity) (PDB: 4BBW). The aromatic stacks are circled (a-c). The protein backbones are shown in cartoon mode, whilst the key residues are highlighted as sticks. This figure was produced using PyMOL.

The active site of NanB is highly similar to that of NanL, with residues Y589 and W674 corresponding to Y643 and W734 in NanL (Figure 5.1b) (Gut *et al*., 2008). A hydrophobic stack is also seen in RgNanH of *R. gnavus* and is formed by Y607 and W698 (Figure 5.1c) (Tailford *et al*., 2015). In contrast, there is no tryptophan corresponding to W734 in NanL in the crystal structure of the dual specific BTSA from *B. thetaiotaomicron* (Figure 5.1d). BTSA has a wide-open entrance exposing R479, one of the triad

members. It is this wide groove and lack of a tryptophan that is thought to be the structural basis for the broad specificity of BTSA. (Park *et al*., 2013).

Similarly to the proposed role of these bulky residues in proximity to the arginine triad, the α2,3-linked sialic acid preference of the *S. typhimurium* sialidase is suggested to be the result of the stacking of Y307 and H278, which stack in a similar location in the β-propeller as seen with the bulky residues in the β-propeller of NanL. Y307 is part of a loop extension that appears to be missing in other bacterial sialidases (Luo *et al.,* 1998).

Chapter 4 briefly introduced the α2,3 specific trans-sialidase TcTS from *T. cruzi* (section 4.3.6). Once again, a tryptophan (W312) and tyrosine (Y119) are proposed to be responsible for this specificity, as when a point mutation was made upon W312, α2,3 specificity was lost (section 5.1.3) (Paris *et al*., 2001). In TcTS, these aromatic residues are arranged differently to those in NanL, NanB and RgNanH (Figure 5.2a). The arrangement of W312 and Y119 result in a narrow entrance to the binding pocket, resulting in α2,3 specificity (Figure 5.2b). W312, like Y307 in the *S. typhimurium* sialidase, is also part of a loop extension that is missing in bacterial sialidases. As discussed in Chapter 3, NanH from *T. forsythia* is a broad specificity sialidase with high sequence similarity to BTSA (65 %). Similarly to BTSA, the active site of NanH has a tryptophan (W516, which corresponds to W508 in BTSA), which forms a low platform with an axial orientation of the side chain and creates a wide-open entrance in NanH and BTSA (Figure 5.2c), which is noticeably different to the binding site entrance of TcTS.

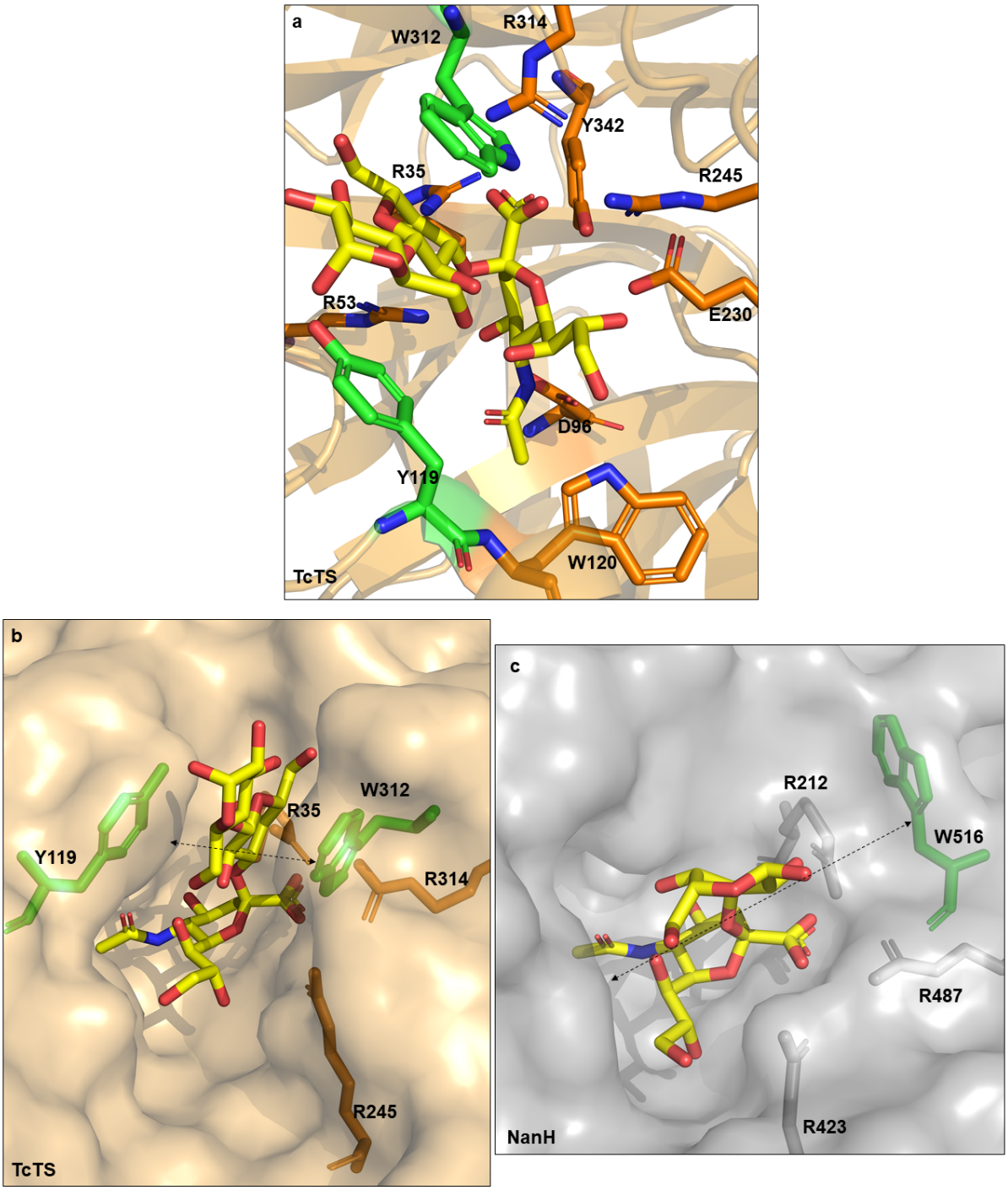


Figure 5.2: The differences between the binding pockets of TcTS-D59A from *T. cruzi* and NanH from *T. forsythia*. (a) TcTS-D59A (PDB: 1S0I) in complex with 3-SL (yellow sticks). Y119 and W312 are highlighted as green sticks. **(b)** A surface representation of the active site of TcTS-D59A (PDB: 1S0I) in complex with 3-SL (yellow sticks), with the narrow entrance indicated by an arrow. **(c)** A surface representation of NanH-D237A in complex with 3-SL (yellow sticks), with the wide entrance to the binding pocket indicated by an arrow. This figure was made using PyMOL.

5.1.3 Sialidase mutagenesis studies

The literature involving bacterial sialidases contains a few studies regarding the mutagenesis of conserved active site residues to determine their functions in the enzyme mechanism. For example, a Tyr/Glu pair are thought to be the catalytic nucleophile involved in catalysis and therefore essential for sialidase activity (Buschiazzo and Alzari, 2008). However, Watson and colleagues showed that point mutants of this conserved tyrosine (Y370A, Y370G and Y370D) in *M. viridifaciens* retained significant activity, arguing against this. This revealed some plasticity of nucleophilic catalysis in sialidases and more data from various other bacterial sialidases would confirm the validity of this finding (Watson *et al*., 2003).

There have been no mutational studies published about whether the specificity of broad specificity sialidases (hydrolytic sialidases) can be altered to become specific to either α2,3- or α2,6-linked sialic acid. As mentioned in section 4.3.6, the strict α2,3 specificity of the protozoan sialidase, TcTS, was probed using single point mutations (Paris *et al*., 2001). It was suggested that the α2,3 specificity of TcTS arises from the aromatic residue W312, as when W312 was mutated to an alanine, the resulting variant (W312A) was capable of cleaving both α2,3- and α2,6-linked sialic acid (Paris *et al*., 2001).

5.1.4 Chapter aims

As stated earlier in this thesis, the overall aim of this project was to gain an understanding of the biochemical and structural properties of microbial sialidases with a view to engineering their specificities for glycoanalytical pipelines. Chapter 3 provided evidence that NanH from *T. forsythia* was a dual specific sialidase, with a slight preference for α2,3-linked sialic acid, and an appropriate candidate for engineering, as it is highly active and can be produced at high yields with evidence from this thesis of long-term storage and shelf life potential. Chapter 4 provided understanding of the crystal structure in complex with substrates containing α2,3- and α2,6-linked sialic acid. Sialidases specific for α2,6-linked sialic acid would be highly useful in glycoanalytics of pharmaceutical compounds. However, there are currently no known α2,6-specific sialidases. This chapter describes the production and analysis of several NanH variants by SDM based upon ideas that have been gained from the literature and what was seen in the crystal structure of NanH in an attempt alter the specificity of NanH to either α2,3- or α2,6-linked sialic acid only and also probe whether 2,6 specificity is possible.

5.2 Results

5.2.1 Mutagenesis of NanH to improve α2,3 specificity

The NanH variants described in this section were constructed in the early stages of this project when the only crystal structure obtained was the NanH-HEPES model and there was a lack of understanding about which residues could perhaps alter the specificity. Section 5.1.2 introduced the structural features behind the α2,3 specificity of IT- and trans-sialidases, which generally involves the stacking of aromatic residues above where the 2-carboxylate of sialic acid would sit in the active site. Bulky aromatic residues are considered responsible for α2,3 specificity in the sialidase from *S. typhimurium* (Y307 and H278) (Figure 5.3a) (Luo *et al*., 1998) were determined to be absent in the same region in *T. forsythia* NanH (Figure 5.3b). Therefore, in attempt to alter the specificity of NanH, which, in the WT form has a wide substrate specificity, single point mutants were made to incorporate these aromatic residues into the protein structure. This also involved making additional mutants prior to the addition of the aromatic residues to generate the missing loop extension mentioned in section 5.1.2.

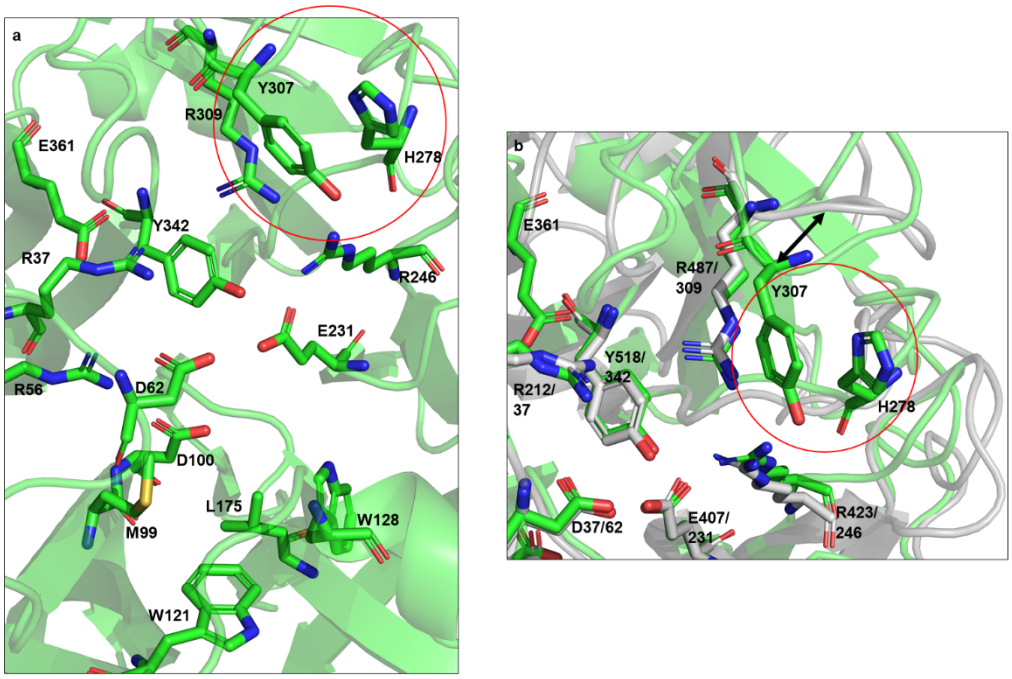


Figure 5.3: The Y307/H278 stack in the sialidase of *S. typhimurium*. (a) The active site region of the sialidase from *S. typhimurium*, with the key residues highlighted as green sticks. The Y307/H278 stack is surrounded by a red circle. **(b)** The active site region of *S. typhimurium* (green) overlaid with that of NanH (grey) to emphasise the lack of aromatic residues in this region in NanH. A black arrow indicates the larger looping region of the *S. typhimurium* sialidase compared to that of NanH. This figure was created using PyMOL.

A histidine and tyrosine (to replicate H278 and Y307 from the *S. typhimurium* sialidase) were inserted into the NanH-pET21a construct using SDM (see section 2.3 for more detail and Table2.3 for list of primers used). The mutagenesis was performed in several stages to eventually build NanH variant 5, which contained both the histidine and tyrosine insertions (Table 5.1). The variant from each step was analysed for protein overexpression, activity and specificity.

Firstly, it was determined that the histidine would be inserted between residues S455 and I456 in NanH to fall in approximately the same position (relative to the conserved active site) as the histidine in the *S. typhimurium* sialidase. The insertion of this histidine produced variant 1 (Table 5.1), with WT NanH DNA being used as the template. Variant 2 involved the addition of a tyrosine, along with a threonine, which were inserted between residues G486 and R487 (variant 2, Table 5.1). The threonine was inserted directly after the tyrosine in the hope that this would extend the loop in NanH to match the loop seen in the *S. typhimurium* sialidase. A further mutation was made to determine whether specificity could be altered without extending the G486-R487 loop. This involved substituting G486 for a tyrosine (G486Y) to mimic in NanH the location of the tyrosine in the *S. typhimurium* sialidase (relative to the conserved active site). This produced variant 3, with variant 1 being used as the DNA template (Table 5.1). Variant 4 involved the insertion of a tyrosine after G486, with variant 1 being used as the template (Table 5.1). The final construct (variant 5) contained both mutations from variant 1 and variant 2.

Template	NanH Variant	Primer set used (see methods, Table 2.3)
NanH-pET21a (WT)	(1) S455, +H, I456 (insertion)	SDM1
NanH-pET21a (WT)	(2) G486, +Y, +T, R487 (insertion)	SDM2
Variant 1-pET21a	(3) S455, +H, I456… G486Y (substitution)	SDM3
Variant 1-pET21a	(4) S455, +H, I456… G486, +Y, R487 (insertion)	SDM4
Variant 1-pET21a	(5) S455, +H, I456… G486, +Y, +T, R487 (variant 1 + variant 2)	SDM2

Table 5.1: NanH variants 1-5 were produced by SDM in attempt to engineer an α2,3 specific sialidase. The '+' symbol and underline represent residues that have been inserted into the protein structure. The residues highlighted in red represent the variant 1 insertion that has been built upon to make further variants.

Once the SDM reactions had finished, the mutants were treated with DpnI for 1 h at 37 °C to remove any remaining template plasmid, before being transformed to *E. coli* XL1-Blue supercompetent cells.

Colony PCR confirmed which colonies contained plasmids with a band of the correct number of base pairs to represent NanH DNA. The plasmids were extracted from the *E. coli* and DNA sequencing provided by GATC Services (Eurofins Genomics) confirmed all 5 variants in Table 5.1 contained the expected mutation.

The plasmid DNA for each variant was transformed into *E. coli* BL21 Origami cells and overexpression trials were carried out under the same conditions in which NanH WT expresses. All NanH variants apart from variant 2 expressed when the *E. coli* was grown O/N at 25 °C, 200 rpm, post induction using IPTG (Figure 5.4). A small-scale protein purification protocol (starting from 100 mL of *E. coli* cells transformed with NanH variants) (section 2.8), using Ni-NTA resin for His tag purification, was optimised to purify the NanH variant protein faster than from the 0.5 L expressions discussed in section 3.2.2. This method was beneficial as it allowed the purification and buffer exchange of multiple proteins and analysis by SDS-PAGE to be carried out within 4 h, rather than several days. NanH variants 1, 2, 3 and 5 had over 95 % purity after the small-scale purification (Figure 5.4). The concentration of each variant obtained from 100 mL of culture was approximately 150 µL at 0.8-1.3 mg mL^{-1}. It was for unknown reasons why variant 2 did not express but since variant 5 expressed and purified (contains both the inserted histidine and tyrosine, as well as the threonine to extend the loop region), it was decided that the project could move on without repeating variant 2 at this stage.

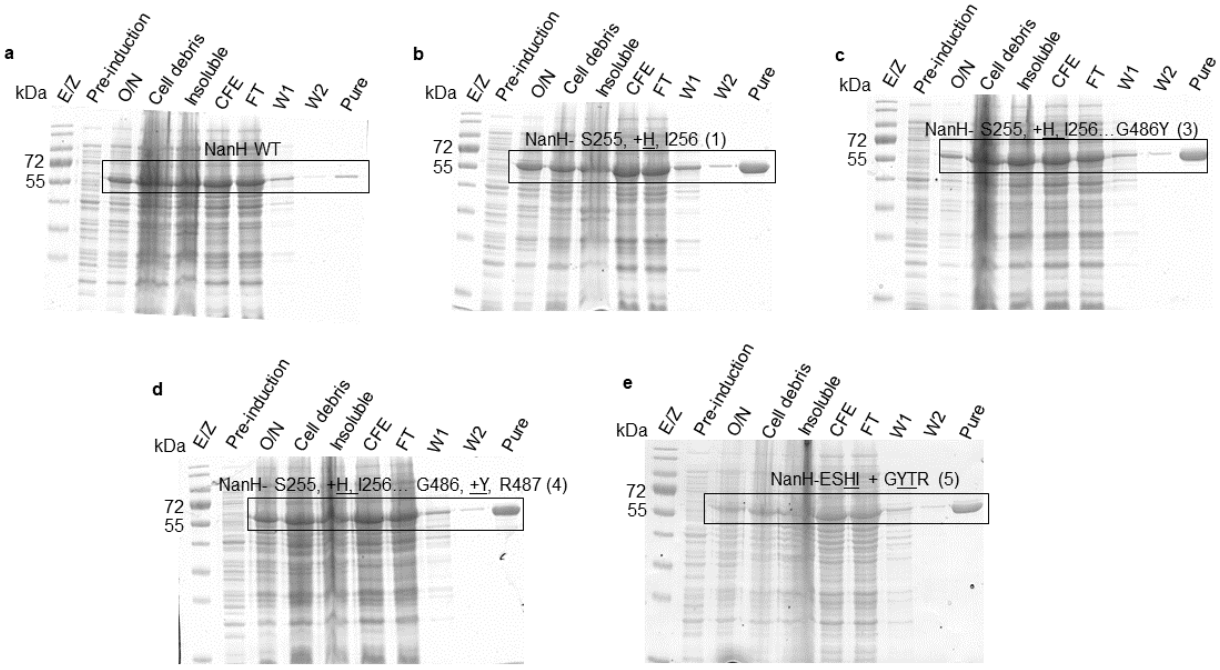


Figure 5.4: SDS-PAGE for the small-scale overexpression trial and purification of NanH variants 1, 3, 4 and 5, alongside a WT NanH positive control. (a) NanH WT positive control, **(b)** S255, +H, I256 (1), **(c)** Variant 1 + G486Y (3), **(d)** Variant 1 + G486, +Y, R487 (4), **(e)** Variant 1 + variant 2 (5). *E. coli* cells

containing NanH variants were induced after reaching OD600 0.6**.** Cell samples were collected prior to induction (**pre-induction**) and 18 h after induction (**O/N**). The *E. coli* were lysed, and the **cell debris** and **insoluble protein** were collected after two rounds of centrifugation. The cell free extract **(CFE)** was loaded to nickel resin for purification. The flow through **(FT)** and wash fractions **(W1, W2)** were collected before eluting the **pure protein** (section 2.8).

Post-purification and buffer exchange into 50 mM sodium phosphate, 200 mM NaCl pH7.4, the variants were assessed for sialidase activity with the synthetic substrate MU-NANA as in section 3.2.4. The data showed that variants 1, 3, 4 and 5 were inactive when increasing concentrations of MU-NANA were exposed to 2.5 nM of each variant, when compared to a WT NanH positive control (Figure 5.5a). The variants were also assessed for activity against both 3- and 6-SL in the TBA assay (used to detect free sialic acid, as in section 3.2.6), to determine whether the mutations had altered the specificity. However, no free sialic acid was detected after increasing concentrations of 3-SL and 6-SL were exposed to 50 nM of each NanH variant, which suggested that the mutations had inactivated the enzyme (Figure 5.5b).

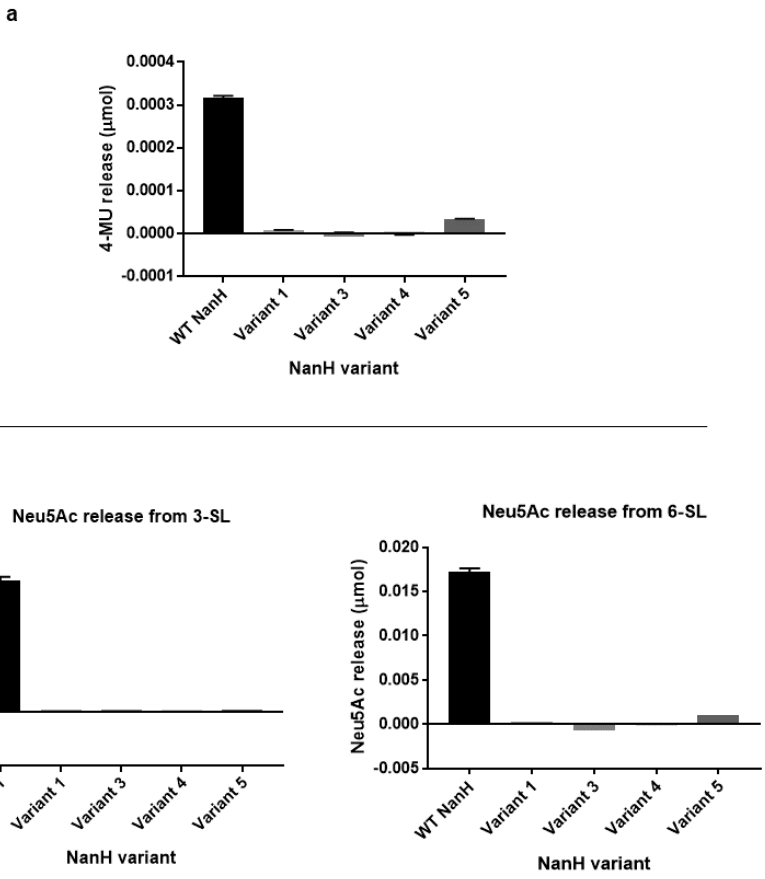


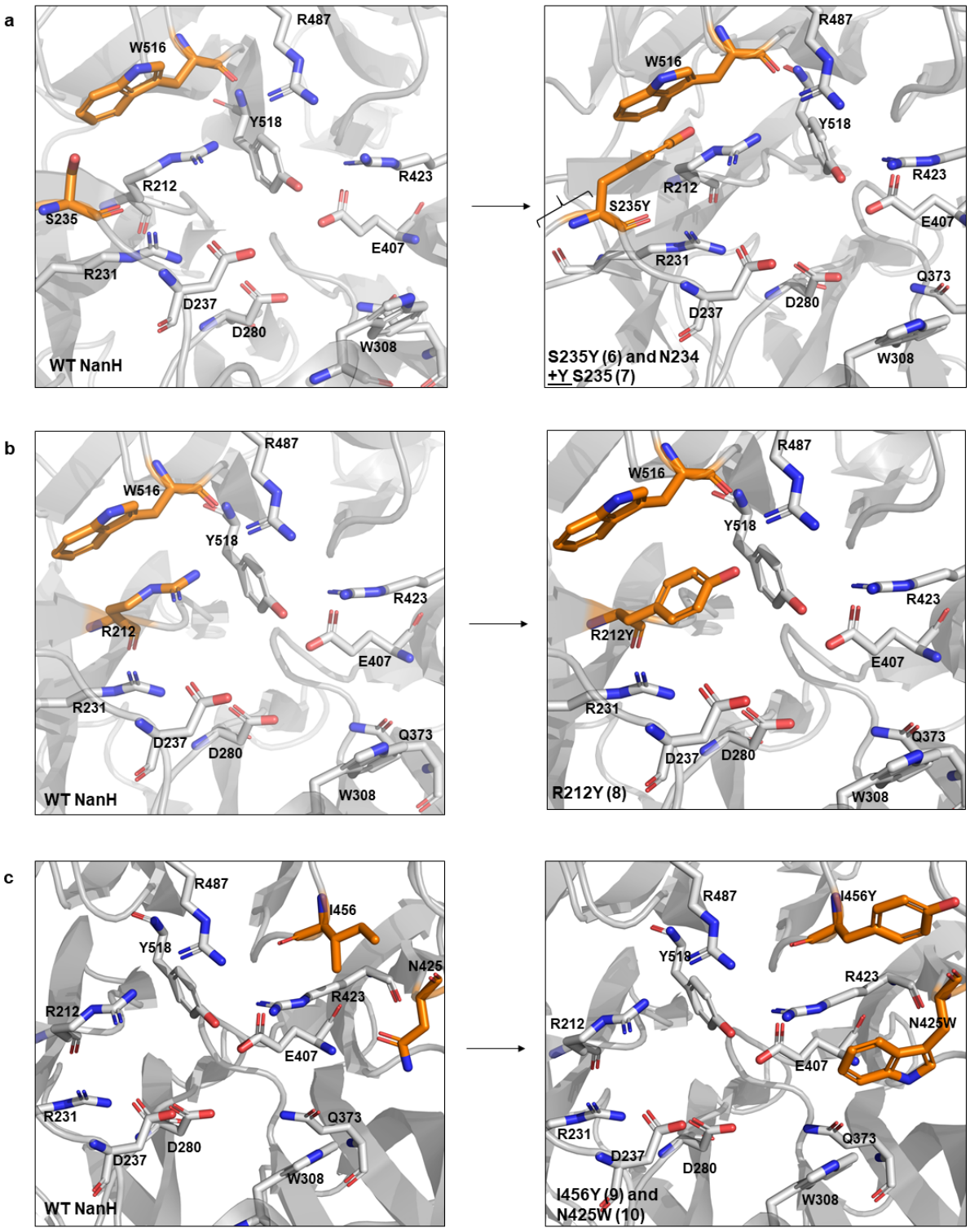
Figure 5.5: The activity of NanH variants 1, 3, 4 and 5 with MU-NANA and 3-/6-SL. (a) Increasing concentrations of MU-NANA were exposed to 2.5 nM NanH/variant. The reaction was stopped at 1, 2- and 3-min intervals and the amount of 4-MU produced (4-MU release, µmol) was determined by applying the raw data to a standard curve of known 4-MU amounts. The µmol release of 4-MU when 200 µM MU-NANA was digested by 2.5 nM NanH variant for 3 min was plotted as a bar chart for each variant. **(b)** Increasing concentrations of 3- and 6-SL were exposed to 50 nM NanH/variant. The reactions were stopped at 1-, 3- and 5-min intervals and the amount of Neu5Ac produced (µmol) was determined by applying the raw data to a standard curve of known Neu5Ac concentrations. The µmol release of Neu5Ac when 1 mM 3- and 6-SL were digested by 50 nM NanH variant for 5 min was plotted as a bar chart.

5.2.2 The additions of aromatic residues in the differing blades of the NanH β-propeller

The next batch of NanH variants constructed were based upon the idea that perhaps creating an aromatic stack like in section 5.2.1 around other blades of the NanH β-propeller would retain enzyme activity and potentially alter the specificity by imposing some kind of steric hindrance on the substrate orientation. The insertions and substitutions made are listed in Table 5.2 and the region they are found in the β-propeller is highlighted in Figure 5.6. It was decided to insert or substitute residues that were in the active site or near to the active site in the hope that these residues were close to the glycosidic linkage of the substrate. For some of the variants, only one aromatic residue was added, as there was already a second aromatic in the model for it to potentially stack with. But for other variants, two aromatic residues were to be added. The 'Mutagenesis Wizard' in PyMOL was used to mutate residues in the NanH model discussed in Chapter 4 (Figure 5.6). Some of the rotamers of the newly mutated aromatic residues were suggested to be unstable and possibly not the best residue to be placed in the chosen position. However, it was hoped that the rest of the protein would compensate for the change *in situ* since the mutagenesis tool is a prediction. Some of the mutations targeted conserved active site residues (D237 and R212) to determine what would happen to the activity as well as the specificity if they were substituted.

Template	NanH variant	Primer set used (see methods, Table2.3)
NanH-pET21a WT	(6) S235Y (substitution)	SDM6
NanH-pET21a WT	(7) N234 +Y S235 (insertion)	SDM7
NanH-pET21a WT	(8) R212Y (substitution)	SDM8
NanH-pET21a WT	(9) I456Y (substitution)	SDM9
Variant 9-pET21a	(10) I456Y, N425W (substitution)	SDM10
NanH-pET21a WT	(11) L238W (substitution)	SDM11
Variant 11-pET21a	(12) L238W, A307Y (substitution)	SDM12
NanH-pET21a WT	(13) D237Y (substitution)	SDM13

Table 5.2: NanH variants 6-13 were produced by SDM in attempt to alter the specificity of NanH through the addition of aromatic residues. The '+' symbol and underline represent residues that have been inserted into the protein structure. The residues highlighted in red represent a previous variant that has been built upon to make a further variant.


a
R487
W516
Y518
R423
R212
S235
E407
R231
D237
D280
W308
WT NanH
R487
W516
Y518
R423
R212
S235Y
E407
R231
Q373
D237
D280
W308
S235Y (6) and N234
+Y S235 (7)
b
R487
W516
Y518
R423
R212
E407
R231
D237
D280
Q373
W308
WT NanH
R487
W516
Y518
R423
R212Y
E407
R231
D237
D280
Q373
W308
R212Y (8)
c
R487
I456
Y518
N425
R423
R212
E407
R231
D237
D280
Q373
W308
WT NanH
R487
I456Y
Y518
R423
R212
N425W
E407
R231
D237
D280
Q373
W308
I456Y (9) and
N425W (10)

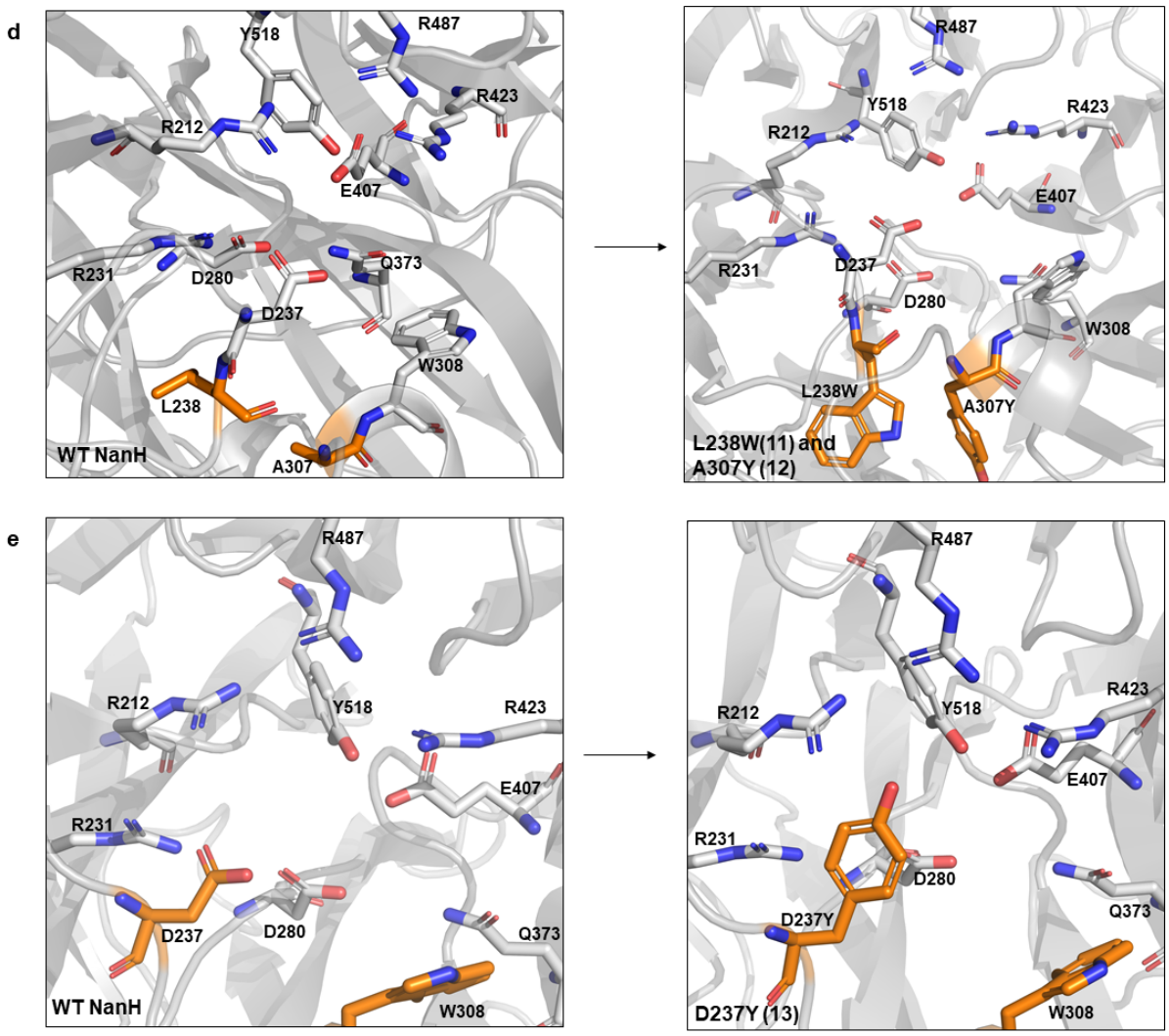


Figure 5.6: The location of where aromatic residues are to be added into the NanH active site and surrounding regions. All figures show the active site with the key conserved residues and the surrounding region in NanH-apo. The figures to the left of the arrow highlight the residue or residues that are to be substituted (orange sticks), along with a second residue in orange that it was hoped it would form a stacking interaction with. The figures to the right of the arrows shows a potential rotamer of the mutated residue based upon a prediction by 'Mutagenesis Wizard' in PyMOL. **(a)** S235Y substitution (6), with the left brace highlighting where the tyrosine would approximately be inserted after N234 (7). **(b)** R212Y substitution (8). **(c)** I456Y and N425W substitutions (9, 10). **(d)** L238W and A307Y substitutions (11, 12). **(e)** D237Y substitution (13). These figures were made in PyMOL.

SDM reactions were performed upon WT NanH-pET21a using the same method used in section 5.2.1 and the plasmid DNA for each NanH variant was transformed into *E. coli*. Plasmids confirmed to contain DNA of the correct number of base pairs to be NanH by colony PCR were extracted from the

E. coli and DNA sequencing provided by GATC Services (Eurofins Genomics) confirmed all variants in Table 5.2 apart from variant 10 contained the expected mutation. All of the NanH variants (apart from variant 10) were expressed in *E. coli* BL21 Origami cells and purified under the same conditions as in section 5.2.1 (Figure 5.7). The production of variants obtained from 0.5 L of culture was 10 mL at concentrations between 0.5 and 4.2 mg mL^{-1}.

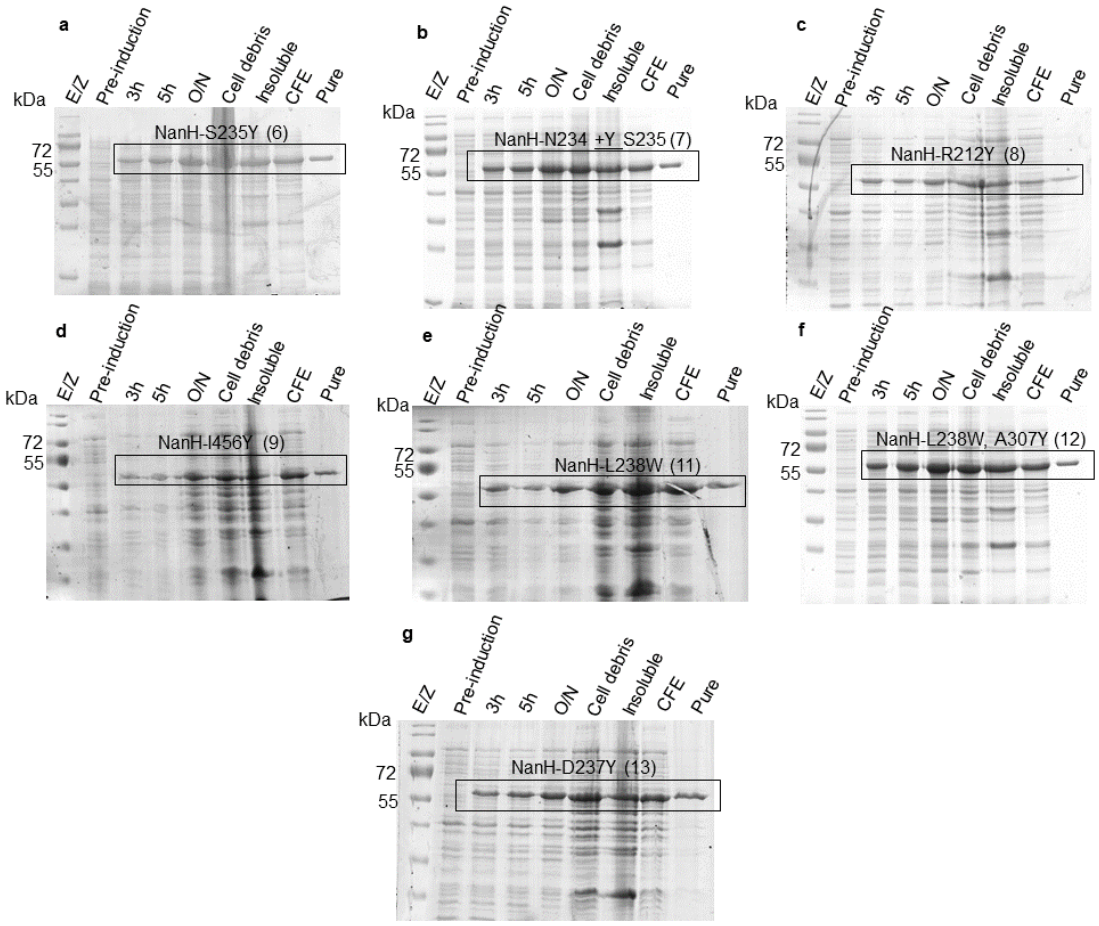


Figure 5.7: SDS-PAGE for the small-scale overexpression trial and purification of NanH variants 6-9 and 11-13. (a) S235Y (6), **(b)** N234, +Y, S235 (7), **(c)** R212Y (8), **(d)** I456Y (9), **(e)** L238W (11), **(f)** L238W, A307Y (12), **(g)** D237Y (13). *E. coli* cells containing NanH variants were induced after reaching OD600 0.6**.** Cell samples were collected prior to induction (**pre-induction**), **3 h**, **5 h** and 18 h after induction (**O/N**). The *E. coli* were lysed, and the **cell debris** and **insoluble protein** were collected after two rounds of centrifugation. The cell free extract **(CFE)** was loaded to nickel resin for purification. The resin was washed to remove impurities before eluting the **pure protein** (see section 2.8).

After purification and buffer exchange into 50 mM sodium phosphate, 200 mM NaCl pH 7.4, the NanH variants were assessed for sialidase activity with the sialidase substrate MU-NANA. Reaction kinetics

for the NanH variants were obtained by quantifying the 4-MU release from increasing concentrations of MU-NANA and applying a standard curve of known amounts of 4-MU, as in section 3.2.4. The data showed that NanH variants 6, 7 and 9 displayed sialidase activity when increasing concentrations of MU-NANA were exposed to 2.5 nM of each variant, however variant 8, 11, 12 and 13 were inactive (Figure 5.8a). NanH variants 6, 7 and 9 had V_{max} values of 7.1, 10.6 and 5.7 µmol min^{-1} mg^{-1} respectively whilst the V_{max} of WT NanH was determined to be 13.8 µmol min^{-1} mg^{-1} in section 3.2.4 (Figure 5.9a, Table 5.3). The K_M values were calculated to be 7.2, 16.6 and 18.8 µM for variants 6, 7 and 9, respectively, whilst the K_M of WT NanH was determined to be 6.2 µM. The k_{cat} values (calculated as described in section 3.2.4) were divided by the K_M (k_{cat}/K_M) to determine the catalytic efficiency. The k_{cat}/K_M values for NanH variants 6, 7 and 9, were determined to be 57.3, 37.0 and 17.5 min^{-1} $µM^{-1}$ respectively, whereas the k_{cat}/K_M for WT NanH was determined to be 129.0 min^{-1} $µM^{-1}$. This showed that these 3 variants remained active but were less efficient than WT NanH.

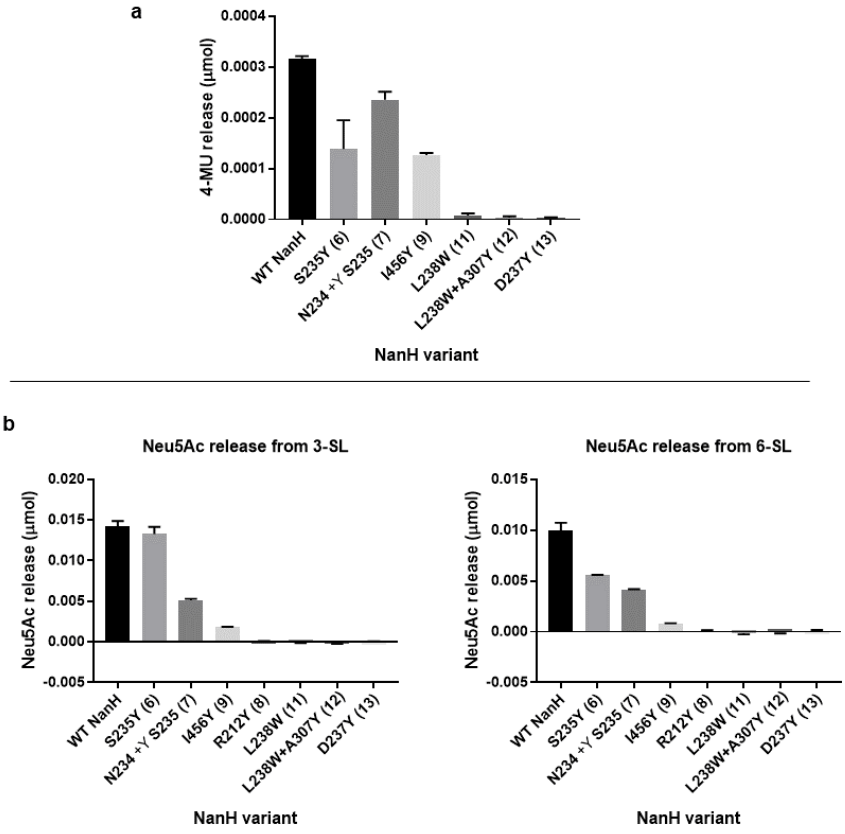


Figure 5.8: The activity of NanH variants 6-13 with MU-NANA and 3-/6-SL. (a) Increasing concentrations of MU-NANA were exposed to 2.5 nM NanH/variant. The reaction was stopped at 1-, 2- and 3-min intervals and the amount of 4-MU produced (4-MU release, µmol) was determined by applying the raw data to a standard curve of known 4-MU amounts. The µmol release of 4-MU when 200 µM MU-NANA was digested by 2.5 nM NanH variant for 3 min was plotted as a bar chart for each variant. **(b)** Increasing concentrations of 3- and 6-SL were exposed to 50 nM NanH/variant. The

reactions were stopped at 1-, 3- and 5-min intervals and the amount of Neu5Ac produced (µmol) was determined by applying the raw data to a standard curve of known Neu5Ac concentrations. The µmol release of Neu5Ac when 0.5 mM 3- and 6-SL were digested by 50 nM NanH variant for 5 min was plotted as a bar chart.

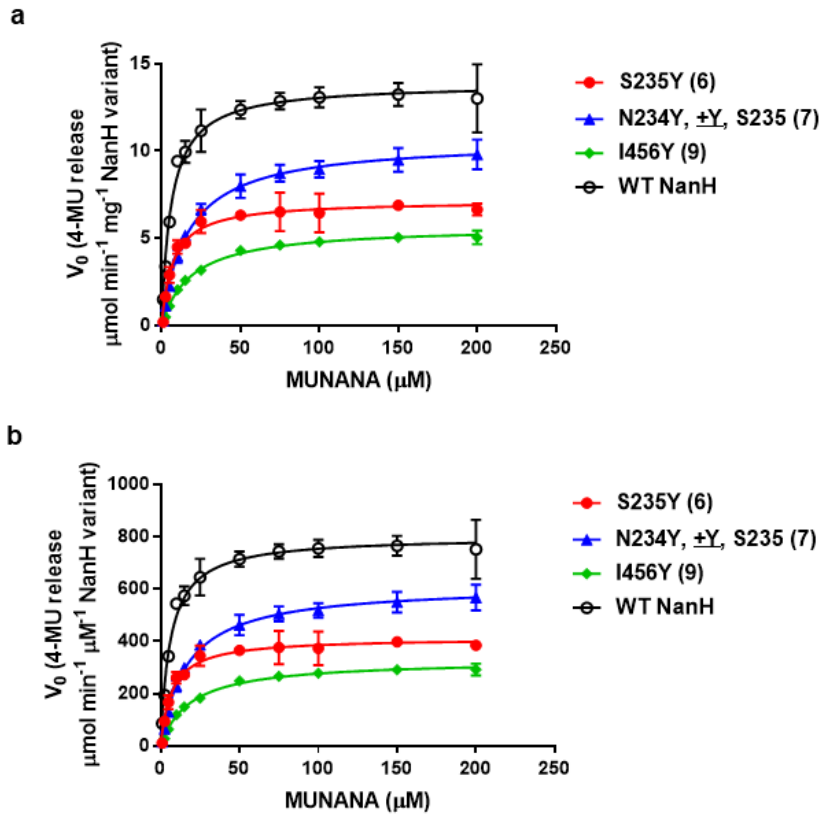


Figure 5.9: The reaction kinetics of NanH variants 6, 7 and 9 with MU-NANA. (a) Increasing concentrations of MU-NANA were exposed to 2.5 nM NanH/variant. The reaction was stopped at 1-, 2- and 3-min intervals and the rate of reaction (4-MU release per mg NanH variant) was determined by applying the raw data to a standard curve of known 4-MU concentrations. The rate was plotted against [MU-NANA] in a Michaelis-Menten curve from which the V_{max} and K_M could be calculated (Table 5.3). **(b)** The 4-MU release (per µM NanH variant) was plotted against [MU-NANA] enabling the k_{cat} to be calculated and in turn the k_{cat}/K_M (Table 5.3). In both cases the data represents the mean of three repeats, with the error bars representing the standard deviation.

	S235Y (6)	N234 +Y S235 (7)	NanH-I456Y (9)	WT NanH
V_{max} (µmol min^{-1} mg^{-1})	7.1 ± 0.2 (6.8 to 7.5)	10.6 ± 0.2 (10.2 to 11.0)	5.7 ± 0.1 (5.5 to 5.9)	13.8 ± 0.2 (13.4 to 14.3)
K_M ([MU-NANA], µM)	7.2 ± 0.9 (5.7 to 9.0)	16.6 ± 1.1 (14.6 to 19.0)	18.8 ± 1.3 (16.5 to 21.4)	6.2 ± 0.5 (5.2 to 7.3)
k_{cat} (min^{-1})	412.7 ± 10.5 (392.3 to 433.9)	613.8 ± 11.2 (592.0 to 636.7)	328.5 ± 6.0 (316.9 to 340.7)	799.6 ± 13.7 (722.7 to 827.3)
k_{cat}/K_M (min^{-1} µM^{-1})	57.3	37.0	17.5	129.0

Table 5.3: Reaction kinetics values obtained from Michaelis-Menten curves of NanH variants 6, 7 and 9 with MU-NANA. The V_{max}, K_M, k_{cat} and k_{cat}/K_M values are presented (± standard error) for the reaction kinetics of MU-NANA with NanH variants. The values in parentheses display the 95 % confidence intervals. The WT NanH data was taken from section 3.2.4.

The TBA assay (used to detect sialic acid release, section 3.2.6) has a laborious protocol and there were several NanH variants to assess for altered specificity with 3- and 6-SL. Therefore, a quick TBA assay analysis was developed and performed on the variants that expressed and purified (Figure 5.7). In this assay, one concentration (0.5 mM) of 3- and 6-SL was exposed to 50 nM of each of the NanH variants and sialic acid release was detected using the same principles as in section 3.2.6. These concentrations were chosen, as it was at these concentrations where WT NanH reached the V_{max} on the Michaelis-Menten curve (Figure 3.10, section 3.2.6), which suggested these concentrations should be high enough to detect activity in the variants. NanH variants 6, 7 and 9 displayed activity with 0.5 mM 3- and 6-SL (Figure 5.8b) and were therefore exposed to increasing concentrations of the substrates to produce Michaelis Menten curves as in section 3.2.6. The data obtained from the TBA assays showed that NanH variants 6, 7 and 9 were active against both 3- and 6-SL, but at reduced activities or increased binding efficiencies compared to the WT (Figure 5.10, Table 5.4). Variants 6, 7 and 9 had k_{cat}/K_M values of 4306.0, 2249.0 and 567.8 min^{-1} mM^{-1} with 3SL respectively and 1672.2, 1340.7 and 348.9 min^{-1} mM^{-1} with 6-SL respectively. All three variants had a reduced k_{cat}/K_M compared to WT NanH, which had k_{cat}/K_M values of 8564.8 and 3920.3 min^{-1} mM^{-1} with 3- and 6-SL respectively (Table 5.4). Variant 6 has a higher ratio of the k_{cat}/K_M with α2,3-linked sialic acid to α2,6-linked sialic acid than that of WT NanH (2.6:1 compared to 2.1:1 respectively). Interestingly, variant 7 had a higher V_{max} with 6-SL than 3-SL (32.4 and 15.5 µmol min^{-1} mg^{-1} respectively) compared to WT NanH which has an almost equal V_{max} with 3- and 6-SL (46.3 and 43.4 µmol min^{-1} mg^{-1} respectively). However, the K_M had increased from 0.6 mM for WT NanH with 6-SL to 1.4 mM for variant 7 with 6-SL.

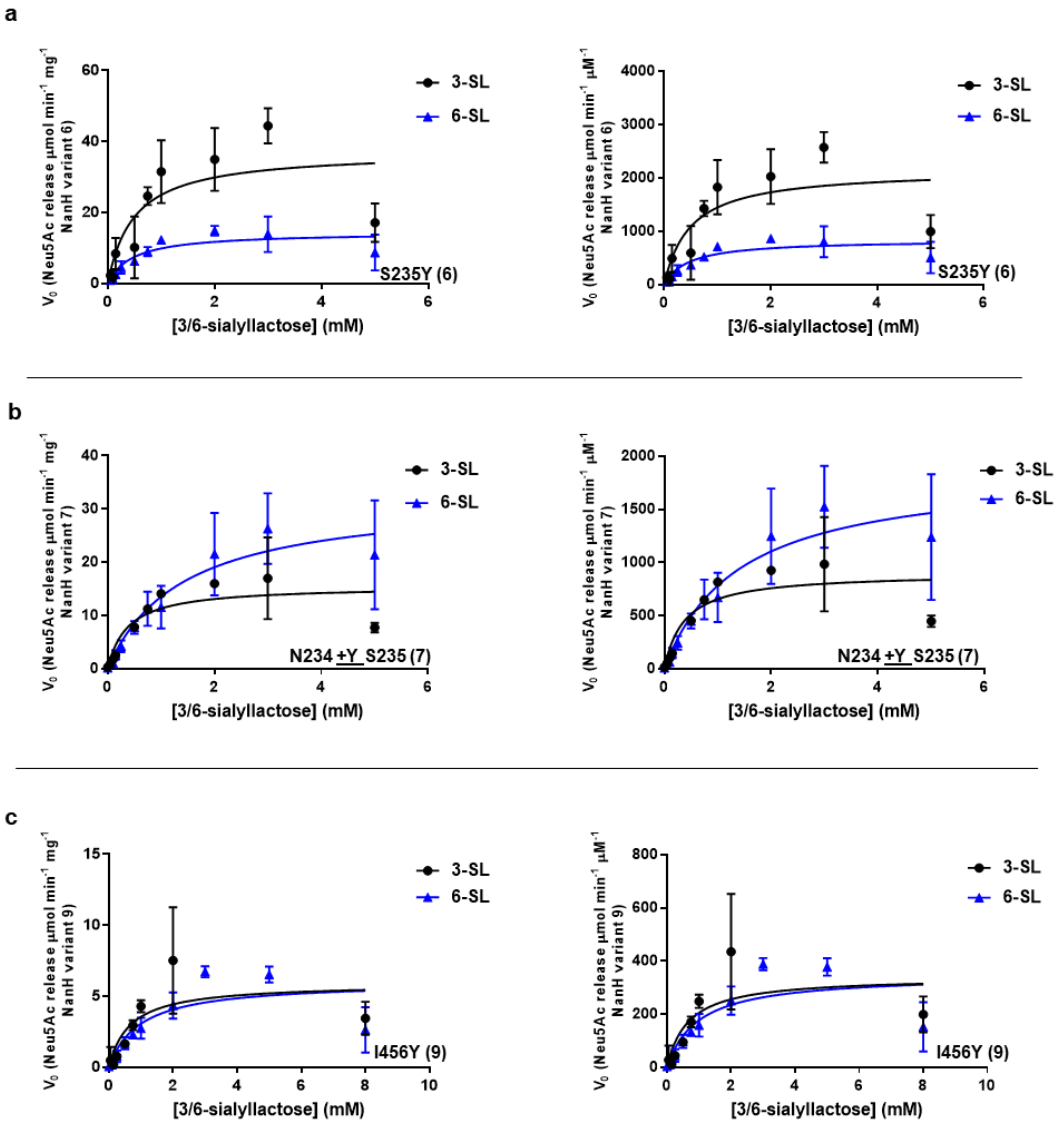


Figure 5.10: The reaction kinetics of 3-/6-SL with NanH variants 6, 7 and 9. (a) S235Y (6), **(b)** N234, +Y, S235 (7), **(c)** I456Y (9). The curves on the left show that increasing concentrations of 3-/6-SL were exposed to 50 nM NanH variant. The reaction was halted at 1-, 3- and 5-min intervals before performing the TBA assay to quantify free sialic acid. Sialic acid was quantified by applying the data to a standard curve of known concentrations of sialic acid that had also been passed through the TBA assay. The rate (Neu5Ac release per mg variant) was plotted against [3-/6-SL] in a Michaelis-Menten curve from which the V_{max} and K_M could be calculated (Table 5.4). The curves on the right show that Neu5Ac release (Neu5Ac per µM variant) was plotted against [3-/6-SL] enabling the k_{cat} to be calculated and in turn the catalytic efficiency (k_{cat}/K_M) (Table 5.4). In both cases the data represents the mean of three repeats, with the error bars representing the standard deviation.

	S235Y (6)		NanH-N234+Y S235 (7)		NanH-I456Y (9)		WT NanH	
	3-SL	6-SL	3-SL	6-SL	3-SL	6-SL	3-SL	6-SL
V_{max} (µmol min^{-1} mg^{-1})	37.2 ± 5.3 (28.8 to 48.4)	14.5 ± 1.5 (11.9 to17.5)	15.5 ± 1.8 (12.6 to 19.2)	32.4 ± 4.7 (25.1 to 43.9)	5.9 ± 1.1 (4.1 to 8.1)	6.0 ± 0.8 (4.8 to 7.6)	46.4 ± 1.2 (44.1 to 49.0)	43.4 ± 1.1 (41.3 to 45.8)
K_M ([3-/6-SL], mM)	0.5 ± 0.2 (0.2 to 1.1)	0.5 ± 0.2 (0.3 to 0.8)	0.4 ± 0.2 (0.2 to 0.8)	1.4 ± 0.5 (0.7 to 2.7)	0.6 ± 0.4 (0.3 to 1.5)	1.0 ± 0.4 (0.5 to 1.9)	0.3 ± 0.0 (0.3 to 0.4)	0.6 ± 0.1 (0.5 to 0.7)
k_{cat} (min^{-1})	2153.0 ± 305.2 (1666.0 to 2802.0)	836.1 ± 86.3 (690.5 to 1014.0)	899.6 ± 103.3 (729.8 to 1109.0)	1877.0 ± 269.6 (1456.0 to 2544.0)	340.7 ± 65.2 (238.4 to 467.6)	348.9 ± 44.1 (277.9 to 439.3)	2683.0 ± 69.7 (2550.0 to 2829.0)	2509.0 ± 63.5 (2226.0 to 2539.0)
k_{cat}/K_M (min^{-1} mM^{-1})	4306.0	1672.2	2249.0	1340.7	567.8	348.9	8654.8	3920.3

Table 5.4: Reaction kinetics values obtained from Michaelis-Menten curves of NanH variants 6, 7 and 9 with 3-/6-SL. The V_{max}, K_M, k_{cat} and k_{cat}/K_M values are presented (± standard error) for the reaction kinetics of 3-/6-SL with NanH variants. The values in parentheses display the 95 % confidence intervals. The WT NanH data was taken from section 3.2.6.

5.2.3 The mutagenesis of catalytic residues to investigate the activity and specificity of NanH

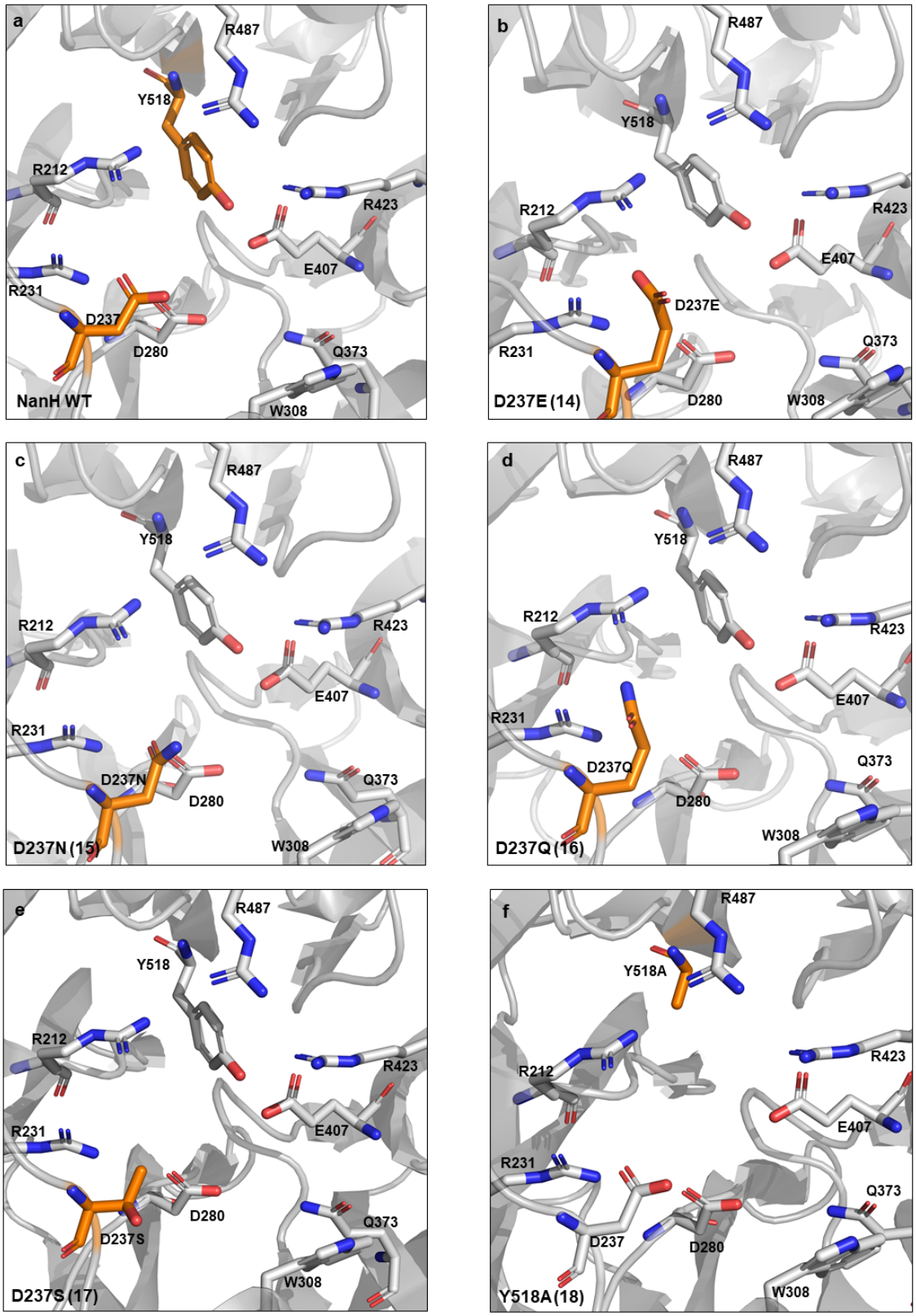
At this point in the project, there was yet to be a structure solved of an inactive NanH variant in complex with substrates containing α2,3- and α2,6-linked sialic acid which would be useful in providing ideas upon how to alter the specificity of NanH (NanH-D237A, section 4.2.5, was solved at a later stage). A study by Zhu and colleagues solved a structure of an inactive variant of an influenza A viral NA in complex with substrates containing α2,3- and α2,6-linked sialic acid (Zhu *et al.*, 2012). The conserved catalytic aspartate (D51) had been mutated to a glycine to produce a D51G variant. Analysis of these models showed that the glycosidic linkages of the α2,3- and α2,6-linked sialic acid sit in almost the same position, close to the D51G mutation (Figure 4.3, section 4.1.4). Therefore one could assume that the glycosidic linkage in α2,3- and α2,6-linked sialic acid could sit close to the conserved aspartate in NanH (D237) since bacterial sialidase active sites are also conserved with the influenza NA active site (Newstead *et al.*, 2008). It was hoped that mutating D237 to either a glutamate, asparagine, glutamine or serine would introduce a small change in this region to alter the specificity of NanH (Table 5.5, Figure 5.11).

Whilst the aim of this chapter was primarily to investigate whether the specificity of NanH could be altered, Y518 from the conserved Tyr/Glu pair was also mutated to determine its importance in catalysis. In section 5.1.3, it was mentioned that Watson and colleagues demonstrated that point mutants upon the tyrosine from the conserved Tyr/Glu nucleophilic pair (Y370A, Y370G and Y370D) in the active site of the sialidase from *M. viridifaciens* retained significant activity (Watson *et al.*, 2003). Y518 in NanH was mutated to an alanine, a glutamine, a serine, a threonine and a phenylalanine to

determine whether NanH has sialidase activity without what is considered to be the nucleophile (Table 5.5, Figure 5.11).

Template	NanH variant	Mutation purpose	Primer set used (see method section xx)
NanH-pET21a WT	(14) D237E	Investigate specificity	SDM14
NanH-pET21a WT	(15) D237N	Investigate specificity	SDM15
NanH-pET21a WT	(16) D237Q	Investigate specificity	SDM16
NanH-pET21a WT	(17) D237S	Investigate specificity	SDM17
NanH-pET21a WT	(18) Y518A	Investigate role in catalysis	SDM18
NanH-pET21a WT	(19) Y518Q	Investigate role in catalysis	SDM19
NanH-pET21a WT	(20) Y518S	Investigate role in catalysis	SDM20
NanH-pET21a WT	(21) Y518T	Investigate role in catalysis	SDM21
NanH-pET21a WT	(22) Y518F	Investigate role in catalysis	SDM22

Table 5.5: NanH variants (14-22) produced by SDM in attempt to alter the specificity of NanH and assess the activity of NanH.


a
R487
Y518
R212
R423
E407
R231
D237
D280
Q373
NanH WT
W308
b
R487
Y518
R212
R423
E407
D237E
R231
Q373
D237E (14)
D280
W308
c
R487
Y518
R212
R423
E407
R231
D237N
D280
Q373
D237N (15)
W308
d
R487
Y518
R212
R423
E407
R231
D237Q
Q373
D280
W308
D237Q (16)
e
R487
Y518
R212
R423
E407
R231
D280
Q373
D237S
W308
D237S (17)
f
R487
Y518A
R212
R423
E407
R231
Q373
D237
D280
W308
Y518A (18)

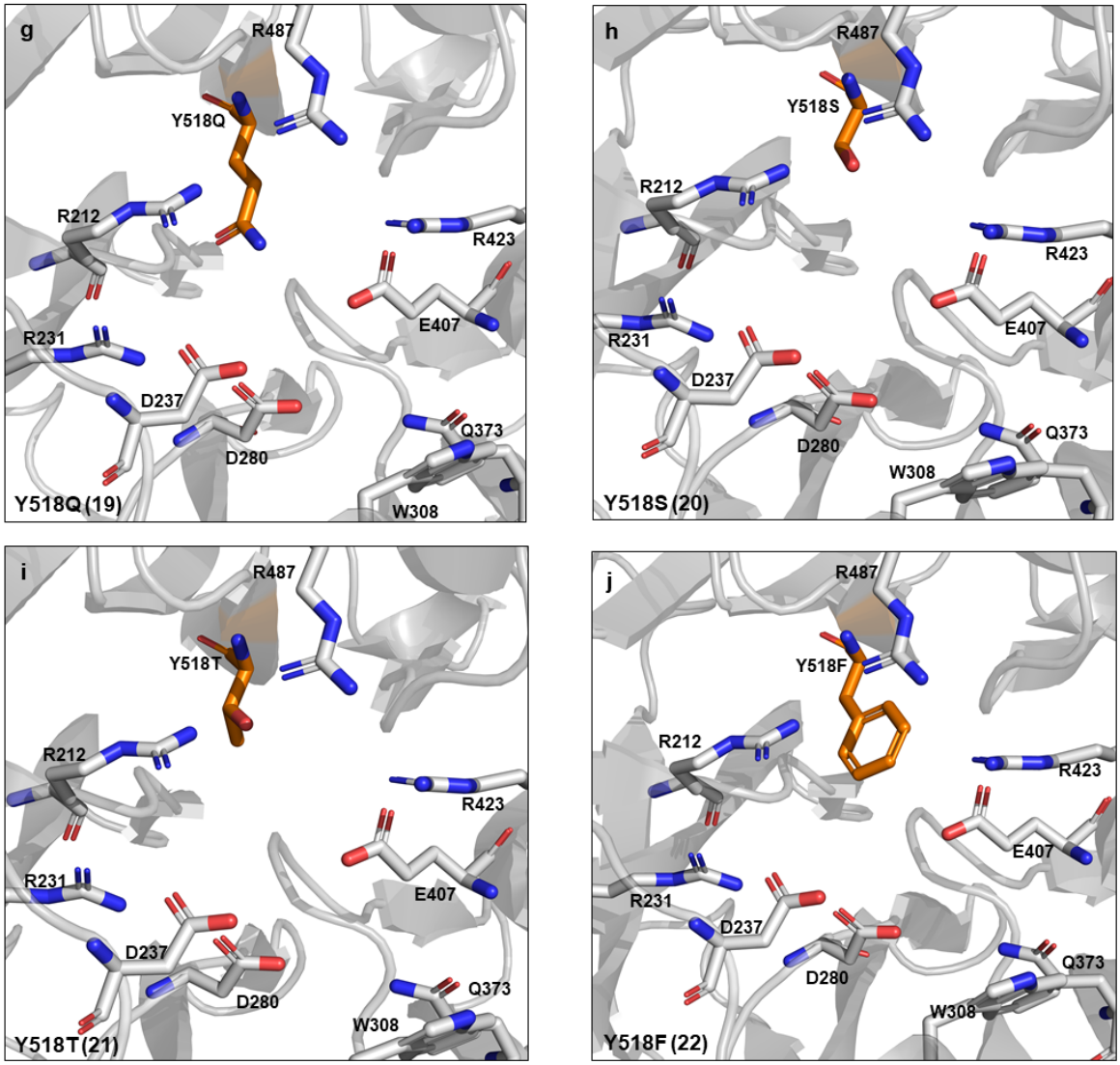


Figure 5.11: Location of residues D237 and Y518 in NanH. All figures show the active site with the key conserved residues and the surrounding region in NanH-apo. Figures **(b-j)** show a potential rotamer of mutated residue based upon a prediction by the 'Mutagenesis Wizard' in PyMOL **(a)** The NanH active site region prior to mutagenesis with D237 and Y518 shown as orange sticks. **(b)** D237E substitution. **(c)** D237N substitution. **(d)** D237Q substitution. **(e)** S237S substitution. **(f)** Y518A substitution. **(g)** Y518Q substitution. **(h)** Y518S substitution. **(i)** Y518T substitution. **(j)** Y518F substitution.

SDM reactions were performed upon WT NanH-pET21a using the same method used in section 5.2.1 and the plasmid DNA for each NanH variant was transformed into *E. coli*. The plasmids confirmed to contain a gene with the correct number of base pairs to be NanH, were extracted from the *E. coli* and DNA sequencing provided by GATC Services (Eurofins Genomics) confirmed all variants in Table 5.5

contained the expected mutation. All of the NanH variants, apart from variant 17 (D237S), expressed in *E. coli* BL21 Origami cells and purified under the same conditions as in section 5.2.1 (Figure 5.12). The amount of NanH variant obtained from a 100 mL culture was 150 µL at a concentration between 0.6 and 2 mg mL^{-1}.

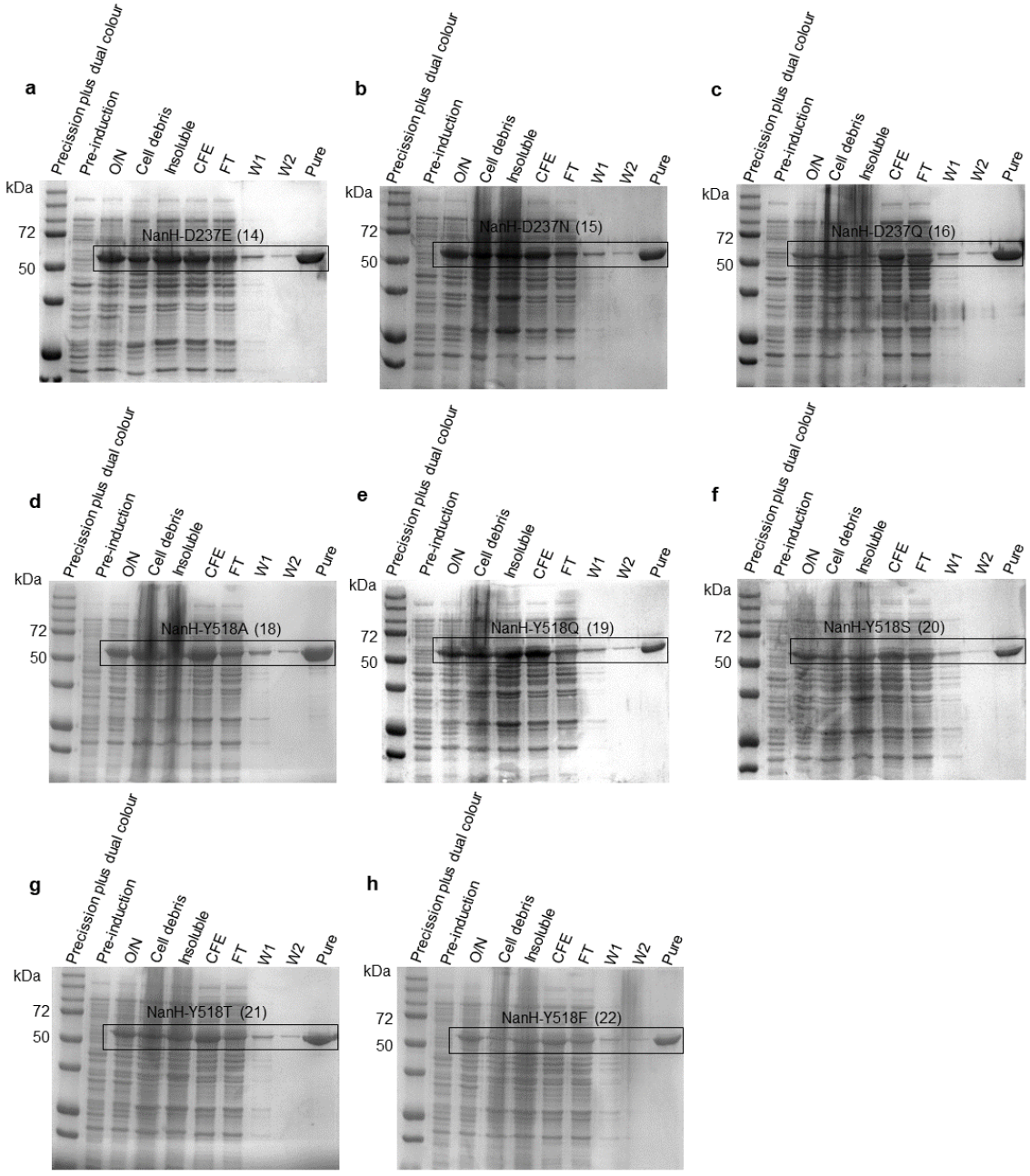


Figure 5.12: SDS-PAGE for the small-scale overexpression trial and purification of NanH variants (14-22). (a) D237E (14), **(b)** S237N (15), **(c)** D237Q (16), **(d)** Y518A (18), **(e)** Y518Q (19), **(f)** Y518S (20), **(g)** Y518T (21), **(h)** Y518F (22). *E. coli* cells containing NanH variants were induced after reaching OD600 0.6**.** Cell samples were collected prior to induction (**pre-induction**) and 18 h after induction (**O/N**). The *E. coli* were lysed, and the **cell debris** and **insoluble protein** were collected after two rounds of

centrifugation. The cell free extract **(CFE)** was loaded to nickel resin for purification. The flow through **(FT)** and wash fractions **(W1, W2)** were collected before eluting the **pure protein** (section 2.8).

Post-purification and buffer exchange into 50 mM sodium phosphate, 200 mM NaCl pH7.4, the variants that expressed (14-16 and 18-22) were assessed for sialidase activity against the synthetic substrate MU-NANA. To screen multiple variants at once, the assay size was reduced compared to that in section 3.2.4. 4 concentrations of MU-NANA (0, 10, 50 and 200 µM), as opposed to 12, were exposed to 2.5 nM of each variant, whilst all other variables remained the same as in section 3.2.4. However, there was no detectable 4-MU release from any of the NanH variants listed in Table 5.5 (Figure 5.13a). To be confident that the variants were indeed inactive, 200 µM MU-NANA was exposed to 25 nM of each variant. Once again, the level of 4-MU release by all variants was negligible compared to WT NanH (data not shown). It was also confirmed that 50 nM of each NanH variant was unable to hydrolyse sialic acid from 3- and 6-SL in a small-scale TBA assay, as in section 5.2.2 (Figure 5.13b). As a final test for activity or altered specificity, 0.5 mM of 3- and 6-SL were exposed to 100 nM of the variants and left to digest for approximately 18 h before the reaction was halted and assessed for sialic acid release in the TBA assay. However, sialic acid release was not detected by any of the NanH variants (data not shown).

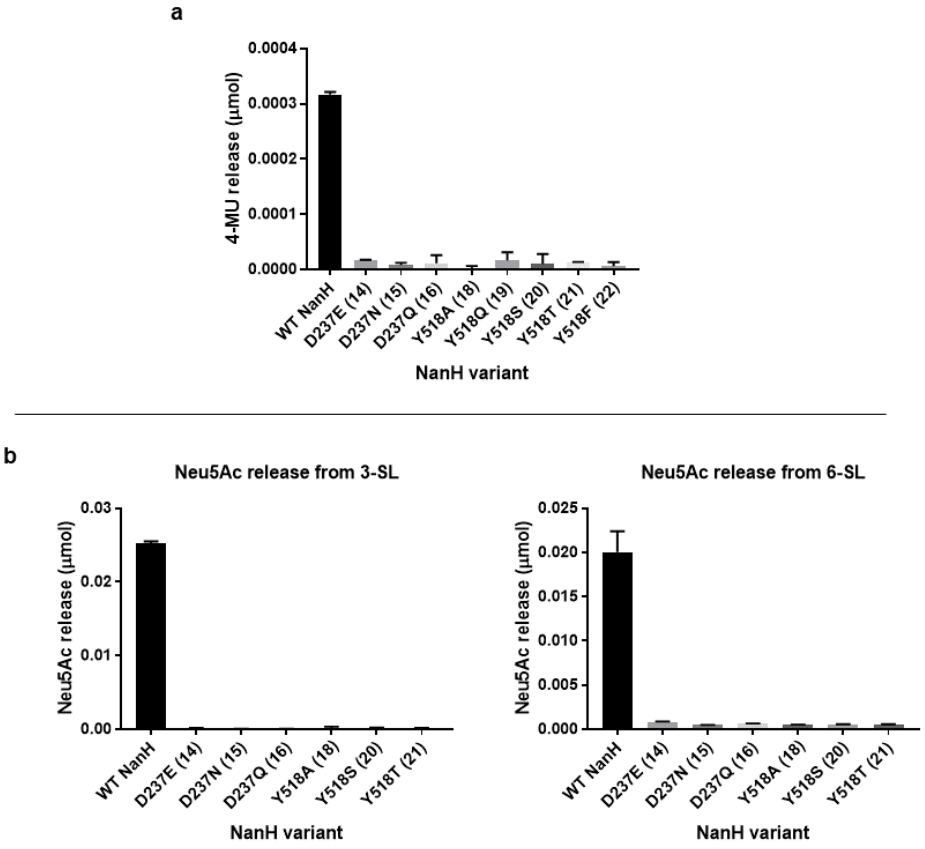


Figure 5.13: The activity of NanH variants (14-22) with MU-NANA and 3-/6-SL. (a) Increasing concentrations of MU-NANA were exposed to 2.5 nM NanH/variant. The reaction was stopped at 1-, 2- and 3-min intervals and the amount of 4-MU produced (4-MU release, µmol) was determined by applying the raw data to a standard curve of known 4-MU amounts. The µmol release of 4-MU when 200 µM MU-NANA was digested by 2.5 nM NanH variant for 3 min was plotted as a bar chart for each variant. **(b)** 0.5 mM of 3- and 6-SL were exposed to 50 nM NanH/variant. The reactions were stopped at 1-, 3-, 5- and 15-min intervals and the amount of Neu5Ac produced (µmol) was determined by applying the raw data to a standard curve of known Neu5Ac concentrations. The µmol release of Neu5Ac when 0.5 mM 3- and 6-SL were digested by 50 nM NanH variant for 15 min was plotted as a bar chart.

5.2.4 The mutagenesis of active site surrounding residues in NanH

Once the crystal structures of 3- and 6-SL in complex with NanH-D237A were solved (Figure 4.20, section 4.2.5), it was confirmed that the glycosidic linkage for both 3- and 6-SL sit in an almost identical position, close to what would be D237 in WT NanH, as predicted in section 5.2.3. Altering the region where the glycosidic linkages fall could alter the specificity of NanH. The data collected in section 5.2.3 suggested that mutations upon the conserved aspartate, D237, are likely to inactivate NanH even if

charge is retained since the D237E variant was inactive. V236 in NanH is not classed as a conserved residue that is crucial for sialidase activity and it is also near to the glycosidic linkages in the substrates. Valine is a hydrophobic neutral residue and it was substituted by SDM to both a glutamine (polar, hydrophilic with a neutral charge) and a tyrosine (polar and hydrophobic) to determine whether these changes would allow NanH to retain activity, whilst altering the specificity (Table 5.6, Figure 5.14). Section 4.2.3 also mentioned that R306 in NanH takes an uncommon rotamer. This arginine is conserved in BTSA of *B. thetaiotaomicron* and the sialidase of *P. distasonis* but its importance is unknown. Therefore R306 was substituted to an alanine in attempt to understand the function of this uncommon rotamer (Table 5.6, Figure 5.14).

Template	NanH variant	Mutation purpose	Primer set used (see method section xx)
NanH-pET21a WT	(23) V236Q (substitution)	Investigate specificity	SDM23
NanH-pET21a WT	(24) V236Y (substitution)	Investigate specificity	SDM24
NanH-pET21a WT	(25) R306A (substitution)	Investigate function	SDM25

Table 5.6: NanH variants 23, 24 and 25 that were produced by SDM in attempt to understand the function of certain residues and alter the specificity of NanH.

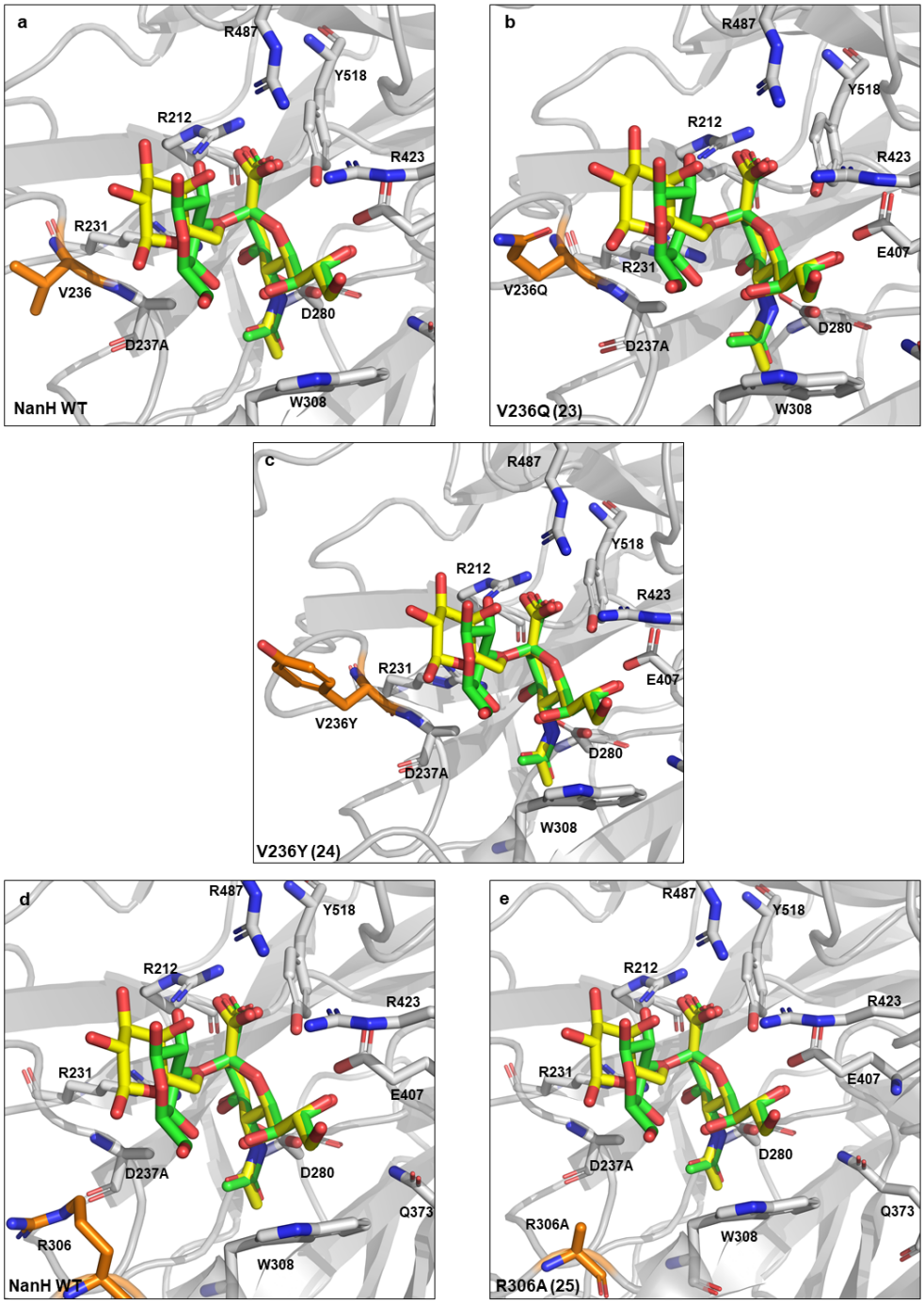


Figure 5.14: Location of residues V236 and R306 in NanH. All figures show the active site with the key conserved residues and the surrounding region of NanH-D237A in complex with 3- and 6-SL (green and yellow sticks respectively, overlaid). Figures (**b, c** and **e**) show a potential rotamer of the mutated residues based upon a prediction by the 'Mutagenesis Wizard' in PyMOL **(a)** The NanH active site region prior to mutagenesis with V236 shown as orange sticks. **(b)** V236Q (23) substitution. **(c)** V236Y (24) substitution. **(d)** The NanH active site region prior to mutagenesis with R306 shown as orange sticks. **(e)** R306A (25) substitution.

SDM reactions were performed upon WT NanH-pET21a using the same method used in section 5.2.1 and the plasmid DNA for each NanH variant was transformed into *E. coli*. The plasmids were extracted from the *E. coli* and DNA sequencing provided by GATC Services (Eurofins Genomics) confirmed all variants in Table 5.6 contained the expected mutation. All of the NanH variants (variant 23, 24 and 25) expressed in *E. coli* BL21 Origami cells and purified under the same conditions as in section 5.2.1 (Figure 5.15). The amount of NanH variant obtained from a 100 mL culture was 150 µL at a concentration between 0.8 and 1.5 mg mL^{-1}.

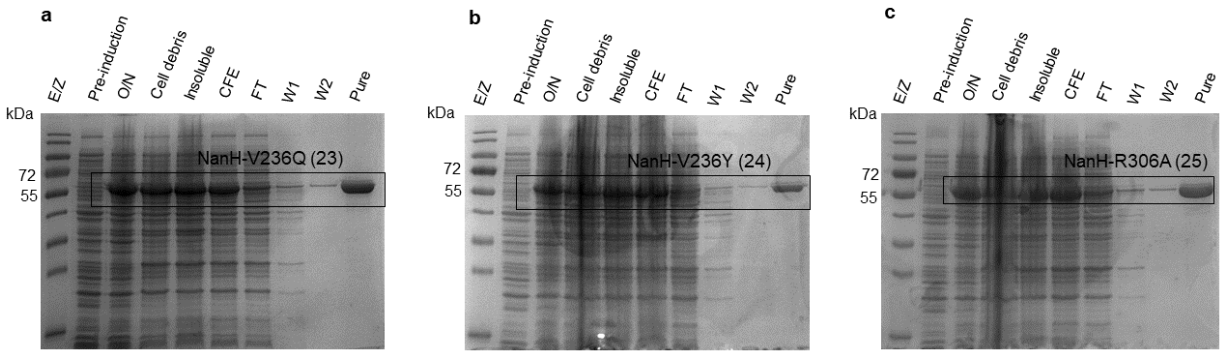


Figure 5.15: SDS-PAGE for the small-scale overexpression trial and purification of NanH variants 23, 24 and 25. (a) R306 (25), **(b)** V236Q (23), **(c)** V236Y (24)). *E. coli* cells containing NanH variants were induced after reaching OD600 0.6**.** Cell samples were collected prior to induction (**pre-induction**) and 18 h after induction (**O/N**). The *E. coli* were lysed, and the **cell debris** and **insoluble protein** were collected after two rounds of centrifugation. The cell free extract **(CFE)** was loaded to nickel resin for purification. The flow through **(FT)** and wash fractions **(W1, W2)** were collected before eluting the **pure protein** (section 2.8).

After purification and buffer exchange into 50 mM sodium phosphate, 200 mM NaCl pH 7.4, the NanH variants were assessed for sialidase activity with MU-NANA as in section 3.2.4. All three NanH variants displayed sialidase activity with MU-NANA (Figure 5.16, Table 5.7). The V_{max} values were 2.4, 4.8 and 4.8 µmol min^{-1} mg^{-1} for variants 23, 24 and 25 respectively whilst the V_{max} of WT NanH was determined to be 13.8 µmol min^{-1} mg^{-1}. The K_M values were determined to be 2.5, 4.4 and 4.0 µM for variant 23, 24 and 25 respectively whilst the K_M of WT NanH was determined to be to 6.2 µM. The k_{cat}/K_M values were calculated to be 56.0, 63.3 and 70.0 min^{-1} μM^{-1} for variants 23, 24 and 25 respectively. This showed that the 3 variants retained approximately 50 % of the efficiency of WT NanH, which had a k_{cat}/K_M of 129.0 min^{-1} μM^{-1}.

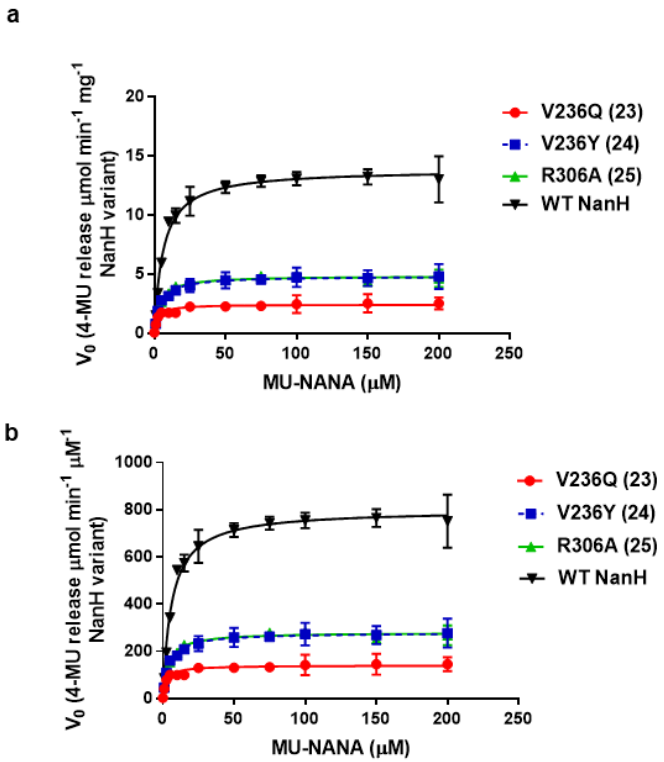


Figure 5.16: The reaction kinetics of NanH variants 23, 24 and 25 with MU-NANA. (a) Increasing concentrations of MU-NANA were exposed to 2.5 nM NanH/variant. The reaction was stopped at 1-, 2- and 3-min intervals and the rate of reaction (4-MU release) was determined by applying the raw data to a standard curve of known 4-MU concentrations. The rate (per mg NanH variant) was plotted against [MU-NANA] in a Michaelis-Menten curve from which the V_{max} and K_M could be calculated (Table 5.7). **(b)** The 4-MU release (per µM NanH variant) was plotted against [MU-NANA] enabling the k_{cat} to be calculated and in turn the catalytic efficiency (Table 5.7). In both cases the data represents the mean of three repeats, with the error bars representing the standard deviation.

	WT NanH	V236Q (23)	V236Y (24)	R306A (25)
V_{max} (µmol min^{-1} mg^{-1})	13.8 ± 0.2 (13.4 to 14.2)	2.4 ± 0.1 (2.2 to 2.6)	4.8 ± 0.1 (4.6 to 5.1)	4.8 ± 0.1 (4.7 to 5.0)
K_M ([MU-NANA], µM)	6.2 ± 0.5 (5.2 to 7.3)	2.5 ± 0.6 (1.5 to 4.1)	4.4 ± 0.6 (3.2 to 5.9)	4.0 ± 0.4 (3.3 to 4.7)
k_{cat} (min^{-1})	799.6 ± 13.7 (772.7 to 827.3)	140.0 ± 5.4 (129.1 to 151.7)	278.6 ± 7.7 (263.2 to 294.8)	279.9 ± 4.8 (270.3 to 289.7)
k_{cat}/K_M (min^{-1} µM^{-1})	129.0	56.0	63.3	70.0

Table 5.7: Reaction kinetics values obtained from Michaelis-Menten curves of NanH variants 23, 24 and 25 with MU-NANA. The V_{max}, K_M, k_{cat} and k_{cat}/K_M values are presented (± standard error) for the reaction kinetics of MU-NANA with NanH variants. The values in parentheses display the 95 % confidence intervals. The WT NanH data was taken from section 3.2.4.

NanH variants 23, 24 and 25 were assessed for altered specificity with 3- and 6-SL in the TBA assay, as in section 3.2.6. All three variants appeared to hydrolyse both α2,3- and α2,6-linked sialic acid (Figure 5.17, Table 5.8). The V_{max}, K_M and k_{cat} values for these variants with 3- and 6-SL were similar to those of WT NanH (Table 5.8). Variant 23 had the largest decrease in V_{max} to 33.1 and 30.4 µmol min^{-1} mg^{-1} with 3-SL and 6-SL respectively from 46.4 and 43.4 µmol min^{-1} mg^{-1} for 3-SL and 6-SL respectively in WT NanH. The k_{cat}/K_M values were analysed to determine differences in specificity with 3- and 6-SL compared to WT NanH. The k_{cat}/K_M values for WT NanH were 8654.8 and 3920.3 min^{-1} mM^{-1} with 3- and 6-SL respectively. Variant 23 had a k_{cat}/K_M of 6366.7 and 4397.5 min^{-1} mM^{-1} with 3- and 6-SL respectively indicating a decrease in efficiency with 3-SL, but also a small increase in efficiency with 6-SL compared to the WT. The k_{cat}/K_M values for variant 24 were 5656.0 and 4746.0 min^{-1} mM^{-1} with 3- and 6-SL respectively, which again shows a decrease in efficiency with 3-SL but an increase in efficiency with 6-SL compared to NanH WT. The analysis of variant 25 (R306A) was to determine how crucial the uncommon rotamer of the arginine was in the activity of NanH. This variant retained its activity and had k_{cat}/K_M values of 5476.0 and 4183.3 min^{-1} mM^{-1} with 3- and 6-SL respectively. The small increases in efficiency of these variants with 6-SL is caused by an increase in K_M with 3-SL and a decrease in the K_M with 6-SL (Table 5.8). Overall, all 3 of the variants remained dual specific, however the preference for α2,3-linked sialic acid has been reduced, whilst the efficiency towards α2,6-linked sialic acid increased slightly.

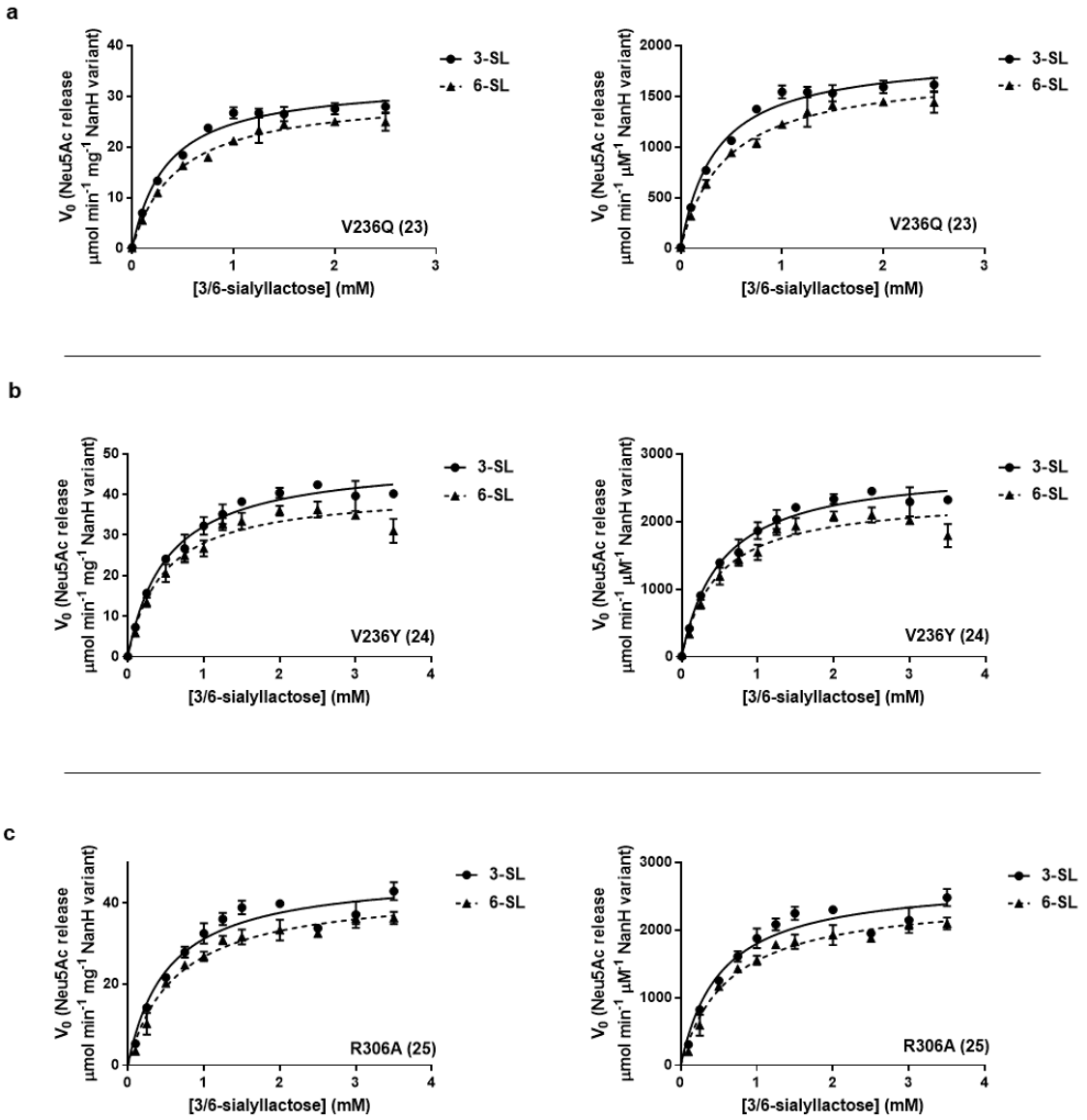


Figure 5.17: The reaction kinetics of 3-/6-SL with NanH variants 23, 23 and 25. (a) V236Q (23), **(b)** V236Y (24), **(c)** R306A (25). The curves on the left show that increasing concentrations of 3-/6-SL were exposed to 50 nM NanH variant. The reaction was halted at 1-, 3- and 5-min intervals before performing the TBA assay to quantify free sialic acid. Sialic acid was quantified by applying the data to a standard curve of known concentrations of sialic acid that had also been passed through the TBA assay. The rate (per mg NanH variant) was plotted against [3-/6-SL] in a Michaelis-Menten curve from which the V_{max} and K_M could be calculated (Table 5.8). The curves on the right show that the rate of Neu5Ac release (per µM NanH variant) was plotted against [3-/6-SL] enabling the k_{cat} to be calculated and in turn the catalytic efficiency (k_{cat}/K_M) (Table 5.8). In both cases the data represents the mean of three repeats, with the error bars representing the standard deviation.

	Variant 23 (V236Q)		Variant 24 (V236Y)		Variant 25 (R306A)		WT NanH	
	3-SL	6-SL	3-SL	6-SL	3-SL	6-SL	3-SL	6-SL
V_{max} (µmol min^{-1} mg^{-1})	33.1 ± 0.7 (31.6 to 34.6)	30.4 ± 0.8 (29.0 to 32.1)	48.8 ± 1.2 (45.5 to 51.3)	41.1 ± 1.4 (38.5 to 44.0)	47.4 ± 1.7 (44.3 to 51.0)	43.5 ± 1.3 (41.1 to 46.0)	46.4 ± 1.2 (44.1 to 49.0)	43.4 ± 1.1 (41.3 to 45.8)
K_M ([3-/6-SL], mM)	0.3 ± 0.0 (0.3 to 0.4)	0.4 ± 0.0 (0.4 to 0.5)	0.5 ± 0.0 (0.4 to 0.6)	0.5 ± 0.1 (0.4 to 0.5)	0.5 ± 0.1 (0.4 to 0.7)	0.6 ± 0.1 (0.5 to 0.8)	0.3 ± 0.0 (0.3 to 0.4)	0.6 ± 0.1 (0.5 to 0.7)
k_{cat} (min^{-1})	1910.0 ± 43.3 (1827.0 to 2000.0)	1759.0 ± 44.4 (1673.0 to 1854.0)	2828.0 ± 69.6 (2686.0 to 2964.0)	2373.0 ± 80.3 (2226.0 to 2539.0)	2738.0 ± 99.8 (2557.0 to 2945.0)	2510.0 ± 72.9 (2376.0 to 2660.0)	2683.0 ± 69.7 (2550.0 to 2829.0)	2509.0 ± 63.5 (2226.0 to 2539.0)
k_{cat}/K_M (min^{-1} mM^{-1})	6366.7	4397.5	5656.0	4746.0	5476.0	4183.3	8654.8	3920.3

Table 5.8: Reaction kinetics values obtained from Michaelis-Menten curves of NanH variants 23, 24 and 25 with 3-/6-SL. The V_{max}, K_M, K_{cat} and k_{cat}/K_M values are presented (± standard error) for the reaction kinetics of 3- and 6-SL with NanH variants. The values in parentheses display the 95 % confidence intervals. The WT NanH data was taken from section 3.2.6.

5.3 Discussion

5.3.1 Improving the α2,3 specificity of NanH

WT NanH was determined to be a dual specific sialidase in section 3.2.6, as it was able to hydrolyse both α2,3- and α2,6-linked sialic acid. A histidine and tyrosine were either inserted or substituted by SDM into NanH at amino acid positions S255 and G486 respectively (Figure 5.3b). The intention of these mutations was to form an aromatic stack, near to where the carboxyl group of sialic acid would sit in the active site, which is supposedly responsible for strong α2,3 preference in the sialidase from *S. typhimurium* (Luo *et al*., 1998). All of the 5 NanH variants produced which incorporated these aromatic residues into the protein (variants 1-5, section 5.2.1) were inactive against the synthetic sialidase substrate MU-NANA and 3- and 6-SL. Perhaps these variants were inactivated because the arginine triad member, R487, follows G486 in the protein sequence. The addition of these large aromatic residues could have caused a change in this region of the protein, such as a shift in the position of the arginine, which could have reduced binding of the sialic acid by the arginine triad. However, this would need to be confirmed by solving the crystal structures of the variants.

5.3.2 The introduction of aromatic residues into differing blades of the NanH β-propeller

Since the aromatic stack introduced into NanH close to R487 in section 5.2.1 resulted in the complete loss of enzyme activity, the decision was made to introduce aromatic residues at different locations in the β-propeller compared to those in section 5.2.1 (Table 5.2, Figure 5.6). It was hoped that these variants would retain enzyme activity and also alter the specificity by imposing some kind of steric hindrance on the substrate orientation. Variants 6, 7 and 9 (Table 5.3, Table 5.4) retained a proportion of sialidase activity after being assayed with MU-NANA and both 3- and 6-SL.

The V_{max}, k_{cat} and k_{cat}/K_M of variant 6 with MU-NANA had all decreased by approximately 50 % when compared to WT NanH (Table 5.3). The K_M of variant 6 increased, but only by 16 %, which suggests that the variant can efficiently bind to MU-NANA. The K_M values for variant 6 with 3- and 6-SL were not too dissimilar to those of WT NanH with 3- and 6-SL (Table 5.4). The k_{cat}/K_M of variant 6 with both 3- and 6-SL had reduced by approximately 50 % when compared to that of the WT (Table 5.4). Overall variant 6 remained dual specific as it was able to hydrolyse sialic acid from both 3- and 6-SL. Variant 6 has a k_{cat}/K_M for α2,3-linked sialic acid which is 2.6-fold that of α2,6-linked sialic acid (Table 5.4). This is comparable to WT NanH which has a k_{cat}/K_M for α2,3-linked sialic which is 2.2-fold that of α2,6-linked sialic acid (Table 5.4). The decrease in activity of variant 6 against MU-NANA and 3- and 6-SL is not disadvantageous, as an increase in α2,3 specificity was seen compared to the WT. Perhaps variant 6 could be used as a starting point for further research into engineering α2,3 specificity in NanH. There

were also signs of product inhibition as the concentration of both 3- and 6-SL increased (Figure 5.10a). However, it is unlikely that it would be necessary to use substrate concentrations as high as 6 mM in the future, nor that this concentration is reached *in vivo*, as it looks like the V_{max} is reached by approximately 4 mM.

The V_{max} of variant 7 with MU-NANA had decreased by 25 % compared to the WT. The K_M on the other hand increased by a factor of 2.3 suggesting that variant 7 is likely to have a weaker binding affinity for the substrate. This raised K_M led to the k_{cat}/K_M of variant 7 being 28 % of that of the WT and therefore less efficient (Table 5.3). Variant 7 also had reduced efficiencies with both 3- and 6-SL with the k_{cat}/K_M values decreasing by approximately 75 and 65 % respectively when compared to WT NanH (Table 5.4). Variant 7 is a dual specific sialidase, as the values in Table 5.4 show it cleaves both α2,3- and α2,6-linked sialic acid. It was noticed that the V_{max} for variant 7 with 6-SL was twice the V_{max} for variant 7 with 3-SL. However, when the k_{cat}/K_M was calculated, variant 7 was found to have a preference for α2,3-linked sialic acid that is 1.7-fold that of α2,6-linked sialic acid (Table 5.4). Not only does variant 7 have a lower catalytic efficiency, it also has decreased specificity for α2,3-linked sialic when compared to the WT (Table 5.4). The insertion of the tyrosine after N234 in variant 7 was beneficial in reducing some of the α2,3-linkage preference in NanH and perhaps once more is known about the structure of NanH in complex with substrates, this region of the protein could be probed more in attempt to improve the α2,6 specificity of NanH. One could also say that describing variant 7 as dual specific is not quite right. Variant 7 has a lower K_M with 3-SL, which means it is more efficient at using 3-SL at low concentrations (Table 5.4). However, the V_{max} measures the rate of product formation when the substrate is in excess. In this case it seems variant 7 can catalyse 6-SL twice as fast as 3-SL (when there is enough substrate to overcome the less efficient binding) (Table 5.4). Unfortunately, with the current glycan concentrations used in sialidase digests at Ludger (approximately 0.2-0.5 µg per digest), it is unlikely that variant 7 will catalyse α2,6-linked sialic acid faster than α2,3-linked sialic acid as the substrate is not in excess.

Variant 9 was the least active variant when compared to WT NanH, variant 6 and variant 7, with both MU-NANA and the 3- and 6-SL substrates (Table 5.3 and Table 5.4). The mutations made in variant 9 reduced the preference for α2,3-linked sialic acid. WT NanH had a 2.2-fold larger k_{cat}/K_M with α2,3-linked sialic acid than α2,6-linked sialic acid, which was only 1.6-fold in variant 9. However, when comparing the k_{cat}/K_M of variant 9 with 3- and 6-SL with those of WT NanH, the efficiency has decreased by approximately 95 and 90 % for 3- and 6-SL respectively (Table 5.4). This lack of efficiency and retention of α2,3 specificity compared to WT NanH suggest that it would not be worth investigating variant 9 further in the future.

Variants 8, 11, 12 and 13 involved mutating an arginine triad member (R212), the conserved catalytic aspartate (D237) or residues close to D237 to aromatic residues (Table 5.2, Figure 5.6). These variants were all completely inactive with MU-NANA and 3- and 6-SL (Figure 5.8). This suggests that R212Y (variant 8) could have abolished binding of sialic acic and in turn catalysis. A leucine at position 238 (precedes D237) was mutated to a tryptophan (variant 11, 12) and one can only presume that this variant lost all sialidase activity due to the size of the aromatic residue possibly shifting D237 away from the active site. Variant 13 (D237Y) could have been inactivated as the conserved aspartate is thought to act as a general base catalyst in sialic acid release (Amaya *et al*., 2004).

5.3.3 The mutagenesis of active site residues in NanH

As stated in section 5.2.3, variants 14-17 were produced to probe the specificity of NanH and variants 18-22 were to investigate the importance of the conserved tyrosine (Y518) in the catalytic mechanism (Table 5.5, Figure 5.11). Zhu and co-workers solved the crystal structure of an inactivated NA in complex with substrates containing α2,3- and α2,6-linked sialic acid (Zhu *et al*., 2012), which showed that the glycosidic linkage in both substrates was positioned near D51G (what would have been the conserved catalytic aspartate). At this point in the project a structure for NanH or an inactive variant had not been solved in complex with a substrate, so it was predicted that α2,3- and α2,6-linked sialic acid would sit in the active site of NanH in a similar way to that seen by Zhu and co-workers. If this was the case in NanH it was thought that a small change in the region of the conserved aspartate (D237) could alter the specificity. Section 5.3.2 has already discussed that it could be disadvantageous to the activity of NanH if the catalytic aspartate (D237) or nearby residues were mutated since variants 11, 12 and 13 (Table 5.2) were inactivated. However, these variants all introduced bulky aromatic residues into this region. Therefore, D237 was substituted for a glutamate, asparagine, glutamine and a serine (variants 14-17) in section 5.3.3 since these amino acids (especially glutamate) have similar properties or sizes to aspartate. Variants 14-16 were found to be inactive with MU-NANA and 3- and 6-SL and variant 17 failed to express (Figure 5.13). This suggests that it is extremely unlikely that specificity could be altered by mutating in the region of D237, as it results in a completely inactive variant.

The mechanism of bacterial sialidases has become better understood in recent years. As stated in section 4.1.3, it is believed that the conserved tyrosine (Y518 in NanH) is the catalytic nucleophile. However, Watson and co-workers showed that this may not be the case for the sialidase from *M. viridifaciens*, as the point mutants they produced that substituted the conserved tyrosine (for either an alanine, glycine or aspartate) remained active with MU-NANA (Watson *et al*., 2003). When Y518 in NanH was substituted for a variety of amino acids with differing properties (alanine, glutamine, serine,

threonine and phenylalanine) all of the variants were inactive with the MU-NANA and 3- and 6-SL. This is a clear indication that in *T. forsythia* NanH, the conserved tyrosine is crucial in some component of the binding or catalysis.

5.3.4 The mutagenesis of residues surrounding the active site

The NanH variants produced so far that have involved residues known to be involved in catalysis or substrate binding (D237, Y518, R212 and R487) have been inactive against MU-NANA and 3- and 6-SL. Once the structure of NanH-D237A was solved in complex with both 3- and 6-SL it was confirmed that the glycosidic linkages of both substrates sit in an almost identical position, close to D237 (Figure 4.20, section 4.2.5). Residue V236 was highlighted as an amino acid that is also close to the glycosidic linkages in both substrates but not considered to be crucial in catalysis or binding of the substrate. Therefore, V236 was substituted for a glutamine (variant 23) and tyrosine (variant 24) in attempt to alter the specificity (Figure 5.14). Both variant 23 and 24 showed activity with MU-NANA however the catalytic efficiency decreased in both variants as the k_{cat}/K_M values reduced by approximately 50 % when compared to WT NanH (Table 5.7). This decrease is primarily due to a large decrease in V_{max} in both variants. Interestingly, there seems to be a slight decrease in K_M, suggesting a more efficient binding of MU-NANA by the variants compared to the WT (Table 5.7). Variants 23 and 24 were the most successful variants produced with regards to their activities with both 3- and 6-SL, which were similar to WT NanH when comparing the V_{max} and k_{cat} values and therefore continued to display dual specificity (Table 5.8). The K_M values for both variant 23 and 24 with 6-SL decreased slightly by 33 and 17 % respectively compared to the WT. There was a decrease in the k_{cat}/K_M for the variants with 3-SL but an increase in the k_{cat}/K_M with 6-SL in comparison to WT NanH, which shows that overall, the preference of NanH for α2,3-linked sialic acid had decreased. Perhaps residue V236 is more important for the binding of α2,3-linked sialic acid than α2,6-linked sialic acid. Variant 7 (section 5.2.2 and section 5.3.2) along with variant 23 and 24 are the most interesting variants produced so far, as small changes in specificity were seen. Although the V_{max} was only highest for variant 7 with 6-SL when the enzyme was saturated (Table 5.4), perhaps this could be improved if this variant was merged with either variant 23 and 24. All three of these variants contain mutations located near to D237 and therefore where the glycosidic linkages sit, which suggests this region should be probed further to alter the specificity of NanH.

Variant 25 involved the substitution of R306 to an alanine. R306 is a residue found on the outskirts of the active site close to D237 and is interesting because the sidechain takes an uncommon rotamer. The R306A variant was produced to determine whether this uncommon rotamer had any involvement in the enzyme activity in NanH since it is only conserved in BTSA of *B. thetaiotaomicron* and the

sialidase of *P. distasonis*. It is unlikely that R306 has any involvement in the sialidase mechanism as the R306A mutant remained highly active, with similar levels of activity as WT NanH (Table 5.8). The K_M for R306A with 3-SL increased slightly, resulting in a decrease in the k_{cat}/K_M with 3-SL compared to the WT, however this variant has remained dual specific for α2,3- and α2,6-linked sialic. R306 is also near to D237 and V236. Perhaps this variant could also be merged with variant 7, 23 or 24. Ultimately, the removal of the large arginine side chain will provide more space in this region of interest, which could perhaps be filled with bulkier residues, which could alter the specificity.

5.3.5 Alternative approaches to engineering an enzyme

Due to the availability of the NanH crystal structure in the early stages of this product, 'rational design' was the method selected to generate the NanH variants discussed in this thesis. Rational design involves using knowledge of the structure and function of an enzyme in attempt to make the desired changes (Steiner and Schwab, 2012). This method benefits from the wide applicability of SDM protocols, extensive development of which means this method is often simple and inexpensive. In addition, it is likely that there will be an increased probability of beneficial mutations and that the library size will be smaller meaning less time and effort can be spent screening the enzyme variants (Steiner and Schwab, 2012). However, the disadvantages of rational design are that for many proteins detailed structural knowledge is a necessity yet may be unavailable and that it is difficult to predict the effect of specific mutations (Steiner and Schwab, 2012).

When structural information in unavailable an alternative engineering approach can be used which is referred to as 'directed evolution'. Directed evolution can be achieved by introducing random changes into a protein sequence using techniques such as error-prone PCR and it advantageous in the fact that not structural information is needed and that mutations at positions distant from the active site can be introduced (Steiner and Schwab, 2012). On the other hand, directed evolution is time and labour consuming and requires fast high-throughput screening methods as the changes are usually small and several rounds of evolution are required, resulting in a large library of variants (Steiner and Schwab, 2012). Additionally, if there is structural knowledge of the protein of interest, 'semi-rational design' could be employed to introduce random changes in specific regions of the protein, which is beneficial in reducing the library size (Lutz, 2010; Steiner and Schwab, 2012). Perhaps a semi-rational design approach could be used in the future engineering of NanH to introduce random mutations into the regions of the β-propeller that have been suggested to be involved in specificity.

Furthermore, perhaps future studies on sialidase specificity could involve a metagenomics approach in attempt produce an α2,6 specific sialidase. For example, a phylogenetic analysis could be performed on all of the known GH33 sialidase sequences, where the aim would be to search for sequences that

are distantly related from sialidases which have dual or α2,3-linked sialic acid specificities. The selected genes could then be produced recombinantly and screened for novel specificities in the hope of finding one with an α2,6-linked sialic acid preference.

5.4 Conclusion

This chapter has provided a stepping stone into engineering the specificity of NanH. Although the variants that retained activity continued to display dual specificity with 3- and 6-SL, it is thought that a region in NanH that could influence the specificity has been located. Variants 7, 23, 24 and 25 all showed a decrease in preference for α2,3-linked sialic acid compared to α2,6. The residues mutated in the four variants are all located near to the conserved aspartate (D237) and therefore where the glycosidic linkages of both 3- and 6-SL sit in the active site of NanH (Figure 4.20, section 4.2.5). Future work should involve the production of variants to investigate the region around D237 more heavily, whilst avoiding any residues that are conserved amongst sialidases and thought to have roles in binding and catalysis. Indeed, mutations of D237, R212 and Y518 resulted in complete loss of activity.

Chapter 6

Summary and concluding remarks

The ultimate aim of this thesis was to produce a sialidase, or sialidases, with specific activity against either α2,3- or α2,6-linked sialic acid using biochemical and structural techniques, with a view to their potential use in glycan-analysis pipelines. The previous results chapters have described the experiments performed with this aim in mind and the purpose of this chapter is to briefly summarise the key findings, discuss the impact of the results and highlight ideas for future experiments.

6.1 Results chapter summaries

6.1.1 Chapter 3 summary

'Biochemical analysis of NanH from *Tannerella forsythia*'

- SEC and SEC-MALLS analysis suggested that NanH is highly likely to be a dimer in solution.
- NanH has a higher k_{cat}/K_M with MU-NANA than some of the other bacterial sialidases reported in the literature.
- NanH was determined to be a dual specific sialidase as it is was found to hydrolyse both α2,3- and α2,6-linked sialic acid when assayed with 3- and 6-SL (disaccharides), although it had a higher k_{cat}/K_M for the α2,3-linkage.
- Reaction kinetics were also determined for NanH with SLe^A and SLe^X, both of which have α2,3-linked sialic acid, and it was found that there was little difference between the kinetics values with both substrates. However, the k_{cat}/K_M was four-fold less with SLe^A and SLe^X than with 3-SL, showing the importance of the underlying sugar in sialidase activity.
- There was no cooperativity arising from the dimeric structure of NanH as evidenced by Figure 3.7.
- The neuraminidase inhibitors DANA and oseltamivir are both capable of inhibiting NanH, with DANA having a lower IC_{50} than both oseltamivir and zanamivir.
- It was found that NanH is capable of hydrolysing sialic acid from bi- and tri-antennary plasma-N-glycans.
- NanH remained stable after storage at 37 °C and RT for 48 h and retained 50 % of its original k_{cat}/K_M after storage in glycerol for up to 9 months.

6.1.2 Chapter 4 summary

'Solving the crystal structure of NanH and NanH variants'

- A NanH inactive variant (NanH-D237A) was produced by SDM and retained the ability to bind both 3- and 6-SL.
- The apo forms of both WT NanH and NanH-D237A crystallised under the C4 condition of the JCSG screen (Appendix).
- The crystal structure of WT NanH showed that the enzyme folds into a 6-bladed β-propeller and has an N-terminal CBM. The first structure of WT NanH to be solved was found to have a molecule of HEPES buffer from the crystallisation condition in complex with the active site.
- The apo form of WT NanH was solved by reducing the concentration of HEPES by 10-fold in the crystallisation condition.
- The structure of the NanH CBM was found to differ to those of the CBM40 family, which suggested that it the founding member of a novel CBM family, in agreement with the biochemical and bioinformatics findings by Frey and colleagues (Frey *et al*., 2018).
- Crystal structures were also solved for WT NanH in complex with the inhibitor oseltamivir and also for NanH-D237A in complex with both 3- and 6-SL. All 3 of these complexes showed the inhibitor/substrates to be orientated in the active site by their carboxyl groups, which interact with the arginine triad.
- Both 3- and 6-SL were found to sit in almost identical positions in the active site with their glycosidic linkages being near V236 and D237.
- The complete set of WT NanH and NanH-D237A structures all had two copies of the protein in the asymmetric unit. Analysis of the interfaces between the two molecules in the asymmetric unit suggested that NanH is likely to be a stable dimer in solution, in agreement with the SEC data in Chapter 3.

6.1.3 Chapter 5 summary

'Engineering the specificity of NanH'

- The structural feature thought to be responsible for α2,3 specificity in known α2,3 specific sialidases is the stacking of aromatic residues above where the 2-carboxylate of sialic acid would sit in the active site. These residues were found to be absent in NanH and therefore point mutants were produced to incorporate aromatic residues into the protein chain in the same region. However, none of these variants showed activity with MU-NANA, 3-SL or 6-SL.

- Another batch of NanH variants produced were based upon the idea that perhaps creating an aromatic stack around other blades in the β-propeller would perhaps retain enzyme activity and potentially alter the specificity. Two of the eight variants produced still had reasonable activity with MU-NANA, 3-SL and 6-SL.
    - Variant 6 remained dual specific and although the k_{cat}/K_M decreased by 50 % with both 3- and 6-SL when compared to WT NanH, an increase in specificity for α2,3-linked sialic acid was seen compared to the WT.
    - Variant 7 remained dual specific and had a decrease in k_{cat}/K_M by approximately 70 % with both 3- and 6-SL compared to WT NanH. It was also noticed that there was a decrease in specificity for α2,3-linked sialic acid compared to the WT. This variant can also catalyse 6-SL twice as fast as 3-SL when the substrate is in excess.
- NanH residue D237 was targeted by SDM, as it was predicted that this residue would be close to where the glycosidic linkage would sit in both α2,3- and α2,6-linked sialic acid, and therefore small alterations in this region could alter the specificity. This idea was based upon what was seen with α2,3- and α2,6-linked sialic acid in a crystal structure of an influenza A NA. Since D237 was expected to have a role in catalysis, mutations which maintained similar charges were made, however all variants were inactive against MU-NANA, 3-SL and 6-SL, which highlighted the importance of D237 in catalysis.
- The conserved tyrosine has been reported to have a role as a nucleophile in catalysis, but discrepancies in the literature suggested that this might not be the case, hence this residue was targeted for mutagenesis with Y518 being the target. All the Y518 variants were inactive with MU-NANA, 3-SL and 6-SL, which suggested that Y518 is likely to be the catalytic nucleophile.
- Once the NanH-D237A crystal structure was solved in complex with 3- and 6-SL it was discovered that V236 was in proximity to the glycosidic linkages in both substrates, but this residue is unlikely to be required in the catalytic mechanism. The point mutations (variants 23 and 24) that targeted V236 were the most successful for retaining activity with 3- and 6-SL and even though they remained dual specific, there was a decrease in the k_{cat}/K_M with 3-SL but an increase with 6-SL in comparison to WT NanH. In both variant 23 and 24, the overall preference for α2,3-linked sialic acid decreased.

6.2 Concluding remarks

6.2.1 Future plans for altering the specificity of NanH

This project has seen the production of 25 NanH variants, all of which have been assessed for alterations in specificity. Interesting alterations have been seen in a small handful of variants (section 6.3), especially in the reduction of α2,3-linkage specificity, however, overall all the variants that retained activity remained dual specific. Fortunately, crystal structures of the NanH inactive variant in complex with 3- and 6-SL were solved, which highlighted a region by V236 and D237 which was near to where the glycosidic linkage sits and could therefore influence the specificity. Variants 7 (N234 +Y S235), 23 (V236Q) and 24 (V236Y) all targeted residues near to where the glycosidic linkages sit, so perhaps further mutation upon these variants could be made, with other residues in this region being mutated to encourage larger alterations in specificity. Perhaps even producing V236 point mutants but using different residues to glutamine and tyrosine could introduce larger changes in specificity. It is difficult to predict which residues could be targeted, but the mutants produced in section 5.2.3 showed that it is not beneficial to mutate residues involved in the catalytic mechanism.

Currently, there are no known α2,6 strict specificity sialidases that have been published in the literature, but it has been reported that there has been an increase in α2,6-linked sialic acid specificity for the human influenza N2 NA over evolutionary time (section 1.3.5). A study by Kobasha and colleagues on influenza N2 NA has suggested that the increase in α2,6 specificity of the human neuraminidase occurred due to a mutation in the residue at position 275 (from isoleucine in avian and early human NA to valine in later human NA) (Kobasa *et al*., 1999). Residue 275 is two residues away from the conserved glutamate (E277) that forms a nucleophilic pair with the conserved tyrosine in the active site (Figure 6.1a). It has been suggested that V275 allows the side chain of E276 to make strong contacts with the glycerol group of sialic acid (Kobasa *et al*., 1999). This leaves the pyranose ring of sialic acid horizontal in the active site allowing the remaining moiety of the α2,3 or α2,6 oligosaccharide to extend freely to the open region of the site. A bulkier residue such as an isoleucine in this position would cause E276 to be lifted slightly, causing the pyranose ring to tilt to the left of the active site. It is then suggested that this tilt is likely to prevent access of α2,6-linked sialic acid because of a steric hindrance that could occur between the α2,6-gal portion of the oligosaccharide and the side chain of the conserved catalytic aspartate (D151) (Kobasa *et al*., 1999). This suggestion by Kobasha and colleagues is interesting, as residue 275 is at the opposite side of the active site to the conserved catalytic aspartate (D151, however D151G in Figure 6.1a), and therefore the region that this study believed to be solely responsible for the specificity of sialidases, as this is where the glycosidic linkage sits.

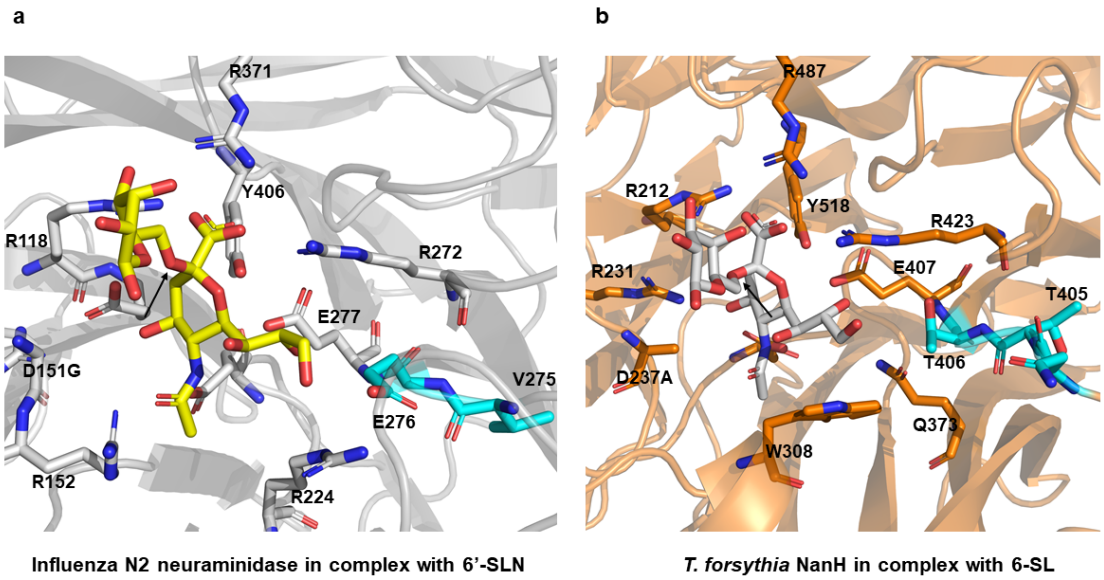


Figure 6.1: The amino acid at position 275 in N2 NA is suggested to influence specificity. (a) The active site of N2 NA (grey) (PDB: 4GZX) in complex with 6'-SLN (yellow sticks). V275 is highlighted as cyan sticks and is at the opposite side of the active site to D151G. **(b)** The active site of *T. forsythia* NanH (orange) in complex with 6-SL (grey sticks). The amino acid that is thought to correspond to V275 in N2 has been highlighted in cyan (T405) and is at the opposite side of the active site to D237A. In both figures, the glycosidic linkage in each substrate is indicated by a black arrow.

When the residues following the conserved glutamate (E407) in NanH were identified, it was found that two threonine's precede E407, with T406 able to make contact with the glycerol group of sialic acid (Figure 6.1b). Perhaps T405 could be substituted to a valine in NanH to determine whether this variant could also introduce an increase in α2,6 specificity and α2,3 specificity could also be probed further if T405 was substituted for a bulkier residue. The work carried out by Kobasha and colleagues has provided ideas for future NanH variants that could be produced, as combinations of V236 and T405 variants could also be studied (Kobasa *et al*., 1999). Both of these residues surround the active site but are not involved in the mechanism. Once analysed in in-house assays, any NanH variants that retain activity would be analysed with complex glycans, such as plasma-N-glycans and EPO glycans, as in section 3.2.7.1 to determine their action with bi-, tri- and tetra- antennary glycans.

This body of work suggests that altering the binding specificity of a sialidase will require one or a combination of mutations in and around the active site. Whether this is achievable with NanH is questionable since it has evolved to be dual specific, but there are certainly several more residues that could be probed in the future. Evolutionarily, it may be that specific sialidases are not necessary. It may be the case that bacteria can utilise either linkage without a preference. With regards to the

structure, the fact that α2,3- and α2,6-linked sialic acid sit in the active site in the same way and share the same mechanism may also be a barrier to achieving specificity and may be why the small alterations in specificity in this work also caused a lower activity.

6.2.2 Insights from the crystal structure of NanH

Solving the crystal structure of WT NanH and NanH variants was the key breakthrough in this project, as not only has it enabled the rational engineering of NanH for use in glycan diagnostics, it also has a wider significance in the sialidase field. Firstly, this project supports that sialidases employ a mechanism that uses a conserved Tyr/Glu (Y518 and E407) pair as a catalytic nucleophile, whilst the conserved aspartate (D237) is likely to act as an acid catalyst in the initial glycosyl-enzyme formation and a general base catalyst during the nucleophilic attack (Amaya *et al*., 2004). This mechanism is believable for NanH since the inhibitor and substrate bound models show that sialic acid is orientated by its carboxyl group in the active site by the arginine triad, which brings the glycosidic linkages close to Y518 and D237. As well as this, point mutants where D237 and Y518 were targeted resulted in the inactivation of the NanH, which highlighted the importance of both of these residues in the mechanism of catalysis.

The crystal models could also be used to design more efficient inhibitors for bacterial sialidases, by computational based design, with the intention of inhibiting NanH in the nanomolar range and therefore be used as a potential therapy for periodontal disease, or other diseases where sialidases assist virulence. Perhaps the CBM domain could be crystallised in isolation from NanH in complex with substrates, such as $SLe^{X/A}$, which were discovered to bind to the CBM by Frey and colleagues (Frey *et al*., 2018). These CBM complexes would locate the binding pocket, which could also lead to the design of inhibitors that bind this domain directly. As of now, the CBM has only been found to have a role in binding a wide range of oligosaccharides (Frey *et al*., 2018), so perhaps it would be useful to determine whether it also has a role in catalysis by analysing catalysis of the catalytic domain in the absence of the CBM.

Since NanH has been found to be predominantly dimeric in solution and the dimeric structure does not appear to provide any biochemical benefits, such as cooperativity, it was suggested that a dimer could increase the stability of the protein *in-situ*. To probe the stability of NanH, the crystal model could be used to select residues at the interface which could be altered by SDM in attempt to abolish the interactions involved in dimerization to produce a monomer. SEC would be used to confirm the multimeric form of the variant and if it is monomeric, the kinetic parameters with MU-NANA would be assessed and compared to the WT dimeric form of NanH. As well as this, stability studies, such as those described in section 3.2.8 and 3.2.9 could be performed where the variant would be stored at

RT, 37 ^{o}C and -20 ^{o}C and the kinetic parameters would be assessed at regular intervals alongside WT dimeric NanH, which would be stored in the same conditions.

6.3 Conclusion

Overall, this project has provided a stepping stone into understanding the structural features behind *T. forsythia* NanH sialidase specificity. Even though a highly specific sialidase was not produced, two regions within the active site (around D237 and also E407) have been highlighted as potentially having importance in increasing α2,6 specificity, a research area that is lacking knowledge in the bacterial sialidase literature. Whether or not NanH can be engineered to produce a true α2,6 specific sialidase is difficult to tell, however several ideas have been laid down for future mutagenesis experiments, which are expected to increase α2,6 specificity at the very minimum.

Bibliography

Aebi, M. (2013) 'N-linked protein glycosylation in the ER.' *Biochimica et Biophysica Acta* - Molecular Cell Research. Elsevier B.V., 1833(11), pp. 2430–2437.

Amaya, M. F., Watts, A. G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A., Paris, G., Frasch, A. C., Withers, S. G. and Alzari, P. M. (2004) 'Structural insights into the catalytic mechanism of Trypanosoma cruzi trans-sialidase.' *Structure*, 12(5), pp. 775–784.

Aminoff, D. (1961) 'Methods for the quantitative estimation of N-acetylneuraminic acid and their application to hydrolysates of sialomucoids.' *The Biochemical journal*, 81(2), pp. 384–392.

Amith, S. R., Jayanth, P., Franchuk, S., Finlay, T., Seyrantepe, V., Beyaert, R., Pshezhetsky, A. V. and Szewczuk, M. R. (2010) 'Neu1 desialylation of sialyl α-2,3-linked β-galactosyl residues of TOLL-like receptor 4 is essential for receptor activation and cellular signaling.' *Cellular Signalling*, 22(2), pp. 314–324.

Aronson, M., Medalia, O., Schori, L., Mirelman, D., Sharon, N. and Ofek, I. (1979) 'Prevention of Colonization of the Urinary Tract of Mice with Escherichia coli by Blocking of Bacterial Adherence with Methyl α-D-Mannopyranoside.' *The Journal of Infectious Diseases,* 139(3), pp. 329–332.

Ashwell, G. and Harford, J. (1982) 'Carbohydrate-Specific Receptors of the Liver.' *Annual Review of Biochemistry*, 51(1), pp. 531–554.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) 'Electrostatics of nanosystems: Application to microtubules and the ribosome.' *Proceedings of the National Academy of Sciences*, 98(18), pp. 10037–10041.

Baum, L. G. and Paulson, J. C. (1991) 'The N2 neuraminidase of human influenza virus has acquired a substrate specificity complementary to the hemagglutinin receptor specificity.' *Virology,* 180(1), pp. 10–15.

Berg, J. M., Tymoczko, J. L. and Stryer, L. (2011) *Biochemistry.* 7th edition. Palgrave MacMillan.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) 'The Protein Data Bank.' *Nucleic acids research,* 28(1), pp. 235–242.

Bertozzi, C. R., Freeze, H. H. and Varki, A. (2009) 'Glycans in Biotechnology and the Pharmaceutical Industry.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 2nd edition, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Bertozzi, C. R. and Rabuka, D. (2009) 'Structural basis of glycan diversity.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 2nd edition, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M. and Parekh, R. B. (1995) 'Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid.' *Analytical Biochemistry,* 230(2), pp. 229–238.

Bodnar, J., Szekrenyes, A., Szigeti, M., Jarvas, G., Krenkova, J., Foret, F. and Guttman, A. (2016) 'Enzymatic removal of N-glycans by PNGase F coated magnetic microparticles.' *Electrophoresis,* 37(10), pp. 1264–1269.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) 'Carbohydrate-binding modules: fine-tuning polysaccharide recognition.' *Biochemical Journal,* 382(3), pp. 769–781.

Boraston, A. B., Ficko-Blean, E. and Healey, M. (2007) 'Carbohydrate recognition by a large sialidase toxin from Clostridium perfringens.' *Biochemistry,* 46(40), pp. 11352–11360.

Bossart-Whitaker, P., Carson, M., Babu, Y. S., Smith, C. D., Laver, W. G. and Air, G. M. (1993) 'Three-dimensional Structure of Influenza A N9 Neuraminidase and Its Complex with the Inhibitor 2-Deoxy 2,3-Dehydro-N-Acetyl Neuraminic Acid.' *Journal of Molecular Biology,* 232(4), pp. 1069–1083.

Braham, P. H. and Moncla, B. J. (1992) 'Rapid presumptive identification and further characterization of Bacteroides forsythus.' *Journal of Clinical Microbiology,* 30(3), pp. 649-654.

Brigham, C., Caughlan, R., Gallegos, R., Dallas, M. B., Godoy, V. G. and Malamy, M. H. (2009) 'Sialic acid (N-acetyl neuraminic acid) utilization by Bacteroides fragilis requires a novel N-acetyl mannosamine epimerase.' *Journal of Bacteriology,* 191(11), pp. 3629–3638.

Briliūtė, J., Urbanowicz, P. A., Luis, A. S., Baslé, A., Paterson, N., Rebello, O., Hendel, J., Ndeh, D. A., Lowe, E. C., Martens, E. C., Spencer, D. I. R., Bolam, D. N. and Crouch, L. I. (2019) 'Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci.' *Nature Microbiology.* In press.

Brockhausen, I., Schachter, H. and Stanley, P. (2009) 'O-GalNAc Glycans.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 2nd edition, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Burmeister, W. P., Henrissat, B., Bosso, C., Cusack, S. and Ruigrok, R. W. H. (1993) 'Influenza B virus neuraminidase can synthesize its own inhibitor.' *Structure,* 1(1), pp. 19–26.

Buschiazzo, A. and Alzari, P. M. (2008) 'Structural insights into sialic acid enzymology.' *Current Opinion in Chemical Biology,* 12(5), pp. 565–572.

Byers, H. L., Tarelli, E., Homer, K. A. and Beighton, D. (2000) 'Isolation and characterisation of sialidase from a strain of Streptococcus oralis.' *Journal of medical microbiology,* 49(3), pp. 235–44.

Byrd-Leotis, L., Cummings, R. D. and Steinhauer, D. A. (2017) 'The interplay between the host receptor and influenza virus hemagglutinin and neuraminidase.' *International Journal of Molecular Sciences,* 18(7), p. 1541.

Byres, E., Paton, A. W., Paton, J. C., Löfling, J. C., Smith, D. F., Wilce, M. C. J., Talbot, U. M., Chong, D. C., Yu, H., Huang, S., Chen, X., Varki, N. M., Varki, A., Rossjohn, J. and Beddoe, T. (2008) 'Incorporation of a non-human glycan mediates human susceptibility to a bacterial toxin.' *Nature,* 456(7222), pp. 648–652.

Byrne, B., Donohoe, G. G. and O'Kennedy, R. (2007) 'Sialic acids: carbohydrate moieties that influence the biological and physical properties of biopharmaceutical proteins and living cells.' *Drug discovery today,* 12(7-8), pp. 319-326.

Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A. and Haslma, S. M. (2008) 'GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans.' *Journal of proteome research,* 7(4), pp. 1650–1659.

Chan, J., Watson, J. N., Lu, A., Cerda, V. C., Borgford, T. J. and Bennet, A. J. (2012) 'Bacterial and viral sialidases: Contribution of the conserved active site glutamate to catalysis.' *Biochemistry,* 51(1), pp. 433–441.

Chang, Y.-C., Uchiyama, S., Varki, A. and Nizet, V. (2012) 'Leukocyte Inflammatory Responses Provoked by Pneumococcal Sialidase.' *MBio,* 3(1) pp. e00220-11.

Chavas, L. M. G., Tringali, C., Fusi, P., Venerando, B., Tettamanti, G., Kato, R., Monti, E. and Wakatsuki, S. (2005) 'Crystal structure of the human cytosolic sialidase Neu2: Evidence for the dynamic nature of substrate recognition.' *Journal of Biological Chemistry,* 280(1), pp. 469–475.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity : all-atom structure validation for macromolecular crystallography.' *Acta Crystallographica Section D: Biological Crystallography,* 66(1), pp. 12–21.

Chong, A. K. J., Pegg, M. S., Taylor, N. R. and von Itzstein, M. (1992) 'Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus.' *European Journal of Biochemistry,* 207(1), pp. 335–343.

Chou, M. Y., Li, S. C. and Li, Y. T. (1996) 'Cloning and expression of sialidase L, a NeuAcα2→3Gal-specific sialidase from the leech, Macrobdella decora.' *Journal of Biological Chemistry,* 271(32), pp. 19219–19224.

Citartan, M., Gopinath, S. C. B., Chen, Y., Lakshmipriya, T. and Tang, T. H. (2015) 'Monitoring recombinant human erythropoietin abuse among athletes.' *Biosensors and Bioelectronics,* 63, pp. 86–98.

Clark, J., Harrison, J. C., Mdegela, R. H. and March, J. B. (2000) 'Extended stability of restriction enzymes at ambient temperatures.' *BioTechniques,* 29(3), pp. 536–542.

Cohen, M. and Varki, A. (2010) 'The Sialome—Far More Than the Sum of Its Parts.' *OMICS: A Journal of Integrative Biology,* 14(4), pp. 455–464.

Colombo, C., Podlipnik, Č., Presti, L. Lo, Niikura, M., Bennet, A. J. and Bernardi, A. (2018) 'Design and synthesis of constrained bicyclic molecules as candidate inhibitors of influenza A neuraminidase.' *Plos One*, 13(2), pp.e0193623.

Colley, K. J., Varki, A. and Kinoshita, T. (2017) 'Cellular Organization of Glycosylation.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 3rd edition, Cold Spring Harbour (NY): Cold Spring Harbor Laboratory Press.

Corfield, A. P. (2015) 'Mucins: A biologically relevant glycan barrier in mucosal protection.' *Biochimica et Biophysica Acta - General Subjects,* 1850(1), pp. 236–252.

Corfield, T. (1992) 'Bacterial sialidases - roles in pathogenicity and nutrition.' *Glycobiology,* 2(6), pp. 509–521.

Crennell, S., Garman, E., Laver, G., Vimr, E. and Taylor, G. (1994) 'Crystal structure of Vibrio cholerae neuraminidase reveals dual lectin-like domains in addition to the catalytic domain.' *Structure,* 2(6), pp. 535–544.

Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R. and Taylor, G. L. (1993) 'Crystal structure of a bacterial sialidase (from Salmonella typhimurium LT2) shows the same fold as an influenza virus neuraminidase.' *Proceedings of the National Academy of Sciences,* 90(21), pp. 9852–9856.

Cross, B. W. and Ruhl, S. (2018) 'Glycan recognition at the saliva – oral microbiome interface.' *Cellular Immunology,* 333, pp. 19-33.

Davies, G. J. and Sinnott, M. L. (2008) 'Sorting the diverse.' *Biochemical Journal,* 30, pp. 26-32.

Davies, G. J., Wilson, K. S. and Henrissat, B. (1997) 'Nomenclature for sugar-binding subsites in glycosyl hydrolases.' *Biochemical Journal,* 321(Pt2), pp. 557-559.

Edinger, T. O., Pohl, M. O. and Stertz, S. (2014) 'Entry of influenza A virus: Host factors and antiviral targets.' *Journal of General Virology,* 95(2), pp. 263–277.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) 'Features and development of Coot.' *Acta Crystallographica Section D: Biological Crystallography,* 66(4), pp. 486–501.

Eneva, R. T., Engibarov, S. A., Petrova, P., Abrashev, R., Strateva, T., Kolyovska, V. and Abrashev, I. (2015) 'High Production of Neuraminidase by a Vibrio cholerae Non-O1 Strain—the First Possible Alternative to Toxigenic Producers.' *Applied Biochemistry and Biotechnology,* 176(2), pp. 412–427.

Esko, J. D., Doering, T. L. and Raetz, C. R. (2009) 'Eubacteria and Archaea.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 2nd edition, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Fernandes, D. (2004) 'Reducing risk in biopharmaceutical production by controlling glycosylation.' *EBR - European Biopharmaceutical Review,* pp. 92–97.

Fersht, A. (1985) *Enzyme structure and mechanism.* 2nd edition. W. H. Freeman and Company.

Fraguas, L. F., Carlsson, J. and Lönnberg, M. (2008) 'Lectin affinity chromatography as a tool to differentiate endogenous and recombinant erythropoietins.' *Journal of Chromatography A,* 1212(1-2), pp. 82-88.

Frey, A. (2016) *Investigation of the Sialidase of Periodontal Pathogens.* Ph.D. thesis, University of Sheffield.

Frey, A. M., Satur, M. J., Phansopa, C., Parker, J. L., Bradshaw, D., Pratten, J. and Stafford, G. P. (2018) 'Evidence for a carbohydrate-binding module (CBM) of Tannerella forsythia NanH sialidase, key to interactions at the host–pathogen interface.' *Biochemical Journal,* 475(6), pp. 1159–1176.

Friedrich, V., Gruber, C., Nimeth, I., Pabinger, S., Sekot, G., Posch, G., Altmann, F., Messner, P., Andrukhov, O. and Schäffer, C. (2015) 'Outer membrane vesicles of Tannerella forsythia: Biogenesis, composition, and virulence.' *Molecular Oral Microbiology,* 30(6), pp. 451–473.

Furmanek, A. and Hofsteenge, J. (2000) 'Protein C-mannosylation: Facts and questions.' *Acta Biochemica Polonica,* 47(3), pp. 781–789.

Goh, J. B. and Ng, S. K. (2018) 'Impact of host cell line choice on glycan profile.' *Critical Reviews in Biotechnology,* 38(6), pp. 851–867.

Gratz, N., Nam Loh, L., Mann, B., Gao, G., Carter, R., Rosch, J. and Tuomanen, E. I. (2017) 'Pneumococcal neuraminidase activates TGF-β signalling.' *Microbiology,* 163(8), pp. 1198–1207.

Grilo, A. L. and Mantalaris, A. (2019) 'The Increasingly Human and Profitable Monoclonal Antibody Market.' *Trends in biotechnology,* 37(1), pp. 9–16.

Grobe, K., Sartori, B., Traving, C., Schauer, R. and Roggentin, P. (1998) 'Enzymatic and molecular properties of the Clostridium tertium sialidase.' *Journal of Biochemistry,* 124(6), pp. 1101–1110.

Gubareva, L. V., Webster, R. G. and Hayden, F. G. (2001) 'Comparison of the activities of zanamivir, oseltamivir, and RWJ-270201 against clinical isolates of influenza virus and neuraminidase inhibitor-resistant variants.' *Antimicrobial Agents and Chemotherapy,* 45(12), pp. 3403–3408.

Gut, H., King, S. J. and Walsh, M. A. (2008) 'Structural and functional studies of Streptococcus pneumoniae neuraminidase B: An intramolecular trans-sialidase.' *FEBS Letters,* 582(23–24), pp. 3348–3352.

Guthrie, E. P. and Magnelli, P. E. (2016) 'Using glycosidases to remove, trim, or modify glycans on therapeutic proteins.' *BioProcess International,* 14(2).

Han, L. and Costello, C. E. (2013) 'Mass spectrometry of glycans.' *Biochemistry (Moscow),* 78(7), pp. 710–720.

Hata, K., Koseki, K., Yamaguchi, K., Moriya, S., Suzuki, Y., Yingsakmongkon, S., Hirai, G., Sodeoka, M., Von Itzstein, M. and Miyagi, T. (2008) 'Limited inhibitory effects of oseltamivir and zanamivir on human sialidases.' *Antimicrobial Agents and Chemotherapy,* 52(10), pp. 3484–3491.

Hata, K., Wada, T., Hasegawa, A., Kiso, M. and Miyagi, T. (1998) 'Purification and characterization of a membrane-associated ganglioside sialidase from bovine brain.' *Journal of Biochemistry,* 123(5), pp. 899–905.

Hattrup, C. L. and Gendler, S.J. (2008) 'Structure and function of the cell surface (tethered) mucins. *Annual Review of Physiology,* 70, pp. 431-457.

Henrissat, B. (1991) 'A classification of glycosyl hydrolases based on amino acid sequence similarities.' *Biochemical Journal,* 280(2), pp. 309-316.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. and Davies, G. (1995) 'Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases.' *Proceedings of the National Academy of Sciences,* 92(15), pp. 7090-7094.

Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., Ng, K. F., Suen, M. W. M. and Tai, H. L. (2003) 'Electrospray ionisation mass spectrometry: principles and clinical applications.' *The Clinical Biochemist Reviews,* 24(1), p.3.

Hofsteenge, J., Müller, D. R., de Beer, T., Löffler, A., Richter, W. J. and Vliegenthart, J. F. (1994) 'New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us.' *Biochemistry,* 33(46), pp. 13524–13530.

Hokke, C. H., Bergwerff, A. A., van Dedem, G. W. K., van Oostrum, J., Kamerling, J. P. and Vliegenthart, J. F. G. (1990) 'Sialylated carbohydrate chains of recombinant human glycoproteins expressed in Chinese hamster ovary cells contain traces of N-glycolylneuraminic acid.' *FEBS Letters,* 275(1–2), pp. 9–14.

Honma, K., Mishima, E. and Sharma, A. (2011) 'Role of Tannerella forsythia NanH sialidase in epithelial cell attachment.' *Infection and Immunity,* 79(1), pp. 393–401.

Hoyer, L. L., Roggentin, P., Schauer, R. and Vimr, E. R. (1991) 'Purification and properties of cloned Salmonella typhimurium LT2 sialidase with virus-typical kinetic preference for sialyl α2→3 linkages.' *Journal of Biochemistry,* 110(3), pp. 462–467.

Hubbard, S. C. and Robins, P. W. (1980) 'Synthesis of the N-linked Oligosaccharide of Glycoproteins. Assembly of the lipid-linked precursor oligosaccharide and its relation to protein synthesis in vivo.' *Journal of Biological Chemistry,* 255(24), pp. 11782–11793.

Ilver, D., Arnqvist, A., Ögren, J., Frick, I. M., Kersulyte, D., Incecik, E. T., Berg, D. E., Covacci, A., Engstrand, L. and Borén, T. (1998) 'Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging.' *Science,* 279(5349), pp. 373–377.

Ishikura, H., Arakawa, S., Nakajima, T., Tsuchida, N. and Ishikawa, I. (2003) 'Cloning of the Tannerella forsythensis (Bacteroides forsythus) siaHI gene and purification of the sialidase enzyme.' *Journal of Medical Microbiology,* 52(12), pp. 1101–1107.

Ito, S., Yamashita, K., Spiro, R. G. and Kobata, A. (1977) 'Structure of a carbohydrate moiety of a unit A glycopeptide of calf thyroglobulin.' *Journal of Biochemistry,* 81(6), pp. 1621–1631.

von Itzstein, M. (2007) 'The war against influenza: Discovery and development of sialidase inhibitors.' *Nature Reviews Drug Discovery,* 6(12), pp. 967–974.

Jacewicz, B. Y. M., Clausen, H., Nudelman, E., Donohue-rolfe, A. and Keusch, G. T. (1986) 'Pathogenesis of shigella diarrhea. XI. Isolation of a shigella toxin-binding glycolipid from rabbit jejunum and HeLa cells and its identification as globotriaosylceramide.' *Journal of Experimental Medicine,* 163(6), pp. 1391–1404.

Jefferis, R. (2005) 'Glycosylation of recombinant antibody therapeutics.' *Biotechnology Progress,* 21(1), pp. 11–16.

Johansen, P. G., Marshall, R. D. and Neuberger, A. (1961) 'Carbohydrates in protein. 3. The preparation and some of the properties of a glycopeptide from hen's-egg albumin.' *The Biochemical Journal,* 78(3), pp. 518–27.

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

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G. W., Holst, M. J., McCammon, J. A. and Baker, N. A. (2018) 'Improvements to the APBS biomolecular solvation software suite.' *Protein Science,* 27(1), pp. 112–128.

Kam, R. K. T. and Poon, T. C. W. (2008) 'The potentials of glycomics in biomarker discovery.' *Clinical Proteomics,* 4(3), pp. 67–79.

Kantardjieff, K. A. and Rupp, B. (2003) 'Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals.' *Protein Science,* 12(9), pp. 1865–1871.

Kati, W. M., Montgomery, D., Maring, C., Stoll, V. S., Giranda, V., Chen, X., Laver, W. G., Kohlbrenner, W. and Norbeck, D. W. (2001) 'Novel alpha- and beta-amino acid inhibitors of influenza virus neuraminidase.' *Antimicrobial agents and chemotherapy,* 45(9), pp. 2563–70.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. (1958) 'A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis.' *Nature,* 181(4610), pp. 662–666.

Kessler, J., Heck, J., Tanenbaum, S. W. and Flashner, M. (1982) 'Substrate and product specificity of Arthrobacter sialophilus neuraminidase.' *Journal of Biological Chemistry,* 257(9), pp. 5056–5060.

Khan, F., Khan, R. H., Sherwani, A., Mohmood, S. and Azfer, M. A. (2002) 'Lectins as markers for blood grouping.' *Medical Science Monitor,* 8(12), pp. RA293–RA300.

Kim, J. Y., Kim, Y. G. and Lee, G. M. (2012) 'CHO cells in biotechnology for production of recombinant proteins: Current state and further potential.' *Applied Microbiology and Biotechnology,* 93(3), pp. 917–930.

Kim, S., Oh, D. B., Kang, H. A. and Kwon, O. (2011) 'Features and applications of bacterial sialidases.' *Applied Microbiology and Biotechnology,* 91(1), pp. 1–15.

Kim, S., Oh, D. B., Kwon, O. and Kang, H. A. (2010) 'Identification and functional characterization of the NanH extracellular sialidase from Corynebacterium diphtheriae.' *Journal of Biochemistry,* 147(4), pp. 523–533.

Kiyohara, M., Tanigawa, K., Chaiwangsri, T., Katayama, T., Ashida, H. and Yamamoto, K. (2011) 'An exo - Sialidase from bifidobacteria involved in the degradation of sialyloligosaccharides in human milk and intestinal glycoconjugates.' *Glycobiology,* 21(4), pp. 437–447.

Kobasa, D., Kodihalli, S., Luo, M., Castrucci, M. R., Donatelli, I., Suzuki, Y., Suzuki, T. and Kawaoka, Y. (1999) 'Amino acid residues contributing to the substrate specificity of the influenza A virus neuraminidase.' *Journal of virology,* 73(8), pp. 6743–51.

Kornfeld, R. and Kornfeld, S. (1985) 'Rosalind Kornfeld and Stuart Kornfeld.' *Annual Review of Biochemistry,* 54(1), pp. 631–664.

Kozak, R. P., Tortosa, C. B., Fernandes, D. L. and Spencer, D. I. R. (2015) 'Comparison of procainamide and 2-aminobenzamide labeling for profiling and identification of glycans liquid chromatography with fluorescence detection coupled to electrospray ionization-mass spectrometry.' *Analytical Biochemistry,* 486, pp. 38–40.

Krissinel, E. and Henrick, K. (2007) 'Inference of Macromolecular Assemblies from Crystalline State.' *Journal of Molecular Biology,* 372(3), pp. 774–797.

Kumlin, U., Olofsson, S., Dimock, K. and Arnberg, N. (2008) 'Sialic acid tissue distribution and influenza virus tropism.' *Influenza and other respiratory viruses,* 2(5), pp. 147-154.

Lai, P.-H., Everett, R., Wangs, F.-F., Arakawa, T. and Goldwassers, E. (1986) 'Structural Characterization of Human Erythropoietin.' *The Journal of Biological Chemistry,* 261(7), pp. 3116–3121.

Lam, S. K. and Ng, T. B. (2011) 'Lectins: Production and practical applications.' *Applied Microbiology and Biotechnology,* 89(1), pp. 45–55.

Lauc, G., Essafi, A., Huffman, J. E., Hayward, C., Knežević, A., Kattla, J. J., Polašek, O., Gornik, O., Vitart, V., Abrahams, J. L., Pučić, M., Novokmet, M., Redžić, I., Campbell, S., Wild, S. H., Borovečki, F., Wang,

W., Kolčić, I., Zgaga, L., Gyllensten, U., Wilson, J. F., Wright, A. F., Hastie, N. D., Campbell, H., Rudd, P. M. and Rudan, I. (2010) 'Genomics meets glycomics-the first gwas study of human N-glycome identifies HNF1A as a master regulator of plasma protein fucosylation.' *PLoS Genetics,* 6(12), p.e1001256.

Lee, J. 'Glycoside hydrolase family 33' in *CAZypedia,* available at URL https://www.cazypedia.org/index.php/Glycoside_Hydrolase_Family_33 , accessed 1st July 2019.

Lewis, A. L. and Lewis, W. G. (2012) 'Host sialoglycans and bacterial sialidases: A mucosal perspective.' *Cellular Microbiology,* 14(8), pp. 1174–1182.

Li, J. and McClane, B. A. (2014) 'The Sialidases of Clostridium perfringens Type D Strain CN3718 Differ in Their Properties and Sensitivities to Inhibitors.' *Applied and Environmental Microbiology,* 80(5), pp. 1701–1709.

Li, Q., Qi, J., Wu, Y., Kiyota, H., Tanaka, K., Suhara, Y., Ohrui, H., Suzuki, Y., Vavricka, C. J. and Gao, G. F. (2013) 'Functional and Structural Analysis of Influenza Virus Neuraminidase N3 Offers Further Insight into the Mechanisms of Oseltamivir Resistance.' *Journal of Virology,* 87(18), pp. 10016–10024.

Li, W., Santra, A., Yu, H., Slack, T. J., Muthana, M. M., Shi, D., Liu, Y. and Chen, X. (2019) '9-Azido-9-deoxy-2,3-difluorosialic acid as a subnanomolar inhibitor against bacterial sialidases.' *The Journal of Organic Chemistry*, 84(11), pp. 6697–6708.

Limsuwat, N., Suptawiwat, O., Boonarkart, C., Puthavathana, P., Wiriyarat, W. and Auewarakul, P. (2016) 'Sialic acid content in human saliva and anti-influenza activity against human and avian influenza viruses.' *Archives of virology,* 161(3), pp. 649-656.

Lin, Y. (2018) 'What's happened over the last five years with high-throughput protein crystallization screening?' *Expert Opinion on Drug Discovery,* 13(8), pp. 691–695.

Lindén, S. K., Sheng, Y. H., Every, A. L., Miles, K. M., Skoog, E. C., Florin, T. H. J., Sutton, P. and McGuckin, M. A. (2009) 'MUC1 limits Helicobacter pylori infection both by steric hindrance and by acting as a releasable decoy.' *PLoS pathogens,* 5(10), p. e1000617.

Ling, H., Boodhoo, A., Hazes, B., Cummings, M. D., Armstrong, G. D., Brunton, J. L. and Read, R. J. (1998) 'Structure of the Shiga-like toxin I B-pentamer complexed with an analogue of its receptor Gb3.' *Biochemistry,* 37(7), pp. 1777–1788.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. (2014) 'The carbohydrate-active enzymes database (CAZy) in 2013.' *Nucleic Acids Research,* 42(D1), pp. D490-D495.

Lutz, S. (2010) 'Beyond directed evolution – semi-rational protein engineering and design.' *Current opinion in biotechnology,* 21(6), pp. 734-743.

Luo, Y., Li, S.-C., Chou, M.-Y., Li, Y.-T. and Luo, M. (1998) 'The crystal structure of an intramolecular trans-sialidase with a NeuAcα2→3Gal specificity.' *Structure,* 6(4), pp. 521–530.

Mahdavi, J., Sonden, B., Hurtig, M., Olfat, F. O., Forsberg, L., Roche, N., Larsson, T., Teneberg, S., Karlsson, K., Altraja, S., Wadstro, T., Kersulyte, D., Berg, D. E., Dubois, A., Petersson, C., Magnusson, K. and Norberg, T. (2002) 'Helicobacter pylori SabA Adhesin in Persistent Infection and Chronic Inflammation.' *Science,* 297(5581), pp. 573–579.

Marcobal, A., Barboza, M., Sonnenburg, E. D., Pudlo, N., Martens, E. C., Desai, P., Lebrilla, C. B., Weimer, B. C., Mills, D. A., German, J. B. and Sonnenburg, J. L. (2011) 'Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways.' *Cell Host and Microbe,* 10(5) pp. 507–514.

McAuley, J. L., Gilbertson, B. P., Trifkovic, S., Brown, L. E. and McKimm-Breschkin, J. L. (2019) 'Influenza virus neuraminidase structure and functions.' *Frontiers in Microbiology,* 10, pp. 39.

McCarter, J. D. and Withers, S. G. (1994) 'Mechanisms of enzymatic glycoside hydrolysis.' *Current Opinion in Structural Biology,* 4(6), pp. 885-892.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) 'Phaser crystallographic software.' *Journal of Applied Crystallography,* 40(4), pp. 658–674.

McDonald, N. D., Lubin, J.-B., Chowdhury, N. and Boyd, E. F. (2016) 'Host-Derived Sialic Acids Are an Important Nutrient Source Required for Optimal Bacterial Fitness In Vivo.' *MBio,* 7(2), pp.e02237-15.

McGuckin, M. A., Lindén, S. K., Sutton, P. and Florin, T. H. (2011) 'Mucin dynamics and enteric pathogens.' Nature reviews. Microbiology. *Nature Publishing Group,* 9(4), pp. 265–78.

Mckimm-Breschkin, J. L. (2013) 'Influenza neuraminidase inhibitors: Antiviral action and mechanisms of resistance.' *Influenza and other Respiratory Viruses,* 7, pp. 25–36.

Mescher, M. F. and Strominger, J. L. (1976) 'Purification and characterization of a prokaryotic glycoprotein from the cell envelope of Halobacterium salinarium.' *Journal of Biological Chemistry,* 251(7), pp. 2005–2014.

Miyagi, T. and Tsuiki, S. (1985) 'Purification and characterization of cytosolic sialidase from rat liver.' *Journal of Biological Chemistry,* 260(11), pp. 6710–6716.

Miyagi, T. and Tsuiki, S. (1984) 'Rat-liver lysosomal sialidase: Solubilization, substrate specificity and comparison with the cytosolic sialidase.' *European Journal of Biochemistry,* 141(1), pp. 75–81.

Miyagi, T. and Yamaguchi, K. (2012) 'Mammalian sialidases: Physiological and pathological roles in cellular functions.' *Glycobiology,* 22(7), pp. 880–896.

Mizan, S., Henk, A., Stallings, A.., Maier, M. and Lee, M. D. (2000) 'Cloning and characterization of sialidases with 2-6' and 2-3' sialyl lactose specificity from Pasteurella multocida.' Journal of Bacteriology, 182(24), pp. 6874–6883.

Moncla, B. J., Braham, P. and Hillier, S. L. (1990) 'Sialidase (neuraminidase) activity among gram-negative anaerobic and capnophilic bacteria.' *Journal of Clinical Microbiology,* 28(3), pp. 422-425.

Monti, E., Bassi, M. T., Papini, N., Riboni, M., Manzoni, M., Venerando, B., Croci, G., Preti, A., Ballabio, A., Tettamanti, G. and Borsani, G. (2000) 'Identification and Expression of NEU3, A Novel Human Sialidase Associated To the Plasma Membrane.' *Biochemical Journal,* 349(1), pp. 343–351.

Monti, E., Preti, A., Rossi, E., Ballabio, A. and Borsani, G. (1999) 'Cloning and characterization of NEU2, a human gene homologous to rodent soluble sialidases.' *Genomics,* 57(1), pp. 137–43.

Monti, E., Preti, A., Venerando, B. and Borsani, G. (2002) 'Recent development in mammalian sialidase molecular biology.' *Neurochemical research,* 27(7–8), pp. 649–663.

Moran, A. P., Gupta, A. and Joshi, L. (2011) 'Sweet-talk: Role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract.' *Gut,* 60(10), pp. 1412–1425.

Moustafa, I., Connaris, H., Taylor, M., Zaitsev, V., Wilson, J. C., Kiefel, M. J., Von Itzstein, M. and Taylor, G. (2004) 'Sialic acid recognition by Vibrio cholerae neuraminidase.' *Journal of Biological Chemistry,* 279(39), pp. 40819–40826.

Mulloy, B., Hart, G. W. and Stanley, P. (2009) 'Structural Analysis of Glycans.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 2nd edition, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Newstead, S. L., Potter, J. A., Wilson, J. C., Xu, G., Chien, C. H., Watts, A. G., Withers, S. G. and Taylor, G. L. (2008) 'The structure of Clostridium perfringens NanI sialidase and its catalytic intermediates.' *Journal of Biological Chemistry,* 283(14), pp. 9080–9088.

Ng, K. M., Ferreyra, J. A., Higginbottom, S. K., Lynch, J. B., Kashyap, P. C., Gopinath, S., Naidu, N., Choudhury, B., Weimer, B. C., Monack, D. M. and Sonnenburg, J. L. (2013) 'Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens.' *Nature,* 502(7469), pp. 96–99.

Nothaft, H. and Szymanski, C. M. (2010) 'Protein glycosylation in bacteria: Sweeter than ever.' *Nature Reviews Microbiology,* 8(11), pp. 765–778.

Owen, C. D., Lukacik, P., Potter, J. A., Sleator, O., Taylor, G. L. and Walsh, M. A. (2015) 'Streptococcus pneumoniae NanC: Structural insights into the specificity and mechanism of a sialidase that produces a sialidase inhibitor.' *Journal of Biological Chemistry,* 290(46), pp. 27736–27748.

Owen, C. D., Tailford, L. E., Monaco, S., Šuligoj, T., Vaux, L., Lallement, R., Khedri, Z., Yu, H., Lecointe, K., Walshaw, J., Tribolo, S., Horrex, M., Bell, A., Chen, X., Taylor, G. L., Varki, A., Angulo, J. and Juge, N. (2017) 'Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont Ruminococcus gnavus.' *Nature Communications,* 8(1), p. 2196.

Owen, C. J. and Stewart, D. A. (2015) 'Obinutuzumab for the treatment of patients with previously untreated chronic lymphocytic leukemia: overview and perspective.' *Therapeutic advances in hematology,* 6(4), pp. 161–170.

Paris, G., Cremona, M. L., Amaya, M. F., Buschiazzo, A., Giambiagi, S., Frasch, A. C. C. and Alzari, P. M. (2001) 'Probing molecular function of trypanosomal sialidases: Single point mutations can change substrate specificity and increase hydrolytic activity.' *Glycobiology,* 11(4), pp. 305–311.

Park, K. H., Kim, M. G., Ahn, H. J., Lee, D. H., Kim, J. H., Kim, Y. W. and Woo, E. J. (2013) 'Structural and biochemical characterization of the broad substrate specificity of Bacteroides thetaiotaomicron commensal sialidase.' *Biochimica et Biophysica Acta - Proteins and Proteomics,* 1834(8), pp. 1510–1519.

Parker, D., Soong, G., Planet, P., Brower, J., Ratner, A. J. and Prince, A. (2009) 'The NanA neuraminidase of Streptococcus pneumoniae is involved in biofilm formation.' *Infection and Immunity,* 77(9), pp. 3722–3730.

Pflugrath, J. W. (2015) 'Practical macromolecular crystallography.' *Acta Crystallographica Section F: Structural Biology Communications,* 71(6), pp. 622-642.

Phansopa, C., Kozak, R. P., Liew, L. P., Frey, A. M., Farmilo, T., Parker, J. L., Kelly, D. J., Emery, R. J., Thomson, R. I., Royle, L., Gardner, R. A., Spencer, D. I. R. and Stafford, G. P. (2015) 'Characterization of a sialate-O-acetylesterase (NanS) from the oral pathogen Tannerella forsythia that enhances sialic acid release by NanH, its cognate sialidase.' *The Biochemical Journal,* 472(2), pp. 157–167.

Pillai, S., Netravali, I. A., Cariappa, A. and Mattoo, H. (2012) 'Siglecs and Immune Recognition.' *Annual Review of Immunology,* 30(3), pp. 357–392.

Prevost, R. R. (1998) 'Recombinant Follicle-Stimulating Hormone: New Biotechnology for Infertility.' *Pharmacotherapy: The Journal of Human Pharmacolohy and Drug Therapy,* 18(5), pp. 1001-1010.

Quosdorf, S., Schuetz, A. and Kolodziej, H. (2017) 'Different inhibitory potencies of oseltamivir carboxylate, zanamivir, and several tannins on bacterial and viral neuraminidases as Assessed in a Cell-Free Fluorescence-Based Enzyme Inhibition Assay.' *Molecules,* 22(11), p. 1989.

Reusch, D., Haberger, M., Maier, B., Maier, M., Kloseck, R., Zimmermann, B., Hook, M., Szabo, Z., Tep, S., Wegstein, J., Alt, N., Bulau, P. and Wuhrer, M. (2015a) 'Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles--part 1: separation-based methods.' *MAbs*, 7(1) pp. 167–179.

Reusch, D., Haberger, M., Falck, D., Peter, B., Maier, B., Gassner, J., Hook, M., Wagner, K., Bonnington, L., Bulau, P., et al. (2015b). Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles-Part 2: Mass spectrometric methods. *MAbs,* 7(4), 732–742.

Rhodes, G. (1999) *Crystallography made crystal clear: a guide for users of macromolecular models.* 2nd edition. Academic Press.

Ribeiro, J. P., Pau, W., Pifferi, C., Renaudet, O., Varrot, A., Mahal, L. K. and Imberty, A. (2016) 'Characterization of a high-affinity sialic acid-specific CBM40 from Clostridium perfringens and engineering of a divalent form.' *Biochemical Journal,* 473(14), pp. 2109–2118.

Rogers, G. N. and Paulson, J. C. (1983) 'Receptor determinants of human and animal influenza virus isolates: Differences in receptor specificity of the H3 hemagglutinin based on species of origin.' *Virology,* 127(2), pp. 361–373.

Roggentin, P., Berg, W. and Schauer, R. (1987) 'Purification and characterization of sialidase from Clostridium sordellii G12.' *Glycoconjugate Journal,* 4(4), pp, 349-359.

Roggentin, P., Rothe, B., Kaper, J. B., Galen, J., Lawrisuk, L., Vimr, E. R. and Schauer, R. (1989) 'Conserved sequences in bacterial and viral sialidases.' *Glycoconjugate Journal,* 6(3), pp. 349–353.

Rossmann, M. G. (1990) 'The molecular replacement method.' *Acta Crystallographica Section A: Foundations of Crystallography,* 46(2), pp. 73–82.

Roth, Z., Yehezkel, G. and Khalaila, I. (2012) 'Identification and Quantification of Protein Glycosylation.' International Journal of Carbohydrate Chemistry, 2012.

Roy, S., Douglas, C. W. I. and Stafford, G. P. (2010) 'A novel sialic acid utilization and uptake system in the periodontal pathogen Tannerella forsythia.' *Journal of Bacteriology,* 192(9), pp. 2285–2293.

Roy, S., Honma, K., Ian Douglas, C. W., Sharma, A. and Stafford, G. P. (2011) 'Role of sialidase in glycoprotein utilization by Tannerella forsythia.' *Microbiology,* 157(11), pp. 3195–3202.

Rupp, B. (2009) *Biomolecular crystallography: principles, practice and application to structural biology.* Garland Science.

Russell, R. J., Haire, L. F., Stevens, D. J., Collins, P. J., Lin, Y. P., Blackburn, G. M., Hay, A. J., Gamblin, S. J. and Skehel, J. J. (2006) 'The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design.' *Nature,* 443(7107), p.45.

Samraj, A. N., Pearce, O. M., Läubli, H., Crittenden, A. N., Bergfeld, A. K., Banda, K., Gregg, C. J., Bingman, A. E., Secrest, P., Diaz, S. L. and Varki, N. M. (2015) 'A red-meat derived glycan promotes inflammation and cancer progression.' *Proceedings of the National Academy of Sciences,* 112(2), pp. 542-547.

Sasaki, H., Bothner, B., Dell, A. and Fukuda, M. (1987) 'Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA.' *Journal of Biological Chemistry,* 262(25), pp. 12059–12076.

Scapin, G. (2013) 'Molecular replacement then and now.' *Acta Crystallographica Section D: Biological Crystallography,* 69(11), pp. 2266–2275.

Schenkman, S., Jiang, M. S., Hart, G. W. and Nussenzweig, V. (1991) 'A novel cell surface trans-sialidase of Trypanosoma cruzi generates a stage-specific epitope required for invasion of mammalian cells.' *Cell,* 65(7), pp. 1117–1125.

Schwerdtfeger, S. M. and Melzig, M. F. (2010) 'Sialidases in biological systems.' *Pharmazie,* 65(8), pp. 551–561.

Sela, D. A., Li, Y., Lerno, L., Wu, S., Marcobal, A. M., Bruce German, J., Chen, X., Lebrilla, C. B. and Mills, D. A. (2011) 'An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides.' *Journal of Biological Chemistry,* 286(14), pp. 11909–11918.

Severi, E., Hood, D. W. and Thomas, G. H. (2007) 'Sialic acid utilization by bacterial pathogens.' *Microbiology,* 153(9), pp. 2817–2822.

Sharon, N. and Lis, H. (2004) 'History of lectins: From hemagglutinins to biological recognition molecules.' *Glycobiology,* 14(11), pp. 53–62.

Shields, B. M., Hicks, S., Shepherd, M. H., Colclough, K., Hattersley, A. T. and Ellard, S. (2010) 'Maturity-onset diabetes of the young (MODY): How many cases are we missing?' *Diabetologia,* 53(12), pp. 2504–2508.

Slack, T. J., Li, W., Shi, D., McArthur, J. B., Zhao, G., Li, Y., Xiao, A., Khedri, Z., Yu, H., Liu, Y. and Chen, X. (2018) 'Triazole-linked transition state analogs as selective inhibitors against V. cholerae sialidase.' *Bioorganic and Medicinal Chemistry,* 26(21), pp. 5751–5757.

Sleytr, U. B. and Thorne, K. J. I. (1976) 'Chemical characterization of the regularly arranged surface layers of Clostridium thermosaccharolyticum and Clostridium thermohydrosulfuricum.' *Journal of Bacteriology,* 126(1), pp. 377–383.

Sokurenko, E. V., Chesnokova, V., Dykhuizen, D. E., Ofek, I., Wu, X.-R., Krogfelt, K. A., Struve, C., Schembri, M. A. and Hasty, D. L. (1998) 'Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin.' *Proceedings of the National Academy of Sciences,* 95(15), pp. 8922–8926.

Stafford, G., Roy, S., Honma, K. and Sharma, A. (2012). 'Sialic acid, periodontal pathogens and Tannerella forsythia: stick around and enjoy the feast!' *Molecular oral microbiology,* 27(1), pp. 11-22.

Stamatos, N. M., Liang, F., Nan, X., Landry, K., Cross, A. S., Wang, L. X. and Pshezhetsky, A. V. (2005) 'Differential expression of endogenous sialidases of human monocytes during cellular differentiation into macrophages.' *FEBS Journal,* 272(10), pp. 2545–2556.

Stanley, P., Schachter, H. and Taniguchi, N. (2009) 'N-Glycans.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 2nd edition, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Stanley, P., Taniguchi, N. and Aebi, M. (2017) 'N-Glycans.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 3rd edition, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Van den Steen, P., Rudd, P. M., Dwek, R. A. and Opdenakker, G. (1998) 'Concepts and principles of O-linked glycosylation.' *Critical Reviews in Biochemistry and Molecular Biology,* 33(3), pp. 151–208.

Steiner, K. and Schwab, H. (2012) 'Recent advances in rational approaches for enzyme engineering.' *Computational and structural biotechnology journal,* 2(3), pp. e201209010.

Stimson, E., Virji, M., Makepeace, K., Dell, A., Morris, H. R., Payne, G., Saunders, J. R., Jennings, M. P., Barker, S., Panico, M. and Blench, I. (1995) 'Meningococcal pilin: a glycoprotein substituted with digalactosyl 2, 4-diacetamido-2, 4, 6-trideoxyhexose.' *Molecular Microbiology,* 17(6), pp. 1201-1214.

Sun, X., Li, Q., Wu, Y., Wang, M., Liu, Y., Qi, J., Vavricka, C. J. and Gao, G. F. (2014) 'Structure of Influenza Virus N7: the Last Piece of the Neuraminidase "Jigsaw" Puzzle.' *Journal of Virology,* 88(16), pp. 9197–9207.

Szymanski, C. M., Ruijin, Y., Ewing, C. P., Trust, T. J. and Guerry, P. (1999) 'Evidence for a system of general protein glycosylation in Campylobacter jejuni.' *Molecular Microbiology,* 32(5), pp. 1022–1030.

Tailford, L. E., Owen, C. D., Walshaw, J., Crost, E. H., Hardy-Goddard, J., Le Gall, G., de Vos, W. M., Taylor, G. L. and Juge, N. (2015) 'Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation.' *Nature communications*, 6, p. 7624.

Tailford, L. E., Crost, E. H., Kavanaugh, D. and Juge, N. (2015b) 'Mucin foraging in the human gut microbiome.' *Frontiers in genetics,* 6(81).

Tangvoranuntakul, P., Gagneux, P., Diaz, S., Bardor, M., Varki, N., Varki, A. and Muchmore, E. (2003) 'Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid.' *Proceedings of the National Academy of Sciences,* 100(21), pp. 12045–12050.

Taylor, G. (2003) 'The phase problem.' *Acta Crystallographica Section D: Biological Crystallography,* 59(11), pp. 1881–1890.

Taylor, N. R. and von Itzstein, M. (1994) 'Molecular Modeling Studies on Ligand Binding to Sialidase from Influenza Virus and the Mechanism of Catalysis.' *Journal of Medicinal Chemistry,* 37(5), pp. 616–624.

Thanabalasingham, G. and Owen, K. R. (2011) 'Diagnosis and management of maturity onset diabetes of the young (MODY). *BMJ,* 343, p.d6044.

Thanabalasingham, G., Huffman, J. E., Kattla, J. J., Novokmet, M., Rudan, I., Gloyn, A. L., Hayward, C., Adamczyk, B., Reynolds, R. M., Muzinic, A., Hassanali, N., Pucic, M., Bennett, A. J., Essafi, A., Polasek, O., Mughal, S. A., Redzic, I., Primorac, D., Zgaga, L., Kolcic, I., Hansen, T., Gasperikova, D., Tjora, E., Strachan, M. W. J., Nielsen, T., Stanik, J., Klimes, I., Pedersen, O. B., Njølstad, P. R., Wild, S. H., Gyllensten, U., Gornik, O., Wilson, J. F., Hastie, N. D., Campbell, H., McCarthy, M. I., Rudd, P. M., Owen, K. R., Lauc, G. and Wright, A. F. (2013) 'Mutations in HNF1A result in marked alterations of plasma glycan profile.' *Diabetes,* 62(4), pp. 1329–1337.

Thobhani, S., Ember, B., Siriwardena, A. and Boons, G. J. (2003) 'Multivalency and the mode of action of bacterial sialidases.' *Journal of the American Chemical Society,* 125(24), pp. 7154–7155.

Thomas, G. H. (2016) 'Sialic acid acquisition in bacteria-one substrate, many transporters.' *Biochemical Society Transactions,* 44(3), pp. 760–765.

Thompson, H., Homer, K. A., Rao, S., Booth, V. and Hosie, A. H. F. (2009) 'An orthologue of Bacteroides fragilis NanH is the principal sialidase in Tannerella forsythia.' *Journal of Bacteriology,* 191(11), pp. 3623–3628.

Thursby, E. and Juge, N. (2017) 'Introduction to the human gut microbiota.' *Biochemical Journal,* 474(11), pp. 1823-1836.

Turner, M. W. (2004) 'The role of mannose-binding lectin in health and disease.' *Netherlands Journal of Medicine,* 62(3), pp. 4–9.

Ueno, K., Koga, T., Kato, K., Golenbock, D. T., Gendler, S. J., Kai, H. and Kim, K. C. (2008) 'MUC1 mucin is a negative regulator of toll-like receptor signaling.' *American Journal of Respiratory Cell and Molecular Biology,* 38(3), pp. 263–268.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. and Murshudov, G. N. (2004) 'REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use.' *Acta Crystallographica Section D: Biological Crystallography*, 60(12) pp. 2184–2195.

Vandekerckhove, F., Schenkman, S., Yoshida, M., Carvalho, L. P. de, Hasegawa, A., Nussenzweig, V., Tomlinson, S. and Kiso, M. (1992) 'Substrate specificity of the Trypanosoma cruzi trans-sialidase.' *Glycobiology,* 2(6), pp. 541–548.

Varghese, J. N. and Colman, P. M. (1991) 'Three-dimensional Structure of the Neuraminidase of Influenza Virus A / Tokyo / 3 / 67 at 2-2 A Resolution receptor.' *Journal of molecular biology,* 221(2), pp. 473–486.

Varki, A. (2008) 'Sialic acids in human health and disease.' *Trends in Molecular Medicine,* 14(8), pp. 351–360.

Varki, A. and Schauer, R. (2009) 'Sialic Acids.' *In* Varki, A., Cummings, R., and Esko, J. *Essentials of Glycobiology*. 2nd edition, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Vavricka, C. J., Liu, Y., Kiyota, H., Sriwilaijaroen, N., Qi, J., Tanaka, K., Wu, Y., Li, Q., Li, Y., Yan, J., Suzuki, Y. and Gao, G. F. (2013) 'Influenza neuraminidase operates via a nucleophilic mechanism and can be targeted by covalent inhibitors.' *Nature Communications*, 4, p. 1491.

Vimr, E. R., Kalivoda, K. A., Deszo, E. L. and Steenbergen, S. M. (2004) 'Diversity of Microbial Sialic Acid Metabolism.' *Microbiology And Molecular Biology Reviews,* 68(1), pp. 132-153.

Walsh, G. (2018) 'Biopharmaceutical benchmarks 2018.' *Nature biotechnology,* 36, pp. 1136–1145.

Watson, J. N., Dookhun, V., Borgford, T. J. and Bennet, A. J. (2003) 'Mutagenesis of the Conserved Active-Site Tyrosine Changes a Retaining Sialidase into an Inverting Sialidase.' *Biochemistry,* 42(43), pp. 12682–12690.

Withers, S. and Williams, S. "Glycoside Hydrolases" in *CAZypedia,* available at URL https://www.cazypedia.org/index.php/Glycoside_hydrolases#Alternative_nucleophiles, accessed 30th April 2019.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011) 'Overview of the CCP4 suite and current developments.' *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 235–242.

Winter, G. (2010) 'Xia2: an Expert System for Macromolecular Crystallography Data Reduction.' *Journal of Applied Crystallography*, 43(1), pp. 186–190.

Withers, S. and Williams, S. 'Glycoside Hydrolases' in *CAZypedia,* available at URL https://www.cazypedia.org/index.php/Glycoside_hydrolases, accessed 17th September 2019).

Wormald, M. R., Petrescu, A. J., Pao, Y. L., Glithero, A., Elliott, T. and Dwek, R. A. (2002) 'Conformational Studies of Oligosaccharides and Glycopeptides: Complementary of NMR, X-Ray Crystallography, and Molecular Modelling.' *Chemical Reviews,* 102(2), pp. 371–386.

Wu, A. M., Csako, G. and Herp, A. (1994) 'Structure, biosynthesis, and function of salivary mucins.' *Molecular and Cellular Biochemistry,* 137(1), pp. 39–55.

Xu, G., Potter, J. A., Russell, R. J. M., Oggioni, M. R., Andrew, P. W. and Taylor, G. L. (2008) 'Crystal Structure of the NanB Sialidase from Streptococcus pneumoniae.' *Journal of Molecular Biology,* 384(2), pp. 436–449.

Yang, L., Connaris, H., Potter, J. A. and Taylor, G. L. (2015) 'Structural characterization of the carbohydrate-binding module of NanA sialidase, a pneumococcal virulence factor.' BMC Structural Biology. *BMC Structural Biology,* 15(1), p. 15.

Zenz, K. I., Roggentin, P. and Schauer, R. (1993) 'Isolation and properties of the natural and the recombinant sialidase from Clostridium septicum NC 0054714.' *Glycoconjugate Journal,* 10(1), pp. 50–56.

Zhao, S., Walsh, I., Abrahams, J. L., Royle, L., Nguyen-Khuong, T., Spencer, D., Fernandes, D. L., Packer, N. H., Rudd, P. M. and Campbell, M. P. (2018) 'GlycoStore: a database of retention properties for glycan analysis.' *Bioinformatics,* 34(18), pp. 3231-3232.

Zhu, X., McBride, R., Nycholat, C. M., Yu, W., Paulson, J. C. and Wilson, I. A. (2012) 'Influenza Virus Neuraminidases with Reduced Enzymatic Activity That Avidly Bind Sialic Acid Receptors.' *Journal of Virology,* 86(24), pp. 13371–13383.

Zunk, M. and Kiefel, M. J. (2014) 'The occurrence and biological significance of the α-keto-sugars pseudaminic acid and legionaminic acid within pathogenic bacteria.' *RSC Advances,* 4(7), pp. 3413–3421.

Appendix

1. NanH gene synthesis - codon optimised for *E. coli*

Nucleotide sequence:

This gene is based upon the *Tannerella forsythia* 92A2 strain. The NdeI and XhoI restriction sites are highlighted in yellow.

CACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGCATCATATGGCCGATAGCGTTTATGTTCAGAATCCGCA
GATTCCGATTCTGGTTGATCGTACCGATAATGTGCTGTTTCGTATTCGTATTCCGGATGCAACCAAAGGTGATGT
TCTGAATCGTCTGACCATTCGCTTTGGCAATGAAGATAAACTGAGCGAAGTTAAAGCAGTGCGTCTGTTTTATGC
AGGCACCGAAGCAGCCACCAAAGGTCGTAGCCGTTTTGCACCGGTTACCTATGTTAGCAGCCATAACATTCGTAA
TACCCGTAGCGCAAATCCGAGCTATAGCATTCGTCAGGATGAAGTTACCACCGTTGCAAATACCCTGACCCTGAA
AACCCGTCAGCCGATGGTTAAAGGTATTAACTATTTTTGGGTGAGCGTGGAAATGGATCGCAATACCAGCCTGCT
GAGCAAACTGACCAGCACCGTTACCGAAGTTGTGATTAATGATAAACCGGCAGTTATTGCCGGTGAACAGGCAGC
AGTTCGTCGTATGGGTATTGGTGTTCGTCATGCCGGTGATGATGGTAGCGCAAGCTTTCGCATTCCGGGTCTGGT
TACCACCAATAAAGGCACCCTGCTGGGTGTTTATGATGTGCGTTATAACAATAGCGTGGATCTGCAAGAACATAT
TGATGTTGGTCTGAGCCGTAGCACCGATAAAGGCCAGACCTGGGAACCGATGCGTATTGCAATGAGCTTTGGTGA
AACCGATGGTCTGCCGAGCGGTCAGAATGGTGTTGGTGATCCGAGCATTCTGGTGGATGAACGTACCAATACCGT
TTGGGTTGTTGCAGCATGGACCCATGGTATGGGTAATGCACGTGCCTGGACCAATAGCATGCCTGGTATGACACC
GGATGAAACCGCACAGCTGATGATGGTTAAAAGCACCGATGATGGTCGTACCTGGTCAGAAAGCACCAATATTAC
CAGCCAGGTTAAAGATCCGAGCTGGTGTTTTCTGCTGCAGGGTCCGGGTCGTGGTATTACCATGCGTGATGGCAC
CCTGGTTTTTCCGATTCAGTTTATTGATAGCCTGCGTGTTCCGCATGCAGGTATTATGTATAGCAAAGATCGCGG
TGAAACCTGGCATATTCATCAGCCTGCACGTACAAATACCACCGAAGCCCAGGTTGCCGAAGTTGAACCGGGTGT
TCTGATGCTGAATATGCGTGATAATCGTGGTGGTAGCCGTGCAGTTAGCATTACCCGTGATCTGGGTAAAAGCTG
GACCGAACATAGCAGCAATCGTAGCGCACTGCCGGAAAGCATTTGTATGGCAAGCCTGATTAGCGTTAAAGCCAA
AGATAACATCATCGGCAAAGATCTGCTGCTGTTTAGCAATCCGAATACCACAGAAGGTCGTCATCATATTACCAT
TAAAGCAAGCCTGGATGGTGGTGTTACCTGGCTGCCTGCACATCAGGTGCTGCTGGATGAAGAAGATGGTTGGGG
TTATAGCTGTCTGAGCATGATTGATCGTGAAACCGTGGGTATCTTTTATGAAAGCAGCGTTGCACACATGACCTT
TCAGGCAGTTAAAATCAAAGACCTGATTCGCCTCGAGCTGGGCCTCATGGGCCTTCCGCTCACTGCCCGCTTTCC
AG

Protein sequence:

The open reading frame as reported by the manufacturer (Life Technologies – Thermo Fisher Scientific).

M A D S V Y V Q N P Q I P I L V D R T D N V L F R I R I P D A T K G D V L N
R L T I R F G N E D K L S E V K A V R L F Y A G T E A A T K G R S R F A P V
T Y V S S H N I R N T R S A N P S Y S I R Q D E V T T V A N T L T L K T R Q
P M V K G I N Y F W V S V E M D R N T S L L S K L T S T V T E V V I N D K P
A V I A G E Q A A V R R M G I G V R H A G D D G S A S F R I P G L V T T N K
G T L L G V Y D V R Y N N S V D L Q E H I D V G L S R S T D K G Q T W E P M
R I A M S F G E T D G L P S G Q N G V G D P S I L V D E R T N T V W V V A A
W T H G M G N A R A W T N S M P G M T P D E T A Q L M M V K S T D D G R T W
S E S T N I T S Q V K D P S W C F L L Q G P G R G I T M R D G T L V F P I Q
F I D S L R V P H A G I M Y S K D R G E T W H I H Q P A R T N T T E A Q V A
E V E P G V L M L N M R D N R G G S R A V S I T R D L G K S W T E H S S N R
S A L P E S I C M A S L I S V K A K D N I I G K D L L L F S N P N T T E G R
H H I T I K A S L D G G V T W L P A H Q V L L D E E D G W G Y S C L S M I D
R E T V G I F Y E S S V A H M T F Q A V K I K D L I R

2. Crystallisation screens

Molecular Dimensions

JCSG-*plus* HT-96 Wells A1- D12 MD1–40

Well #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
A1	0.2 M	Lithium sulfate	0.1 M	Sodium acetate	4.5	50 % w/v	PEG 400
A2		None	0.1 M	Sodium citrate	5.5	20 % w/v	PEG 3000
A3	0.2 M	Ammonium citrate dibasic		None		20 % w/v	PEG 3350
A4	0.02 M	Calcium chloride dihydrate	0.1 M	Sodium acetate	4.6	30 % v/v	MPD
A5	0.2 M	Magnesium formate dihydrate		None		20 % w/v	PEG 3350
A6	0.2 M	Lithium sulfate	0.1 M	Phosphate/citrate	4.2	20 % w/v	PEG 1000
A7		None	0.1 M	CHES	9.5	20 % w/v	PEG 8000
A8	0.2 M	Ammonium formate		None		20 % w/v	PEG 3350
A9	0.2 M	Ammonium chloride		None		20 % w/v	PEG 3350
A10	0.2 M	Potassium formate		None		20 % w/v	PEG 3350
A11	0.2 M	Ammonium phosphate monobasic	0.1 M	Tris	8.5	50 % v/v	MPD
A12	0.2 M	Potassium nitrate		None		20 % w/v	PEG 3350
B1	0.8 M	Ammonium sulfate	0.1 M	Citrate	4.0		None
B2	0.2 M	Sodium thiocyanate		None		20 % w/v	PEG 3350
B3		None	0.1 M	BICINE	9.0	20 % w/v	PEG 6000
B4		None	0.1 M	HEPES	7.5	10 % w/v	PEG 8000
B5		None	0.1 M	Sodium cacodylate	6.5	40 % v/v	MPD
						5 % w/v	PEG 8000
B6		None	0.1 M	Phosphate/citrate	4.2	40 % v/v	Ethanol
						5 % w/v	PEG 1000
B7		None	0.1 M	Sodium acetate	4.6	8 % w/v	PEG 4000
B8	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	7.0	10 % w/v	PEG 8000
B9		None	0.1 M	Citrate	5.0	20 % w/v	PEG 6000
B10	0.2 M	Magnesium chloride hexahydrate	0.1 M	Sodium cacodylate	6.5	50 % v/v	PEG 200
B11	1.6 M	Sodium citrate tribasic dihydrate pH 6.5		None			None
B12	0.2 M	Potassium citrate tribasic monohydrate		None		20 % w/v	PEG 3350
C1	0.2 M	Sodium chloride	0.1 M	Phosphate/citrate	4.2	20 % w/v	PEG 8000
C2	1.0 M	Lithium chloride	0.1 M	Citrate	4.0	20 % w/v	PEG 6000
C3	0.2 M	Ammonium nitrate		None		20 % w/v	PEG 3350
C4		None	0.1 M	HEPES	7.0	10 % w/v	PEG 6000
C5	0.8 M	Sodium phosphate monobasic monohydrate	0.1 M	Sodium HEPES	7.5		None
	0.8 M	Potassium phosphate monobasic		None			
C6		None	0.1 M	Phosphate/citrate	4.2	40 % v/v	PEG 300
C7	0.2 M	Zinc acetate dihydrate	0.1 M	Sodium acetate	4.5	10 % w/v	PEG 3000
C8		None	0.1 M	Tris	8.5	20 % v/v	Ethanol
C9		None	0.1 M	Sodium/potassium phosphate	6.2	25 % v/v	1,2-Propandiol
						10 %v/v	Glycerol
C10		None	0.1 M	BICINE	9.0	10 % w/v	PEG 20,000
						2 % v/v	1,4-Dioxane
C11	2.0 M	Ammonium sulfate	0.1 M	Sodium acetate	4.6		None
C12		None		None		10 % w/v	PEG 1000
						10 % w/v	PEG 8000
D1		None		None		24 % w/v	PEG 1500
						20 % v/v	Glycerol
D2	0.2 M	Magnesium chloride hexahydrate	0.1 M	Sodium HEPES	7.5	30 % v/v	PEG 400
D3	0.2 M	Sodium chloride	0.1 M	Sodium/potassium phosphate	6.2	50 % v/v	PEG 200
D4	0.2 M	Lithium sulfate	0.1 M	Sodium acetate	4.5	30 % w/v	PEG 8000
D5		None	0.1 M	HEPES	7.5	70 % v/v	MPD
D6	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	8.5	20 % w/v	PEG 8000
D7	0.2 M	Lithium sulfate	0.1 M	Tris	8.5	40 % v/v	PEG 400
D8		None	0.1 M	Tris	8.0	40 % v/v	MPD
D9	0.17 M	Ammonium sulfate		None		25.5 % w/v	PEG 4000
						15 % v/v	Glycerol
D10	0.2 M	Calcium acetate hydrate	0.1 M	Sodium cacodylate	6.5	40 % v/v	PEG 300
D11	0.14 M	Calcium chloride dihydrate	0.07 M	Sodium acetate	4.6	14 % v/v	2-Propanol
						30 % v/v	Glycerol
D12	0.04 M	Potassium phosphate monobasic		None		16 % w/v	PEG 8000
						20 % v/v	Glycerol

Molecular Dimensions

Soluble Proteins

JCSG-*plus* HT-96 Wells E1 – H12 MD1–40

Well #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
E1	1.0 M	Sodium citrate tribasic dihydrate	0.1 M	Sodium cacodylate	6.5		None
E2	2.0 M	Ammonium sulfate	0.1 M	Sodium cacodylate	6.5		None
	0.2 M	Sodium chloride					
E3	0.2 M	Sodium chloride	0.1 M	HEPES	7.5	10 % v/v	2-Propanol
E4	1.26 M	Ammonium sulfate	0.1 M	Tris	8.5		None
	0.2 M	Lithium sulfate					
E5		None	0.1 M	CAPS	10.5	40 % v/v	MPD
E6	0.2 M	Zinc acetate dihydrate	0.1 M	Imidazole	8.0	20 % w/v	PEG 3000
E7	0.2 M	Zinc acetate dihydrate	0.1 M	Sodium cacodylate	6.5	10 % v/v	2-Propanol
E8	1.0 M	Ammonium phosphate dibasic	0.1 M	Sodium acetate	4.5		None
E9	1.6 M	Magneisum sulfate heptahydrate	0.1 M	MES	6.5		None
E10		None	0.1 M	BICINE	9.0	10 % w/v	PEG 6000
E11	0.16 M	Calcium acetate hydrate	0.08 M	Sodium cacodylate	6.5	14.4 % w/v	PEG 8000
						20 % v/v	Glycerol
E12		None	0.1 M	Imidazole	8.0	10 % w/v	PEG 8000
F1	0.05 M	Cesium chloride	0.1 M	MES	6.5	30 % v/v	Jeffamine® M-600
F2	3.2 M	Ammonium sulfate	0.1 M	Citrate	5.0		None
F3		None	0.1 M	Tris	8.0	20 % v/v	MPD
F4		None	0.1 M	HEPES	7.5	20 % v/v	Jeffamine® M-600
F5	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	8.5	50 % v/v	Ethylene glycol
F6		None	0.1 M	BICINE	9.0	10 % v/v	MPD
F7	0.8 M	Succinic acid pH 7.0		None			None
F8	2.1 M	DL-Malic acid pH 7.0		None			None
F9	2.4 M	Sodium malonate dibasic monohydrate pH 7.0		None			None
F10	1.1 M	Sodium malonate dibasic monohydrate	0.1 M	HEPES	7.0	0.5 % v/v	Jeffamine® ED-2003
F11	1.0 M	Succinic acid	0.1 M	HEPES	7.0	1 % w/v	PEG 2000 MME
F12		None	0.1 M	HEPES	7.0	30 % v/v	Jeffamine® M-600
G1		None	0.1 M	HEPES	7.0	30 % v/v	Jeffamine® ED-2003
G2	0.02 M	Magnesium chloride hexahydrate	0.1 M	HEPES	7.5	22 % w/v	Poly(acrylic acid sodium salt) 5100
G3	0.01 M	Cobalt(II) chloride hexahydrate	0.1 M	Tris	8.5	20 % w/v	Polyvinylpyrrolidone
G4	0.2 M	TMAO	0.1 M	Tris	8.5	20 % w/v	PEG 2000 MME
G5	0.005 M	Cobalt(II) chloride hexahydrate	0.1 M	HEPES	7.5	12 % w/v	PEG 3350
	0.005 M	Cadmium chloride hemi(pentahydrate)					
	0.005 M	Magnesium chloride hexahydrate					
	0.005 M	Nickel(II) chloride hexahydrate					
G6	0.2 M	Sodium malonate dibasic monohydrate		None		20 % w/v	PEG 3350
G7	0.1 M	Succinic acid		None		15 % w/v	PEG 3350
G8	0.15 M	DL-Malic acid		None		20 % w/v	PEG 3350
G9	0.1 M	Potassium thiocyanate		None		30 % w/v	PEG 2000 MME
G10	0.15 M	Potassium bromide		None		30 % w/v	PEG 2000 MME
G11	2.0 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5		None
G12	3.0 M	Sodium chloride	0.1 M	BIS-Tris	5.5		None
H1	0.3 M	Magnesium formate dihydrate	0.1 M	BIS-Tris	5.5		None
H2	1.0 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5	1 % w/v	PEG 3350
H3		None	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H4	0.2 M	Calcium chloride dihydrate	0.1 M	BIS-Tris	5.5	45 % v/v	MPD
H5	0.2 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	45 % v/v	MPD
H6	0.1 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	17 % w/v	PEG 10,000
H7	0.2 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H8	0.2 M	Sodium chloride	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H9	0.2 M	Lithium sulfate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H10	0.2 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H11	0.2 M	Magnesium chloride hexahydrate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H12	0.2 M	Ammonium acetate	0.1 M	HEPES	7.5	45 % v/v	MPD

Abbreviations: Bis Tris; Bis-(2-hydroxyethyl)imino-tris(hydroxymethyl)methane, **CAPS**; N-Cyclohexyl-3-aminopropanesulfonic acid, **CHES**; 2-(N-Cyclohexylamino)ethane Sulfonic Acid, **HEPES;** 2-(4-(2-Hydroxyethyl)-1-piperazinyl)ethanesulfonic Acid, **Na HEPES**; 2-(4-(2-Hydroxyethyl)-1-piperazinyl)ethanesulfonic Acid Sodium Salt, **MES**; 2-(N-morpholino)ethanesulfonic acid, **MPD;** 2,4-methyl pentanediol, **PEG**; Polyethylene glycol (2K, 6K , 8K and 10K correspond to the molecular weight, in thousands of Daltons, of PEG), **TMAO:** Trimethylamine N-oxide, **Tris**; 2-Amino-2-(hydroxymethyl)propane-1,3-diol.

N.B. Jeffamine ED-2001 has been superseded with Jeffamine ED-2003. Polyvinylpyrolidone K15 is called Polyvinylpyrolidine.

Molecular Dimensions

Soluble Proteins

PACT *premier* HT-96 Conditions A1 – D12 MD1-36

Well #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
A1			0.1 M	SPG	4.0	25 % w/v	PEG 1500
A2			0.1 M	SPG	5.0	25 % w/v	PEG 1500
A3			0.1 M	SPG	6.0	25 % w/v	PEG 1500
A4			0.1 M	SPG	7.0	25 % w/v	PEG 1500
A5			0.1 M	SPG	8.0	25 % w/v	PEG 1500
A6			0.1 M	SPG	9.0	25 % w/v	PEG 1500
A7	0.2 M	Sodium chloride	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A8	0.2 M	Ammonium chloride	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A9	0.2 M	Lithium chloride	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A10	0.2 M	Magnesium chloride hexahydrate	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A11	0.2 M	Calcium chloride dihydrate	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A12	0.01 M	Zinc chloride	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
B1			0.1 M	MIB	4.0	25 % w/v	PEG 1500
B2			0.1 M	MIB	5.0	25 % w/v	PEG 1500
B3			0.1 M	MIB	6.0	25 % w/v	PEG 1500
B4			0.1 M	MIB	7.0	25 % w/v	PEG 1500
B5			0.1 M	MIB	8.0	25 % w/v	PEG 1500
B6			0.1 M	MIB	9.0	25 % w/v	PEG 1500
B7	0.2 M	Sodium chloride	0.1 M	MES	6.0	20 % w/v	PEG 6000
B8	0.2 M	Ammonium chloride	0.1 M	MES	6.0	20 % w/v	PEG 6000
B9	0.2 M	Lithium chloride	0.1 M	MES	6.0	20 % w/v	PEG 6000
B10	0.2 M	Magnesium chloride hexahydrate	0.1 M	MES	6.0	20 % w/v	PEG 6000
B11	0.2 M	Calcium chloride dihydrate	0.1 M	MES	6.0	20 % w/v	PEG 6000
B12	0.01 M	Zinc chloride	0.1 M	MES	6.0	20 % w/v	PEG 6000
C1			0.1 M	PCTP	4.0	25 % w/v	PEG 1500
C2			0.1 M	PCTP	5.0	25 % w/v	PEG 1500
C3			0.1 M	PCTP	6.0	25 % w/v	PEG 1500
C4			0.1 M	PCTP	7.0	25 % w/v	PEG 1500
C5			0.1 M	PCTP	8.0	25 % w/v	PEG 1500
C6			0.1 M	PCTP	9.0	25 % w/v	PEG 1500
C7	0.2 M	Sodium chloride	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C8	0.2 M	Ammonium chloride	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C9	0.2 M	Lithium chloride	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C10	0.2 M	Magnesium chloride hexahydrate	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C11	0.2 M	Calcium chloride hexahydrate	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C12	0.01 M	Zinc chloride	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
D1			0.1 M	MMT	4.0	25 % w/v	PEG 1500
D2			0.1 M	MMT	5.0	25 % w/v	PEG 1500
D3			0.1 M	MMT	6.0	25 % w/v	PEG 1500
D4			0.1 M	MMT	7.0	25 % w/v	PEG 1500
D5			0.1 M	MMT	8.0	25 % w/v	PEG 1500
D6			0.1 M	MMT	9.0	25 % w/v	PEG 1500
D7	0.2 M	Sodium chloride	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D8	0.2 M	Ammonium chloride	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D9	0.2 M	Lithium chloride	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D10	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D11	0.2 M	Calcium chloride dihydrate	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D12	0.002 M	Zinc chloride	0.1 M	Tris	8.0	20 % w/v	PEG 6000

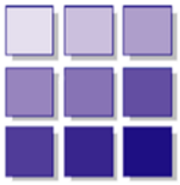
Molecular Dimensions

Soluble Proteins

PACT *premier* HT-96 **Conditions E1 – H12** **MD1-36**

Well #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
E1	0.2 M	Sodium fluoride				20 % w/v	PEG 3350
E2	0.2 M	Sodium bromide				20 % w/v	PEG 3350
E3	0.2 M	Sodium iodide				20 % w/v	PEG 3350
E4	0.2 M	Potassium thiocyanate				20 % w/v	PEG 3350
E5	0.2 M	Sodium nitrate				20 % w/v	PEG 3350
E6	0.2 M	Sodium formate				20 % w/v	PEG 3350
E7	0.2 M	Sodium acetate trihydrate				20 % w/v	PEG 3350
E8	0.2 M	Sodium sulfate				20 % w/v	PEG 3350
E9	0.2 M	Potassium sodium tartrate tetrahydrate				20 % w/v	PEG 3350
E10	0.02 M	Sodium/potassium phosphate				20 % w/v	PEG 3350
E11	0.2 M	Sodium citrate tribasic dihydrate				20 % w/v	PEG 3350
E12	0.2 M	Sodium malonate dibasic monohydrate				20 % w/v	PEG 3350
F1	0.2 M	Sodium fluoride	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F2	0.2 M	Sodium bromide	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F3	0.2 M	Sodium iodide	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F4	0.2 M	Potassium thiocyanate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F5	0.2 M	Sodium nitrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F6	0.2 M	Sodium formate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F7	0.2 M	Sodium acetate trihydrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F8	0.2 M	Sodium sulfate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F9	0.2 M	Potassium sodium tartrate tetrahydrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F10	0.02 M	Sodium/potassium phosphate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F11	0.2 M	Sodium citrate tribasic dihydrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F12	0.2 M	Sodium malonate dibasic monohydrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
G1	0.2 M	Sodium fluoride	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G2	0.2 M	Sodium bromide	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G3	0.2 M	Sodium iodide	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G4	0.2 M	Potassium thiocyanate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G5	0.2 M	Sodium nitrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G6	0.2 M	Sodium formate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G7	0.2 M	Sodium acetate trihydrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G8	0.2 M	Sodium sulfate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G9	0.2 M	Potassium sodium tartrate tetrahydrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G10	0.02 M	Sodium/potassium phosphate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G11	0.2 M	Sodium citrate tribasic dihydrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G12	0.2 M	Sodium malonate dibasic monohydrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
H1	0.2 M	Sodium fluoride	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H2	0.2 M	Sodium bromide	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H3	0.2 M	Sodium iodide	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H4	0.2 M	Potassium thiocyanate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H5	0.2 M	Sodium nitrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H6	0.2 M	Sodium formate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H7	0.2 M	Sodium acetate trihydrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H8	0.2 M	Sodium sulfate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H9	0.2 M	Potassium sodium tartrate tetrahydrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H10	0.02 M	Sodium/potassium phosphate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H11	0.2 M	Sodium citrate tribasic dihydrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H12	0.2 M	Sodium malonate dibasic monohydrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350

4

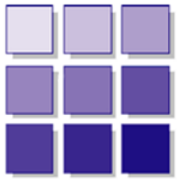


moleculardimensions.com

Targeted Sparse Matrix

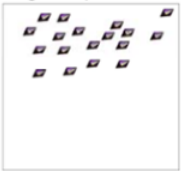
ProPlex Screen HT-96 **Rows A - D** **MD1-42**

HT-96 Well	Salt	Buffer	pH	Precipitant	Additive
A1	None	0.1 M Tris	8.0	25 % v/v PEG 350 MME	None
A2	0.1 M calcium acetate	0.1 M MES	6.0	15 % v/v PEG 400	None
A3	0.1 M lithium Chloride	0.1 M Na HEPES	7.5	20 % v/v PEG 400	None
A4	None	0.1 M Tris	8.0	25 % v/v PEG 400	None
A5	None	0.1 M MES	6.5	15 % v/v PEG 550 MME	None
A6	0.2 M sodium chloride	0.1 M Na/K phosphate	6.5	25 % w/v PEG 1000	None
A7	0.1 M ammonium sulfate	0.1 M Tris	7.5	20 % w/v PEG 1500	None
A8	0.2 M ammonium sulfate	0.1 M sodium acetate	5.5	10 % w/v PEG 2000 MME	None
A9	0.2 M sodium chloride	0.1 M MES	6.0	20 % w/v PEG 2000 MME	None
A10	0.1 M potassium chloride	0.1 M Tris	8.0	15 % w/v PEG 2000 MME	None
A11	None	0.1 M Na HEPES	7.5	25 % w/v PEG 2000 MME	None
A12	0.2 M sodium acetate	0.1 M sodium citrate	5.5	5 % w/v PEG 4000	None
B1	0.2 M lithium sulfate	0.1 M Tris	7.5	5 % w/v PEG 4000	None
B2	0.1 M calcium acetate	0.1 M sodium acetate	4.5	10 % w/v PEG 4000	None
B3	0.2 M sodium acetate	0.1 M sodium citrate	5.5	10 % w/v PEG 4000	None
B4	0.2 M sodium chloride	0.1 M MES	6.5	10 % w/v PEG 4000	None
B5	0.1 M magnesium chloride	0.1 M Na HEPES	7.5	10 % w/v PEG 4000	None
B6	None	0.1 M Na HEPES	7.0	10 % w/v PEG 4000	10 % v/v 2-propanol
B7	0.2 M ammonium acetate	0.1 M sodium acetate	4.0	15 % w/v PEG 4000	None
B8	0.1 M magnesium chloride	0.1 M sodium citrate	5.0	15 % w/v PEG 4000	None
B9	None	0.1 M sodium cacodylate	6.0	15 % w/v PEG 4000	None
B10	0.15 M ammonium sulfate	0.1 M MES	6.0	15 % w/v PEG 4000	None
B11	None	0.1 M Na HEPES	7.0	15 % w/v PEG 4000	None
B12	0.1 M magnesium chloride	0.1 M Na HEPES	7.0	15 % w/v PEG 4000	None
C1	0.15 M ammonium sulfate	0.1 M Tris	8.0	15 % w/v PEG 4000	None
C2	None	0.1 M sodium citrate	4.5	20 % w/v PEG 4000	None
C3	0.2 M ammonium acetate	0.1 M sodium acetate	5.0	20 % w/v PEG 4000	None
C4	0.2 M lithium sulfate	0.1 M MES	6.0	20 % w/v PEG 4000	None
C5	None	0.1 M Tris	8.0	20 % w/v PEG 4000	None
C6	0.15 M ammonium sulfate	0.1 M Na HEPES	7.0	20 % w/v PEG 4000	None
C7	None	0.1 M sodium citrate	5.6	20 % w/v PEG 4000	20 % v/v 2-propanol
C8	0.2 M sodium chloride	0.1 M Tris	8.0	20 % w/v PEG 4000	None
C9	None	0.1 M sodium cacodylate	5.5	25 % w/v PEG 4000	None
C10	0.15 M ammonium sulfate	0.1 M MES	5.5	25 % w/v PEG 4000	None
C11	None	0.1 M sodium cacodylate	6.5	25 % w/v PEG 4000	None
C12	0.2 M potassium iodide	0.1 M MES	6.5	25 % w/v PEG 4000	None
D1	0.2 M sodium chloride	0.1 M Na HEPES	7.5	25 % w/v PEG 4000	None
D2	None	0.1 M MES	6.5	10 % w/v PEG 5000 MME	12 % v/v 1-propanol
D3	0.1 M potassium chloride	0.1 M Na HEPES	7.0	15 % w/v PEG 5000 MME	None
D4	0.2 M ammonium sulfate	0.1 M Tris	7.5	20 % w/v PEG 5000 MME	None
D5	0.1 M magnesium chloride	0.1 M MES	6.0	8 % w/v PEG 6000	None
D6	0.15 M sodium chloride	0.1 M Tris	8.0	8 % w/v PEG 6000	None
D7	None	0.1 M sodium citrate	5.5	15 % w/v PEG 6000	None
D8	0.1 M magnesium acetate	0.1 M sodium cacodylate	6.5	15 % w/v PEG 6000	None
D9	None	0.1 M MES	6.5	15 % w/v PEG 6000	5 % v/v MPD
D10	0.1 M potassium chloride	0.1 M Na HEPES	7.5	15 % w/v PEG 6000	None
D11	None	0.1 M Tris	8.5	15 % w/v PEG 6000	None
D12	None	0.1 M Tris	8.5	20 % w/v PEG 6000	None



moleculardimensions.com

Targeted Sparse Matrix



ProPlex Screen HT-96 Rows E - H MD1-42

HT-96 Well	Salt	Buffer	pH	Precipitant	Additive
E1	0.1 M magnesium acetate	0.1 M sodium acetate	4.5	8 % w/v PEG 8000	None
E2	None	0.1 M sodium citrate	5.0	8 % w/v PEG 8000	None
E3	0.2 M sodium chloride	0.1 M sodium cacodylate	6.0	8 % w/v PEG 8000	None
E4	None	0.1 M Na HEPES	7.0	8 % w/v PEG 8000	None
E5	None	0.1 M Tris	8.0	8 % w/v PEG 8000	None
E6	0.1 M calcium acetate	0.1 M sodium cacodylate	5.5	12 % w/v PEG 8000	None
E7	None	0.1 M sodium phosphate	6.5	12 % w/v PEG 8000	None
E8	0.1 M magnesium acetate	0.1 M MOPS	7.5	12 % w/v PEG 8000	None
E9	0.2 M sodium chloride	0.1 M Na HEPES	7.5	12 % w/v PEG 8000	None
E10	0.2 M ammonium sulfate	0.1 M Tris	8.5	12 % w/v PEG 8000	None
E11	None	0.1 M sodium citrate	5.0	20 % w/v PEG 8000	None
E12	0.2 M ammonium sulfate	0.1 M MES	6.5	20 % w/v PEG 8000	None
F1	None	0.1 M Na HEPES	7.0	20 % w/v PEG 8000	None
F2	0.2 M lithium chloride	0.1 M Tris	8.0	20 % w/v PEG 8000	None
F3	0.1 M magnesium acetate	0.1 M MES	6.5	10 % w/v PEG 10 000	None
F4	None	0.1 M Na HEPES	7.0	18 % w/v PEG 12 000	None
F5	0.1 M sodium chloride	0.1 M Tris	8.0	8 % w/v PEG 20 000	None
F6	None	0.1 M Na HEPES	7.0	15 % w/v PEG 20 000	None
F7	None	0.1 M MES	6.5	0.5 M ammonium sulfate	None
F8	None	0.1 M sodium acetate	5.0	1 M ammonium sulfate	None
F9	None	0.1 M MES	6.5	1 M ammonium sulfate	None
F10	None	0.1 M Tris	8.0	1 M ammonium sulfate	None
F11	None	0.1 M sodium acetate	5.0	1.5 M ammonium sulfate	None
F12	None	0.1 M Na HEPES	7.0	1.5 M ammonium sulfate	None
G1	None	0.1 M Tris	8.0	1.5 M ammonium sulfate	None
G2	None	0.1 M sodium acetate	5.0	2 M ammonium sulfate	None
G3	None	0.1 M Na HEPES	7.0	2 M ammonium sulfate	None
G4	None	0.1 M Tris	8.0	2 M ammonium sulfate	None
G5	1 M potassium chloride	0.1 M Na HEPES	7.0	1 M ammonium sulfate	None
G6	None	0.1 M sodium acetate	5.0	2 M sodium formate	None
G7	None	0.1 M Tris	7.5	3 M sodium formate	None
G8	None	None	7.5	0.8 M Na/K hydrogen phosphate	None
G9	None	None	7.0	1.3 M Na/K hydrogen phosphate	None
G10	None	None	6.5	1.6 M Na/K hydrogen phosphate	None
G11	None	0.1 M Na HEPES	7.5	1 M sodium acetate	None
G12	None	0.1 M Na HEPES	7.0	1 M sodium citrate	None
H1	None	0.1 M sodium citrate	6.0	2 M sodium chloride	None
H2	None	0.1 M MES	6.5	1 M lithium sulfate	None
H3	None	0.1 M Tris	8.0	1.6 M lithium sulfate	None
H4	None	None	6.0	1.4 M sodium malonate	None
H5	None	0.1 M Tris	8.0	1.2 M Na/K tartrate	None
H6	None	0.1 M MES	6.5	1.6 M magnesium sulfate	None
H7	None	0.1 M sodium acetate	5.0	15 % v/v MPD	2 % w/v PEG 4000
H8	0.05 M calcium acetate	0.1 M sodium cacodylate	6.0	25 % v/v MPD	None
H9	None	0.1 M imidazole	7.0	50 % v/v MPD	None
H10	0.05 M magnesium Chloride	0.1 M MES	6.5	10 % v/v 2-propanol	5 % w/v PEG 4000
H11	0.2 M ammonium acetate	0.1 M Na HEPES	7.5	25 % v/v 2-propanol	None
H12	0.1 M sodium chloride	0.1 M Tris	8.0	15 % v/v ethanol	5 % v/v MPD

Abbreviations:
HEPES; N-(2-hydroxyethyl)-piperazine-N'-2-ethanesulfonic acid, **MES**; 2-(N-morpholino)ethanesulfonic acid, **MME**; Monomethylether, **PEG**; Polyethylene glycol, **Tris**; 2-Amino-2-(hydroxymethyl)propane-1,3-diol, **MOPS;** 3-(N-Morpholino)-propanesulfonic acid;