Crystallographic studies on the bacterial sialidases from *Clostridium perfringens* and *Micromonospora viridifaciens*.

Simon Newstead

A thesis submitted for the degree of
Doctor of Philosophy

University of St Andrews

November 2004

Declaration

I, Simon Louis Newstead certify that the thesis, which is approximately 49,300 words and which has been written by me, is a record of work carried out by myself and that it has not been submitted in any previous application for a degree.

Date 8/12/2004. Signature of candidate

I was admitted as a research student to St Andrews University in October of 2001 and as a candidate for the degree of Ph.D in September of 2002.

Date 8/12/04. Signature of candidate

I herby certify that the candidate has fulfilled the conditions of the resolution and regulations appropriate for the degree of Ph.D in the university of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 8/12/4 Signature of supervisor

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the university library for the time being in force, subject to any copyright vested in the work not being directly affected thereby. I also understand that the title and abstract will be published and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date 8/12/04. Signature of candidate

Abstract.

Sialidases catalyse the removal of terminal sialic acids from a range of glycoproteins, glycolipids and oligosaccharides. They have been found in bacteria, viruses, and parasites, where they play important roles in pathogenesis and/or microbial nutrition, and in mammalian cells, where they modulate cell-surface glycosylation associated with a range of cellular activities. *Clostridium pefringens*, a causative agent of gas gangrene and peritonitis in humans, possess three sialidases: nanH, nanI, and nanJ, with molecular weights of 42, 77 and 129 kDa, respectively. The two larger enzymes are secreted by the bacterium and are involved in the pathogenesis and nutrition of *Clostridium*. As part of a study to examine the structural and mechanistic features of these enzymes, the crystallisation of the of the 77 kDa *nanI* isoenzyme was attempted. The expressed full-length protein was found to degrade easily; a stable 50 kDa fragment of the enzyme was therefore subcloned, from amino acids 243-694, which retained full catalytic activity. Crystals of this fragment were produced and diffraction data collected to 0.97 Å resolution for the apo and sialic acid bound complex. Using a novel di-fluoro sialic acid derivative a covalent sialyl-enzyme intermediate was also trapped, the first in a bacterial sialidase, and diffraction data collected to 1.6 Å on our in house generator. The covalent complex supported earlier structural studies on the *T.cruzi* enzyme, where the conserved catalytic tyrosine was identified as the catalytic nucleophile in the hydrolysis reaction. The 50 kDa fragment of the nanI sialidase was found to adopt the canonical six-bladed β-propeller fold observed in all previous sialidase structures reported to date. The active site is located in the same position as that observed previously in the viral sialidase structures, formed by a small cavity of the top face of the β-propeller. In addition the structures of three active site mutants of the sialidase from *M.viridifaciens* were solved in complex with both a transition state analogue inhibitor and sialic acid. Their structures are discussed with respect to the results of kinetic and biochemical studies carried out on these mutants.

Acknowledgements

I would like to thank my supervisor, Professor Garry Taylor, for all his help and tutoring over the past three years and for introducing me to protein crystallography. It has been a very rewarding experience.

I would like to thank my parents and family, without whose support this degree would have been much harder to achieve. Thank you.

My thanks also go to my friends and colleagues in St Andrews who have made the last three years a joy to live and whose friendship I hope to keep for many more years to come.

Professor Andrew Bennet and Dr. Jacki Watson at the Simon Fraser University in British Columbia, Canada provided the *M.viridifaciens* mutants and compounds for this project.

The α2,3 di-fluoro sialic acid compound was kindly provided by Professor Stephan Withers and Dr. Andrew Watts at the University of British Columbia, BC, Canada.

This project was funded through a BBSRC studentship.

Abbreviations

Amino Acids

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartate	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamate	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

CCP4	Collaborative Computing Project Number 4 in Protein Crystallography
CC_F	The correlation coefficient between the observed amplitudes for the data and the calculated amplitudes for model.
CC_I	The correlation coefficient between the observed intensities from the data and the sum of the calculated intensities for all symmetry equivalents of the model.
CC_P	The Patterson correlation coefficient between the data and the model Patterson's evaluated within the defined sphere centred on the Patterson origin.
CDC	Centres for Disease Control and Prevention
CNS	Crystallography and NMR System
C.perfringens	*Clostridium perfringens*
DANA	2-deoxy-2,3-didehydro-5-N-acetylneuraminic acid
$F_o - F_c$	Maximum likelihood / σ_A-weighted $F_o - F_c$ electron density map
$2F_o - F_c$	Maximum likelihood / σ_A-weighted $2F_o - F_c$ electron density map

FOM	Figure of Merit
Gal	Galactose
GalNAc	N-actetylgalactosamine
GlcNAc	N-acetylglucosamine
HA	Haemagglutinin surface protein
HN	Haemagglutinin Sialidase (Neuraminidase)
IPTG	Isopropyl-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
Lac	Lactose
M.decora	*Macrobdella decora*
M.viridifaciens	*Micromonospora viridifaciens*
MVNA	*Micromonospora viridifaciens* Sialidase (Neuraminidase)
MR	Molecular Replacement
Neu5Ac	N-acetylneuraminic acid
Neu5Ac2en	DANA; 2-deoxy-2,3-didehydro-5-N-acetylneuraminic acid
NMR	Nuclear Magnetic Resonance Spectroscopy
PAGE	Polyacrylamide Gel Electrophoresis
PEG	Polyethylene Glycol
PDB	Protein Data Bank depository
RF	The correlation coefficient between squared normalised structure factors.
RF_F	The classic R factor between the observed amplitudes for the data and the calculated amplitudes for the model.
rmsd	Root mean square deviation
SDS	Sodium Dodecyl Sulphate
S.typhimurium	*Salmonella typhimurium*
T.cruzi	*Trypanosoma cruzi*
T.rangeli	*Tyrpanosoma rangeli*
V.cholerae	*Vibrio cholerae*
Wt	Wild type enzyme/protein

Contents

List of figures

showing lactose bound in the C-terminal CBM domain.

List of tables

Chapter 1

Introduction.

Structure and Function of Sialidases.

1.1. Overview.

Sialidases (N-acetylneuraminosyl glycohydrolsaes, neuraminidases, EC 3.2.1.18) belong to a family of exoglycosidases that catalyse the cleavage of non-reducing sialic acid residues linked to mono- or oligosaccharide chains of glycoconjugates (Siaito and Yu, 1995). They are widely distributed in nature, and have been found in viruses, bacteria, fungi, mycoplasma and protazoa as well as avian and mammalian species (Achyuthan and Achyuthan, 2001; Corfield *et al* ., 1981; Rosenberg and Schengrund, 1976).

Among the various sialidases, those from viral and bacterial species have been studied most extensively, due to their role in the pathogenesis of a number of important human and animal diseases. The most important of these diseases in humans is that caused by the influenza A and B viruses, which account for an average of about 36, 000 deaths in the USA each year due to complications arsing from the infection (source CDC; www.cdc.gov). The influenza type A and B viruses have two membrane-bound glycoproteins; haemagglutinin (HA), which recognises the carbohydrate sialic acid for host cell attachment and fusion, and a sialidase (NA) (Siaito and Yu, 1995).

The main role of the sialidase is to process progeny virus particles when they bud from an infected cell; it specifically cleaves the glycosidic bond, usually an α2,3 or α2,6 linkage, from terminal sialic acid to galactose or galactosamine, halting self-agglutination and enhancing infectivity. A secondary role is to process the mucin layer of the upper respiratory tract to aid viral entry into the cell (Matrosovich *et al* ., 2004). The most devastating outbreak of infectious disease in recorded history was the 'Spanish' influenza pandemic of 1918-19. The enhanced virulence was caused by a mutation in the Haemagglutinin (HA) surface protein of the influenza A virus. This mutation allowed attachment of the

virus onto sialic acid linked α2,3 to a sub-terminal galactose sugar, a linkage common to the respiratory tracts of humans and higher primates (Kobasa *et al* ., 2004). The disease killed approximately 20 million people (Johnson and Mueller, 2002).

The link between virulence and sialidase activity by the NA protein led to a structure-based drug design initiative based on the three-dimensional crystal structure of the tetrameric influenza A and B virus sialidases (Burmeister *et al* ., 1993; Varghese and Colman, 1991). Using the structure of the enzyme bound to the sialic acid derivative and micro-molar inhibitor 2-deoxy-2,3-dehydro-N-acetylneuraminic acid (Neu5Ac2en; DANA), two highly potent influenza sialidase inhibitors were designed (von Itzstein *et al* ., 1993), with K_i's in the sub nano-molar range, marketed as Relenza and Tamiflu. This project is still one of the most successful rational drug design initiatives to date and was pioneering in using the three-dimensional crystal structure of a target enzyme and computational studies to design an effective prophylactic enzyme inhibitor.

The successful study of the sialidases from the influenza viruses opened the way for a wave of interest in using structural biology to understand the function and mechanism of action of these enzymes in micoorganisms, such as bacteria, fungi and parasites, reviewed in (Taylor, 1996; Taylor *et al* ., 1999). Many pathogenic trypanosome species, such as *T.cruzi* and *T.brucei* possess a unique trans-sialidase on their surfaces. In contrast with sialidases, which are strict hydrolyses, trans-sialidases preferentially catalyse the transfer of sialic acids from sialoglyconjugates to β-galactosyl acceptor sugars (Ferrero-Garcia *et al* ., 1993). The activity of the trans-sialidase is vital for the pathogenesis of the trypanosome species (Schenkman *et al* ., 1991), and thus represents another potential sialidase drug target. The recent determination of the three-dimensional crystal structure (Buschiazzo *et al* ., 2002) and insightful ligand studies (Amaya *et al* ., 2004), have now paved the way for the rational design of inhibitors for this enzyme.

1.2. Bacterial sialidases: roles in pathogenicity and nutrition.

Sialidases are found in a variety of pathogenic and non-pathogenic bacterial strains, including *Clostridium, Vibrio, Corynebacterium, Bacteroides, Salmonella, Streptococcus and Streptomyces* species (Corfield, 1992). Many of these bacteria can use sialic acid as a carbon and energy source, possessing both the permeases to transport the sugar inside the cell, and a cascade of lyases for its catabolism (Corfield, 1992; Vimr and Troy, 1985). The terminal non-reducing position of sialic acids in sialo-glyconjugates demands the removal of this monosaccharide before stepwise degradation of the remaining oligosaccharide chain can occur.

Many of these bacterial species however, have utilised the function of sialidases to enhance their pathogenic capability. Many of the most devastating bacterial infections in humans are characterised by the release of highly active sialidase protein, including septicaemia, gas gangrene and bacterial meningitis (Corfield, 1992). Table 1.1 lists a number of bacterial diseases were sialidase secretion is a hallmark of the infection process. The results of sialidase action on cellular and soluble glycoconjugate substrates throughout the body can be dramatic, leading to both general and specific functional changes, including haemolytic anaemia, erythrocyte panagglutination and the onset of autoimmune diseases. These severe changes are a result of the important function of sialic acid and its derivatives in the biochemistry of higher eukaryotes (Rosenberg, 1995).

The specific role of these enzymes in the pathogenicity of these bacteria, however, is still not fully understood. In certain cases, the enzyme has a defined role in the molecular basis of pathogenesis. For example, during infection by *V.cholerae,* the effects of the cholera toxin (CT) is greatly enhanced by the action of the secreted sialidase, which acts to cleave the terminal sialic acid sugars 'masking' the toxin receptors on the surface of the host cells (Dirita, 2001). In the majority of cases, however, a specific role has still to be found. That sialidases are important components of the virulence factors associated with bacterial

infection is not in dispute. The enzyme has been shown to exert a direct toxic effect on mammalian host tissue (Milligan *et al* ., 1977) and interfere with host immunologic and other defence mechanisms, such as loss of mucus viscoelasticity in the respiratory and digestive tracts (Klein *et al* ., 1977; Puchelle *et al* ., 1975).

Further evidence that sialidases are important components of the infection process come from a multitude of biochemical, bioinformatic and *in vivo* experiments. It is apparent from these studies that a number of typical properties can be identified that apply to most cases and, which can be used as a general indication of sialidase association with the infectious capability of the bacteria listed in Table 1.1. These are the high specific activity of the seceted enzymes, in some instances as much as 600 U mg^{-1} of protein (Roggentin and Schauer, 1987), large quantities of the enzyme are produced and secreted and the substrate specificity often reflects the site of colonisation (Demuth *et al* ., 1990). In *V.cholerae*, the major virulence genes required for pathogenesis are in clusters, termed vibrio pathogenicity islands (VPI's) in the genome, and can apparently propagate laterally and disperse among difference strains. The VPI-2 gene cluster encodes several genes for amino acid catabolism and the sialidase enzyme (Jermyn and Boyd, 2002).

The two main functions of bacterial sialidases, nutrition and pathogenesis, are in many ways intimately interlinked. Removing the outer sialic acid coating on mammalian cell surfaces has the added benefit of exposing both hidden binding sites for cell attachment and invasion, and allowing access for the other glcyosidases and lipases that are secreted. Whether the sialidase is utilised in a specific pathogenic role, such as in *V.cholerae* or not, for example in *C.perfringens*, its effect on the hosts biochemistry will still be disruptive to the point of being adventitious to the bacterium.

Table 1.1. Sialidase producing bacteria in disease.
Taken from (Corfield, 1992).

Disease	Bacteria	Site of infection	Secreted into blood stream?
Gas gangrene	Clostridia	Wounds	Yes
Septicaemia	Streptococcus Pneumococcus Bacteroides Corynebacteria	Various	Yes
Pneumonia	Streptococcus	Respiratory tract	Yes
Peritonitis	Clostridia Bacteroides Enterococcus Escherichia	Peritoneum	Yes
Meningitis	Streptococcus Group B Type III	Brain	Yes
Haemolytic-uraemic syndrome	Pneumococcus	Kidney	Yes
Glomerulonephritis	Streptococcus Type A		
Periodontal disease	Streptococcus Haemophilus Actinomyces	Oral cavity	No
Otis media	Streptococcus	Middle ear	No
Cholera	Vibrio cholerae	Gastrointestinal tract	No

1.3. Sialidase structures: the sialidase superfamily.

The crystal structure of the influenza virus sialidase was solved in 1983 (Varghese *et al* ., 1983). This structure revealed a tetrameric association of identical monomers whose fold has been described as a superbarrel or β-propeller. Each monomer is made from six four-stranded antiparallel β-sheets arranged as the blades of a propeller around a pseudo sixfold axis (Figure 1.3.1). Structural studies on sialidases from all organisms determined to date have shown that this fold in highly conserved, even at sequence identities < 30 %, and forms the catalytic domain of each enzyme (Taylor *et al* ., 1999). Table 1.2. lists a representative set of all currently published sialidase structures, with a representative PDB accession code.

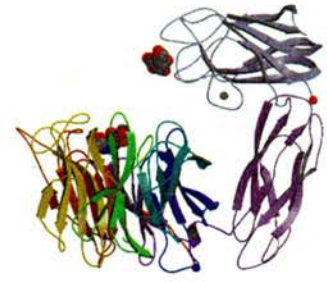
Table 1.2. The current list of sialidase/trans-sialidase structures determined from different organisms.

Organism:	PDB code:	Resolution (Å):	Reference:
Influenza A (subtype N2)	2QWA	1.70	(Varghese *et al* ., 1998)
Influenza B	1NSC	1.80	(Burmeister *et al* ., 1993)
V.cholerae	1KIT	2.30	(Crennell *et al* ., 1994)
S.typhimurium	2SIL	1.60	(Crennell *et al* ., 1993)
M.viridifaciens	1EUU	2.50	(Gaskell *et al* ., 1995)
T.rangeli	1N1S	1.64	(Amaya *et al* ., 2003)
T.cruzi (TS)*	1S0K	1.60	(Buschiazzo *et al* ., 2002)
Newcastle disease virus (HN)	1E8T	2.50	(Crennell *et al* ., 2000)
M.decora	1SLL	2.00	(Lou *et al* ., 1998a)
Human Parainfluenza virus type III (HN)	1V2I	2.20	(Lawrence *et al* ., 2004)
H.sapiens (Neu2)	1SNT	1.75	(Chavas *et al* ., 2004)
Bacteriphage endosialidase	1V0E	Coordinates on hold	No reference for structure given.

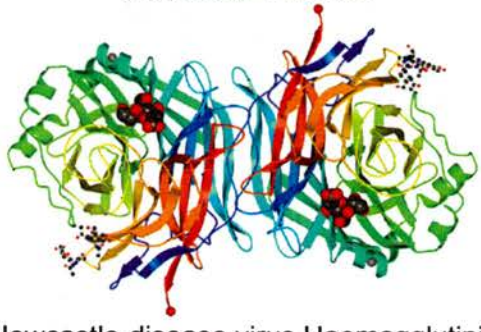
M.viridifaciens 'small' sialidase

S.typhimurium sialidase

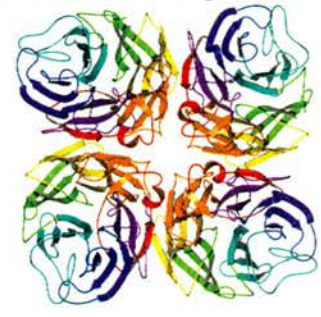
V.cholerae sialidase



M.viridifaciens 'large' sialidase



Newcastle-disease virus Haemagglutinin sialidase



Influenza virus sialidase

H.sapiens (Neu2) sialidase

T.cruzi trans-sialidase

Figure 1.3.1. The three-dimensional crystal structures of representative members of the sialidase superfamily. The canonical β-propeller fold can be seen in all the structures. The bound inhibitor DANA is shown as a space-filling model in some pictures, identifying the active site. In the *V.cholerae* structure, sialic acid can be seen bound to one of the lectin 'wings' of this enzyme.

Structural studies on the bacterial sialidases have revealed a superfamily of multi-domain enzymes built around the canonical catalytic β-propeller fold (Figure 1.3.1) (Taylor *et al* ., 1999). The resent structural studies published on the trypanosomal sialidases from *T.rangeli* (Amaya *et al* ., 2003) and the trans-sialidase from *T.cruzi* (Buschiazzo *et al* ., 2002) have confirmed that they too belong to this family of multi-domain sialidases (Figure 1.3.1).

The sequence identity between the influenza enzymes and the bacterial/trypansome enzymes is very low, ~15 %, and even between the non-viral enzymes the sequence identities are only ~ 30%. Nevertheless, the topology of the catalytic domain is conserved, and the active sites share many of the same features. These are discussed in chapters 4 and 5 of this thesis when discussing the active site of the newly solved structure of the *C.perfringens* nanI sialidase. Despite such low sequence identities, bioinformatic and structural studies on the early structures of the bacterial sialidases identified a number of sequence fingerprints. These have allowed both the identification and modelling of other non-viral sialidases are more bacterial genomes become available. The first is the RIP/RLP motif, which contains one of the conserved arginines that interact with the sialic acid carboxylate group. The second are the Asp-box motifs, that occur downstream of the RIP/RLP motif. The Asp-box (S/T-X-D-[X]-G-X-T/W/F) appears in topologically equivalent positions in the β-propeller of all currently solved non-viral sialidase structures, repeating three to five times downstream of the RIP/RLP motif (Taylor, 1996). Its biological function is unknown, but it allows the structural biologist to align newly identified sialidases with current structures to identify novel inserted domains in this superfamily (Figure 1.3.2).

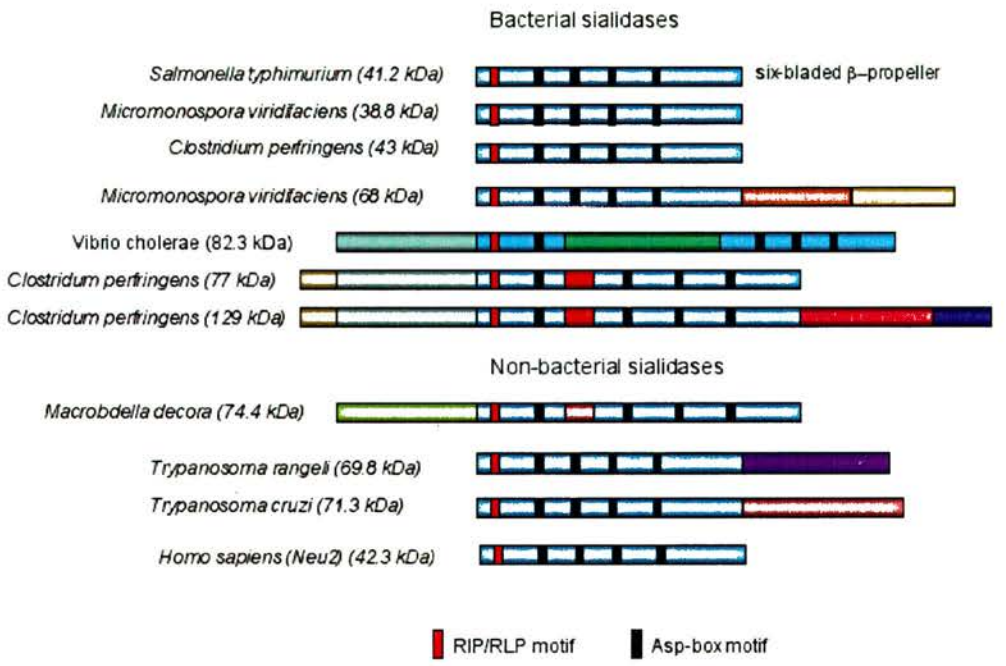


Figure 1.3.2. A schematic representation of the amino acid sequences of a representative set of non-viral sialidases. The figure shows how the primary structures can be aligned using the RIP/RLP and Asp-box motifs. The β-propeller domain is represented in marine blue and by the location of the sequence motifs. The additional domains present in these sialidases are shown in different colours, respectively.

One of the most striking features of the bacterial sialidases noted in the early biochemical work was the large size variation observed. The structural data now available have shown that this variation is due the addition of a variety of extra domains appended to the β-propeller structure. The small sialidases, for example, those from *S.typhimurium* and the small form of the *M.viridifaciens* enzyme are approximately 40 kDa, and consist of just the β-propeller domain. Whilst the large sialidases, for example the large form of the *M.viridifaciens* sialidase is 68 kDa and consists of three separate domains attached to the C-terminus of the β-propeller (Figure 1.3.1 and 1.3.2). The structure of the large 77

kDa sialidase, nanI, from *C.perfringens* is presented in chapters 5 and 6 of this thesis. Whilst the structure is missing the N-terminal domain, it does show a small β-barrel insertion within the β-propeller, reinforcing this common structural theme.

Where present, these additional domains have been shown to be involved in carbohydrate recognition, and many of them show the characteristic β-barrel / jelly-roll fold of plant lectins (Amaya *et al* ., 2003; Buschiazzo *et al* ., 2000; Crennell *et al* ., 1994; Gaskell *et al* ., 1995). A recent structural study on the *V.cholerae* sialidase has identified the target ligand for one of these extra domains as the substrate sugar, sialic acid (Moustafa *et al* ., 2004). The role of this domain has been ascribed to targeting the enzyme to the substrate and attachment to the cell surface of the small intestine. The function of the second domain of this enzyme has still to found, although it too is likely to be carbohydrate recognising domain.

In chapter 8 of this thesis, the structure of the large form of the *M.viridifaciens* sialidase is presented in complex with the sugar lactose. This structure is the first high-resolution structure showing the interaction of the C-terminal 'galactose-binding' domain with a disaccharide, and the complex is discussed in the relation to the function of the sialidase.

What is clear from these previous structural studies, is that the bacterial and trypanosome species have modulated the use of the 'basic' sialidase unit, the β-propeller, by inserting different domains that infer an advantage for the evolutionary niche inhabited by the organism.

1.4. Sialidase substrates: Sialic acid.

The term sialic acid, derived from the fact that it was first discovered in saliva, encompasses all derivatives of neuraminic acid that share a basic 9-carbon carboxylated skeleton, illustrated in Figure 1.4.1. The chemical backbone can be modified by many substitutions, making sialic acids the most structurally diverse of any naturally occurring monosaccharide (Rosenberg, 1995; Schauer, 2000). The most abundant sialic acid is 5-N-acetylneuraminic acid (Neu5Ac), which is shown in Figure 1.4.1. Together with 2-keto-3-deoxy-nonolusonic acid (KDN) these two sugars are believed to form the precursors of all naturally occurring sialic acids (Brooks *et al* ., 2002).

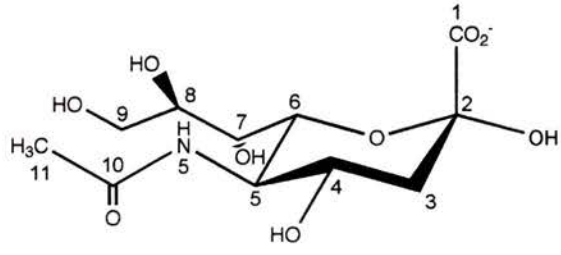


Figure 1.4.1. The structure of α-D-Neu5Ac. Sialic acids are based on a 9 carbon structure; the nine carbons atoms are labelled (1) to (9). The only difference between the structure of Neu5Ac and KDN is a substitution at the 5 carbon position to a hydroxyl group in the axial position in KDN.

Sialic acids are mainly found as the terminal components of cell surface glycoproteins and glycolipids, attached either α2,3 or α2,6 linked to D-Galactose or α2,6 linked to N-acetylglucosamine, α2,6 linked to N-acetylgalactosamine, or in some instances α2,8 linked to other sialic acids, in the form of polysialic acid (Brooks *et al* ., 2002; Rosenberg, 1995). Sialic acids are relatively large, hydrophilic and acidic molecules that exert direct physiochemical effects on the glycoconjugates to which they are bound. An example of this is the greater thermal stability and protection from proteolysis provided to a number of glycoconjugates in an *in vitro* study (Schauer, 1982).

Cell-type specific expression of sialyltransferases leads to specific sialylation patterns of oligosaccharides, which are key determinants of the cell biology and biochemistry of cells (Paulson *et al* ., 1989). Striking differences have been found in the sialoglycosylation patterns of cells during development, activation, aging and apoptosis, implicating sialic acids in a number of important biological processes. Aberrent sialylation in cancer cells is thought to be a characteristic feature associated with malignant properties, including invasiveness and metastatic potential; this area of sialic acid research has been recently reviewed in (Miyagi *et al* ., 2004).

The main biological functions of sialic acids can be summarised into four important areas; influencing conformation, stability, charge and viscosity, particularly of soluble glycoproteins; protection of cells and molecules from attack and degradation; modulation of cell adhesion, aggregation and agglutination; and finally as ligands for cell type specific receptors. These varied functions are reviewed in (Rosenberg, 1995; Schauer, 2000).

The ability of sialic acid to mediate such a diverse range of functions comes mainly from two features unique to this sugar. The first is the negative charge of all sialic acids. Under physiological conditions, they dissociate fully into their component ions and then act as anions, binding to organic cations. The pKa of Neu5Ac is 2. This is relevant in the binding and subsequent uptake of cations at the cell surface. Furthermore, the juxtaposition of negatively charged sialic acid residues on cell surfaces results in electrostatic repulsion, preventing cell-cell or membrane-membrane contact and/or adhesion (Varki, 1992).

The second feature is the location of the sugar at the terminal position of cell-surface glycoconjugates. This external position clearly implies a very significant role in cell biology, and this has been alluded to above. The ability of sialic acid to act as a molecular 'mask' is one that has received much attention in the literature (Schauer, 1985), and in vivo experiments have shown that desialylated erythrocytes are quickly cleared from the blood stream of rabbits (Jancik and Schauer, 1974).

In the context of this thesis, however, the role of sialylation as a protection against infection and cellular degradation is very relevant. Pathogenic bacteria have developed the ability to secrete large quantities of sialidases into the site of infection with the aim of severely crippling the immune system of the host and revealing the underlying glycoconjugate structures (Corfield, 1992). The importance of this activity during infection can be inferred by the evolutionary diversity seen within the bacterial sialidase structures discussed in section 1.3 above. The sialidases of bacteria and certain pathogenic trypanosome species have evolved to have very specific targeting 'domains' to enhance the sialidase activity at the correct time and place during the infection cycle.

1.5. Sialidase Mechanism: The eccentric glycosidase.

O-glycoside hydrolyses (EC 3.2.1-X) are a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. Glycosidases are amongst the most efficient enzymes being studied, catalysing rate enhancements up to 10^{17}-fold over the spontaneous hydrolysis of the glycosidic bond in water (Wolfenden *et al* ., 1998). This extraordinary efficiency, combined with the important roles played by glycosidases in biological systems, which range from the degradation of polysaccharides for food sources through to manipulation of the structures of glycoconjugates, have led to a concerted effort to understand these enzymes at both a mechanistic and structural level. A great deal of research has now been published on these enzymes, and a number of excellent general reviews are available (Davies and Henrissat, 1995; Davies *et al* ., 2003; Rye and Withers, 2000; Zechel and Withers, 2001). This section will focus on the current ideas proposed for the mechanism utilised by the sialidase superfamily of enzymes, which due to the unusual biophysical properties of sialic acid (see section 1.4) have evolved a unique method of cleaving the glycosidic bond in these substrates.

Members of the sialidase superfamily (EC 3.2.1.18) have been classified into three distinct families based on amino acid sequence similarities (Coutinho and Henrissat, 1999a; Coutinho and Henrissat, 1999b), these are: glycosidase family 33 (GH-33) that includes most bacterial and simple eukaryotic sialidases and trans-sialidases, GH-34 that includes the viral sialidases, and GH-83 that contain the haemagglutinin-sialidases. Despite low sequence similarities between members of this superfamily, all sialidases have a common active site with eight strictly conserved residues at the reactions centre (Taylor *et al* ., 1999). This observation strongly favours a common evolutionary origin for these enzymes and a similar catalytic mechanism.

Sialidases hydrolyse sialic acids from glycoconjugates with retention of configuration at the anomeric centre (Chong *et al* ., 1992a). The early mechanistic and structural studies on the influenza virus sialidase led the proposal that the mechanism involved distortion of the sialic acid towards an oxocarbonium ion transition-state (TS). Structures of the influenza sialidase complexed with the inhibitor Neu5Ac2en (DANA), led the theory that the positive charge on the TS could be stabilised by the strictly conserved tyrosine residue sitting directly beneath the C2 atom of this compound (Burmeister *et al* ., 1993). Complexes of bacterial and trypanosomal sialidases solved since then have seemed to confirm this idea, with the conserved tyrosine sitting close to the C2 atom of DANA, with a mean distance of ~ 2.5 Å (PDB 1EUS, 1KIT, 1MZ6) (Buschiazzo *et al* ., 2000; Taylor *et al* ., 1999). The proposed mechanism was one in which a conserved aspartic acid acted as a general acid catalyst, facilitating acid-catalysed cleavage of the glycosidic bond between the sialic acid and the sub-terminal aglycon. The subsequent oxocarbonium transition state adopts a half-chair configuration, similar to that seen in the Neu5Ac2en (DANA) compound, and is stabilised by interactions with the conserved tyrosine. In sialidases, a water molecule then acts as the catalytic nucleophile, attacking the positive charge on the C2 atom and facilitated by base catalysis from the conserved aspartic acid, which regains its proton (Burmeister *et al* ., 1993; Chong *et al* ., 1992a).

This mechanism is subtlety different from that proposed for other retaining glycosidases, which have been shown to operate via a covalent glycosyl-enzyme intermediate, involving the action of a protein carboxylate nucleophile and an acid/base catalyst (Rye and Withers, 2000). The structures of sialidases showed that no carboxylate group was favourably positioned to form a covalent sialyl-enzyme intermediate during catalysis. It was this observation that prompted much of the earlier researches to favour an ionic rather than covalent stabilisation of the reaction intermediate (Burmeister *et al* ., 1993; Chong *et al* ., 1992b; Lou *et al* ., 1998b).

Recent kinetic work carried out on the trans-sialidase from *T.cruzi* using a novel fluoro sugar derivatives of sialic acid led to the trapping of a catalytically competent sialyl-enzyme covalent intermediate on the conserved tyrosine residue (Watts *et al* ., 2003). This led to the proposal that the conserved catalytic tyrosine was in fact the nucleophile for this family of glycosidases. This elegant kinetic study was further supported by an extensive structural investigation, in which the structures of the Michaelis complex and the covalent sialyl-enzyme intermediate were solved (Amaya *et al* ., 2004). These structures, combined with the original DANA complex solved earlier (Buschiazzo *et al* ., 2002), have provided new and exciting information on a number of key features utilised by these enzymes.

Prominent among these are substrate distortion and electronic manipulation within the active site, with a proposed charge relay system between the conserved tyrosine and a glutamic acid that allows the former to act as the nucleophile (Amaya *et al* ., 2004). In chapter 5 of this thesis, the first structure of the covalent sialyl-enzyme intermediate in a bacterial sialidase is presented, along with the first complex with the substrate, sialic acid (Neu5Ac) at atomic resolution. These complexes are discussed in relation to this new information on the mechanism of sialidases. Figure 1.5.1. shows the current model for catalysis of glycosidic bond cleavage in the sialidases and trans-sialidases.

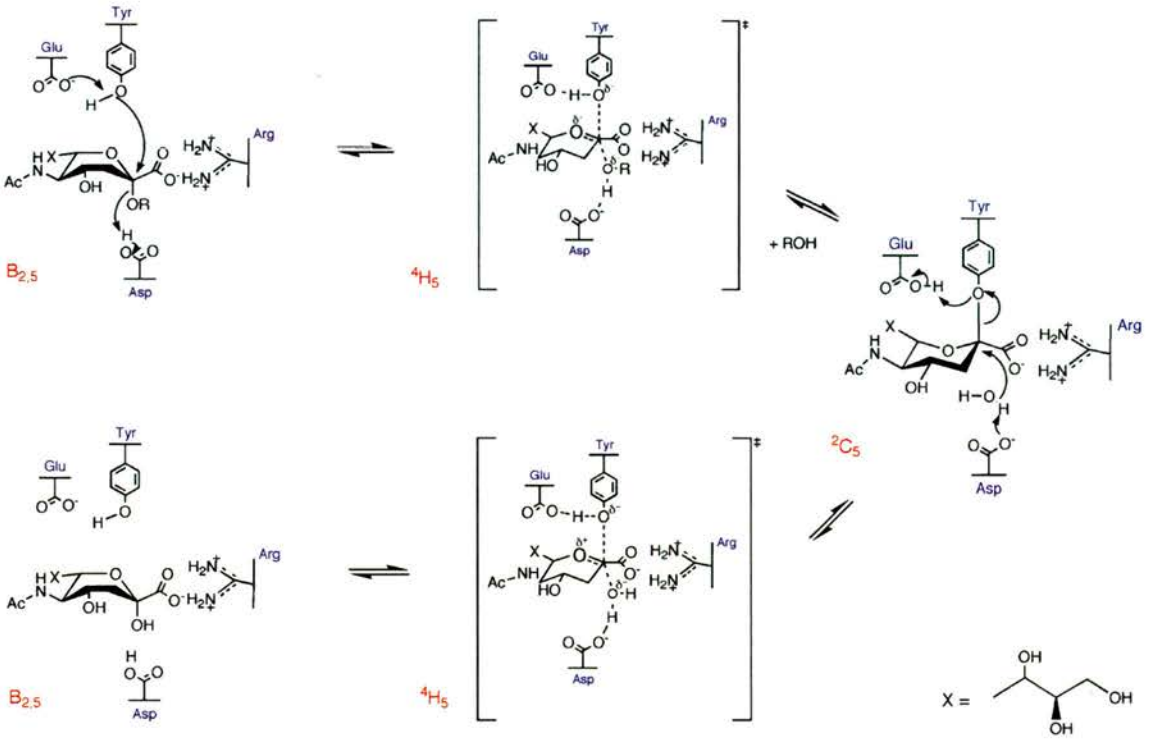


Figure 1.5.1. The currently proposed mechanism of glycosidic bond cleavage catalysed by the sialidase superfamily of glycosidases (GH33, 34 and 83). In the mechanism water is shown attacking the C2 atom of the TS, in the trans-sialdase mechanism this would be the 3-hydroxyl of an acceptor sugar (Amaya *et al* ., 2004). The acid/base catalyst (Asp), nucleophile couple (Glu and Tyr) are shown. Only one of the conserved arginine residues is shown for simplicity. The conformation adopted by the sugar ring is shown throughout the reaction.

In this mechanism the reaction proceeds through two transition states, with the sugar ring adopting the half chair $^4H_5$ configuration. The conformation of the transition state has not been unequivocally deduced, and recent work on other glycosidases suggests that other ring conformations can be utilised to stabilise the TS (Davies *et al* ., 2003). At present, the high affinity of the sialidase active site for the compound DANA, which adopts the $^4H_5$ half chair configuration due to the double bond between the C2 and C3 atom, have led to this conformation being displayed in the mechanism.

The mechanism used by sialidases is still open to investigation, and in chapter 8 of this thesis a number of site directed mutants of the bacterial sialidase from *M.viridiaciens* are explored in complex with DANA and sialic acid.

These studies show that the active site of sialidases is quite distinct from other glycosidases studied to date, and shows a remarkable ability to function when either the acid/base catalyst or nucleophile are removed.

1.6. The *Clostridium perfringens* sialidases.

Clostidium perfringens is a Gram-positive anaerobic spore forming bacterium that causes life-threatening gas gangrene and mild enterotoxaemia in humans and animals (Rood, 1998). *C.perfringens* infections are characterised by the fast growth rate of the bacterium (8 to 10 minutes generation time) and the release of large amounts of toxins and enzymes that cause massive destruction of the host tissue. Such destruction can quickly lead to systematic toxaemia, shock and eventually death unless prompt antibiotic and surgical treatment is given (Rood, 1998).

The complete *C.perfringens* strain 13 genome was recently mapped and analysed in an attempt to more fully understand its pathogencity and physiology (Shimizu *et al* ., 2002). This analysis revealed that the genome encoded three sialidases, two of which had previously been characterised: the nanH gene product with a molecular weight of 43 kDa, which is not secreted, and the nanI gene product, which has a molecular weight of 77 KDa and is secreted (Kruse *et al* ., 1996; Traving *et al* ., 1994).

These two sialidases have been studied extensively and shown to exhibit very different kinetic and biochemical properties (Roggentin *et al* ., 1995). The most prominent of these was substrate specificity, which was found to differ markedly between the two sialidases; the *nanI* gene product hydrolyses sialic acids from low and high molecular weight substrates in a similar and effective way. The nanH sialidase was mostly limited to oligosaccharides and it hydrolysed sialic acids of colominic acid, glycoproteins, mucins and gangliosides only very slowly (Roggentin *et al* ., 1995).

The third putative sialidase, nanJ, has a predicted molecular weight of 129 kDa. So far this sialidase has not been biochemically or kinetically characterised.

Figure 1.6.1. shows a schematic overview of the three sialidases based on a multiple sequence alignment of their amino acid sequences using TCOFFEE (Notredame *et al* ., 2000), which is shown in Figure 1.6.2. The smallest isoenzyme, nanH, contains only 19 % sequence identity with nanI and nanJ. These two latter isoenzymes are more closely related, having 57 % sequence identity, and both contain a number of additional domains besides the sialidase domain (β-propeller). The domain directly preceding the sialidases domain in both isoenzymes has been assigned as a lectin domain, as judged by homology with the leech sialidase, the structure of which is known (Lou *et al* ., 1998a). The largest sialidase in *C.perfringens*, nanJ, also possesses additional domains: a putative F5/8 C-type or discoiden domain (Baumgartner *et al* ., 1998) upstream of the lectin domain and a domain of unknown function following the sialidase domain, with a putative fibronectin type 3 domain at its C-terminus.

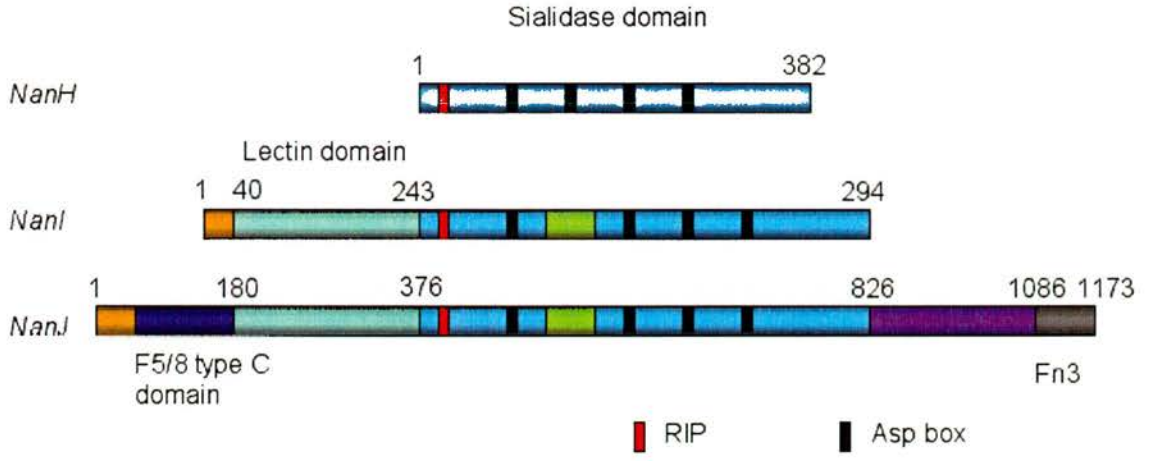
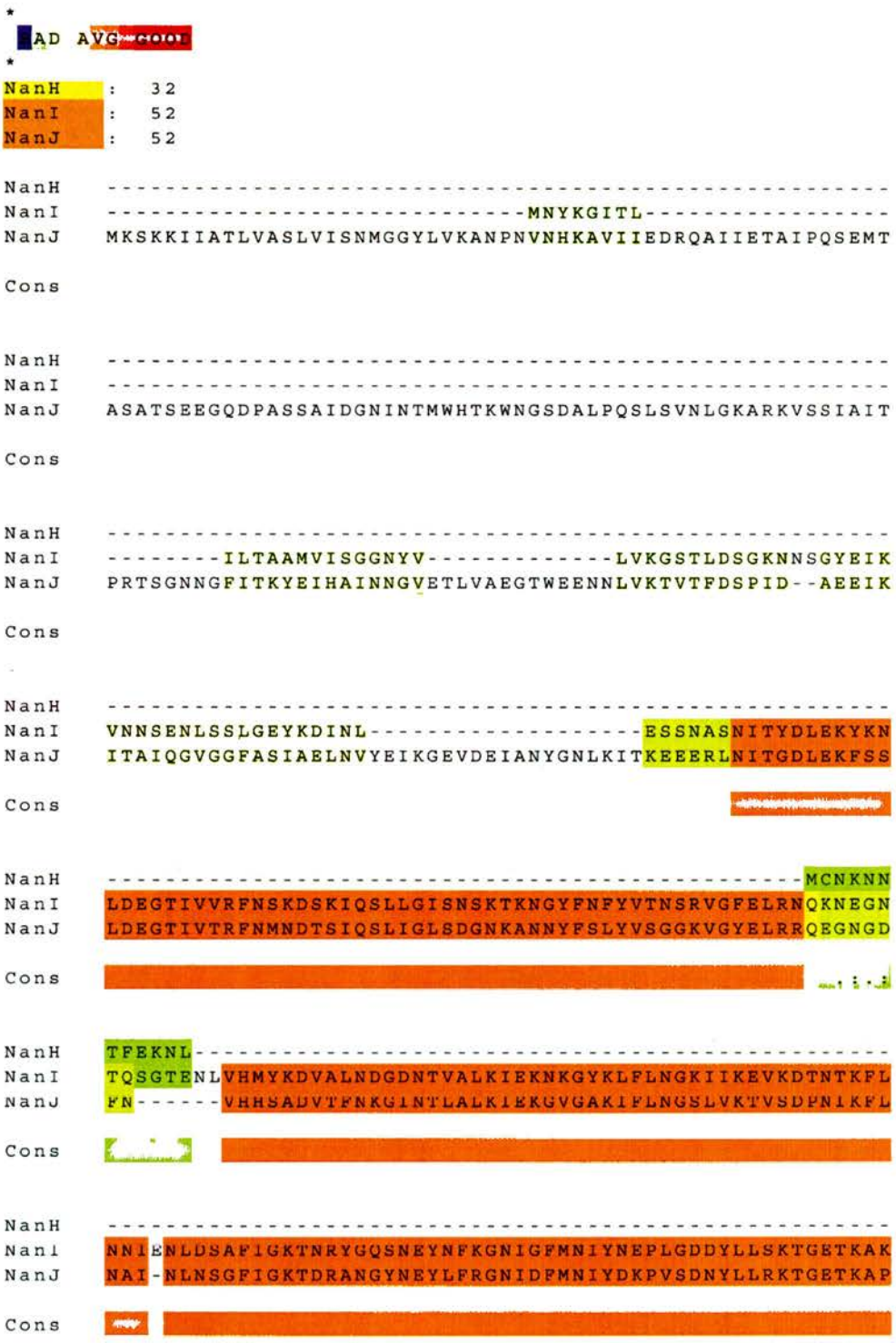
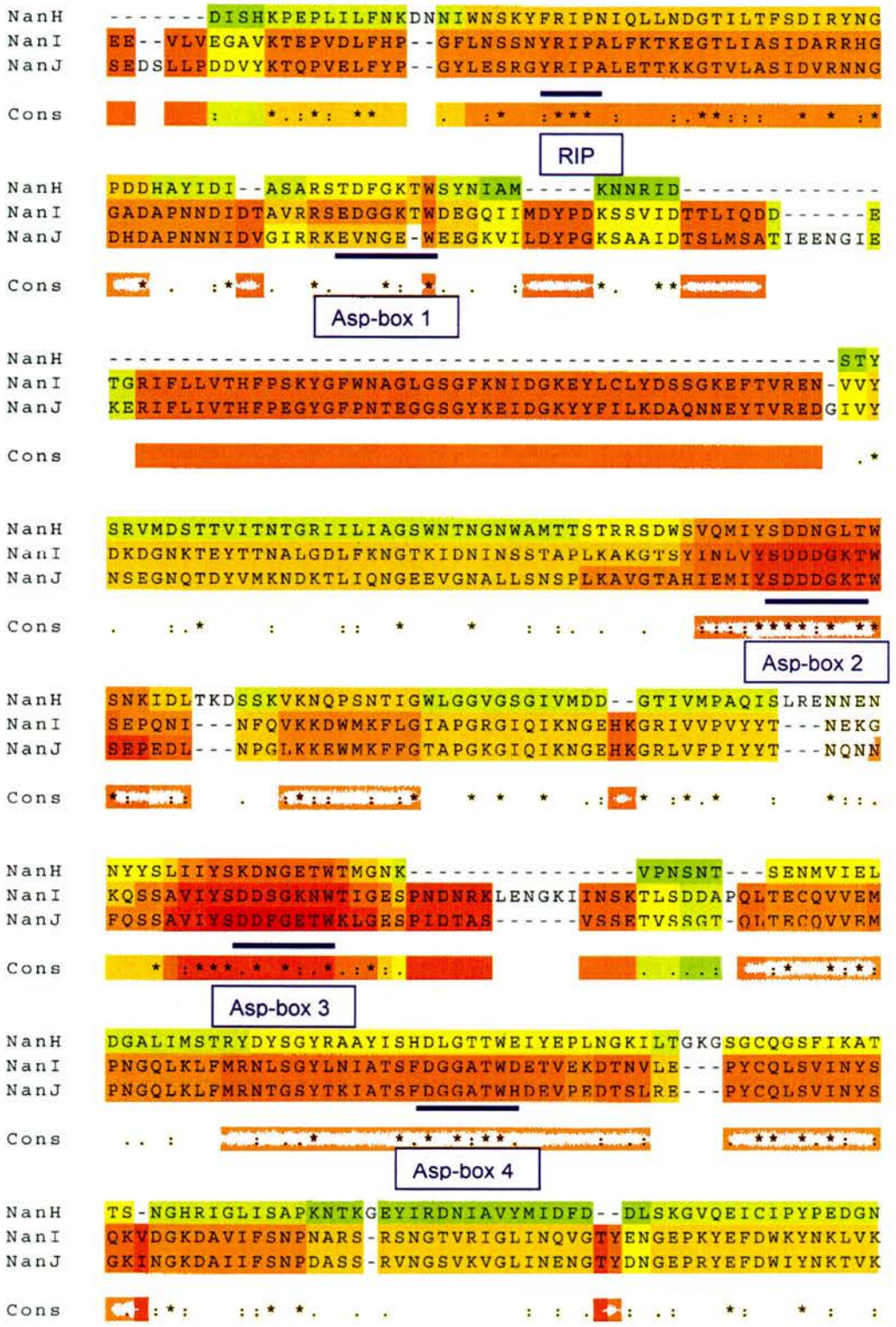
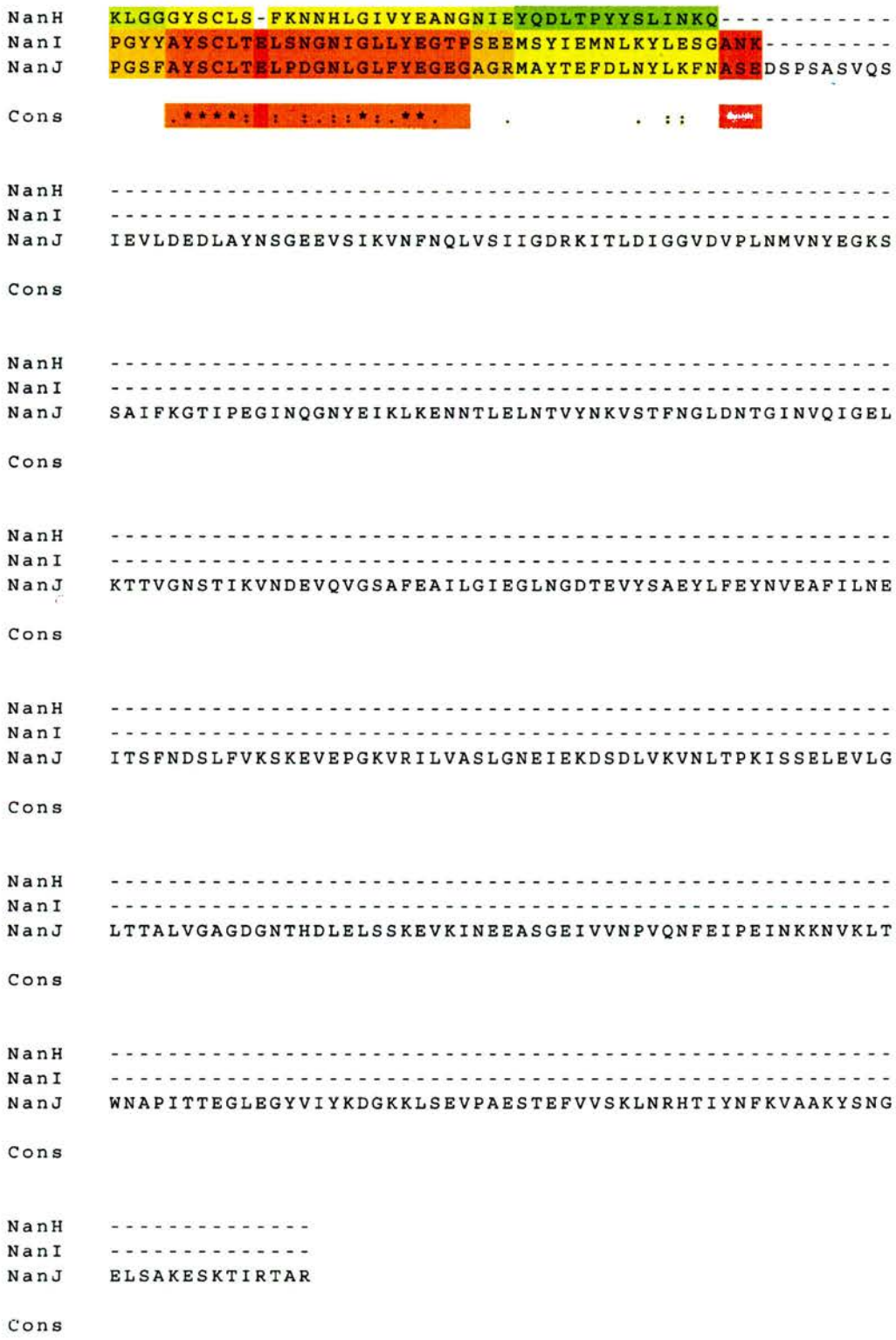


Figure 1.6.1. Schematic view of the three *C.perfringens* sialidases. The locations of the RIP and Asp-box motifs are indicated. Both the nanI and nanJ have a signal peptide at the N-terminus (gold) that results in the secretion of nanI, and possibly of nanJ. This figure is based on the multiple sequence alignment presented in Figure 1.6.2. below.

Figure 1.6.2. Multiple sequence alignment of the three *C.perfringens* sialidase primary structures. The alignment was carried in TCOFFEE (Notredame *et al* ., 2000).




NanH
NanI
NanJ
Cons
RIP
Asp-box 1
Asp-box 2
Asp-box 3
Asp-box 4


NanH KLGGGYSCLS-FKNNHLGIVYEANGNIEYQDLTPYYSLINKQ----------
NanI PGYYAYSCLTELSNGNIGLLYEGTPSEEMSYIEMNLKYLESGANK---------
NanJ PGSFAYSCLTELPDGNLGLFYEGEGAGRMAYTEFDLNYLKFNASEDSPSASVQS
Cons
NanH ----------
NanI ----------
NanJ IEVLDEDLAYNSGEEVSIKVNFNQLVSIIGDRKITLDIGGVDVPLNMVNYEGKS
Cons
NanH ----------
NanI ----------
NanJ SAIFKGTIPEGINQGNYEIKLKENNTLELNTVYNKVSTFNGLDNTGINVQIGEL
Cons
NanH ----------
NanI ----------
NanJ KTTVGNSTIKVNDEVQVGSAFEAILGIEGLNGDTEVYSAEYLFEYNVEAFILNE
Cons
NanH ----------
NanI ----------
NanJ ITSFNDSLFVKSKEVEPGKVRILVASLGNEIEKDSDLVKVNLTPKISSELEVLG
Cons
NanH ----------
NanI ----------
NanJ LTTALVGAGDGNTHDLELSSKEVKINEEASGEIVVNPVQNFEIPEINKKNVKLT
Cons
NanH ----------
NanI ----------
NanJ WNAPITTEGLEGYVIYKDGKKLSEVPAESTEFVVSKLNRHTIYNFKVAAKYSNG
Cons
NanH ----------
NanI ----------
NanJ ELSAKESKTIRTAR
Cons

The relationship and evolutionary distribution of microbial sialdases has highlighted some surprising observations (Schauer *et al* ., 1995). Pairwise comparison of sialidase sequences from a variety of different microorganisms revealed that the occurrence of the sialidases is frequently not in accordance with the phylogenetic relationship of the bacterial species or strains. Some of the sialidases studied were found to be related in accordance with the phylogenetic distances of their producers, for example, the sialidases from *M.viridifaciens* and the large isoenzyme, nanI, from *C.perfringens*. However, the 'small' sialidases of lower molecular weight (42 – 44 kDa) produced by *S.typhimurium*, *C.perfringens* isoenzyme nanH, and *C.sordellii* exhibit a higer similarity than is expected from the relationship of the bacterial species (Schauer *et al* ., 1995). The gram positive (Clostridia) or gram negative (salmonella) bacteria are quite distinct from an evolutionary standpoint, but are found to inhabit the same ecological niche, the intestines of vertebrates. This has prompted many researches to conclude that these bacteria through exchange of genetic information have acquired the genes encoding the sialidase enzymes (Roggentin *et al* ., 1993). The participation of phages in this process has been indicated by the location of the 'small' isoenzyme of *C.perfringens*, nanH, near a phage insertion site in the genome (Canard and Cole, 1990).

This horizontal gene transfer phenomenon has been attributed to the cytoplasmic location of the 'small' sialidases from *C.perfringens* and *S.typhimurium*, which lack a signal peptide and so are unable to make contact with their extracellular substrates. If a gene is acquired from a foreign source, it usually has to be adapted for bacterial functions, or is eliminated (Weaver and Hedrick, 1997). The presence of non-adapted, but active gene products, such as the 'small' sialidases, indicated that gene transfer may have occurred relatively recently, on an evolutionary time scale. Following this logic, the large sialidases seen in the *C.perfringens* genome would have been acquired earlier in the evolution, and have been adapted through the addition of secretion signal sequences at the N-termini and of extra domains. This theory is supported by

the higher sequence similarity between the large sialidases of *C.perfringens* than to the small isoenzyme.

1.7. Project Aims.

The aim of this project is the structural solution of the nanI gene product, which is one of the most well characterised microbial sialidase enzymes (Roggentin *et al* ., 1995; Roggentin *et al* ., 1993; Rood, 1998; Sheu *et al* ., 2002; Shimizu *et al* ., 2002). Using the novel sialyl-fluoro compounds kindly provided by Prof. S. Withers and Dr. A. Watts, Dept. Chemisty, University of British Columbia, Canada, we aim to provide the first structures of the covalently bound sialidase. It is hoped that this work will further aid in understanding the catalytic mechanism of these enzymes and should enable an interesting comparison to be made with the structures already published on the *T.cruzi* enzyme.

In addition to this, a number of active site mutants of the large 68 kDa form of the bacterial sialidase from *M.virdifaciens* have been provided for structural study by our collaborators Prof. A. Bennet and Dr. J. Watson, at Simon Fraser University, Vancouver, Canada. Combined with detailed kinetic and product studies being carried in their laboratory, the structures of the mutant complexes with both substrate analogues and transition state mimics are expected to both aid and underpin these biochemical studies.

Chapter 2

Expression, purification and sub-cloning of the large sialidase, nanI, from *Clostridium perfringens*.

2.1. Summary

Crystallographic structural studies on any protein require a reproducible supply of pure, homogenous protein, for which crystallisation conditions can be found. The structural investigation of nanI began with the kind gift of the cloned, full-length gene in a bacterial expression vector, pQE-70 (Qiagen). The initial cloning and expression of this construct has already been reported (Sheu *et al* ., 2002). The pQE-70 expression vector encodes a six-histidine affinity tag at the carboxy terminus of the protein to aid in purification. The nanI gene consists of 2082 nucleotides, which encode a protein of 694 amino acids with a molecular mass of 77 KDa. The primary structure reveals the location of four 'Asp-boxes' and one RIP region, formed by amino acids 266-268. The *C.perfringens* genome encodes two additional sialidases, *nanH* and *nanJ* with which *nanI* shares 27 and 57 % sequence identity, respectively.

This chapter describes the purification of the full-length protein using a modified method to that described previously (Sheu *et al* ., 2002). Due to the unstable nature of the full-length protein, subcloning and purification of a truncated form, amino acids 243 – 694, is also described. It is this truncated form, which eventually produced crystals suitable for X-ray diffraction analysis.

2.2 Recombinant nanI Expression.

The full-length nanI gene was given as a gift from Dr. Chien, National Yang-Ming University, Taiwan, ROC, as a 1 ml glycerol stock frozen at –80 °C. This glycerol stock was used to inoculate a number of 10 ml overnight cultures of Luria-Bertani (LB) medium supplemented with 100 μg/ml of carbenicillin. This start-up culture was grown at 37 °C overnight in a shaking incubator running at 200 rpm. The following morning, 1L of prewarmed LB media containing 100 μg/ml carbenicillin was incoculated with a 10 ml overnight culture. These 1L cultures were grown in 2L-baffled flasks at 37 °C in a shaking incubator at 190 rpm. Cells were grown to an optical density of A_{600} nm = 0.5-0.8, induced with 1 mM isopropyl-D-thiogalactopyranoside (IPTG) and grown for a further 13 hours under the same conditions. The medium was then spun at 10,000 g for 15 minutes at 4 °C and the supernatant discarded. There was typically six to eight grams of wet cell paste per litre, which was transferred to a 50 ml falcon tube for storage at –20 °C until purification.

2.3. Purification of recombinant nanI.

The cell pellets were resuspended in buffer A (50 mM Sodium Phosphate, pH 7.0, 0.5 M Sodium Chloride (NaCl), 20 mM Imidazole, 2X complete Roche protease inhibitor – EDTA tablets), to a final volume of 6 ml/g cell paste. 20 μg/ml DNase 1 were added to the lysis buffer to degrade the bacterial DNA in the lysate. The cells were lysed by sonication and then centrifuged at 20,000 rpm for 20 minutes at 4 °C. The supernatant was then transferred to ultra-centrifuge tubes and spun a further time at 39,000 rpm for 1 hour, again at 4 ^{0}C. This second centrifugation helps to remove any lipids transferred from the first centrifuge step and helps clarify the lysate before applying to the chromatography columns.

A simple three-step purification procedure was then employed to separate the nanI protein. An initial capture step using nickel affinity was followed by

anion exchange, with a final polishing step using size-exclusion chromatography. All chromatography, unless otherwise stated, was performed on a Perceptive BioSystems BioCAD 700E workstation at 21 °C. For the nickel affinity chromatography, a 7 ml metal chelate POROS column (Perceptive BioSystems) was used, charged with 200 mM nickel sulphate and pre-equilibrated with five column volumes of buffer A.

The cell lysate was syringed through a 0.2 μM filter (Millipore) and applied to the column at a flow rate of 10 ml/min. 10 column volumes of buffer A were then applied to the column in order to remove any unbound protein. To elute the nanI protein, a linear gradient of imidazole was used over ten column volumes to a final concentration of 0.5 M in buffer A. 5 ml fractions were collected over the gradient and eluted protein monitored using the absorbance of aromatic amino acid side chains at 280 nm.

Those fractions containing the nanI protein, as shown by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), were pooled and diluted with buffer B (20 mM Tris, pH 8.0) to a final NaCl concentration of 20 mM. Anion exchange chromatoghraphy was performed using a 7 ml anion POROS column pre equilibrated with buffer B. The diluted sample was then loaded onto the anion column at a flow rate of 10 ml/min.

Following application of the protein sample, the column was washed with 10 column volumes of buffer B, or until the absorbance reading at 280 nm returned to baseline. The nanI protein was eluted using a linear gradient, identical to that employed in the first step, but going to 0.5 M NaCl and collecting 2 ml fractions over the gradient. The fractions were again assayed for the correct sized protein using SDS-PAGE with a 4-12 % Bis-Tris gel (Invitrogen).

Those fractions containing the nanI protein were pooled and concentrated to a volume of 2 ml for size-exclusion chromatography. An XK16/60 S200 Hi-prep gel-filtration column (Amersham BioScience) was equilibrated with two column volumes of Buffer C (20 mM Tris, pH 8.5, 200 mM NaCl). The protein sample was again filtered and loaded onto the column. The separation was performed at a flow rate of 2 ml/min. One column volume of buffer C was applied to elute the

protein. Fractions were collected and analysed by SDS-PAGE to check the purity of the enzyme (Figure 2.1).

Fractions containing the nanI enzyme, as identified by tryptic digestion and mass spectrometric analysis, were pooled and concentrated to 20 mg/ml for crystallisation. Protein concentration was determined using the Bicinchoninic Acid (BCA) protein assay (Smith, 1985) (Pierce).

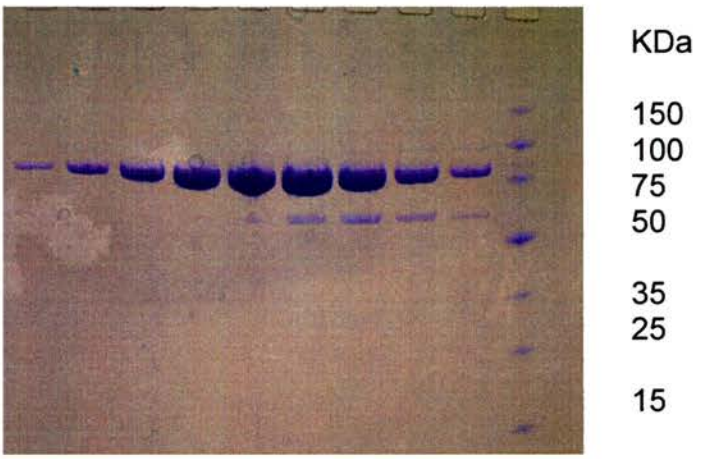


Figure 2.1. SDS-PAGE analysis of the gel filtration step of the full-length nanI purification. L-R, 20 μL fractions of nanI protein eluted from the XK16/60 S200 Hi-prep gel filtration column.

The gel shows two bands in the final sample of nanI. Mass spectrometry identification showed the lower band, at approximately 50 KDa, to contain an N-terminally truncated form of the protein. Efforts to remove this contaminate proved unsuccessful, due to the similar biophysical properties of the two species, and the inability of our gel filtration columns to resolve the two on mass difference. It was therefore decided to continue with the crystallisation trials with this sample.

2.4. Crystallisation of recombinant nanI.

The concentrated protein (20 mg ml^{-1}) in 20 mM Tris, pH 8.5, 200 mM NaCl, was screened for suitable crystallisation conditions using the standard sparse matrix and grid screens from Hampton Research, Decode Genetics, SIGMA and Molecular Dimensions. Initial crystal trials were performed at 21 ^{0}C under oil using a 96-well microbatch system and a Douglas Instruments robot with 2 μL of protein solution and 2 μL of precipitant solution. Crystals were obtained in a Hampton 1 screen; condition 42: 20 % w/v Polyethylene Glycol (PEG) 8000 and 0.05 M KH_2PO_4. The crystals were extracted from under the oil, cryoprotected in the precipitant with the addition of 20 % v/v glycerol for a minute, before being flash frozen at 100 K in a liquid nitrogen stream. Diffraction to 2.5 Å was obtained on an in-house X-ray source, and the crystals belonged to space group 19, $P2_12_12_1$ with unit cell dimensions a = 96.8 Å, b = 69.2 Å, c = 72.8 Å. Assuming one 77 kDa monomer per asymmetric unit, a Mathews coefficient (Mathews, 1968) of V_M = 1.58 $Å^3Da^{-1}$ and a solvent content of 22 % can be calculated. This unusually tight packing led us to check the crystal contents. Several crystals were washed extensively in precipitant solution and dissolved in 10 μL of 0.1 % Triflouro Acetic Acid (TFA) prior to analysis by mass spectrometry and N-terminal sequencing. These results revealed a 50.3 KDa protein with an N-terminus beginning VEGA (Val-Glu-Gly-Ala), corresponding to residues 243-694 of nanI. Sequence alignment with other bacterial sialidases suggested that this would form the C-terminal, catalytic domain of the enzyme. The full-length protein was easily prone to degradation (Sheu *et al* ., 2002), most probably due to the exposure of a protease site between the two putative domains of the protein. The formation of crystals by the N-terminally truncated nanI species, and problems in their reproducibility, led us to sub-cloning this domain for further structural studies.

2.5. Sub-cloning the catalytic domain of the nanI sialidase.

The fragment of the gene encoding the putative C-terminal catalytic domain (amino acids 243-694) was amplified by PCR from the pQE-70 construct encoding the full-length nanI sialidase gene (Sheu *et al* ., 2002). Forward and reverse primers were designed to add the appropriate restriction sites for subcloning into various expression vectors. An *NcoI* site was introduced at the front of the gene, to allow an inframe initiation codon to be inserted before the VEGA sequence,

(Forward primer 5' – TAAA**CCAATGG**TAGAGGGAGCTGTAAAG – 3'). A *PstI* site was incorporated following the natural stop codon from the *nanI* gene (Reverse primer 5' – CCAT**CTGCAG**TTATTTATTAGCTCC – 3'). The Amplification reactions were carried out using 1 μM of each primer and *pfu* DNA polymerase (Promega). This polymerase was chosen for its 3'-5' exonuclease (proofreading) activity. Base misinsertions that may occur during polymerisation are rapidly excised by the proofreading activity of the polymerase. The 1359 bp PCR product was checked by electrophoresis on a 1 % Agarose gel (Figure 2.2).

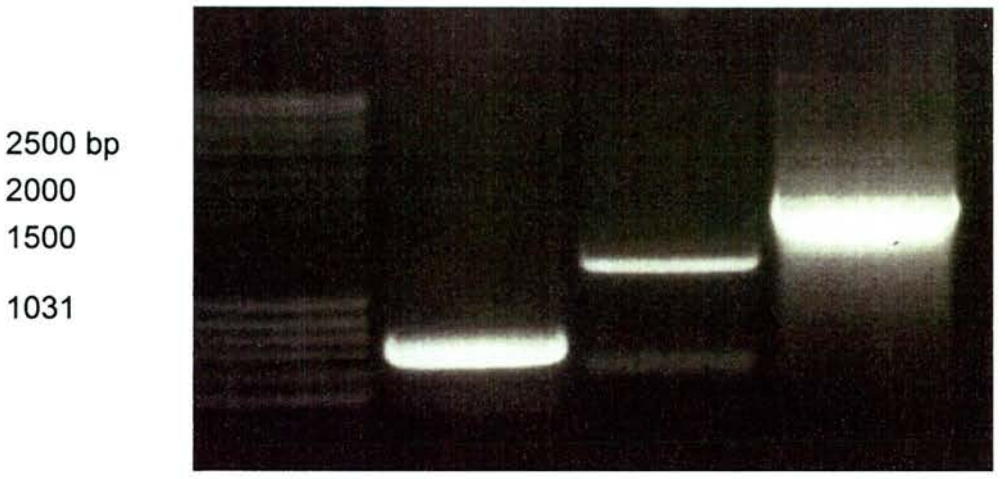


Figure 2.2. PCR amplification of the C-terminal domain from the full-length nanI gene. L-R, DNA ladder, 726 bp control, 1359 bp C-terminal gene fragment, 2085 bp full length nanI gene.

The 1359 bp product was excised from the 1 % agarose gel following the protocol given in the QIAquick™ Gel Extraction Kit (Qiagen). The DNA was eluted from the spin column in 30 μL of buffer EB (20 mM Tris, pH 8.0, 1 mM EDTA). The eluted DNA fragment was then digested with *NcoI* and *PstI* for two hours at 37 °C in a water bath.

The double digest product was then solution cleaned, using the same procedure as described above for the gel extraction, to remove the restriction enzymes and buffers. The DNA fragment was then ligated into the pHIS-MalC2_Tev (St Andrews) (Figure 2.3) expression vector, linearised using the same restriction enzymes, *NcoI* and *PstI*. This vector confers ampicillin/carbenicillin resistance. The ligation reaction was performed overnight at 16 ^{0}C in the presence of 6 units of T4 DNA Ligase (Promega). The ligation mixture was transformed into TAM1 *E.coli* cells (ActiveMotif) and plated out on agar plates supplemented with 100 μg/ml ampicillin. PCR using the forward and reverse amplification primers was used to identify those colonies containing the nanI-pHIS_MalC2_Tev vector.

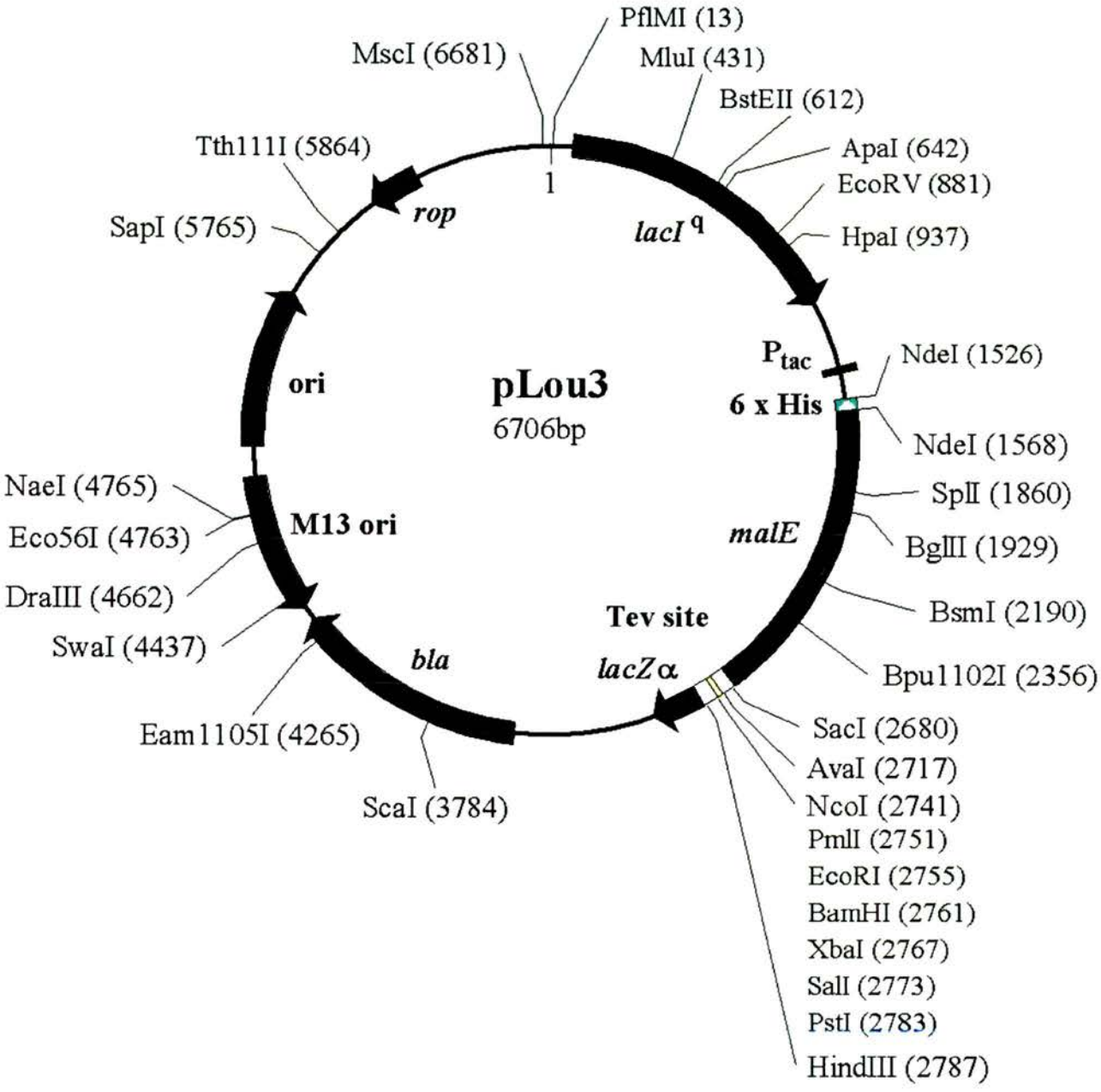


Figure 2.3. Vector map showing restriction sites for pHIS_MalC2_Tev. The MCS (Multiple Cloning Site) is shown in yellow, the six-histidine tag at the N-terminus of the Maltose Binding Protein (MBP) is shown in blue.

Plasmid construction was confirmed by restriction analysis and DNA sequencing using the M13 and MalE primers (NEB). The inserted nanI gene fragment was completely sequenced to ensure no mutations had been introduced. The nanI- pHIS_MalC2_Tev construct was purified from 10 ml overnight cultures of the TAM1 transformed *E.coli* cells grown in LB media supplemented with 100 μg/ml carbenicillin. The construct DNA was purified using the protocol contained within the Wizard® Plus SV Miniprep Purification System (Promega).

2.6. Recombinant Expression of the catalytic domain of nanI.

pHIS_MalC2_Tev is an *E.coli* plasmid cloning vector designed by Dr. Louise Major (St Andrews) for recombinant protein expression and purification using the pMAL protein fusion and purification system (NEB #E8000S) (Maina *et al* ., 1988). The multiple cloning site (MCS) is positioned to allow translational fusion of the *E.coli* maltose binding protein (MBP, encoded by the *malE* gene) to the N-terminus of the cloned target protein. The MBP can be subsequently removed using protease targeted at a specific cleavage site between the two proteins.

The pHIS_MalC2_Tev plasmid is a modified version of the pMAL™-p2X plasmid (NEB), and incorporates a number of significant features to aid in the purification process. Principally, these are a tobacco etch virus (TEV) cleavage site instead of Factor Xa, and a six-Histidine affinity tag at the N-terminus of the MBP. These changes were made to facilitate economical protein purification with 'in-house' TEV protease and the advantages of a double tag system respectively.

Purified nanI- pHIS_MalC2_Tev plasmid DNA was used to transform Rosetta (DE3) pLysS *E.coli* cells (Novagen) for overexpression of the nanI catalytic domain. Single colonies of these cells were used to inoculate a number of 10 ml overnight cultures of LB medium supplemented with 100 μg/ml of carbenicillin. These start-up cultures were grown at 37 °C overnight in a shaking incubator running at 200 rpm.

The following morning, 1L of pre-warmed LB media containing 100 μg/ml carbenicillin was inoculated with a 10 ml overnight culture. These 1L cultures were grown in 2L-baffled flasks at 37 °C in a shaking incubator at 190 rpm. Cells were grown to an optical density of A600 nm = 0.5-0.8, induced with 1 mM IPTG and grown for a further 13 hours at a reduced temperature of 25 °C. The medium was then spun at 10,000 g for 15 minutes at 4 °C and the supernatant discarded. There was typically 6 grams of wet cell paste per litre, which was then transferred to a 50 ml falcon tube before being stored at –20 °C until purification.

2.7. Purification of the catalytic domain of recombinant nanI.

The cell pellets were resuspended in buffer A (50 mM Sodium Phosphate, pH 7.0, 0.5 M Sodium Chloride (NaCl), 20 mM Imidazole, 2X complete Roche protease inhibitor – EDTA tablets), to a final volume of 6 ml/g cell paste. 20 μg/ml DNAse 1 was added to the lysis buffer to degrade the bacterial DNA in the lysate. The cells were lysed by sonication and then centrifuged at 20,000 rpm for 20 minutes at 4 °C. The supernatant was then transferred to ultra-centrifuge tubes and spun a further time at 39,000 rpm for 1 hour, again at 4 °C.

The availability of a double-tagged fusion protein presents a range of different options during purification. In this instance it was decided to use nickel affinity purification as an initial capture step. This would provide a rapid separation of the fusion protein from any proteases in the cell lysate. The cell lysate was syringed through a 0.2 μM filter (Millipore) and applied to a 7 ml metal chelate POROS column at a flow rate of 10 ml/min.

The column was pre-equilibrated with Buffer A and charged with two column volumes of 200 mM nickel sulphate prior to sample application. Ten column volumes of Buffer A were then applied to the column to remove any unbound protein and other contaminates. To elute the nanI-fusion protein, imidazole was added to the column in buffer A to a final concentration of 0.5 M in 100 mM steps. The nanI-MBP fusion protein would typically elute in the 200 mM step. 5 ml

fractions were collected, and those fractions containing the nanI-MBP fusion protein, as assayed by SDS-PAGE, were pooled.

The second stage of the purification employed the characteristic affinity of MBP for the polysaccharide amylose. The pooled fractions containing the nanI-MBP fusion protein were applied to a 20 ml gravity flow amylose column (NEB) at 21 °C. The binding of MBP to amylose is very specific and strong, requiring no special buffer conditions, thus the nanI-MBP fusion protein was loaded in the same buffer it eluted in from the nickel column. The amylose column was used to buffer exchange the sample. For optimal cleavage of the fusion protein by TEV protease, the fusion protein should be in a quite specific buffer. The protein was washed with ten column volumes of Buffer C (50 mM Tris, pH 7.5, 0.3 M NaCl, 0.5 mM EDTA, 1 mM DTT, 20 mM Imidazole, in phosphate buffered saline (PBS)). The nanI-fusion protein was then eluted using 10 mM maltose in Buffer C and 5 ml fractions collected.

Fractions containing the fusion protein were pooled and incubated with 20 μg/ml TEV protease (St Andrews) at 21 °C overnight. Complete digestion was checked by SDS-PAGE the following morning. The digestion mixture was then dialysed overnight at 4 °C in 4 L of Buffer D (20 mM Sodium Phosphate, pH 7.5, 200 mM NaCl) to remove the DTT from the sample. Following dialysis, the sample was applied to a 5 ml HisTrap metal chelate column (Amersham BioSciences) using a hand held syringe.

The column was pre-equilibrated with Buffer D and charged with 2.5 column volumes of 200 mM nickel sulphate. The six-His tagged MBP and TEV protease would bind the column, and the cleaved nanI protein would pass through. The purified nanI was checked for purity using SDS-PAGE on a 4-12 % Bis –Tris gel (Figure 2.4) and concentrated to 20 mg/ml for crystallisation.

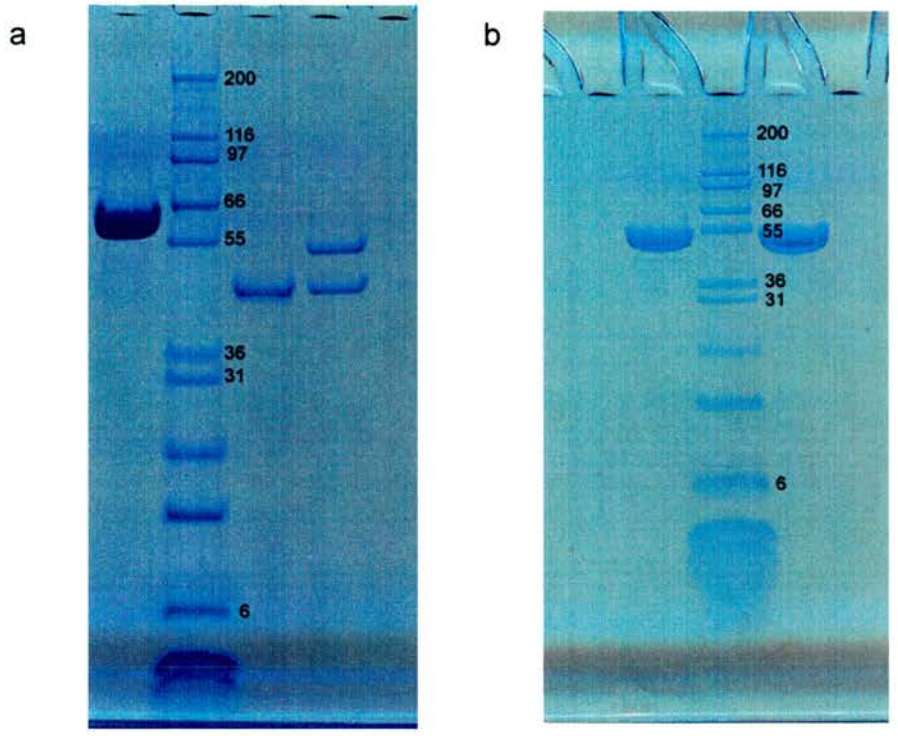


Figure 2.4. SDS-PAGE analysis of final steps in the purification of the nanI catalytic domain, amino acids 243-694. a L-R, nanI-MBP fusion protein, molecular weight markers (kDa), cleaved MBP protein, nanI catalytic domain and MBP protein. b L-R, nanI catalytic domain, molecular weight markers, nanI catalytic domain.

2.8. Crystallisation of the recombinant nanI catalytic domain.

All subsequent crystallisation experiments were performed using the sitting-drop vapour diffusion method in Douglas Instruments 96-well crystallisation plates (Douglas Instruments) at 21 °C. The drops were made up of equal volumes (2 μl) of mother liquor and protein solution (20 mg/ml in Buffer D). The protein was tested for crystallisation conditions in a range of sparse matrix and grid screens from Hampton Research, Decode Genetics, SIGMA and Molecular Dimensions.

Initial crystals were obtained in the PEG/Ion screen from Hampton Research, condition 18 (20 % PEG 3350, 0.2 M Potassium Nitrate). Good quality crystals grew in 4-6 days (Figure 2.5), with no refinement of conditions required. When needed, seeding overnight with a crushed crystal could grow new crystals. The crystals were cryoprotected in mother liquor containing 25 % PEG 400, with an initial step of 12 % PEG 400, before being flash-frozen in a liquid nitrogen stream at 100 K.

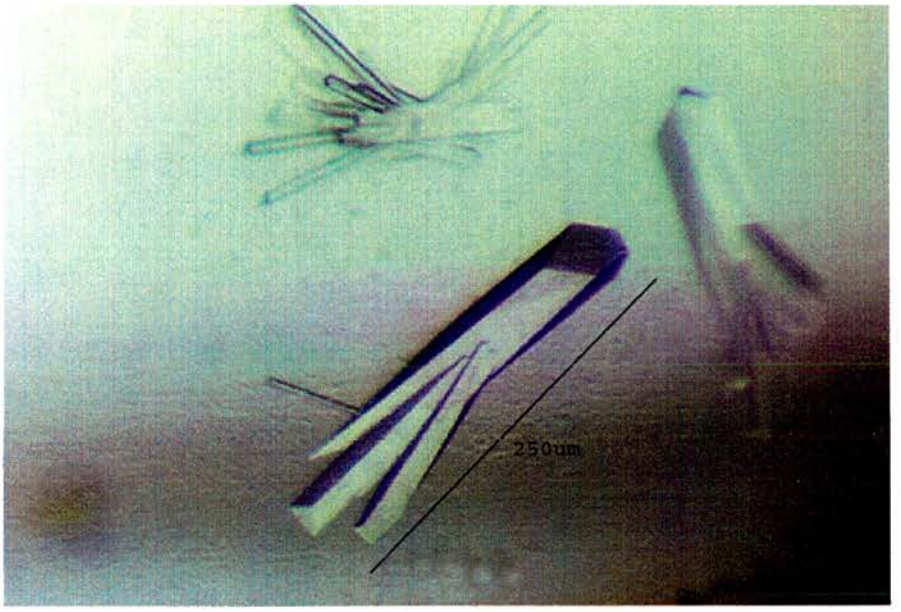


Figure 2.5. Crystals of the catalytic domain of *C.perfringens* nanI.

2.9. Discussion.

Persistent breakdown of the full-length 77kDa *C. perfringens* nanI sialdiase led us to subclone the stable catalytic domain from residues 243 to 694. Previous studies had subcloned residues 216 -694, and found that this domain exhibited a similar specific activity to the full-length protein (Sheu *et al* ., 2002). From the positions of the conserved RIP and Asp-box motifs, and the conservation of other conserved active site residues (Taylor, 1996), it was assumed that all of the catalytic activity resides in the 243-694 domain. This proved correct during preliminary NMR studies on this domain that showed rapid turnover of the natural substrate analogue α2,3-sialyllactose (Chapter 6). The rapid degradation of the full-length protein suggests that the linker between the lectin and sialidase domains is susceptible to protease digestion and that the lectin domain is perhaps flexibly linked. Our main interest is in the catalytic domain, for which single crystals have been obtained and diffraction data collected (Chapter 3).

Chapter 3

Data collection, Processing, Molecular Replacement and Refinement of the nanI catalytic domain.

3.1. Summary.

This chapter will describe the steps involved in determining the three dimensional structure of the nanI sialidase catalytic domain. Data on two ligand complexes are also included; a crystal grown in the presence of 10 mM Neu5Ac2en (DANA) and a crystal soaked for two minutes in 5 mM α-2,3-difluoro Neu5Ac. This last compound was designed to trap the hypothesised covalent glycosyl-enzyme intermediate on the reaction pathway. Neu5Ac2en is a proposed transition-state analogue in the reaction pathway of sialidases (Miller *et al* ., 1978). This compound has been used extensively in previous structural studies on sialidases to identify the active site and provide information on the catalytic mechanism (Taylor, 1996).

Complete X-ray datasets were collected at the European Synchrotron Radiation Facility (ESRF) beamline ID14-1 to 0.97 Å for both the apo and DANA co-crystals. The crystals were grown as described in Chapter 2. The diffraction data were processed and the structure factor amplitudes calculated. The nanI sialidase structure had not previously been solved, but due to the reasonable sequence identity (37 %) between the catalytic domain of the nanI sialidase and the equivalent catalytic domain of the intra-molecular trans-sialidase from the leech, *Macrobdella decora*, it was decided to use this model to calculate a set of initial phase estimates using the technique of Molecular Replacement (MR). Automated protein refinement (ARP) and manual building was then employed to generate a complete model of the nanI catalytic domain, amino acids 243-694. Full high-resolution anisotropic refinement was then carried out using the program SHELXL97, a modified version of the small molecule refinement

program. The final model was checked and validated using the standard geometric dictionaries as implemented in PROCHECK.

Data on the covalent glycosyl-enzyme intermediate was collected in house on a Rigaku micromax 007 rotating anode generator fitted with an MSC R-AXIS IV ++ image plate detector. Diffraction data were collected to 1.7 Å.

All of the nanI crystals were isomorphous, belonging to space group 19 and with negligible differences in unit cell parameters.

3.2. Data collection and processing.

3.2.1. Data collection.

Before data collection could begin, the crystals were cryoprotected and frozen to minimise any radiation damage that would occur during the experiment. Radiation damage is a particular problem when collecting diffraction data at third and fourth generation synchrotrons, due to their higher energy X-rays (Garman, 1999). The crystals were transferred to a cryoprotectant solution containing 25% PEG 400 in mother liquor before being flash-frozen in a nitrogen stream at 100K. Diffraction data were collected to a resolution of 0.92Å on ID14-EH1 at the ESRF, Grenoble. The data were collected at the resolution limit of this station, using an ADSC Q4R CCD detector system (Fig. 3.2.1). Data were collected in 0.5° oscillations with a 3s exposure. A high-resolution pass provided the 0.92Å data, and a subsequent 1.6Å resolution pass with a 2x attenuator in the beam provided measurements of data overloaded on the first pass. The program MOSFLM (Leslie, 1992) was used to index the first image and determine the unit cell and possible spacegroup of the crystal. The Neu5Ac2en complex was collected using a nanI crystal grown in the presence of 10 mM Neu5Ac2en at 21 °C on the same station.

The apo crystals belonged to space group $P2_12_12_1$ with unit-cell parameters a = 96.89 Å, b = 69.02 Å, c = 72.81 Å. The data have an overall completeness of 92% to 0.97Å, and 87% to 0.92Å. Data-collection and

processing statistics are summarised in Table 3.1. The dataset was subsequently cut to 0.97 Å for refinement due to the incompleteness of the data in the 0.97-0.92 Å shell. The Neu5Ac2en and 2-3-difluro Neu5Ac crystals were isomorphous with the apo crystals (Table 3.2.1).

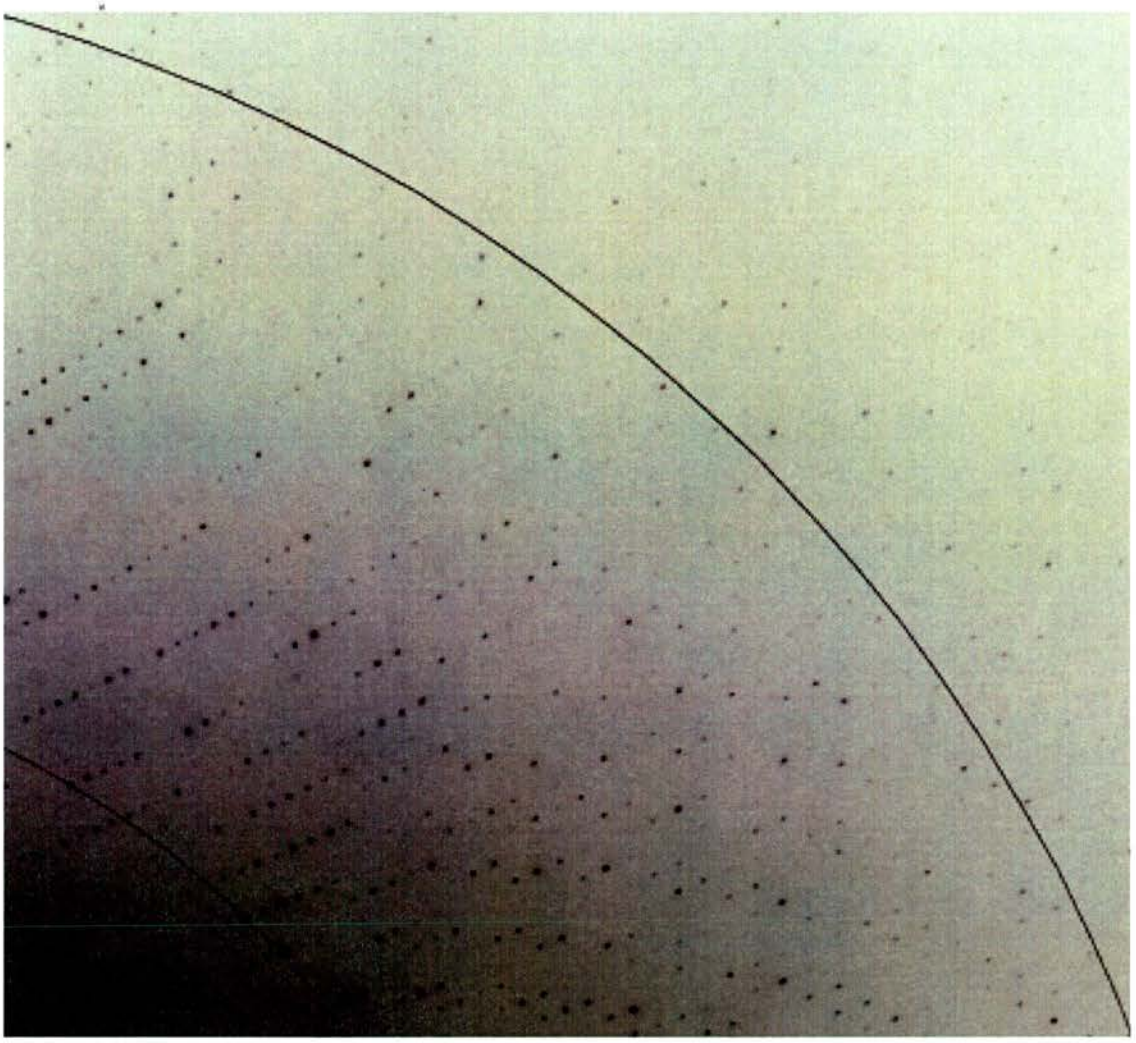
Table 3.2.1. Data collection statistics for the nanI crystals.

Values in parentheses are for the highest resolution shells.

Complex	Apo	Neu5Ac2en	Intermediate
X-ray source	ID14-1 ESRF	ID14-1 ESRF	In house
Wavelength (Å)	0.934	0.934	1.541
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 96.77, b = 69.24, c = 72.81	a = 96.89, b = 69.0, c =72.81	a = 96.98, b = 69.57, c = 72.49
Resolution range (Å)	20-0.97 (1.02-0.97)	20-0.97 (1.03-0.97) (0.97-0.92)	58-1.7 (1.79-1.70)
Mosaicity (°)	0.28	0.32	0.34
Unique reflections	281 317	290 992	54 619
Data completeness (%)	98 (95)	92 (87) (54)	92 (83)
Redundancy	4.6 (3.1)	3.8 (3.1) (2.5)	2.9 (2.6)
R_{merge}†	0.072 (32.9)	0.095 (0.186) (0.272)	0.065 (27.3)
I/σI	15.6 (3.6)	11.8 (5.0) (2.8)	15 (3.8)

† $R_{merge} = \Sigma \mid I(k) - \langle I \rangle \mid / \Sigma I(k)$, where I (k) is the value of the k_{th} measurement of the intensity of a reflection, ‹I› is the mean value of the intensity of that reflection and the summation is over all measurements.

Figure 3.2.1. Part of a 0.5° oscillation X-ray diffraction image of a crystal of nanI collected at ID14-1 ESRF. Inner circle is at 1.3Å, outer circle at 1.0 Å.



3.2.2. Data processing.

For all datasets the integrated intensities output from MOSLFM were scaled and merged using the program SCALA (Evans, 1997) from the CCP4 package (CCP4, 1994). The statistical analysis of the scaling is output to the log file generated by the program. The parameters R_{merge}, $I/\sigma(I)$ and completeness were checked after 90 ° of data had been collected. Tables 3.2.2.1 – 3.2.2.3 show the summary of the log file tables from the output of SCALA for the three datasets. Ideally, the intensities of 100 % of the Bragg reflections should be measured and the data should be significant (≥ 2) in terms of the $I/\sigma(I)$ ratio throughout the resolution range (Wilson, 2001). This was achieved for all data collected from the nanI crystals.

The most striking feature of this diffraction data is the resolution that was achieved. The beamline, ID14-EH1, has a maximum resolution of 0.97 Å for a circle inscribed inside the square ADSC detector. As can be seen from Table 3.2.2.1 however, the diffraction spots measured in the corners of the detector have significant $I/\sigma(I)$ ratios and certainly suggest that these crystals have not been collected at their maximum resolution. The data were cut to 0.97 Å due to the relatively low completeness in the outer resolution shell (0.97 – 0.92) of 53 %.

The definition of atomic resolution data is that diffraction can be measured at < 1.2 Å and that 50 % or more of reflections in the outer range have $I/\sigma(I)$ ratios > 2 (Sheldrick, 1990). Atomic resolution data provides the crystallographer with a number of significant advantages when compared to data > 1.2 Å resolution. These include a more accurate view of hydrogen bonds (Anderson *et al* ., 1997; Wang *et al* ., 1997), protonation states of key residues (Longhi *et al* ., 1997) and statistical or thermal disorder. The quality of electron density maps at atomic resolution and the anisotropic treatment of thermal motion make it possible both to model alternative conformations and obtain information on the precise positions of the amino acids, as well as any ligands or complexes in the structure.

Table 3.2.2.1 Statistics of data processing in the ten resolution shells for the nanI apo dataset, collected at ID14-EH1.

Resolution shell (Å)	Av_I	I/σ(I)	R_{merge} (%)	Completeness (%)
∞ - 3.07	4102	35.3	5.0	98.3
3.07 – 2.17	1759	36.4	5.8	99.5
2.17 – 1.77	978	30.6	7.9	100
1.77 – 1.53	401	22.6	11.7	100
1.53 – 1.37	230	17.8	12.8	99.8
1.37 – 1.25	165	15.0	8.7	99.2
1.25 – 1.16	134	13.1	9.0	98.4
1.16 – 1.08	105	10.1	10.7	97.3
1.08 – 1.02	58	6.7	16.2	96.4
1.02 – 0.97	29	3.6	32.9	95.6

Table 3.2.2.2 Statistics of data processing in the ten resolution shells for the nanI Neu5Ac2en dataset, collected at ID14-EH1.

Resolution shell (Å)	Av_I	I/σ(I)	R_{merge} (%)	Completeness (%)
∞ - 2.91	89,248	20.0	8.0	86.2
2.91 – 2.06	42,963	22.5	9.4	91.5
2.06 – 1.68	21,569	20.0	10.1	94.1
1.68 – 1.45	9,719	16.4	10.4	95.2
1.45 – 1.30	5,749	14.0	9.9	95.3
1.30 – 1.19	4,401	12.5	8.9	94.5
1.19 – 1.10	3,669	10.9	9.9	93.1
1.10 – 1.03	2,199	8.1	12.0	91.6
1.03 – 0.97	1,130	5.0	18.6	87.0
0.97 – 0.92	682	2.8	27.2	53.6

Table 3.2.2.3 Statistics of data processing in the ten resolution shells for the nanI covalent glycosyl-enzyme intermediate dataset, collected in house on a Rigaku MicroMax-007 rotating anode generator.

Resolution shell (Å)	Av_I	I/σ(I)	R_{merge} (%)	Completeness (%)
∞ - 5.38	5842	33.2	3.4	86.5
5.38 – 3.80	10 930	35.0	3.3	92.3
3.80 – 3.10	8761	29.1	4.4	94.1
3.10 – 2.69	4710	22.1	6.4	95.3
2.69 – 2.40	2949	17.7	5.5	95.6
2.40 – 2.19	2590	15.4	7.2	95.3
2.19 – 2.03	2151	12.3	8.1	94.4
2.03 – 1.90	1533	8.5	13.8	93.2
1.90 – 1.79	889	6.0	18.5	91.7
1.79 – 1.70	521	3.8	27.3	82.9

3.2.3. Solvent content.

The Mathews Coefficient and solvent content are calculated from the unit cell and the molecular weight of the molecules in the unit cell according to the following equation:

$$V_m = \frac{\text{Volume of unit cell (a x b x c) } Å^3}{\text{Mr of 1 asymmetric unit x no. a.s.u. Da}}$$

$$\text{Solvent fraction} = 1 - (1.23/ V_m)$$

A Matthews coefficient (V_m) of 2.5 $Å^3$ Da^{-1}, corresponding to a 49% solvent content, suggested one molecule of the nanI sialiadase in the asymmetric unit (Mathews, 1968).

3.3. Structure solution of the catalytic domain of nanI.

The method of molecular replacement, as implemented in the program CNS (Brunger *et al* ., 1998b) was used to estimate an initial set of phases from which the structure of the nanI catalytic domain was solved. Rossman and Blow (Rossman and Blow, 1962) pioneered this method in the 1960's as a novel solution to The Phase Problem (Taylor, 2003). The term molecular replacement is somewhat misleading because nothing is actually 'replaced'. The conventional understanding of what molecular replacement encompasses is the placement of one or more known protein models in the unit cell of the crystal under study. The search or starting models are usually extracted from the Protein Data Bank (Berman *et al* ., 2003; Berman *et al* ., 2000) or different crystal forms that were solved previously. The initial structure of nanI was solved using the data from the Neu5Ac2en co-crystals.

Placing a molecular model in a unit cell constitutes a six-dimensional search problem. The six degrees of freedom are most conveniently parameterised as the three rotation angles and three translations along the basis vectors of the coordinate system. Conventionally, an asymmetric unit (the volume of the search space which is unique under symmetry) is sampled on a uniform grid. For a typical macromolecular unit cell, the product of angular and translational sampling points is usually too large to carry out an exhaustive six-dimensional search in a reasonable amount of time using current computing power. However, a new procedure for molecular replacement has been presented, in which an efficient six-dimensional search is carried out using an evolutionary optimisation algorithm (Kissinger *et al* ., 1999; Kissinger *et al* ., 2001).

Rossmann and Blow (Rossmann and Blow, 1962) showed that the six-dimensional problem presented in molecular replacement can be broken down into two three-dimensional searches: the rotation search to find the angular orientation of the model and the translation search, to find its position within the

unit cell. This greatly reduces the demand for computing resources and is the method of choice in most molecular replacement programs.

In the *Crystallography and NMR system* (CNS) (Brünger *et al* ., 1998a), a third powerful procedure is usually introduced between the rotation search and the translation search: Patterson correlation (PC) refinement of the molecular orientation. The goal of Patterson correlation refinement is to improve the overall orientation of the search model after the rotation function. By improving the accuracy of the search model for the correct rotation, PC refinement improves the discrimination between correct and incorrect orientations. This enables the location of the correct peak in a noisy rotation function.

3.3.1. Starting model.

In order for molecular replacement to have any chance of success, a number of key parameters should be met for the search model in relation to the unknown structure. These include a sequence identity > 25 % and an r.m.s. deviation of < 2 Å between the C_α atoms of the model and the final new structure, although there are exceptions to this. The search for a suitable starting model for the nanI structure began with a look at the current sialidase structures deposited in the PDB (Berman *et al* ., 2000). A comparison of their amino acid sequences using the FASTA (Pearson and Lipman, 1988) program LALIGN as implemented at the EMBnet server, using the default parameters, was made Table 3.3.1.1.

Table 3.3.1.1. Sequence comparison of available bacterial, eukaryotic and viral sialidase sequences whose structures had been deposited in the PDB.

Organism	PDB code	% Identity to nanI Amino acid sequence
Salmonella typhimurium	1DIL	30.7
Micromonospora virdifaciens	1EUS	29.5
Vibrio cholerae	1KIT	23.3
Macrobdella decora	1SLL	35.0
Trypanosoma cruzi	1S0I	32.3
Trypanosoma rangeli	1N1S	33.0
Influenza A virus N2 subtype	1INH	19.0

The intramolecular trans-sialidase from the leech *Macrobdella decora* (PDB 1SLL) was chosen, as it showed the greatest sequence identity, 35 %, to the nanI sialidase structure. The sequence alignment of the leech sialidase with nanI was studied to identify which regions of the two proteins shared identity, Figure 3.3.1.2. When trying to molecular replace an unknown structure, it is important to make the model as accurate a representation of the unknown structure as possible. A number of guesses must inevitably be made. In this case, the regions of the leech sialidase, which were identical to nanI at the primary structure level, were kept and all other amino acids were changed to alanine.

The starting model consisted of just the β-propeller domain, with the N-terminal lectin domain and irregular β-strand domains removed. Figure 3.3.1.3. shows a cartoon representation of the leech IT sialidases with the catalytic

domain shown in orange that was used in the molecular replacement of the nanI structure. The N-terminal lectin domain was removed from the model due to the lack of any such domain in the nanI crystals and the OB fold domain between blades three and four was also excluded due to lack of reasonable sequence identity with nanI. These domains are coloured grey in Figure 3.3.1.3.

Figure 3.3.1.2. Sequence alignment of *C.perfringens* nanI with *M.decora* IT sialidase (1SLL). The regions of the structure used as the starting model are shown in orange. The cleavage site of the nanI sialidase is indicated by a triple asteryx coloured red sitting above the cleavage site.

```
1SLL__          ---------IPEGILMEKNNVDIAEG---------QGYSLDQEAG--------------- 27
nanI__          MNYKGITLILTAAMVISGGNYVLVKGSTLDSGKNNSGYEIKVNNSENLSSLGEYKDINLE 60
                        :. .:::. .*  :.:*         .**.:. : .

1SLL__          ----------AKYVKAMTQGTIILSYKSTSENGIQSLFSVGNSTAGNQDRHFHIYITNSG 77
nanI__          SSNASNITYDLEKYKNLDEGTIVVRFNSK-DSKIQSLLGISNSKTKNG--YFNFYVTNSR 117
                          :  * : :***:: ::*. :. ****:.:.**.: *   :*::*:***

1SLL__          GIGIELRNTDGVFNYTLDRPASVRALYKGERVFN---TVALKADAANKQCRLFANGELLA 134
nanI__          -VGFELRNQKNEGN-TQSGTENLVHMYKDVALNDGDNTVALKIEKN-KGYKLFLNGKIIK 174
                 :*:**** ..  * * . . .:  :**.  : :   ***** :   *  :** **:::

1SLL__          TLDKDAFKFISDITGVDNVTLGGTKRQGK-IAYPFGGTIGDIKVYSNALSDEELIQATGV 193
nanI__          EVKDTNTKFLNNIENLDSAFIGKTNRYGQSNEYNFKGNIGFMNIYNEPLGDDYLLSKTGE 234
                 :..   **:.:* .:*.. :* *:* *:   * * *.** :::*.:.*.*: *:. **
                       ***
1SLL__          TTYGENIFYAGDVT------------ESNYFRIPSLLTLSTGTVISAADAR-YGGTHDSK 240
nanI__          TKAKEEVLVEGAVKTEPVDLFHPGFLNSSNYRIPALFKTKEGTLIASIDARRHGGADAPN 294
                *.  *::::  * *.            :*. :***:*:. . **:*:: *** :**:. .:

1SLL__          SKINIAFAKSTDGGNTWSEPTLPLKFDDYIAKNIDWPRDSVGKNVQIQGSASYIDPVLLE 300
nanI__          NDIDTAVRRSEDGGKTWDEGQIIMDYPD--------------------KSSVIDTTLIQ 333
                ..*: *. :* ***:**.*  : :.: *                     .:* **..*::

1SLL__          DKLTKRIFLFADLMPAGIGSSNASVGSGFKEVNGKKYLKLRWHKDAGRAYDYTIREKGVI 360
nanI__          DDETGRIFLLVTHFPSKYGFWNAGLGSGFKNIDGKEYLCLYDSSGK----EFTVREN-VV 388
                *. * ****:.  :*:  *  **.:*****:::**:** *   ..     ::*:**: *:

1SLL__          YNDATNQPTEFRVDGEYNLYQHDTNLTCKQYDYNFSGNNLIESKTDVDVNMNIFYKNSVF 420
nanI__          YDKDGNK-TEYTTNALGDLFKNGTKIDN------------------------INSSTAPL 423
                *:.  *: **: .:.  :*:::.*::                          *  ..: :

1SLL__          KAFPTNYLAMRYSDDEGASWSD-LDIVSSFKPEVSKFLVVGPGIGKQISTGENAGRLLVP 479
nanI__          KAKGTSYINLVYSDDDGKTWSEPQNINFQVKKDWMKFLGIAPGRGIQIKNGEHKGRIVVP 483
                **  *.*: : ****:* :**:  :*  ..* :  *** :.** * **..**: **::**

1SLL__          LYS---KSSAELGFMYSDDHGDNWTYVEADN----LTGG-------------ATAEAQIV 519
nanI__          VYYTNEKGKQSSAVIYSDDSGKNWTIGESPNDNRKLENGKIINSKTLSDDAPQLTECQVV 543
                :*    *.. . ..:**** *.***  *: *    * .*               :*.*:*

1SLL__          EMPDGSLKTYLRTGSNCIAEVTSIDGGETWSDRVPLQGISTTSYGTQLSVINYSQPIDGK 579
nanI__          EMPNGQLKLFMRNLSGYLNIATSFDGGATWDETVEKDTNVLEPY-CQLSVINYSQKVDGK 602
                ***:*.** ::*. *. :  .**:*** **.: *  :     .*  ********* :***

1SLL__          PAIILSSPNATNGRKNGKIWIGLVNDTGNTGIDKYSVEWKYSYAVDTPQMGYSYSCLAEL 639
nanI__          DAVIFSNPNARS-RSNGTVRIGLINQVGTYENGEPKYEFDWKYNKLVKPGYYAYSCLTEL 661
                 *:*:*.*** . *.**.: ***:*:.*.   .: . *:.:.*   .    *:****:**

1SLL__          PDGQVGLLYEKYDSWSRNELHLKDILKFEKYSISELTGQA 679
nanI__          SNGNIGLLYEGTPSEEMSYIEMN--LKYLESGANK----- 694
                .:*::*****   * . . :.::  **: : . .:
```

Figure 3.3.1.3. A cartoon representation of the leech IT-sialidase three-dimensional crystal structure (1SLL). The region of the structure used in the molecular replacement of the nanI structure is shown in orange and corresponds the equivalent colour scheme in Figure 3.2. N and C-termini are coloured blue and red respectively.

3.3.2. Molecular Replacement.

A molecular replacement solution was obtained using the program CNS (Brunger *et al* ., 1998a). The structure of the intramolecular trans-sialidase from the leech, Macrobdella decora (PDB code 1SLL) was used as a starting model, as described in section 3.3.1. Using data in the range 15 – 4 Å, a clear cross-rotation solution was found with a height of 0.089, 5.9 σ above the mean. The next highest rotation peak was 0.046, 2.3 σ above the mean. The top ten solutions of the cross-rotation function are displayed in Table 3.3.2.1. The rotations identified in the rotation search are applied to the search molecule and the translational position in the unit cell is determined. The top translation function, using the same data limits, gave a monitor value (correlation coefficient) of 0.305, 5 σ above the mean. The next solution had a monitor value of 0.121, 1.09 σ above the mean, see Table 3.3.2.2. Using the molecular visualisation program O, no unfavourable molecular contacts were observed for this solution in the crystal packing. The MR solution was then applied to the coordinates of the search model to work as the starting model in the refinement stage, section 3.3.3.

Table 3.3.2.1. Top ten cross-rotation function solutions in MR for the nanl search model. The top two solutions were passed onto the cross-translation function for testing.

Solution	θ_1	θ_2	θ_3	[a]RF
1	289.03	80.76	176.56	0.089
2	204.62	90.00	196.52	0.046
3	210.45	95.76	190.01	0.043
4	348.37	84.24	302.08	0.043
5	302.23	69.24	184.97	0.042
6	278.30	27.12	158.30	0.039
7	331.69	75.00	336.49	0.038
8	156.81	57.12	74.01	0.036
9	284.25	89.40	179.34	0.036
10	291.43	66.36	202.97	0.036

[a]RF is the correlation coefficient between squared normalised structure factors.

Table 3.3.2.2. Top two translation function solutions in MR for the nanl search model.

Solution	θ_1	θ_2	θ_3	TX	TY	TZ	monitor
1	289.03	80.76	176.56	20.16	0.00	0.00	30.5
2	204.62	90.00	196.52	16.13	21.80	6.74	12.1

3.3.3. Model Building and Refinement.

The solution was applied to the nanI search model and was subsequently used in the model building and refinement process. The initial dataset used to build and refine the nanI model was that collected for the Neu5Ac2en co-crystals. Subsequently the apo and covalent intermediate datasets were refined against this model using SHELXL97 and REFMAC5 respectively. The isomorphous character of the crystals allowed immediate refinement against the newly built nanI model without any need to re-place the model in the unit cell.

Refinement is an iterative process whereby the difference between experimentally observed structure factors (F_{obs}) and model-derived structure factors (F_{calc}) is minimised. This minimisation involves the refinement of the x, y and z coordinates, with the temperature factor (B-factor), of each atom in the model. This is achieved by both manual adjustment in a graphics package, for example O or COOT, to match the electron density and by use of refinement programs such as CNS (Brünger *et al* ., 1998a), SHELXL97 (Sheldrick and Schneider, 1997b) and REFMAC (Murshudov, 1997). The quality of the resultant model can be judged by a number of qualitative and quantitative factors. Usually, the model is judged by comparing the standard crystallographic Refinement Factors, R and free R, as well as checking that all torsion angles, bond-lengths and B-factors are within normal ranges. The free R value is entirely diffraction data based and as such is a particularly powerful tool in structure validation (Brünger, 1997).

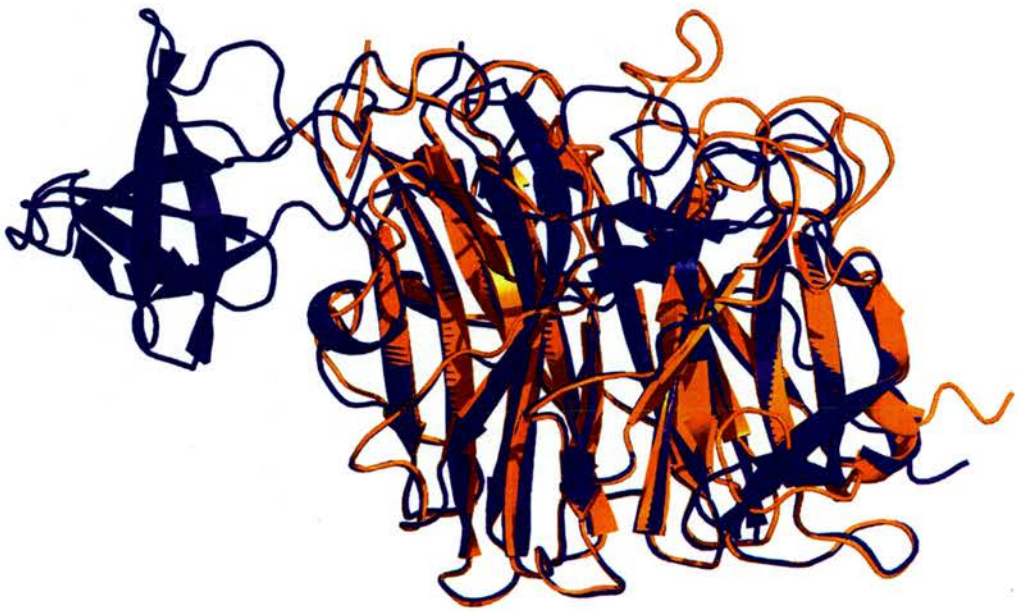
CNS was used to refine the initial nanI model using rigid body and combined simulated annealing and energy minimisation refinement. Data in the range 500 – 2.0 Å was used. Unfortunately, no interpretable electron density for the missing parts of the model was observed in the calculated maps. Most notably the unknown domain formed by amino acids 406 – 505, between strands two and three of the second blade in the nanI β-propeller. After ten rounds of refinement, the final R and free R-values from the simulated annealing

refinement were 42.5 and 46.7 %, respectively. These values indicated that the model was essentially correct, but that the phase information at this stage was too poor to provide any information on the location of amino acids 406-505.

The automated protein model building and structure refinement package, ARP/wARP version 6.1 (Perrakis *et al* ., 1999) was employed to extend the phases of the model and build in the missing amino acids. The warpNtrace protocol was used, which entails automated model building starting from an existing model. The experimentally measured structure factor amplitudes are input into the program along with an initial model, from which starting phases are calculated. Data in the range 40 – 1.6 Å was used with the nanI model from the combined simulated annealing refinement. WarpNtrace was able to build 445 of a possible 448 residues within two chains. The connectivity index was 0.99 and using the sidedock option in warpNtrace, side chains were assigned to these residues with a 98 % confidence level. The starting R and R_{free} were 43.5 and 42.4 % respectively. By the end of 50 cycles, these values had dropped to 16.0 and 19.5 % and an essentially complete model with solvent was output, Figure 3.3.3.1. The domain formed by amino acids 406-505 had been completely built and this acted as a control, confirming the validity of the new phase information contained within the model.

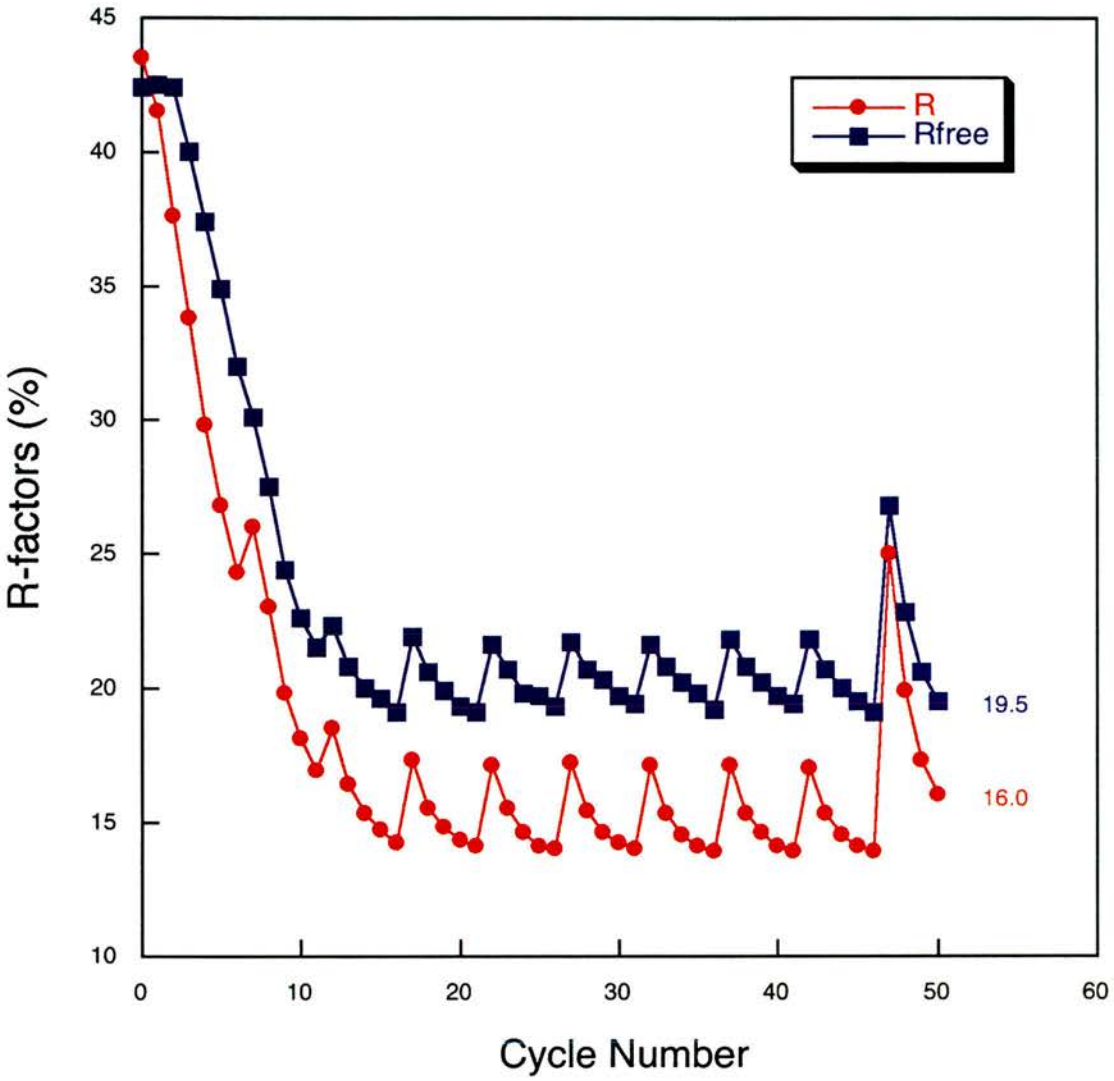
Figure 3.3.3.1. Superimposition of the nanI structures before and after auto building in ARP/wARP. (a) The initial model is shown in orange, which corresponds to the same colour scheme in Figure 3.2. The blue model shows the result of auto building anf refinement in ARP/wARP. (b) Image rotated 90°.

(a)



(b)

Figure 3.3.3.2. Summary of the ARP/wARP auto building and refinement.



From this point the model was refined with SHELX-97 (Sheldrick and Schneider, 1997b) against 92 % complete, 0.97 Å resolution, native data to a final R factor of 11.3 and free R 13.0 %. SHELXL97 enabled the inclusion and refinement of anisotropic B-factors and the calculation of atom positional esd's (estimated standard deviations) more recently referred to as standard uncertainties (su's) (Schwarzenbach *et al* ., 1995). In total, 15 cycles of conjugate gradient least squares (CGLS) refinement were carried out between rounds of model building using the program O (Jones *et al* ., 1990; Jones *et al* ., 1991) and sigma A weighted $2F_O$-F_C and un-weighted F_O-F_C electron density maps. A final round of blocked, full-matrix least squares refinement was carried out on the model including all data to obtain su's on all geometric parameters. The refinement statistics are shown in Table 3.3.3.3.

The initial R and free R were 26.0 and 29.7 % respectively, at the start of the refinement in SHELXL97 . This contrasts with the final values from ARP/wARP of 19.5 and 16 % respectively. This discrepancy can be explained due to the removal of the solvent atoms built in by ARP/wARP and the different methods of refinement between REFMAC and SHELXL97. In particular, SHELXL97 refines against F^2 rather than F. This enables all data to be used in the refinement with weights that include contributions from experimental uncertainties, rather than having to reject F values below a preset threshold (Sheldrick and Schneider, 1997b). In essence, this allows more experimental information to be employed in the refinement. As can be seen from Figure 3.3.3.4., the R and free R quickly return to levels one would expect from such a complete model.

The auxiliary program, SHELXWAT (Sheldrick and Schneider, 1997b) , was used to model the solvent shell around the protein. Cycle 3 and 8 in Figure 3.3.3.4. show the effects of modelling in the solvent on the refinement. The first round of solvent divining found 400 waters, with a concomitant decrease in the free R of 2.2 %. The second round in cycle 8 introduced a further 48 waters, with a decrease in free R of 0.8 %.

The inclusion of all data to the full resolution of 0.97 Å produced a drop in the free R of 2.2 %. The electron density maps after refinement against these data clearly showed anisotropc vibrations throughout the molecule and a switch to an anisotropic model of atom vibrations was made in cycle 6. The transition from isotropic to anisotropic refinement roughly doubles the number of parameters to refine against and almost always results in an appreciable reduction in the R factor. However, this represents an improvement in the model only when a significant drop in the free R accompanies it. Since the free R factor is itself subject to uncertainty, due to the small sample used, a drop of at least 1 % is needed to justify moving to anisotropic refinement (Sheldrick and Schneider, 1997b). In the nanl refinement, moving to anisotropic refinement resulted in a drop in free R of 2.0 %, which clearly justified this move.

The geometry of the model in SHELXL97 was maintained by imposing Engh & Huber restraints (Engh and Huber, 1991) on the bond angles and lengths. The stringency of the restraints can be optimised empirically by adjusting the standard deviations that are applied to the restraint target values. During the refinement of nanl, the list of disagreeable restraints in the output .lst file showed no bad violations and so these values were not changed from the default ones generated. Towards the end of refinement, if the data is of sufficient resolution, SHELXL97 allows the refinement of 'riding' hydrogen's on the atoms in the model. The H coordinates are not directly refined, but they 'ride' on their supporting atoms, providing a small amount of scattering material approximating to where the H atoms would be.

It was clear during the refinement that the diffraction data contained information on the location of certain protons in the molecule. Inspection of the F_O-F_C electron density maps, contoured at 3 σ, showed clear peaks of density at positions one would expect H atoms to be sitting (Figure 3.3.3.3). Not all residues showed this fine detail, the protons on the aromatic side chains of tyrosine and phenylalanine where particularly prominent. It is recommended that the OH hydrogen's and those on histidine residues be excluded from the refinement, as these can cause significant problems in the refinement. The

improvement attributed to the addition of H atoms was a drop in free R of 1.3 %, cycle 14 Figure 3.3.3.4.

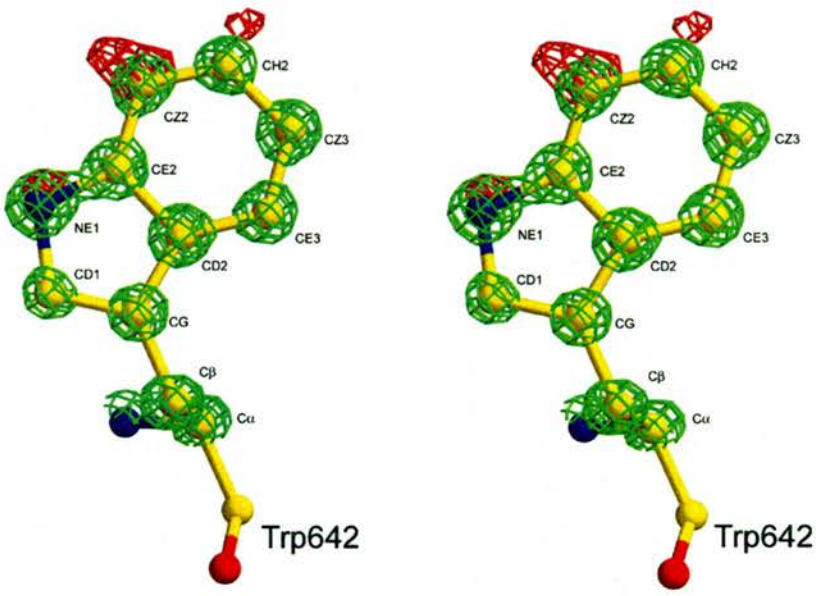


Figure 3.3.3.3. Stereo view showing the electron density around Trp642 during refinement. The $2F_O$-F_C electron density map (green) is contoured at 2 σ. The F_O-F_C difference density map (red) is contoured at 3 σ. Density can be seen around atoms NE1, CZ2 and CH2 in the F_O-F_C map, corresponding to protons not present in the model.

The final round of refinement was used to calculate the standard uncertainties, or estimated standard deviations (e.s.d's), of the atomic positions. This was achieved by refining the nanl model in SHELXL97 using blocked matrix least squares mode (Sheldrick and Schneider, 1997b). All restraints are removed from the refinement, however refinement of atomic positions is turned off so that the program cannot move the atoms from their final restrained positions. Refinement of atomic positions should be complete before estimating the standard uncertainty of their coordinates! In addition, blocked matrix least-squares refinement was also carried out on the restrained model in order to allow a comparison to be made. The values obtained and their significance is discussed in Chapter 4 on the analysis of the nanl structure.

The geometry of the final nanI model was in good agreement with the average geometry of well-refined structures as judged by the values of rms (root mean square deviation) of both bond lengths and bond angles (Rhodes, 2000). Deviations of bond lengths < 0.02 Å and angle deviations of < 4° are expected for good quality models. In addition, the atomic displacement parameters (ADPs), or B-factors, in $Å^2$ were also investigated. Analysis of the main-chain and side-chain B-factors is another measure of the quality of the model (Kleywegt, 2001; Kleywegt and Brunger, 1996). Higher B-factors reflect larger fluctuations of the atoms about their mean positions, which results in less accurate determination of atomic coordinates. The average B-factor for the whole nanI chain was 8.3 $Å^2$, as calculated by the CCP4 program BAVERAGE (CCP4, 1994). Such a low value for the average temperature factor of the protein suggests a very rigid molecule. The significance of this on the biochemistry of the nanI sialidase will be discussed further in Chapter 4.

The final nanI model was built using the data collected from the Neu5Ac2en co-crystals. From the start of the refinement it was clear that we had trapped what we thought was Neu5Ac2en in the active site. The difference map, F_O - F_C , showed clear density for the compound. However, after putting a description of the compound into the SHELXL97 refinement, a significant peak of density was observed approximately 1.2 Å from the C2 carbon. This distance corresponds to a carbon – oxygen single bond and immediately suggested that α-Neu5Ac had been trapped, not the dehydrated transition state analogue Neu5Ac2en. The appropriate changes were made to the SHELXL97 input file to refine the structure of α-Neu5Ac in the Neu5Ac2en co-crystal data. The significance of this finding is discussed in Chapter 5, CPNA Ligand Binding Studies.

Following the refinement of the final model for the nanI catalytic domain from the newly discovered α-Neu5Ac complex, the model was used to refine against both the atomic resolution apo and the covalent glycosyl-enzyme intermediate datasets. The apo dataset was refined using SHELXL97 in a similar way to that already described for the Neu5Ac2en data and summarised in Table

3.3.3.3. The data collected in house for the covalent glycosyl-enzyme intermediate was refined using the CCP4 program REFMAC5 (v5.2.005) (Murshudov, 1997). Lower resolution data is better refined using the Maximum Likelihood methods employed in REFMAC5 than the Least-squares method used in SHELXL97. The refinement of all three datasets is summarised in Table 3.3.3.3. The description of the covalent bond between the oxygen of the catalytic tyrosine (655) and C2 carbon of the 3F Neu5Ac compound was identical to that used in the refinement of a similar intermediate in the *T.cruzi* trans-sialidase (Amaya *et al* ., 2004) PDB code: 1S0K. The significance of this structure and the information it provides are discussed in Chapter 5 along with a detailed comparison with the *T.cruzi* complex.

Figure 3.3.3.4. Summary of the nanl refinement history in SHELXL97. This graph shows the refinement for the first model using the data from the Neu5Ac2en co-crystals.

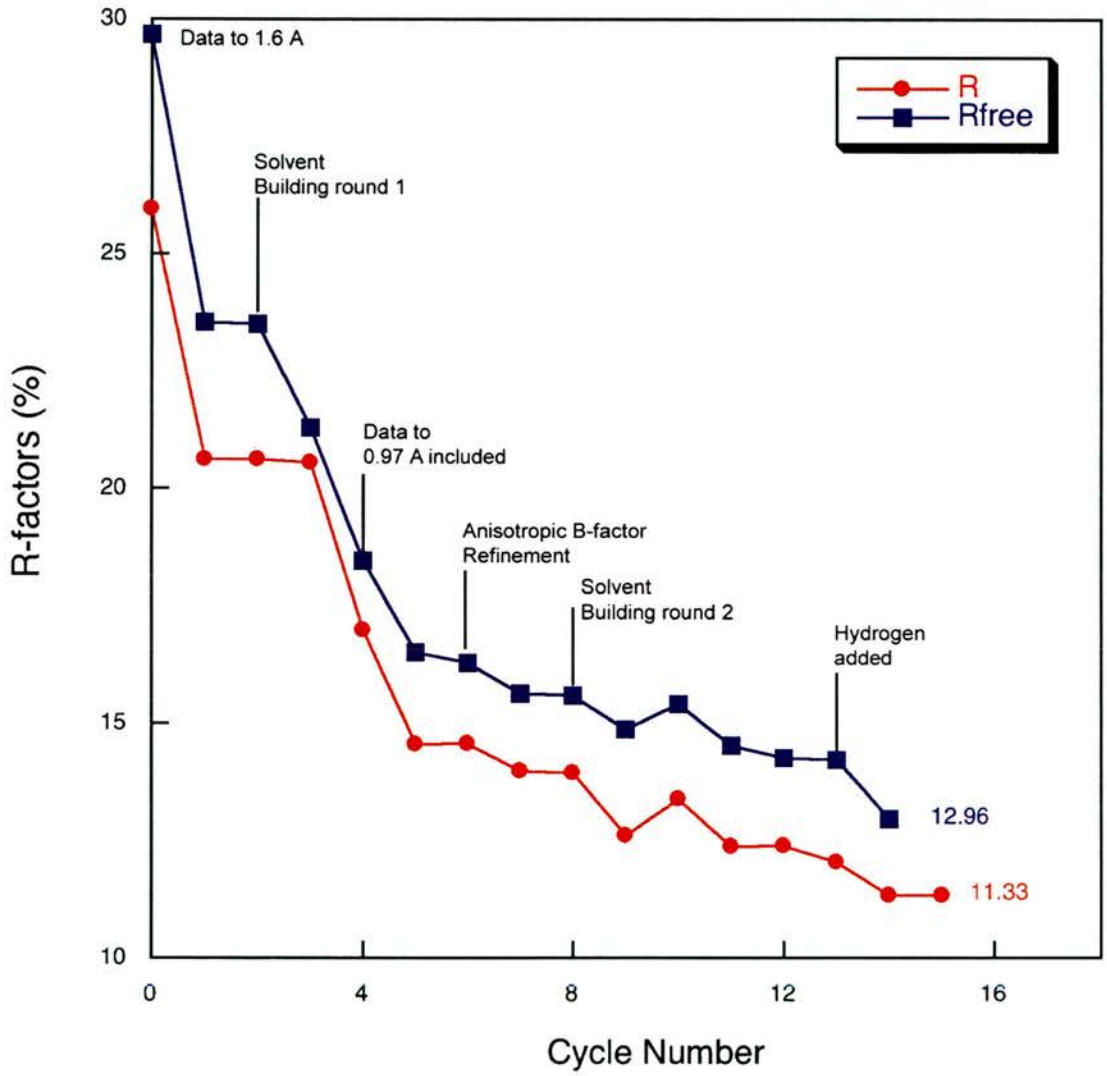


Table 3.3.3.3. Refinement statistics for the nanI apo and complex structures.

Structure	nanI-apo	nanI - Neu5Ac2en	nanI-3Fneu5Ac
Resolution range (Å)	20-0.97	20-0.97	58-1.7
R-factor[1] (%)	11.52	11.33	16.3
R-free[2] (%)	12.60	12.96	19.9
Rms deviation in bond length (Å)	0.017	0.016	0.021
Rms deviation bond angles (°)	2.048	2.086	1.715
Number of protein atoms	3575	3575	3575
Number of water molecules	935	484	581
Average B-factor all atoms (Å^2)	8.9	8.3	10.1
Main-chain atoms	7.2	6.6	9.1
Side-chain atoms	10.5	10.0	11.1
α Neu5Ac	---	10.4	---
3F-Neu5Ac	---	---	8.5
Glycerol		10.2	---

[1]R-factor = $\sum_{hkl} ||F_{obs} - |F_{calc}|| / \sum_{hkl} |F_{obs}|$.
The final R factor for the atomic resolution structures was calculated using all the data from the experiment.
[2]R-free is calculated in the same way as the R-factor, but for a 5 % test set which is excluded from the refinement.

3.3.4. Structure Validation.

The refined structures of the apo enzyme and the two complexes were assessed using the programs PROCHECK (Laskowski *et al* ., 1993) and WHATIF (Vriend, 1990). Figures 3.6-3.8 display the Ramachandran plots of the main chain torsion angles (ϕ,ψ) as well as the parameters for the main and side chains. In the three structures modelled, none of the residue torsion angles existed in the disallowed region of the Ramachandran plot. Comparing the Ramachandran plots for the three models shows that all have the same residues in the generously allowed regions of the plot. These are, ILE 267, ASP 414, TYR 587, ALA 654 and ASP 328. The last residue is only present in the plots for the atomic resolution models. Closer inspection of the Ramachandran plot for these models shows that the residue is on the very edge of the additionally allowed region I. The φ and ψ angles for the covalent complex were 73.5 and 93.7 ° respectively. These compared to the Apo and Neu5Ac complex models of 78.9, 89.4 and 75.3, 84.7 °. There was no obvious explanation for the differences from the electron density maps for these models. It is more probably that the true φ, ψ torsion angles for this residue lie in the regions highlighted above for the atomic resolution models due the increased accuracy of these models over the lower resolution 1.7 Å model.

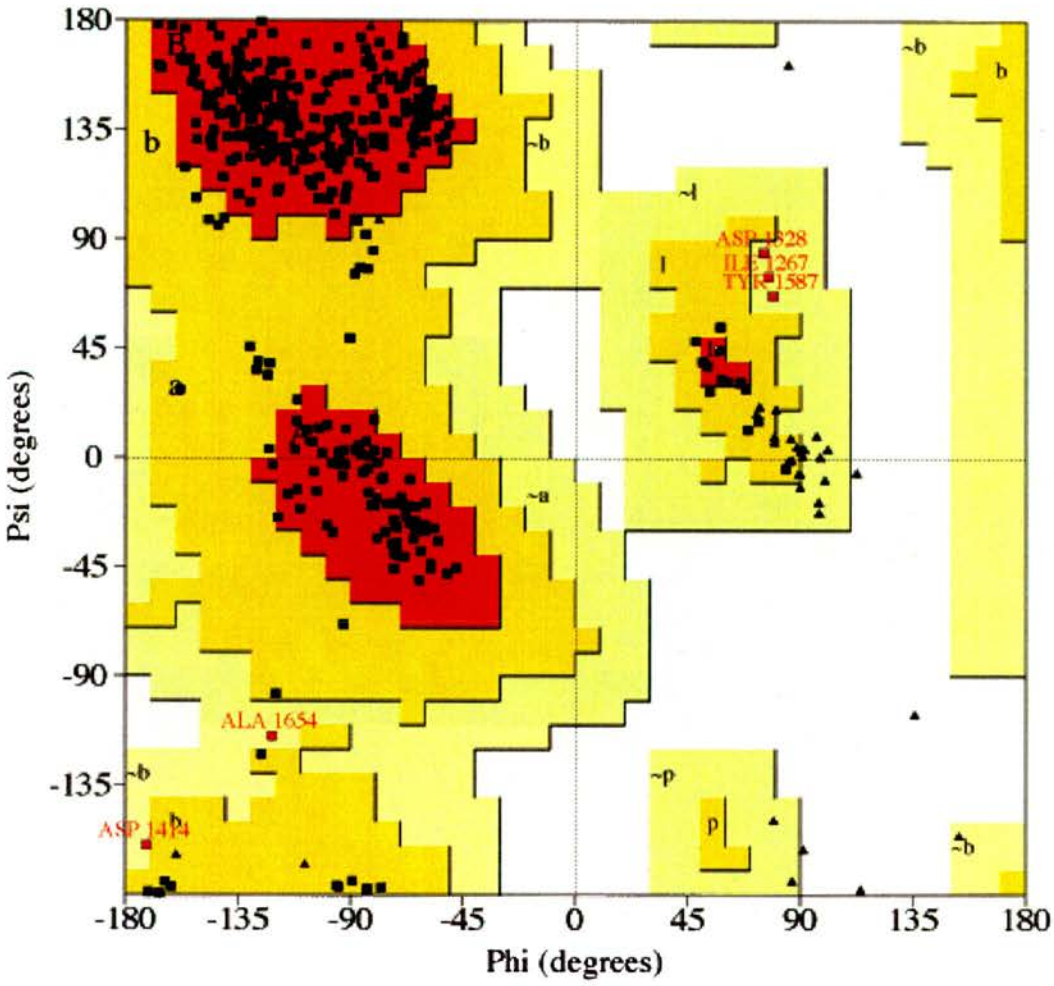
Coordinate uncertainty is a statistic that deserves attention when considering the accuracy and precision of atomic positions within an experimentally determined protein model (Cruickshank, 2001). It is well documented that at the conclusion and full convergence of a least-squares or equivalent refinement, the estimated variances and covariences of the parameters may be obtained through the inversion of the least-squares full matrix. This is possible when refining a model using atomic resolution data and is implemented in the program SHELXL97. This is not always possible when dealing with diffraction data that does not extend this far. In such cases a

simplified error estimation is used, called the Diffraction Precision Index, DPI, and is implemented in REFMAC5 (Murshudov, 1997; Murshudov and Dodson, 1997).

Following the convergence and completion of conjugate gradient least squares refinement for the two atomic resolution models, the unrestrained full matrix (coordinates only) was computed and then inverted. This led to su's $\sigma(x)$, $\sigma(y)$, $\sigma(z)$ and $\sigma(r)$ for all atoms and to $\sigma(l)$ and $\sigma(\theta)$ for all bond length and angles. $\sigma(r)$ is defined as the vector between the atom and the centre of the coordinate system (0,0,0) and is written as $[\sigma^2(x) + \sigma^2(y) + \sigma^2(z)]^{1/2}$.

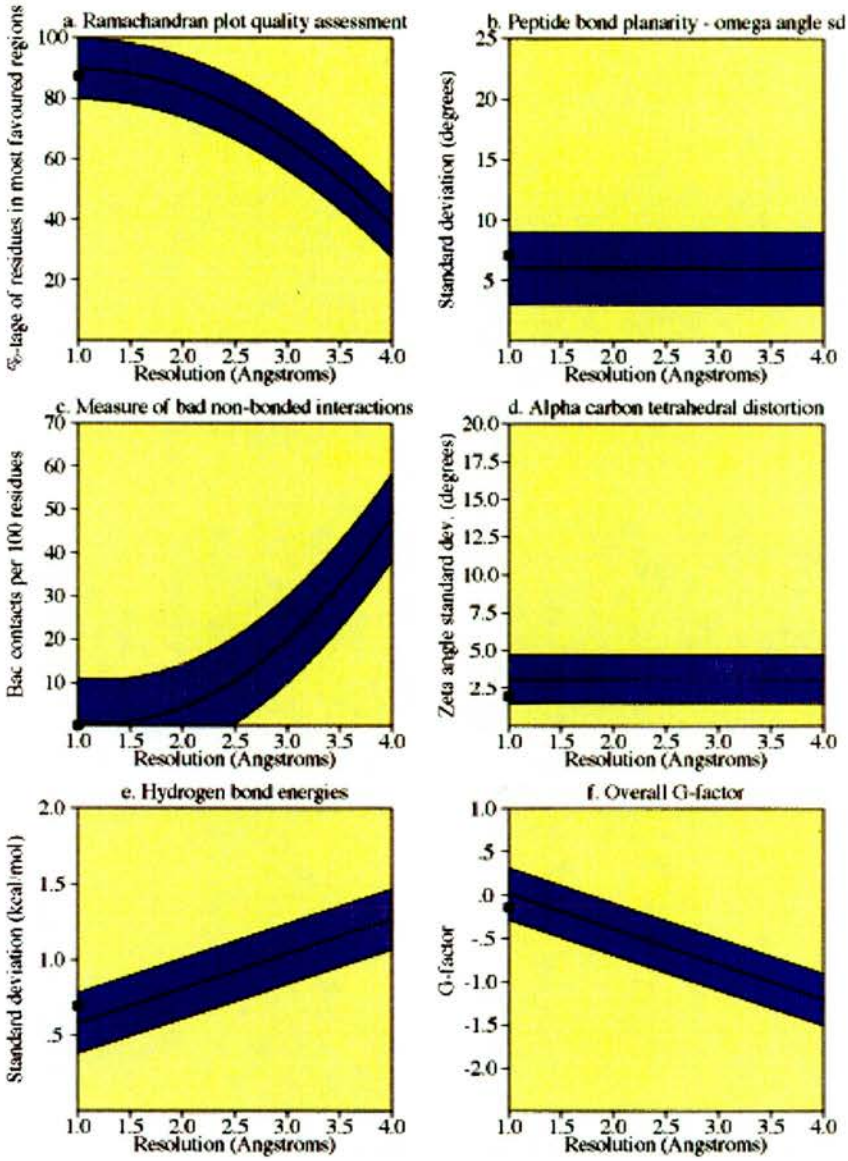
The results for the inversion of the coordinate matrices for the nanI apo atomic resolution structure are shown in Figures 3.9 and 3.10. Figure 3.9 shows σ (r) *versus* B_{eq} for the fully occupied atoms of the protein. The points are colour-coded black for carbon, blue for nitrogen and red for oxygen. Superposed on these data points are least-squares quadratic fits determined with weights $1/B^2$. These plots are generated from the output of the full-matrix least-squares refinement by the auxiliary program SHELXPRO (Sheldrick and Schneider, 1997a). For B < 10 Å^2, the better precision of oxygen as compared with nitrogen, and of nitrogen as compared to carbon, can be clearly seen. For B < 10 Å^2 the σ (r) falls below 0.02 Å are around 0.01 Å at B = 5 Å^2.

Figure 3.10 shows σ (l) *versus* B_{eq} for the bond lengths in the protein. The points are colour-coded black for C-C, blue for C-N and red for C-O. For B < 10 Å^2, the distribution falls to around 0.02 Å, which is the standard uncertainty of the applied restraint for 1-2 bond lengths (Engh and Huber, 1991). It can be seen from Figure 3.10 that many bond lengths with average B < 8 Å^2 have $\sigma_{diff}(l)$ < 0.015 Å. For these bonds the diffraction data have greater weight than the stereochemical dictionary used to describe them. This situation is a consequence of having diffraction data to such high resolution.



Residues in most favoured regions [A, B, L]	339	87.4 %
Residues in additional allowed regions [a, b, l, p]	44	11.3 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	5	1.3 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	388	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	42	
Number of proline residues	18	
Total number of residues	449	

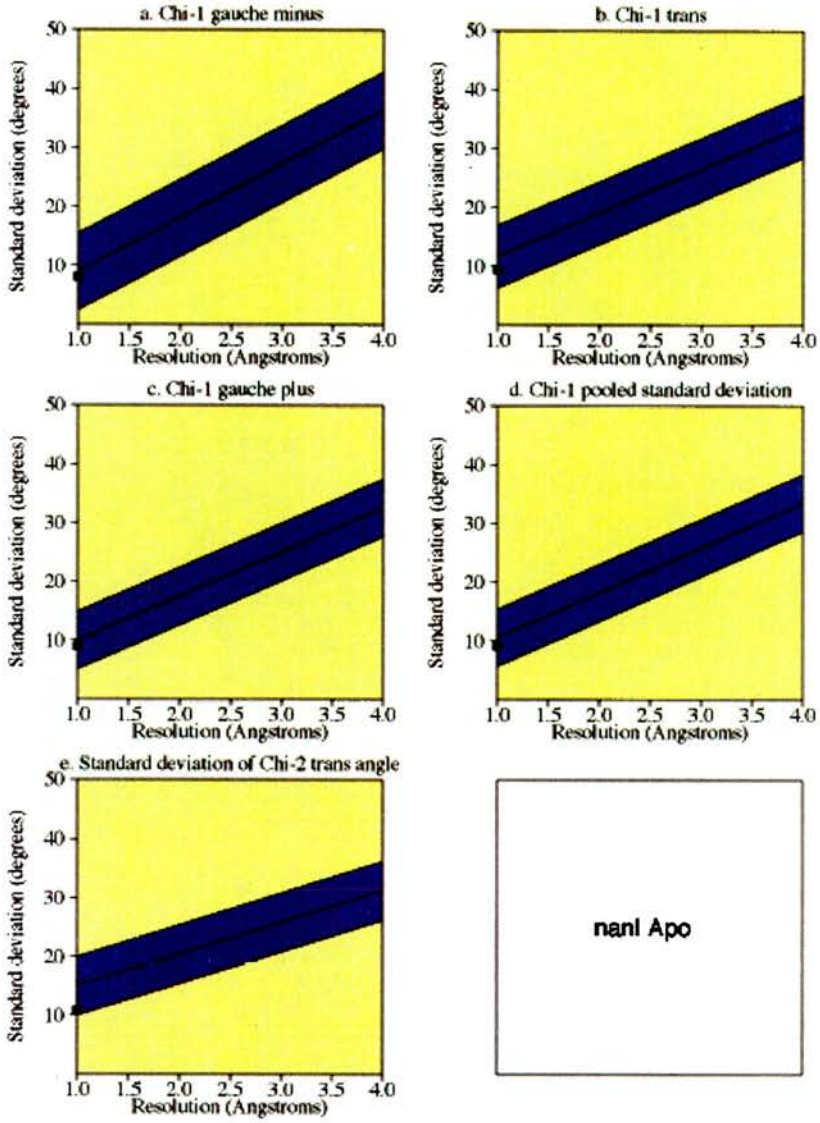
Figure 3.6a. Ramachandran plot for nanI Apo structure.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	388	87.4	89.7	10.0	-.2	Inside
b. Omega angle st dev	446	7.1	6.0	3.0	.4	Inside
c. Bad contacts / 100 residues	0	.0	1.0	10.0	-.1	Inside
d. Zeta angle st dev	407	2.0	3.1	1.6	-.7	Inside
e. H-bond energy st dev	244	.7	.6	.2	.6	Inside
f. Overall G-factor	449	-.1	.0	.3	-.5	Inside

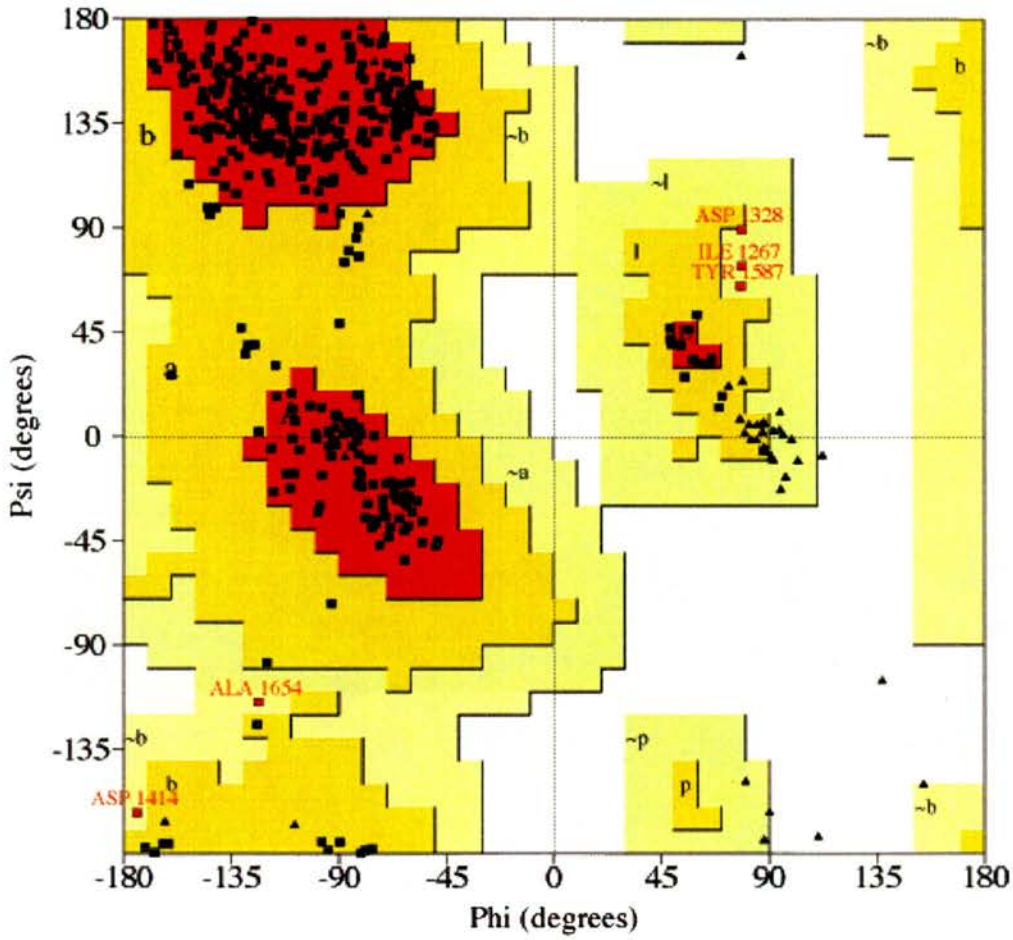
Figure 3.6b. Main-chain parameters for nanl Apo structure.



Plot statistics

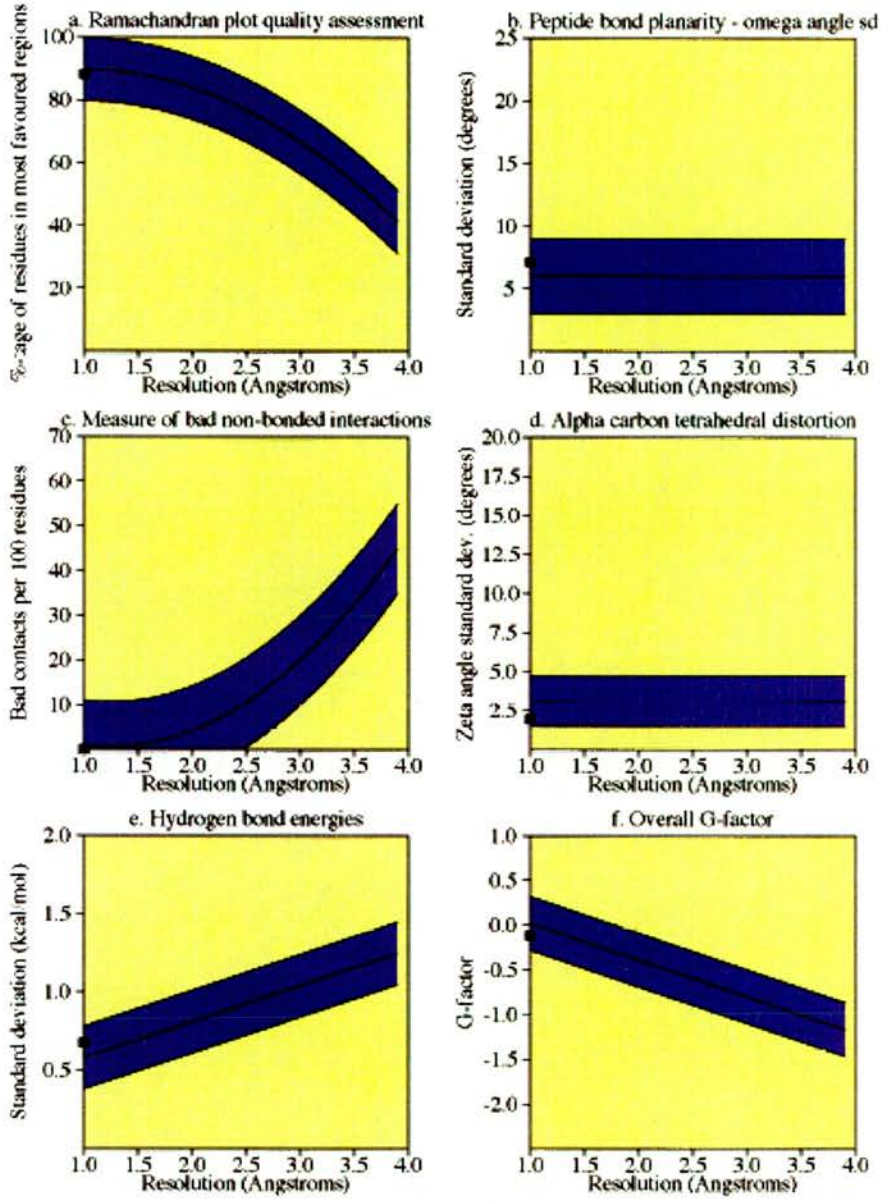
Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	81	8.1	9.0	6.5	-.1	Inside
b. Chi-1 trans st dev	109	9.5	11.6	5.3	-.4	Inside
c. Chi-1 gauche plus st dev	180	9.0	10.0	4.9	-.2	Inside
d. Chi-1 pooled st dev	370	9.2	10.5	4.8	-.3	Inside
e. Chi-2 trans st dev	115	10.8	15.0	5.0	-.8	Inside

Figure 3.6c. Side-chain parameters for nanl Apo structure.



Residues in most favoured regions [A, B, L]	342	88.1 %
Residues in additional allowed regions [a, b, l, p]	41	10.6 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	5	1.3 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	388	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	42	
Number of proline residues	18	
Total number of residues	449	

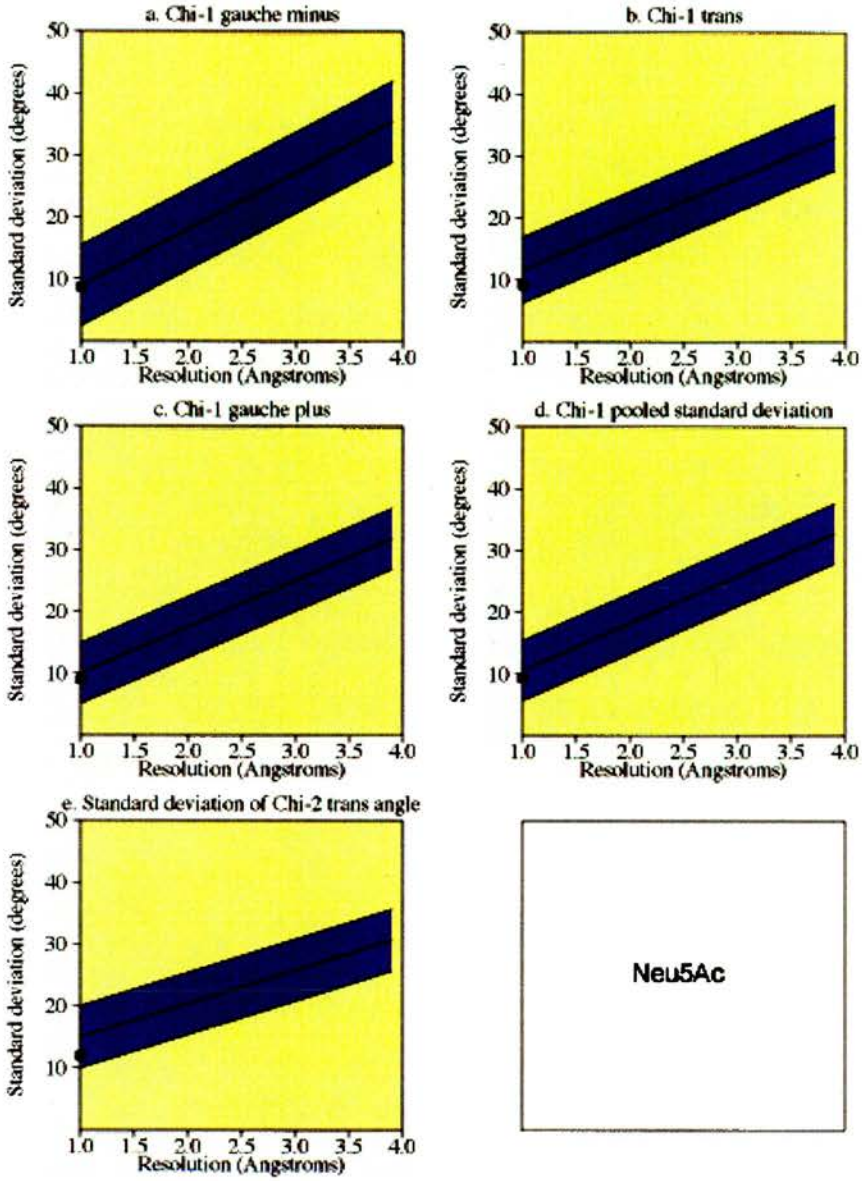
Figure 3.7a. Ramachandran plot for nanI Neu5Ac complex.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	388	88.1	89.7	10.0	-0.2	Inside
b. Omega angle st dev	446	7.1	6.0	3.0	0.4	Inside
c. Bad contacts / 100 residues	0	0.0	1.0	10.0	-0.1	Inside
d. Zeta angle st dev	407	1.9	3.1	1.6	-0.7	Inside
e. H-bond energy st dev	243	0.7	0.6	0.2	0.5	Inside
f. Overall G-factor	449	-0.1	0.0	0.3	-0.4	Inside

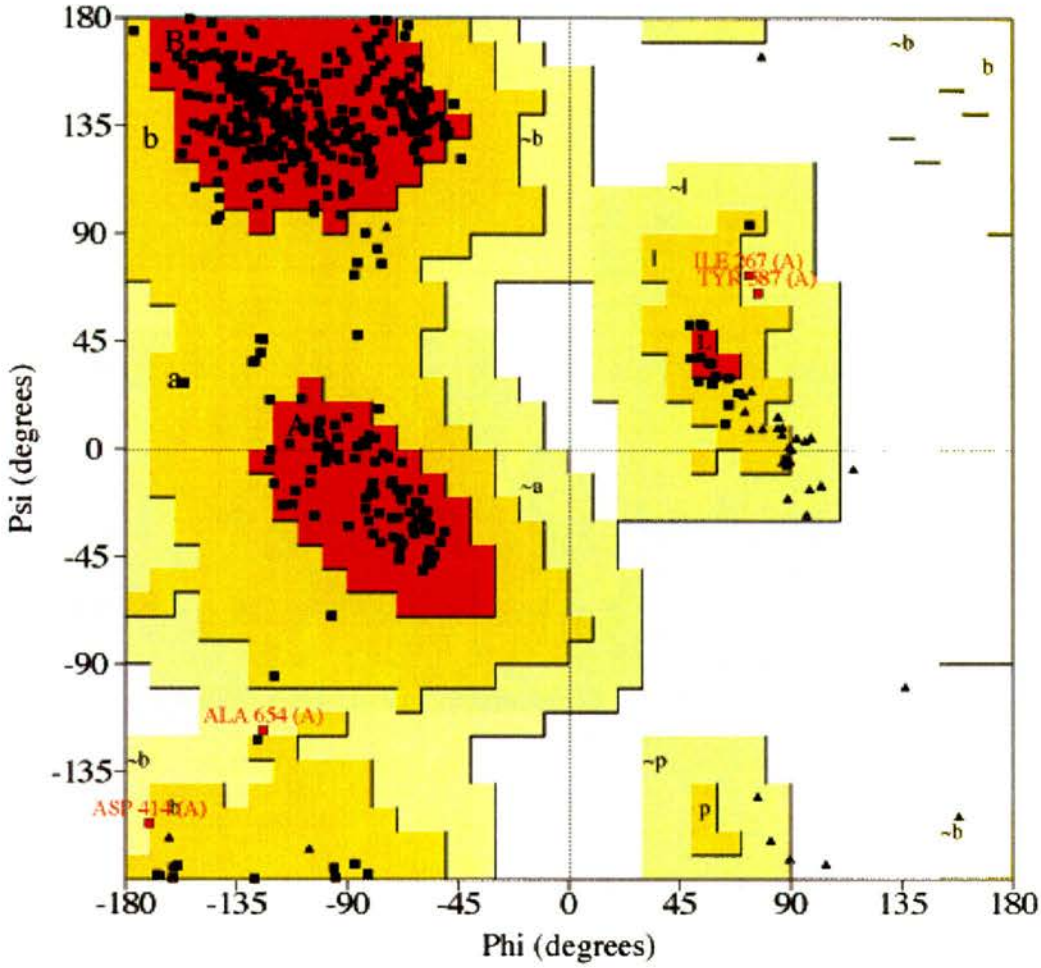
Figure 3.7b. Main-chain parameters for nanI Neu5Ac complex.



Plot statistics

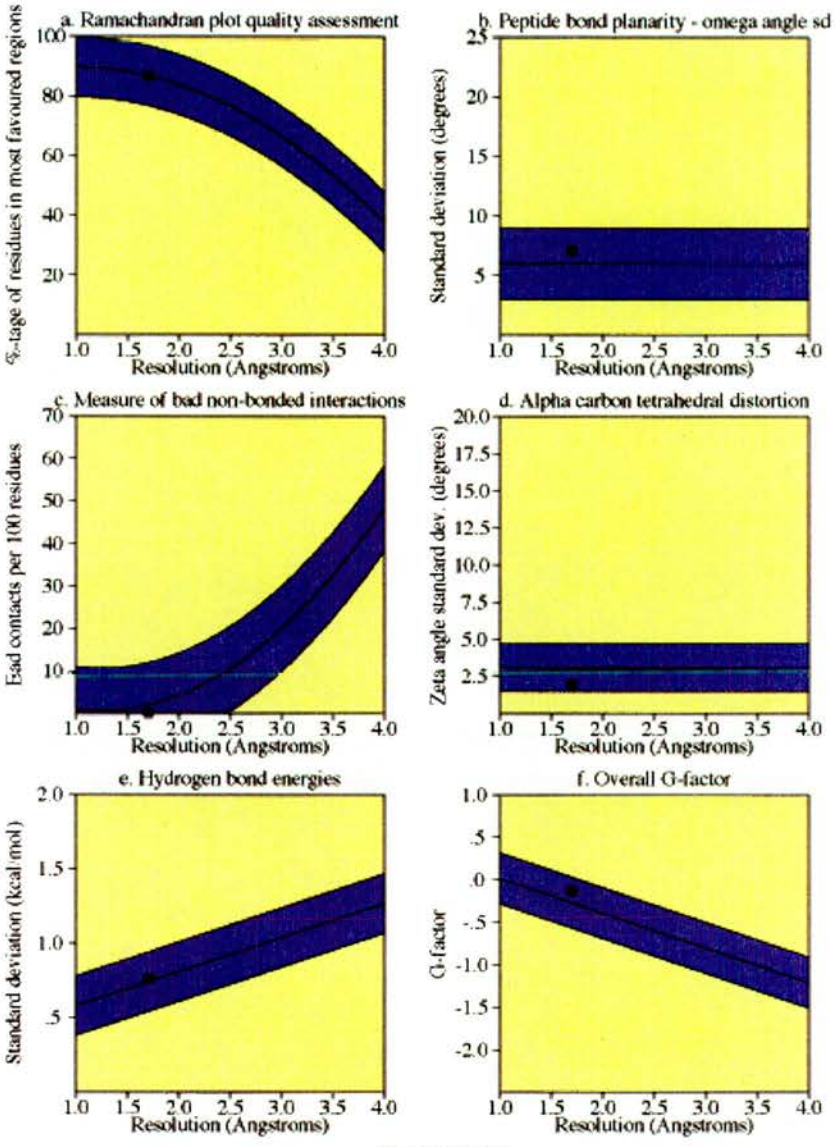
Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	81	8.7	9.0	6.5	0.0	Inside
b. Chi-1 trans st dev	107	9.2	11.6	5.3	-0.5	Inside
c. Chi-1 gauche plus st dev	182	9.0	10.0	4.9	-0.2	Inside
d. Chi-1 pooled st dev	370	9.3	10.5	4.8	-0.2	Inside
e. Chi-2 trans st dev	112	12.0	15.0	5.0	-0.6	Inside

Figure 3.7c. Side-chain parameters for nanI Neu5Ac complex.



Residues in most favoured regions [A, B, L]	336	86.6 %
Residues in additional allowed regions [a, b, l, p]	48	12.4 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	4	1.0 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	388	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	42	
Number of proline residues	18	
Total number of residues	449	

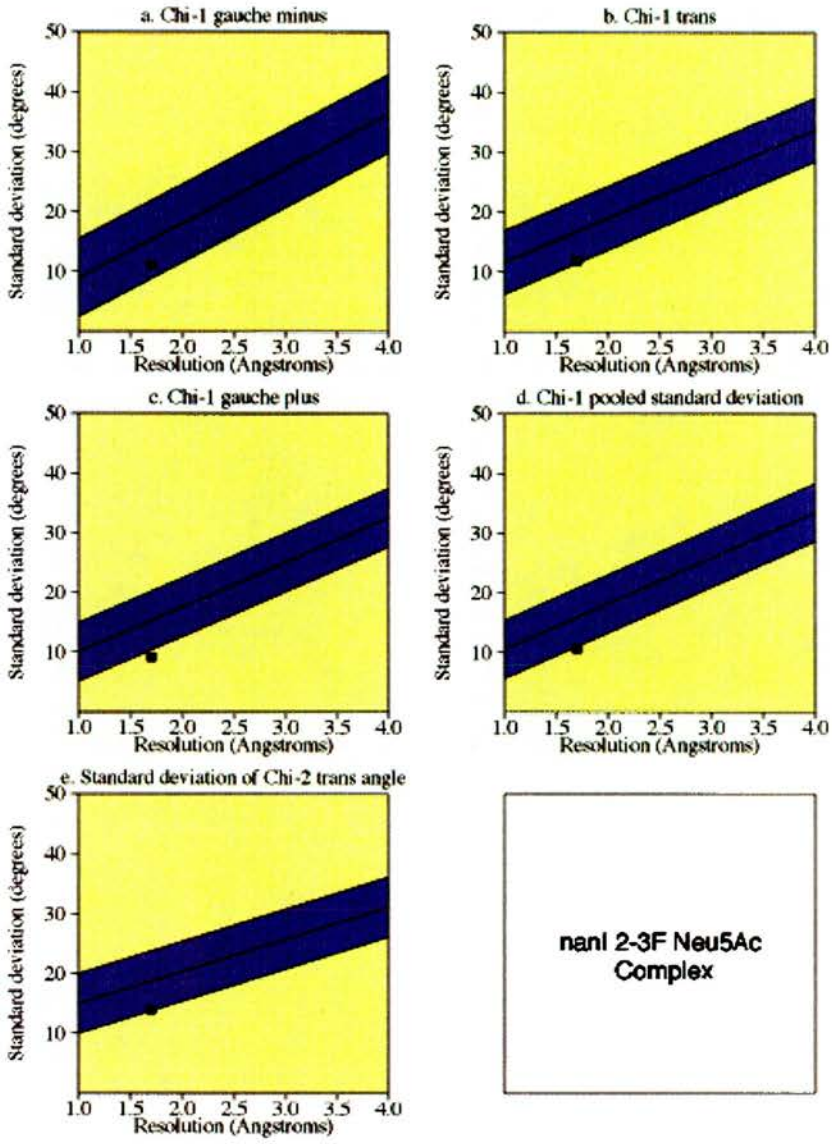
Figure 3.8a. Ramachandran plot for the nanI α2-3difluro-Neu5Ac complex.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Comparison values Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	388	86.6	86.8	10.0	.0	Inside
b. Omega angle st dev	446	7.1	6.0	3.0	.4	Inside
c. Bad contacts / 100 residues	0	.0	1.9	10.0	-.2	Inside
d. Zeta angle st dev	407	2.0	3.1	1.6	-.7	Inside
e. H-bond energy st dev	251	.8	.7	.2	.1	Inside
f. Overall G-factor	449	-.1	-.3	.3	.5	Inside

Figure 3.8b. Main-chain parameters for the nanI α2-3difluro Neu5Ac complex.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	81	11.1	15.4	6.5	-.7	Inside
b. Chi-1 trans st dev	110	11.9	16.8	5.3	-.9	Inside
c. Chi-1 gauche plus st dev	179	9.1	15.3	4.9	-1.3	BETTER
d. Chi-1 pooled st dev	370	10.6	15.9	4.8	-1.1	BETTER
e. Chi-2 trans st dev	112	13.9	18.8	5.0	-1.0	Inside

Figure 3.8c. Side-chain parameters for the nanl α2-3difluro Neu5Ac complex.

Figure 3.9a. Atom positional ESD analysis.

Plot of σ (r) *versus* B_{eq} for nanI Apo structure with 0.97 Å data. σ_{diff} (r) calculated from inversion of unrestrained full-matrix from least-squares refinement. Carbon black, nitrogen blue, oxygen red.

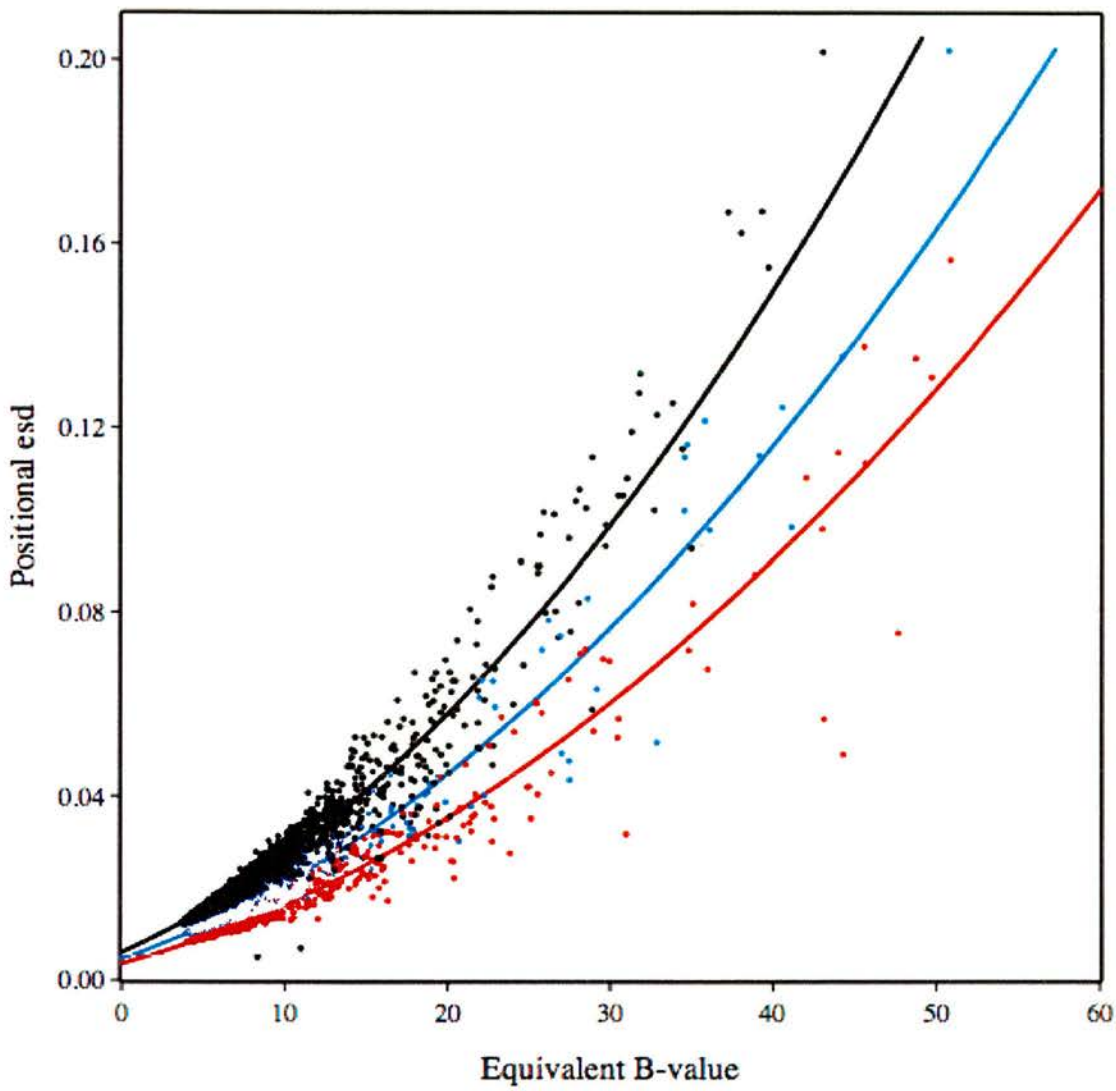
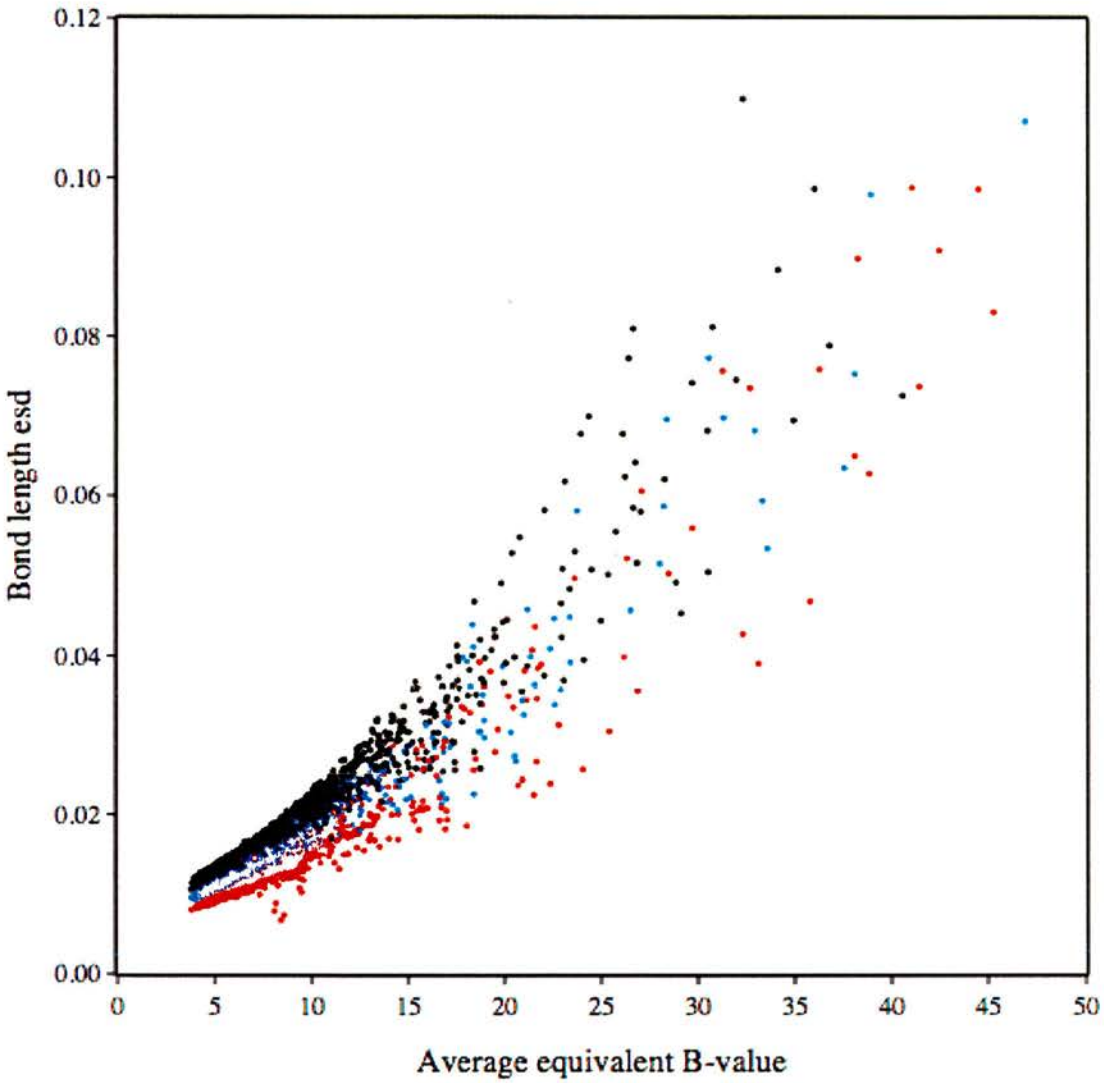


Figure 3.9b. Bond length ESD analysis.
Plot of σ (*I*) *versus* B_{eq} for nanI Apo structure with 0.97 Å data. σ_{diff} (*I*) calculated from inversion of unrestrained full-matrix least-squares refinement. C-C black, C-N blue, C-O red.



3.4. Discussion.

Crystals of the nanI catalytic domain diffracted to atomic resolution at the ESRF beamline ID14-EH1. A complete dataset was obtained at 0.97 Å for the apo and Neu5Ac2en co-crystals and a lower resolution dataset was obtained on an in-house X-ray generator to 1.7 Å for the covalent intermediate. Previous studies on refining protein structures at such high resolution have made a point of stating the importance of data quality and completeness (Dauter *et al* ., 1997). There are at present no clearly defined 'rules' for assessing data quality and the debate is still very much ongoing as to what criteria to use in ranking data in terms of quality. Zbigniew Dauter has suggested that as a guide, the data in the outer shell should have a merging R factor of about 30 %, corresponding to a I/σ(I) ratio of ~2 (Dauter *et al* ., 1997).

The two datasets collected on ID14-EH1 meet these criteria and satisfy what we believe to be good quality atomic resolution data. The crystal space group, 19, was adventitious during data collection and meant that very little user intervention was required at the beamline. Given an appropriate starting orientation, a straight 90 ° rotation range at 0.5 ° oscillations was in many instances sufficient to measure 100 % of the unique Bragg reflections. The low mosiacity of the crystals and good, well-defined diffraction spots, made indexing and integrating the diffraction data straightforward in the program MOSFLM. The result of integrating the diffraction data using MOSFLM is an MTZ file containing a list of all measured reflections. These are subsequently scaled and merged in the CCP4 program SCALA, which allows the analysis of the data through various statistical measurements and outputs a list of scaled reflections reduced to the asymmetric unit of the reciprocal cell of the crystal. It has been suggested that the I/σ(I) ratio is a better criterion than the merging R value, provided that the σ(I) is correctly estimated by the indexing program (Dauter *et al* ., 1997). The estimation of σ(I) is a non-trivial problem and has subsequent repercussions both in data evaluation and refinement, where σ(I) is used to derive optimum weighting schemes.

The choice of model for molecular replacement of the nanI data was based on the sequence alignment of all known sialidase structures in the PDB at that time. The sequences were ranked in order of percentage identity to the catalytic domain of nanI. The equivalent domain of the intra-molecular trans sialidase from the leech, *Macrobdella decora*, was chosen as this had the highest percent identity. It was interesting that molecular replacement of this model in the unit cell of the nanI crystals failed using the CCP4 program AMoRe. It was subsequently accomplished using the crystallography suite of programs, CNS. The reasons for this are unclear, although it must in part be due to the different way in which CNS operates in Patterson space and improves the result of the rotation search using Patterson Correlation refinement. This can help to discriminate between correct and incorrect peaks in a noisy rotation function. Of course, it may also have been due to incorrect estimation of rotation function search sphere and/or resolution limits in AMoRe; subsequent use of the program has found these parameters to be quite influential in obtaining successful solutions.

The initial phases of the model were used successfully in ARP/wARP to iteratively build and refine the nanI catalytic domain. This initial model was then used to refine against the atomic resolution data using the program SHELXL97. This program, initially designed for use in small molecule crystallography applications, provides a versatile and powerful tool for studying these systems. Even at such high resolution as 0.97 Å though, an extensive set of restraints must still be implemented in the refinement of protein models. The great advantage of atomic resolution data lies in the hugely increased ratio of experimental observations to refined parameters. At 1 Å resolution, there are approximately five observations for every parameter, even with an anisotropic model (Dauter *et al* ., 1997). The nanI refinement at 0.97 Å had a data/parameter ratio of seven, clearly demonstrating this advantage. The result of this situation is that errors for each individual atomic parameter can be estimated rigorously by inversion of the least-squares matrix (Cruickshank, 2001; Sheldrick and Schneider, 1997b). The high degree of accuracy for atomic

positions, a rigorous estimation of there error and anisotropic modelling of atomic displacement parameters (ADPs) are usually accompanied by a significant drop in the free R value. This in turn leads to a reduction in the noise level of the subsequent electron density maps. This allows the modelling of a range of features normally hidden in lower resolution models. These can be flexible regions of the protein, hydrogen atoms and refinement of water molecule occupancies.

Coordinate uncertainty is currently a topic of great interest in the crystallography community. The ability to accurately determine the significance of amino acid or rigid body movements in different protein models is still being developed. The work by Cruickshank, Schneider and Read in developing methods and programs to calculate what constitutes a significant movement have been instrumental in this area of structure validation and analysis. The atomic resolution structures of nanI in both the apo and substrate bound states allows the most thorough study of sialidase catalytic mechanisms to date. For lower resolution structures, between 1.3 and 2.5 Å, Cruickshank's DPI provides a reliable indication of the significance in the movement between two related atoms, be they in the same molecule or between free and complexed structures (Schneider, 2004). The atomic resolution data for the nanI catalytic domain provides the means to assess the statistical significance of any amino acid movement as a result of accommodating the substrate, Neu5Ac, or the covalent intermediate.

The refinement of the model against the data collected from the nanI crystals grown in the presence of 5 mM Neu5Ac2en revealed a surprising result. The difference density map, $F_O - F_C$, revealed the presence of α-Neu5Ac and not that of Neu5Ac2en. This significance of this result is discussed in Chapters 5 and 6. The availability of atomic resolution data allowed the refinement of this ligand without any user-defined library, a condition usually necessary at lower resolution. The difference density map showed a clear peak of postive density for the C2 hydroxyl group and provided the first evidence that the nanI sialidase has a very unique characteristic, the ability to hydrate the C2-C3 double bond of

the transition state analogue, Neu5Ac2en. This complex has therefore been labelled the Neu5Ac complex, even though it was grown in the presence of Neu5Ac2en.

Chapter 4

Structure of the *Clostridium perfringens* nanI catalytic domain

4.1. Summary

The structure of the catalytic domain of nanI from *Clostridium perfringens* is described here to a resolution of 0.97 Å in its uncomplexed apo form. The nanI structure consists of 32 β-strands and 2 small α-helices and folds into two distinct domains; a six-bladed β-propeller formed by residues 243-359 and 429-691 and a small β-barrel domain formed by residues 360-428. The nanI β-propeller domain is topologically equivalent to all previously reported sialidase structures. The overall r.m.s deviation between the leech trans sialidase and the refined nanI model is 1.22 Å for 400 C_α positions considered equivalent. The small β-barrel domain in nanI however, shows a striking similarity to the equivalent domain in the leech trans sialidase, which was excluded from the MR model. This domain showed no sequence similarity to the nanI sialidase. The root mean square (r.m.s) deviation for 63 C_α positions considered equivalent is 1.43 Å.

The structure of four conserved sequence motifs, known as 'Asp boxes' (Ser/Thr-Xaa-Asp-[Xaa]-Gly-Xaa-Thr-Trp/Phe), can be clearly seen in the β-propeller domain forming the turns between strands three and four of the first four β-sheets. They sit in topologically similar positions in all currently studied bacterial sialidases. The function of these motifs is still not fully understood, although a role in protein secretion has been proposed (Crennell *et al* ., 1994).

The electrostatic potential on the surface of nanI shows a very distinct asymmetric charge distribution. The surface of the protein opposite the catalytic site is very negative, whereas the opposite surface is fairly neutral. This asymmetry has been observed in both the leech intra molecular trans-sialidase and the sialidase from *Vibrio cholorae*. The location of the Asp-boxes within this

very negative region of the protein may help contribute to this asymmetry, although this has not been tested.

The active site residues of all currently reported sialidase structures are highly conserved and often used to identify novel sialidases from the genome sequencing projects. In common with the viral, bacterial and eukaryotic sialidases/trans-sialidases the nanI active site is highly conserved. It contains an arginine triad, a hydrophobic pocket, a tyrosine and a glutamic acid residue at the base of the site and an aspartic acid at the entrance. Interestingly, the nanI sialidase also has a tryptophan at the entrance to the active site, forming a sort of lid. A comparison with all previously deposited structures of sialidases in the PDB shows this feature to be unique to nanI.

The protein contains two calcium ions coordinated within the β-propeller fold by six and seven oxygen atoms respectively. It is currently unknown whether nanI is dependent on calcium ions for either activity or thermostability. One of the calcium sites is in a very similar location to a key calcium-binding site in the sialidase from *Vibrio cholerae*, a calcium dependent sialidase. The exact residues coordinated by the calcium ion in nanI are not identical to those in *Vibrio cholerae*, which serve to stabilise the positions of important active site residues (Crennell *et al* ., 1994). The calcium ions in nanI do not appear to function in the same way, in that they play no obvious role in stabilising active site residues. However, their location on the surface of the protein may indicate roles in modifying the electrostatic potential of the enzyme or simply in stabilising the β-propeller fold.

4.2. Overall structure.

The enzymatically active 50 KDa fragment from amino acids 243 to 694 of the full length nanI sialidase consists of 32 β-strands and 2 small α-helices and folds into two distinct domains (Figure 4.1.1); a six-bladed β-propeller formed by residues 243-359 and 429-691 and a small β-barrel domain formed by residues 360-428. The overall dimensions of the structure are 62 Å x 45 Å x 45 Å. The canonical six bladed β-propeller fold is a trademark signature of the sialidase superfamily (Taylor, 1996), seen in all previously reported sialidase structures from the influenza virus neuraminidases to the bacterial sialidases from *S.typhimurium* and *V.cholerae*. The secondary structure assignment was determined using the program promotif3 (Hutchinson and Thornton, 1996).

The β-propeller fold seen in the nanI structure consists of six β-sheets, and forms the catalytic domain of the sialidase. Each sheet consists of four antiparallel β-strands running from inside to outside, arranged in a similar way to the blades of a propeller, around an approximate sixfold axis. The catalytic site is located in a pocket at one end of the propeller axis. The small β-barrel domain inserts between the two central β strands of the second β sheet of the β-propeller domain, with which it has a number of hydrophobic interactions. Analysis of the main chain B-factors for the nanI model (Figure 4.1.2) show that the 50 KDa fragment is relatively rigid, with apparently little movement of the β-barrel domain relative to the β-propeller fold. The most flexible parts of the structure are the loop regions connecting each sheet, with the most flexible loop being made by residues 390 to 394, coloured red in figure 4.1.2. This loop is solvent exposed and makes no significant packing interactions with the β-barrel.

Two small α-helices are present, but contribute little to the overall fold of the protein. The first helix is formed by the loop region between strand four of sheet two and strand one of sheet three and consists of a single turn. The second small helix consists of two turns and located at the C-terminus of the protein, packing closely against the sixth sheet.

The four Asp-boxes of the nanI sialidase occur in topologically equivalent positions as observed in the other three bacterial sialidases (Crennell *et al* ., 1994; Crennell *et al* ., 1993; Gaskell *et al* ., 1995), at the turn between the third and fourth strands of the first four sheets (Figure 4.1.1). The four Asp-boxes have very similar conformations (see Section 4.3), maintained by a hydrogen-bonding network between the conserved residues of the motif and with the tryptophan plugging into a hydrophobic pocket between the sheet to which the Asp-box belongs and the preceding sheet.

Sequence alignment of residues 243-694 of the nanI sialidase with the trans sialidase from *M.decora*, with which it shares 32 % sequence identity, suggested that nanI would fold into two distinct domains, in a similar way to the leech protein. A surprising result from the initial analysis of the nanI structure was the striking similarity of this second domain in nanI to the equivalent domain in the leech trans sialidase, which was excluded from the MR model (see section 4.5). Both domains superposed have an r.m.s.d. of 1.42 Å for 67 C_α positions considered equivalent, but share only 33 % sequence identity. It is also interesting to note that the equivalent domain in the *M.decora* structure is considerably larger, being composed of 100 amino acids compared to 68 in nanI.

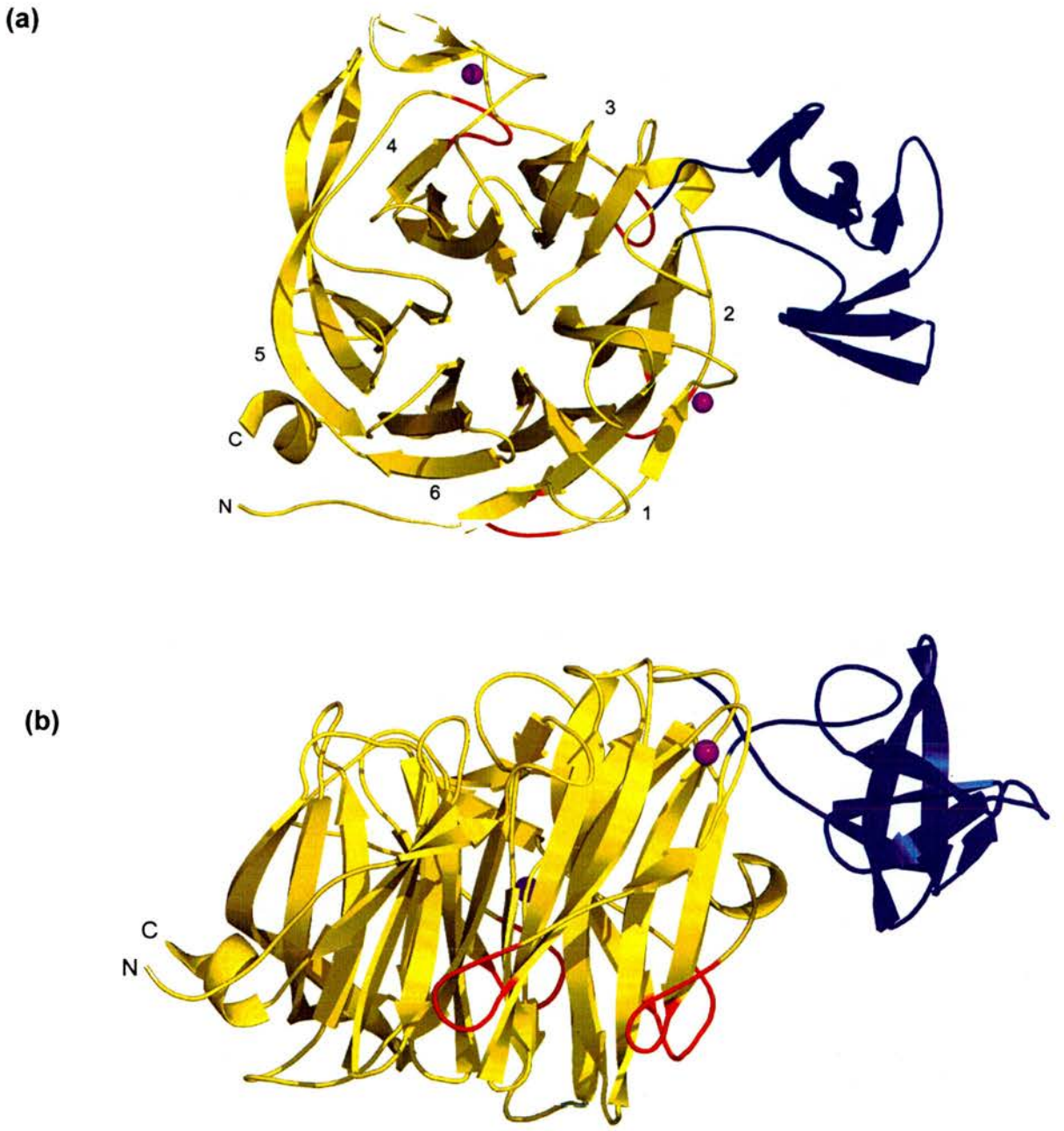


Figure 4.2.1. Multidomain architecture of the *Clostridium perfringens* nanI catalytic domain. (a) Ribbon representation of the 50 KDa fragment of the *Clostridium perfringens* nanI catalytic domain. The view is looking down into the active site pocket. The β-sheets are numbered 1 through 6. The β-propeller catalytic domain is shown in yellow, the small β-barrel domain in blue. The N and C termini are labelled respectively. The four Asp-boxes are shown in red and the two calcium ions in magenta. (b) View turned 90 °.

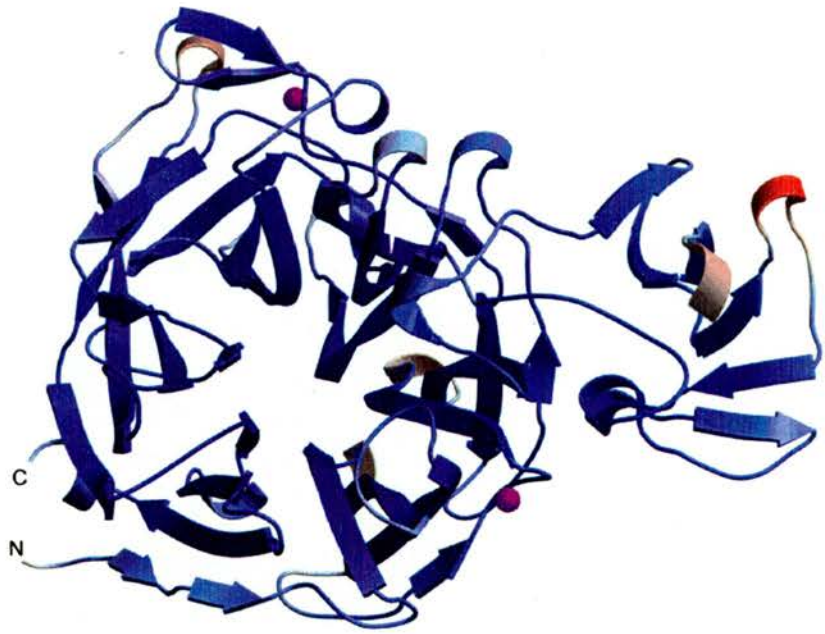


Figure 4.2.2. The Secondary structure of nanI coloured according to B-factor. Colouring is from dark blue (B = 0 $Å^2$) to deep red (B = 30 $Å^2$). The two calcium ions are shown in magenta. The N and C-termini are labelled respectively.

4.3. Comparison of nanI to the intra-molecular trans sialidase from the leech, *Macrobdella decora*; the model used in Molecular Replacement.

Superposition of the nanI structure onto other sialidases, using the program LSQMAN (Kleywegt and Jones, 1997), reveals that it has greatest similarity to the sialidase from the bacterium *Micromonospora viridifaciens* (PDB code 1EUS), Table 4.2.1. The overall root mean square deviation (r.m.s.d) between the *M.viridifaciens* sialidase and the 50 KDa nanI fragment is 1.18 Å for 331 C_α positions considered equivalent. However, the intra-molecular trans sialidase from the leech *Macrobdella decora* (PDB code 1SLL) is also very similar, with an r.m.s.d of 1.22 Å for 400 C_α positions considered equivalent. This high similarity is reflected in the successful use of the *M.decora* catalytic domain

in the molecular replacement of the nanI structure. This analysis suggests that the *M.viridifaciens* β-propeller domain would also be successful in the molecular replacement of the nanI structure. This example serves to highlight the conservation of fold over sequence in protein structure evolution.

The superposition of the nanI structure onto that of the leech trans sialidase highlights the similarity between the two enzymes, in particular between the β-propeller domain and small β-barrel (Figure 4.2.1). The differences between the two superposed structures are mainly in the length and arrangement of the loop regions connecting the β strands. There is considerable rearrangement at the interface between the β-propeller and β-barrel domains, due mainly to the absence of the extended hairpin loop on the *M.decora* β-barrel. This hairpin packs extensively with the β-propeller domain (Lou *et al* ., 1998a). The absence of this hairpin in the nanI structure results in a different arrangement of β-propeller loops in this region of the structure. This is analysed and discussed in more detail in Section 4.5. of this chapter. The *M.decora* enzyme also has a lectin domain at its N-terminus, which is likely to be very similar to the equivalent domain in *nana*.

Table 4.3.1. Comparison between the nanI structure and other sialidases.

Sialidase	PDB	r.m.s.d (Å)	Equivalent Cα	% Sequence Identity
Vibrio cholerae	1W0P	1.43	311	23.3
Micromonospora virdifaciens	1EUS	1.18	331	29.5
Macrobdella decora	1SLL	1.22	400	32.0
Salmonella typhimurium	3SIL	1.49	291	30.7
Typanosoma cruzi	1S0K	1.45	300	32.3
Typanosoma rangeli	1N1S	1.43	299	33.0
Influenza A subtype *N2*	1INH	1.95	201	19.0

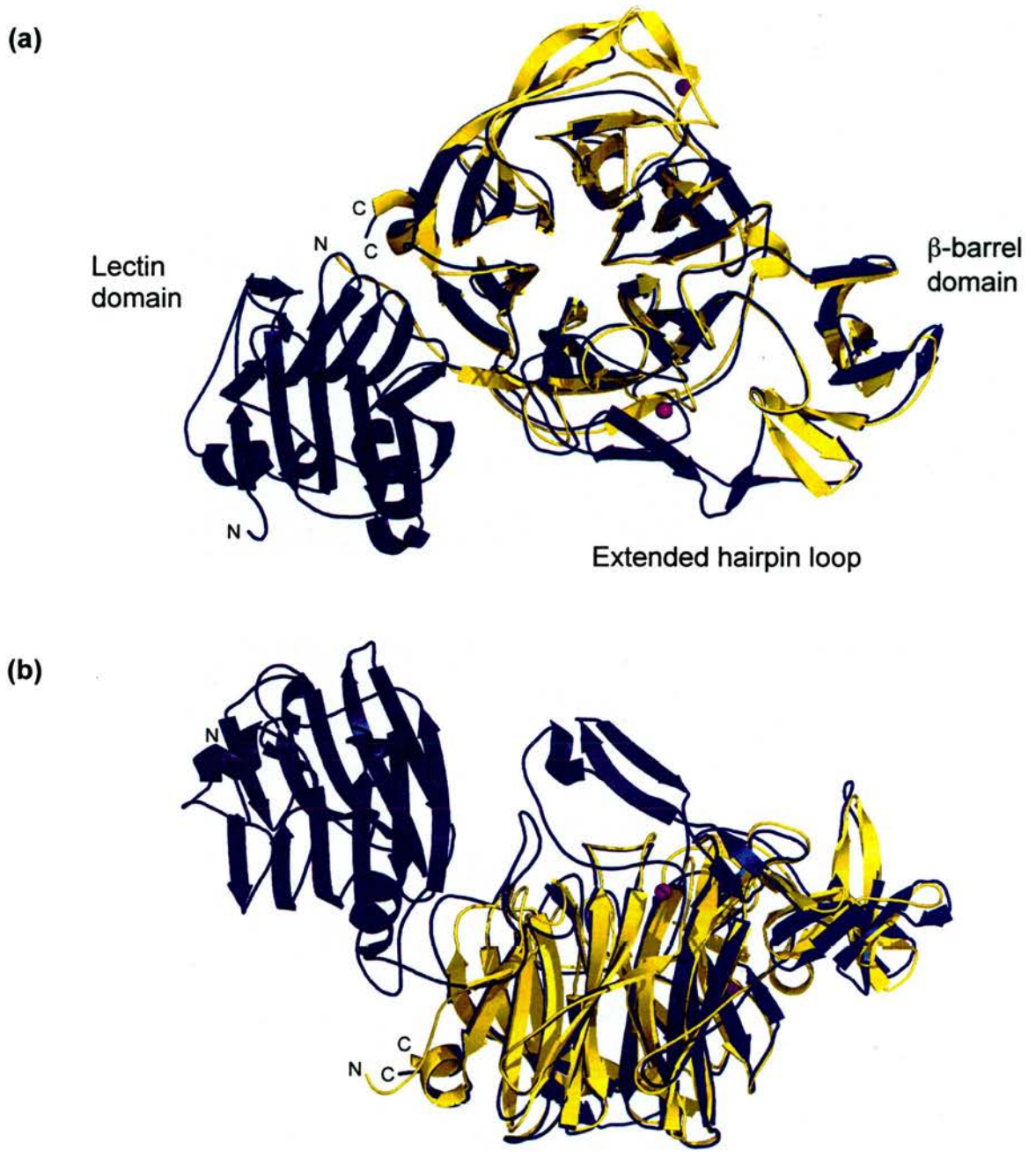


Figure 4.3.1. (a) Superposition of the nanI structure (yellow) onto the trans-sialidase from the leech, *M.decora* (blue) (PDB code 1SLL). The two structures superpose with an r.m.s.d of 1.22 Å for 400 C_α positions considered equivalent. The small β-barrel and lectin domains in the *M.decora* structure were not used in the MR model. (b) Image rotated 90 °.

4.4. Active site of the nanI sialidase.

The residues involved in binding and turnover of substrate were identified through crystals of the nanI catalytic domain grown in the presence of 10 mM 2-deoxy-2,3-didehydro-5-N-acetylneuraminic acid (Neu5Ac2en) or DANA, as discussed in Chapter 5. The active site discussed here is that of the uncomplexed apo structure.

The active sites of all currently studied sialidases are very similar in design, due mainly to the function of this site in binding α-Neu5Ac (sialic acid) and catalysing the hydrolysis of the glycosidic bond to the sub-terminal aglycon. Although there are a number of significant differences between the active sites of viral, bacterial and eukaryotic sialidases, a number of key features seem to have been retained throughout their evolution (Taylor, 1996). These include a tri-arginal cluster (Arg 266, Arg 555, Arg 615), used to form a very strong interaction in a plane with the carboxylate group at C-1 of α-Neu5Ac. The position of the first arginine (Arg 266) is stabilised by a conserved glutamic acid (Glu 671) through a hydrogen bond between the N^{ε} of the arginine and the $O^{\delta}2$ of the glutamic acid. A tyrosine (Tyr 655) and a glutamic acid (Glu 539) hydrogen bond with each other and sit beneath and close to the C1-C2 bond of the substrate.

A recent study into the mechanism of the trans-sialidase from *Trypanosoma cruzi* trapped a covalent intermediate on this tyrosine by use of a fluorinated substrate analogue (Amaya *et al* ., 2004; Watts *et al* ., 2003). They showed that in the *T.cruzi* enzyme, the tyrosine acts as the catalytic nucleophile in the reaction and that the glutamic acid acts as a base catalyst for the attack of the tyrosine on the anomeric centre (Amaya *et al* ., 2004). This paper also proposes that the formation of a covalent glycosyl-enzyme intermediate is a common species on the reaction pathway of all sialidases. This hypothesis is further supported in Chapter 5, where the structure of the covalent intermediate in the nanI pathway is described using the same fluorinated sialic acid substrate.

It was previously thought that the tyrosine-glutamic acid dyad served only to stabilise an oxycarbonium ion intermediate (Burmeister *et al* ., 1993) and could

not form a covalent bond to the anomeric carbon of the substrate. The choice of tyrosine as the catalytic nucleophile, over the more commonly seen carboxylate group in other retaining glycosidases (Wicki *et al* ., 2002), is likely a consequence of the need to minimise repulsive electrostatic interactions between the nucleophile on the protein and the carboxylate group on the substrate. The positive charge generated by the tri-arginyl cluster is primarily a way of neutralising the carboxylate group and 'locking' the substrate firmly in the active site at this end.

The final conserved feature of sialidase active sites is the acid/base catalyst, Asp 291. The structural work on the *T.cruzi* enzyme has clearly pointed to this aspartic acid as the acid catalyst for the initial step involving glycosyl-enzyme formation (Amaya *et al* ., 2004). Further, within the glycosyl-enzyme intermediate it is well positioned to function as the general base catalyst, facilitating attack by the water in sialidases, such as nanI and the 3-hydroxyl of lactose in the trans-sialidase.

It was previously argued that the high solvent exposure of this residue combined with the pH optima of the enzymes meant the pKa of this aspartic acid was too low for it to function as an acid catalyst (Burmeister *et al* ., 1993; Chong *et al* ., 1992). The trapping of the Michaelis complex within the *T.cruzi* enzyme revealed this solvent exposure to be very much reduced by both a tyrosine side chain and the binding of the aglycone (Amaya *et al* ., 2004). Within the active site of the nanI sialidase there is a very prominent aromatic side chain, Trp354, which sits above the active site pocket forming a hydrophobic cap or lid, in an equivalent position to the tyrosine in the *T.cruzi* active site. Comparison with other sialidases from bacteria, viral and eukaryotic sources mark this feature as unique to the nanI enzyme. The effect of having this hydrophobic residue in this location must restrict the access of solvent to the active site pocket and may play a role in regulating the pKa of the catalytic aspartic acid, as proposed for the tyrosine in the *T.cruzi* enzyme. Investigating the effect of mutating this residue on the activity of the sialidase would be informative in identifying a function.

A further characteristic of sialidase active sites is the hydrophobic pocket used to accommodate the N-acetyl group of the substrate (Taylor, 1996). The exact residues used to form this pocket are generally not conserved. In the nanI structure, this pocket is made up of mostly aromatic side chains, Phe 347, Phe 353, Phe 460, Thr 345 and Ile 327 with the Trp 354 forming the lid above this pocket.

In contrast to the features described above, the interactions between the protein and other chemical groups on sialic acid differ between the sialidases. These are discussed in Chapter 5, where the structures of several bacterial and eukaryotic sialidase structures are compared and contrasted to the nanI complexes.

The large nanI sialidase from *C.perfringens* has broad substrate specificity, which is markedly different to that of the small sialidase, nanH. The study by Roggentin *et al* . (Roggentin *et al* ., 1995) compared the properties of the large nanI sialidase against the small nanH from the same species, DSM 756 and A99, and concluded that the large sialidase hydrolyses sialic acids from low and high molecular weight substrates in a similar and effective manner with higher affinity for α2,3 and α2,8 over α2,6. The *S.typhimurium* sialidase is unusual in that it has a 260-fold kinetic preference for α2,3 linked sialic acid over α2,6, similar to the influenza enzyme (Hoyer *et al* ., 1991). This similarity to the nanI enzyme is discussed in Chapter 5 where modifications to the active site of these sialidases could account for this characteristic.

It has recently been shown that sialidases containing carbohydrate-binding domains are able to hydrolyse polyvalent substrates with much greater catalytic efficiency than their monovalent counterparts (Thobhani *et al* ., 2002). Many of the non-viral sialidases have additional carbohydrate-recognising domains and the authors of this study suggest that these domains have been sequestered by the sialidases to increase their catalytic efficiency for siayloglycoconjugates. Indeed, it seems likely that the N-terminal domains of both the *M.decora* and *C.perfringens* nanI sialidases have the ability to recognise and bind carbohydrates. The identity of these carbohydrates is currently

unknown, however, a recent study on the sialidase from *V.cholerae* has identified α-Neu5Ac (sialic acid) as the binding partner for the first 'lectin-like' wing of this enzyme (Moustafa *et al* ., 2004).

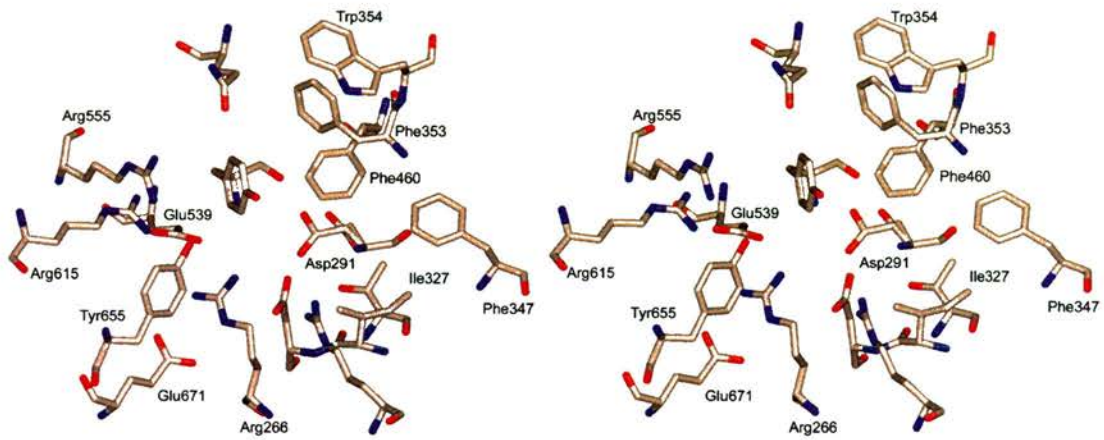


Figure 4.4.1. Stereo view of the *C.perfringens* nanI active site.

4.5. The Asp-boxes of nanI.

A conserved sequence motif named the Asp-box (Ser/Thr-Xaa-Asp-[Xaa]-Gly-Xaa-Thr/Trp/Phe) has been identified in bacterial, mammalian and protozoan sialidases but strangely absent from the viral enzymes (Lou *et al* ., 1998b; Taylor, 1996). This motif, when found, usually repeats three to five times downstream of the RIP/RLP signature, with a separation of 60 to 70 residues (Roggentin *et al* ., 1989). The exact role of these motifs is currently unknown. It has been proposed that they may play a role in protein secretion or fold stabilisation (Crennell *et al* ., 1994; Crennell *et al* ., 1993). The latter suggestion seems unlikely as all currently studied viral sialidases lack these motifs yet seem fully capable of folding correctly, with r.m.s. deviations of 2-3 Å to their homologues in other organisms.

The Asp-box turns in the nanI structure are of well-defined conformation; the four turns when superimposed using LSQMAN on one another have r.m.s. fit values of less than 0.25 Å for the eight C_α atoms involved (Figure 4.3.1). The third Asp-box has an asparagine in place of the usually conserved threonine in position seven of the motif. Interestingly, the conserved aspartic acid side chains point away from the solvent and make extensive hydrogen-bonding interactions with the β-propeller fold. This contrasts with the other bacterial sialidase Asp-boxes, where the conserved aspartic acid point out into the solvent.

Looking at the position of the motifs in relation to the overall fold of the β-propeller (Figure 4.1.1) and taking into account the absence of such motifs in the viral and trypanosomal sialidases and subsequent charge asymmetry, it may be worth considering the role of these motifs in contributing to this unique surface characteristic. This point is discussed further in Section 4.6 of this chapter.

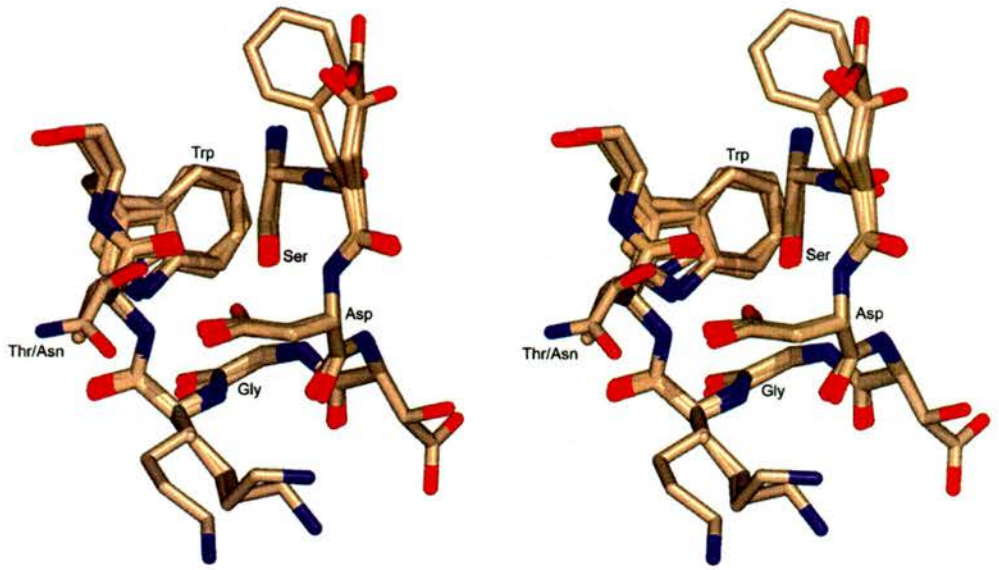


Figure 4.5.1. Stereo view of the superposed four Asp-boxes in the nanI β-propeller fold. The superimposed turns have r.m.s. fit values of less than 0.25 Å for the eight C_α atoms involved. The conserved residues from the motif are labelled. The third Asp-box has an Asparagine in place of the usually conserved Threonine.

4.6. Metal ions in the nanI structure.

Metal ions, in the form of calcium, have been found in only three reported sialidase structures, those from the bacterium *V.cholerea* , the influenza virus neuraminidase and the paramyxovirus hemagglutinin-neuraminidase. The *V.cholerea* sialidase has a total of four calcium ions coordinated within the catalytic domain (Moustafa, 2004), where they play significant roles in stabilising key catalytic side chains. The calcium ion in the influenza enzyme is in a different location to that found in the cholera sialdase, and is not essential for activity in the active site (Chong *et al* ., 1991). However, it does play an essential role in protein thermostability (Burmeister *et al* ., 1994), as it is located in the four-fold axis of the tetramer and stablises the oligomer. The paramyxovirus hemagglutinin-neuraminidase on the other hand requires calcium to stabilise a

helix that contains residues that interact with the glycerol group of sialic acid (Crennell *et al* ., 2000). Removal of this calcium completely abolishes enzyme activity in a similar way to the *V.cholerae* enzyme (Munoz Barroso *et al* ., 1994). All other reported sialidases do not contain calcium ions, so their requirement is not a general characteristic of the sialidase superfamily

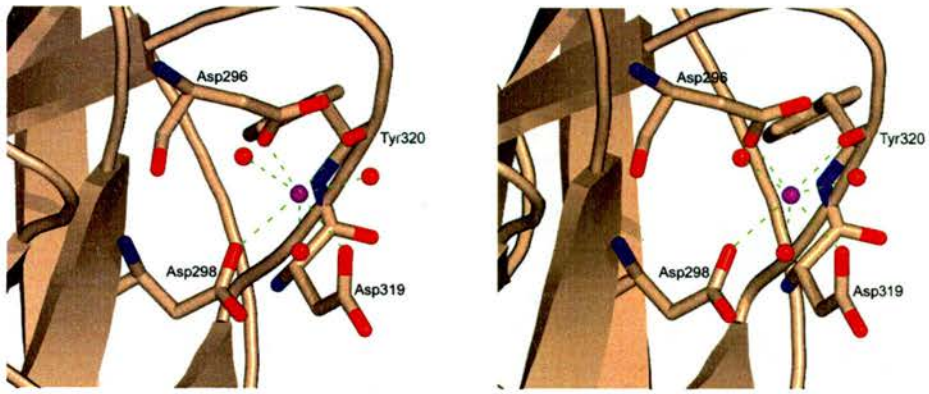
The difference Fourier electron density map (F_O-F_C) has two clearly identifiable metal ions coordinated within the β-propeller fold of the enzyme, with sigma levels of around 12 σ (Figure 4.4.1). These ions have been tentatively assigned as calciums, due to their coordination geometries and the requirement of calcium ions in other members of the sialdiase superfamily. For oxygen donor ligands, the most common coordination numbers for Ca^{2+} range from six to nine (Katz *et al* ., 1996) and there is nothing unusual in the geometry of these ions to suggest they are not calcium.

The first calcium identified (Ca^{2+}-1, B = 5.8 $Å^2$) sits in the loop connecting the first and second β-sheets and coordinated by seven O-atoms, with coordinate bond distances in the range 2.31 – 2.49 Å (Table 4.4.1). The second calcium (Ca^{2+}-2, B = 7.4 $Å^2$) sits between sheets four and five and coordinated by six O-atoms in an octahedral configuration, with coordinate bond distances between 2.58 and 2.79 Å. Four of the O-atoms coordinating the ion are from solvent, compared to only three for the first, with slightly longer bond lengths.

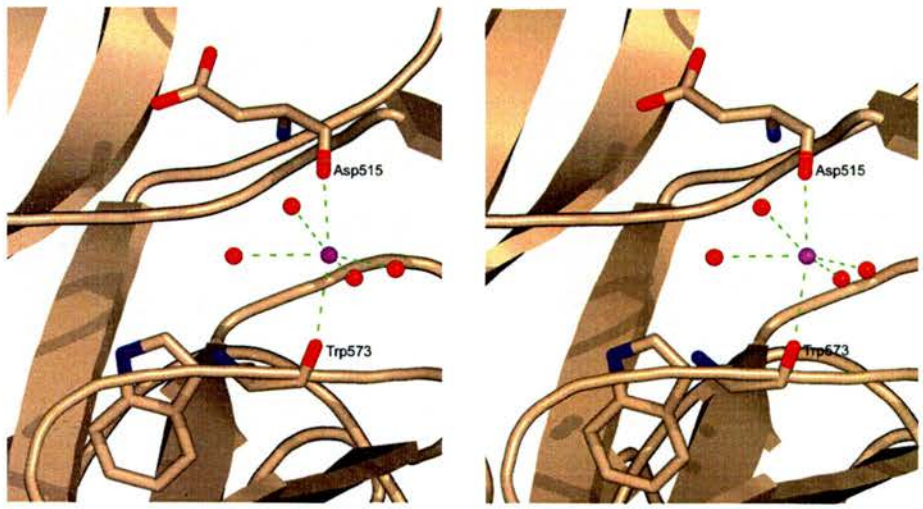
It is interesting that in all reported cases of sialidases having a calcium ion(s), this ion is required to stabilise important residues in the active site. NanI appears to be different, in that there appears to be no obvious role for the two calcium ions in this enzyme. The role of calcium ions in the biochemistry of the nanI sialidase is currently unknown, although it has been shown that they play no part in stimulating enzymatic activity (Nees *et al* ., 1975). Ca^{2+}-1 sits in a similar position to the calcium ion in the *V.cholerae* sialidase responsible for stabilising the conserved aspartic acid at the mouth of the active site pocket. However, the Ca^{2+}-1 ion in the nanI stucture does not coordinate in a similar manner, but sits further along this loop region, a full six residues from the topologically equivalent residue, Asp291.

The second calcium ion, Ca^{2+}-2, sits at the opposite end of the β-propeller to Ca^{2+}-1. This calcium might help to stabilise a 13 residue hairpin formed by amino acids 514-527 connecting sheets three and four of the β-propeller (Figures 4.1.1. and 4.4.1.). Crucially, the Ca^{2+}-2 can be seen forming a bridge between the carbonyl oxygen's of Asp515 on the hairpin to the carbonyl oxygen of Trp573 on the β-propeller.

It is possible that these calcium ions, in a similar fashion to the influenza virus neuraminidase, influence the thermostability of the nanI enzyme. Further studies will need to be carried out to investigate this possibility. Alternatively, their position on the surface of the enzyme could indicate a contributing role in electrostatic character of the protein; discussed in Section 4.6.



Calcium ion 1, B = 5.8 Å^2.



Calcium ion 2, B = 7.4 Å^2.

Figure 4.6.1. Coordination of the two calcium ions in the *C.perfringens* nanI structure in stereo view. The calcium ions are shown in magenta with oxygen's and waters shown as red. Coordinating residues are labelled.

Table 4.6.1. Lists of the coordinate atoms for the two calcium ion's and the corresponding bond distances.

Ca^{2+}-1		Ca^{2+}-2	
Coordinate atom	Distance (Å)	Coordinate atom	Distance (Å)
Asp296-O^{δ}	2.31	Trp573-O	2.63
Asp298-O	2.40	H_2O1704	2.79
H_2O1699	2.44	H_2O1797	2.78
Asp319-O	2.33	Asp515-O	2.58
H_2O1731	2.44	H_2O1744	2.64
H_2O1734	2.49	H_2O1749	2.73
Tyr320-O	2.37	---	---

4.7. β-barrel domain in the nanI structure.

The second of the two domains making up the current nanI structure consists of a small β-barrel formed by residues 360 – 428 between the second and third strands of the second sheet of the β-propeller. The barrel is made up of eight β-strands arranged in an irregular way, as can be seen from the topology diagram in Figure 4.5.1. The similarity of this domain to the homologous domain in the *M.decora* sialidase is striking, both domains superposed have an r.m.s.d. of 1.42 Å for 67 C_{α} positions considered equivalent (Figure 4.5.1).

It can be seen from the superimposition however, that much of this extra size in the *M.decora* structure can be found in a hairpin loop, which extends back along the main catalytic domain. Such a loop in absent in the nanI structure and this has a major impact on the way the two domains interact with the catalytic domain in each structure respectively.

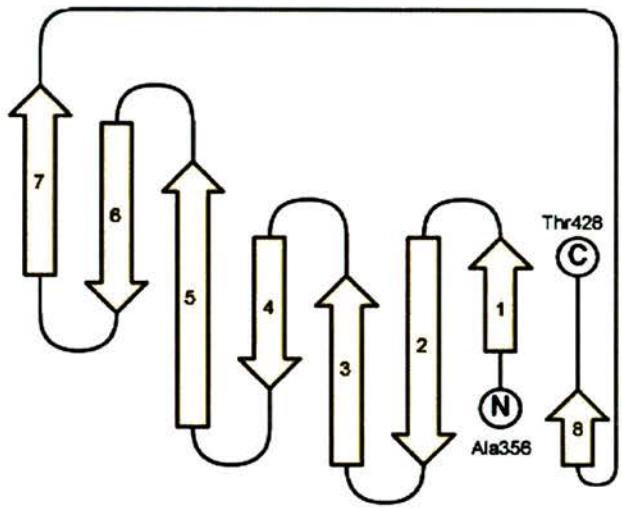
Using the program GRASP (Nicholls *et al* ., 1991), the occluded accessible surface area between the catalytic β propeller and the β barrel domains were calculated for both proteins and compared. The results show that

for the *M.decora* sialidase, the occluded accessible surface area is 2713 $Å^2$ whereas for the nanI sialidase this value is only 877 $Å^2$. This large reduction in the occluded surface area between the two domains in nanI can be attributed to two differences between the structures. The first is the absence of the hairpin loop seen in the *M.decora* structure; the second the presence of a large solvent channel running between the β propeller and β-barrel domains. This channel is absent in the *M.decora* sialidase, due mainly to the different side chains which pack against the β propeller.

The function or relevance of the β barrel domain in both structures is currently unknown. The authors of the *M.decora* paper do not speculate a function and could find no obvious binding pockets for ligand recognition. In an attempt to assign a possible function the PDB coordinates and amino acid sequence of the nanI β barrel domain, from residues 360-428, were submitted to both the Dali (Holm and Sander, 1998) and BLAST (Altschul *et al* ., 1990) databases.

The Dali server could find no structurally similar proteins in the Protein Data Bank, other than the *M.decora* sialidase. The BLAST search however, did provide some interesting leads. It showed that the β barrel domain seen the nanI structure could be present in the sialidases from *Clostridium septicum* (Rothe *et al* ., 1991) and the large sialidase nanJ, from *Clostridium perfringens (Shimizu et al ., 2002)*. Both of these sialidases have stretches of amino acid sequences with 44 % identity to the sequence in the nanI sialidase and occur in topologically equivalent locations Figure 4.5.2. It is too early to speculate on a possible function for this domain in either the *C.perfringens* nanI or *M.decora* sialidases. However, the structure of the sialidase from *C.septicum* and of the nanJ sialidase from *C.perfringens* may well shed more light on this intriguing domain and its function in these enzymes.

(a)



(b)

Image rotated 90 °

Figure 4.7.1. (a) Topology diagram of the β-barrel domain in the *C.perfringens* nanI sialidase. The diagram is drawn from residue 356 (Alanine) to 428 (Threonine). The topology was calculated by promotif (Hutchinson and Thornton, 1996) and drawn in Topdraw (Bond, 2003). (b) The two β-barrel domains from *C.perfringens* nanI (yellow) and *M.decora* trans-sialidase (blue) superposed. The two domains superimpose with an r.m.s.d. of 1.42 Å for 67 C_α positions considered equivalent.

```
NANI                                                        SSNYRIPALF 271
NANJ                                                        SRGYRIPALE 404
NANH                                                        SNNFRIPALY 401
                                                            * .:*****

NANI.          KTKEGTLIASIDARRHGGADAPNNDIDTAVRRSEDGGKTWDEGQIIMDYPDKSSVIDTTL 331
NANJ.          TTKKGTVLASIDVRNNGDHDAPNNNIDVGIRRKEVNG-EWEEGKVILDYPGKSAAIDTSL 463
NANH.          TTKDGTVLASIDVRKGGGHDAPNN-IDTGIKRSTDGGVTWDEGKIILDYPGASSAIDTSL 460
               .**.**::****.*. *. ***** **..::*.  .*  *:**::*:***. *:.***:*

NANI.          IQD------DETGRIFLLVTHFPSKYGFWNAGLGSGFKNIDGKEYLCLYDSSGKEFTVRE 385
NANJ.          MSATIEENGIEKERIFLIVTHFPEGYGFPNTEGGSGYKEIDGKYYFILKDAQNNEYTVRE 523
NANH.          LQD------DETGRIFLIVTHFAEGYGFGNSKTGSGYVEIEGKRYLKLLGANDTIYTVRE 514
               :.       *. ****:****.. *** *:  ***: :*:** *: * .:... :****

NANI.          N-VVYDKDGNKTEYTTNALGDLFKNGTKIDNINSSTAPLKAKGTSYINLVYSDDDGKTWS 444
NANJ.          DGIVYNSEGNQTDYVMKNDKTLIQNGEEVGNALLSNSPLKAVGTAHIEMIYSDDDGKTWS 583
NANH.          G-VVYDSNGEATNYTVDNNNELYENGNRIGNVLLSNSPLKVMGTSFLSLIYSDDDGQTWS 573
               . :**:.:*: *:*. .    * :** .:.*   *.:***. **:.:.::******:***
```

Figure 4.7.2. Clustal W (1.82) multiple sequence alignment of the β-barrel domain present in the *C.perfringens* nanI sialidase (NANI) and the putative β-barrel domains present in the *C.perfringens* nanJ (NANJ) and *C.septicum* (NANH) sialidases, as identified through BLAST. The RIP and Asp-box sequence motifs are highlighted in red and the β-barrel domain in yellow. This alignment suggests that the β-barrel domain is in the same topological position in all three sialidases. The consensus line shows identical residues marked with an asterisk (*).

4.8. Surface Charge: Implications for sialic acid recognition and adaptation to the environment.

The distribution of charge on the surface of nanI was analysed using GRASP and found to be asymmetrical, with a high negative charge on the face of the protein opposite the catalytic site (Figure 4.6.1). Similar charge distributions were observed for the sialidase from *M.viridifaciens* (Gaskell *et al* ., 1995), *V.cholerae* (Moustafa, 2004) and the trans sialidase from *M.decora* (Lou *et al* ., 1998a). It was suggested that the trans sialidase from *M.decora* might utilise

such a feature to increase the enzymes efficiency by avoiding incorrect orientations on a sialoglycoconjugate rich cell surface (Lou *et al* ., 1998a).

The electrostatic character of sialidase active sites is usually conserved, due to mainly to the arginine triad, which dominates with a highly positive potential on one side of the pocket. This potential is used to orientate the sialic acid correctly by locking the carboxylate group at C1 into position. In addition to this, the calculated dipole moment of the protein points directly through the arginine triad, thus drawing the substrate in the correct orientation.

The charge asymmetry seen in the bacterial sialidases may be a characteristic peculiar to secreted sialidases. If the theory proposed by Luo *et al* were correct, then those sialidases that remain tethered to the cell surface would not require such an intrinsic 'ballast' to orientate them. The surface charge of the trans-sialidase from the human pathogen, *Trypanosoma cruzi*, (TcTS) was analysed to see if this was the case. The TcTS enzyme is a developmentally expressed, GPI-anchored, surface trans-sialidase that is distantly related to both the viral and bacterial sialidases, and whose structure was recently determined (Buschiazzo *et al* ., 2002). The overall charge on the surface of TcTS is not similar to any of the bacterial sialidases, showing in fact the opposite type of charge asymmetry, a highly positive surface containing the catalytic site contrasting with the relatively uncharged opposite face (Figure 4.6.2). It would be interesting to consider the charge asymmetry seen in the nanI sialidase to have originated from a requirement to avoid being trapped in an unfavourable orientation. It certainly seems plausible, given the absence of such asymmetry in the tethered sialidases, that some evolutionary pressure has been applied to cause this characteristic to have established itself in the bacterial sialidase proteins.

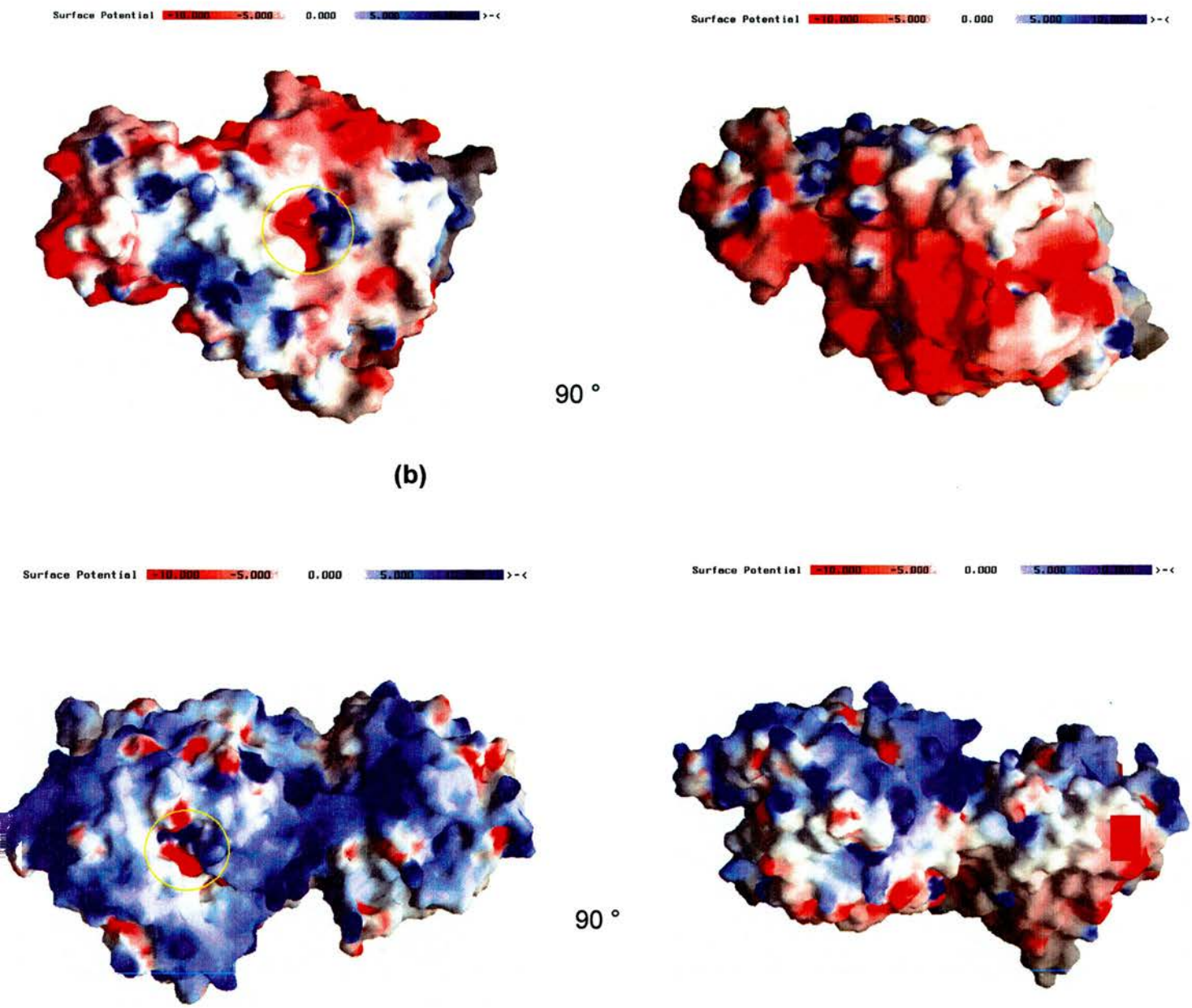


Figure 4.8.1. (a) The electrostatic potential mapped onto the surface of *C.perfringens* nanI and (b) the *T.cruzi* trans sialidase (PDB 1S0K). The potentials are colour coded – 10.0 KT in red to 10.0 KT in blue. The active site pockets are enclosed within the yellow circle on both enzymes.

4.9. Discussion.

The crystal structure of the 50 kDa catalytic domain from the *C.perfringens* nanI sialidase has been determined to 0.97 Å. It was found to consist of two domains, a six bladed β-propeller domain into which a small β-barrel fold had been inserted. The structure has underlined the previously observed modular architecture of the sialidase superfamily and diversity of the folds found associated with catalytic β-propeller. The overall architecture of the β-propeller fold is very similar to the previously reported bacterial sialidase structures, showing greatest similarity to the catalytic domains of the sialidase from the soil bacterium *M.viridifaciens* and the leech trans sialidase from *M.decora*. The former has 29 % sequence identity and the latter 36 %, so this was chosen as the initial molecular replacement model. From this analysis, it appears that both could have been used in this way to phase the initial diffraction data. This result serves to highlight the apparent conservation of fold over sequence in protein folding space.

It was extremely interesting to observe the overall similarity of the 50 kDa nanI structure to that of the leech trans sialidase. The relatively good r.m.s.d. values for the β-propeller domain were expected, but the similarity in the small β-barrel domain was not. It has been speculated for some time that the bacterial sialidases have acquired these extra domains in addition to the canonical catalytic β-propeller via horizontal gene transfer mediated most likely through phages (Roggentin *et al* ., 1993).

The occurrence of two very similar folds in sialidases from different evolutionary lineages seems to indicate that bacterial sialidases are still very active in picking up subsidiary domains from other proteins and organisms. The current structural diversity within the sialidase superfamily also shows a remarkable adaptation of the basic sialidase unit, the six bladed β-propeller. From a purely engineering perspective, the adoption of the almost circular β-propeller as the hub of these multi-domain proteins seems a very fortuitous

event. The β-propeller can protect the integrity of the active site, while seeming able to accommodate a diverse range of bulky domains, not only at the N and C terminus but also within the propeller itself.

The highly asymmetrical distribution of charge on the surface of the nanI sialidase has been seen previously in other sialidases, most notably in the leech enzyme (Lou *et al* ., 1998a) and the sialidase from *V.cholerae* (Moustafa, 2004). The suggestion, put forward by Lou *et al* . (Lou *et al* ., 1998a) that this feature could help orientate the protein more efficiently of the anionic sialyl -rich mucin or epithelial cell surfaces is plausible. Comparing the charge distribution on nanI to that of the trans-sialidase from *T.cruzi* (TcTS) shows the latter to have no noticeably negative surface. In fact, this sialidase has a highly positive charge distribution on the face containing the active site pocket. The TcTS enzyme is a GPI-anchored protein and so does not have the same freedom of movement as the secreted bacterial and leech enzymes. It seems possible therefore, that the secreted sialidases have evolved this charge asymmetry in response to the anionic siaylglycoconjugate rich cell surfaces they encounter.

The N-terminal domain of the nanI sialidase was removed prior to this structural study due to its susceptibility to cleavage by an unknown protease. A recent kinetic study performed on the full-length nanI sialidase indicates that this domain has a significant impact on the kinetic efficiency of the enzyme towards polyvalent substrates, such as a sialylated oligosaccharides (Thobhani *et al* ., 2002). Indeed, this study suggests that the majority of bacterial sialidases have acquired these extra domains for this purpose. The similarity between the C-terminal domains of both the nanI and leech trans sialiadase make it extremely likely that the N-terminal domain of nanI looks and sits in a similar position. Structural studies on the sub-cloned N-terminal domain of nanI are underway; if proved correct, this would further underline the similarity between these two enzymes.

One of the remarkable extended conclusions from structural studies on the sialidase superfamily is that all sialidase catalytic sites include the same set of at least seven conserved amino acid residues (Taylor, 1996). These include a

tri-arginine cluster that interacts with the sialyl C-1 carboxylate, a tyrosine-glutamic acid dyad that form a covalent intermediate with the substrate through the O^{η} of the tyrosine and an aspartic acid at the entrance to the pocket acting as an acid/base catalyst. These features are conserved in the nanI structure. Also conserved is the usually present hydrophobic pocket for the N-acetyl group at C-5, which in the nanI structure is made up by mostly aromatic side chains. The most interesting feature to have been exposed in the active site of the nanI structure is the presence of a tryptophan side chain extending over the ative site pocket directly above the hydrophobic pocket. A comparison of all sialidase structures to date shows this feature to be unique to the nanI sialidase. It is impossible to understand the significance of this unique alteration in the apo structure. The structures of two complexes have been determined to aid in understanding the role played by this chemical group, in the nanI reaction mechanism.

Chapter 5

Structure of the nanI ligand complexes: Structural Insights into a novel catalytic mechanism for a bacterial sialidase.

5.1. Summary

The structure of the catalytic domain of nanI in complex with the reaction product, α-Neu5Ac (sialic acid), and a covalent sialyl-enzyme intermediate, at 0.98 Å and 1.70 Å respectively are described. This work represents the first structural study showing the presence of both α-sialic acid and a glycosyl-enzyme intermediate in a bacterial sialidase. These structures expand the existing studies detailing the mechanism and nature of reaction intermediates in these enzymes.

The trapping of the reaction product, α-sialic acid, was the result of uncovering a hitherto unknown chemical reaction catalysed by nanI; the hydration of the C2=C3 double bond of the transition state analogue Neu5Ac2en (DANA). DANA is generally known as a micromolar inhibitor of bacterial and viral sialidases. Crystals of nanI grown in the presence of 10 mM DANA revealed unambiguous electron density for α-sialic acid in the active site pocket. The presence of α-sialic acid in the active site and refinement of this complex at atomic resolution, has allowed the analysis of this compound for the first time in a bacterial sialidase. This structure, coupled with that of the covalent intermediate provide a very detailed illustration of the role played by substrate distortion in the reaction mechanism of nanI.

The trapping of a covalent intermediate using a novel fluorinated sialic acid derivative supports earlier studies on the trans-sialidase from *Trypanosoma cruzi*. The structure of the nanI intermediate clearly shows a covalent bond between the conserved tyrosine (Tyr655) and the anomeric carbon of the sialic acid derivative. This structure has important implications for understanding the reaction intermediates in the catalytic mechanism of this enzyme. A number of fundamental questions still remain concerning the role played by a number of

conserved catalytic residues, including the tyrosine nucleophile, the aspartic acid, and the conserved glutamic acid. Two different views are currently being debated about the choice of tyrosine as the catalytic nucleophile and these are discussed in light of the new structures presented here.

Comparison of the active site of nanI with that of the intra molecular trans-sialidase from the leech *Macrobdella decora*, show that these two enzymes have different active site chemistries. Rearrangement of a loop region close to the glycerol group of the substrate serves to make the leech sialidase an intra molecular trans-sialidase, releasing 2,7-anhydro-sialic acid instead of α-sialic acid (Lou *et al* ., 1998b). The presence of a tryptophan residue above the conserved arginine triad also gives this enzyme a strict Neu5Acα2-3Gal specificity (Lou *et al* ., 1998a). NanI has no such bulky hydrophobic groups in this position, and subsequently has broad substrate specificity, capable of accepting sialic acid bound either α2-3, α2-6 or α2-8 to the subterminal aglycon (Roggentin *et al* ., 1995).

The role of a tryptophan residue (Trp354), which sits above the active site pocket of nanI but at the opposite end to the tryptophan in the leech sialidase, is currently unknown. It is possible, given its location above the glycerol group of sialic acid, that it may exclude water from the active site and make the resulting hydrogen bonds to the glycerol hydroxyls stronger. This shielding may have the effect of amplifying the hydrogen bond rearrangements seen between the two nanI complexes, and could account for the observed conformation changes seen in the sugar ring of the sialic acid ligand.

5.2. Atomic resolution structure of the nanI - α-Neu5Ac complex.

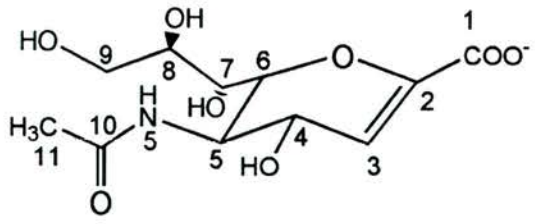
The crystal structure of the nanI catalytic domain in complex with α-Neu5Ac (sialic acid) is reported here to a resolution of 0.98 Å. This complex was the unexpected result of trying to co-crystalise the nanI sialidase with the transition state analogue Neu5Ac2en (DANA) (Figure 5.2.1).

DANA has been used with great success in the structural characterisation of the active site in the sialidase superfamily. It was used initially to identify and map the active sites of the influenza A and B virus neuramindases (Varghese *et al* ., 1992). Given the higher affinity for DANA over free sialic acid, it was proposed that DANA acted as a transition state analogue in the reaction pathway (Miller *et al* ., 1978). The planer arrangement around C2 of DANA suggested that sialic acid went through a similar configuration via an oxocarbonium ion intermediate, and introduced the possibility of substrate distortion in the reaction mechanism.

DANA is an inhibitor of bacterial and viral sialdiases, with K_i values ~ 10^{-6} M range. All previously studied bacterial and viral sialidases have been crystallised in complex with this compound (Crennell *et al* ., 1994; Crennell *et al* ., 1993; Gaskell *et al* ., 1995; Lou *et al* ., 1998b), as have the tyrpanosomal sialidases from *T.rangeli (Buschiazzo et al ., 2000)* and the trans-sialidase from *T.cruzi (Buschiazzo et al ., 2002)*. These complexes have helped characterise the active sites of these enzymes, and aided in the understanding of the hydrogen bonding interactions made by the transition state intermediate.

Initial refinement of the nanI diffraction data included DANA modelled into the F_O-F_C density in the active site. It became clear, however, that DANA did not fit the difference density around the anomeric carbon, which was sitting above the plane of the ring. A strong positive peak (5 σ) was also seen ~1.4 Å from this atom, corresponding to the O2 hydroxyl of α-sialic acid. As a result of this density, α-sialic acid was modelled in. The F_O-F_C and $2F_O$-F_C electron density maps are shown in Figure 5.2.3.

(a)



(b)

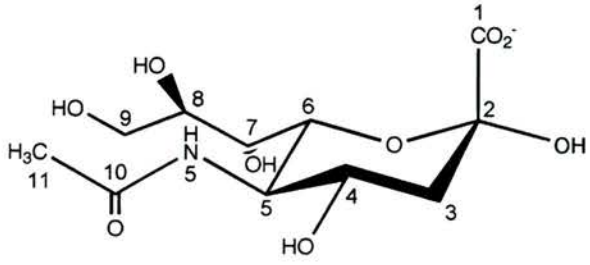


Figure 5.2.1. (a) Chemical structure of Neu5Ac2en (DANA); a transition state analogue of sialidase catalysed reactions. (b) Chemical structure of α-Neu5Ac (α-sialic acid).

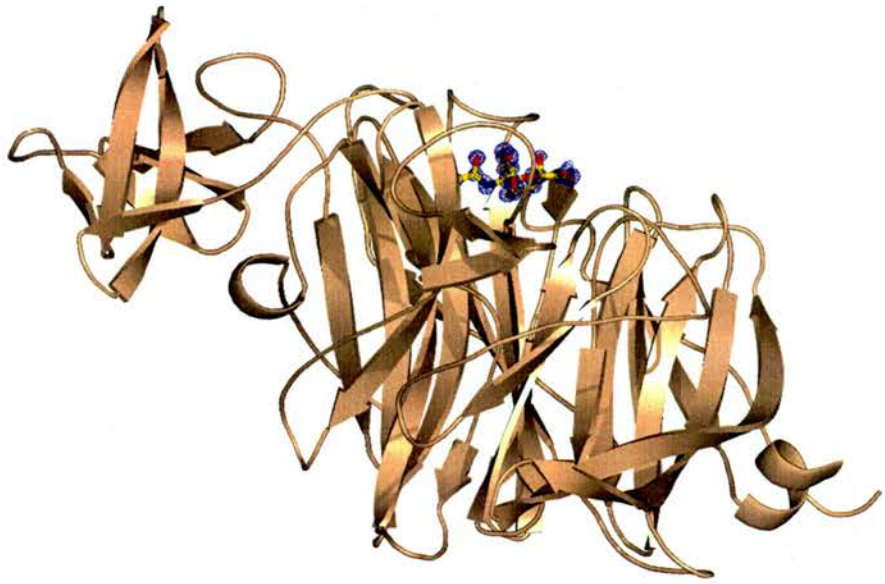
This finding was wholly unexpected and has two obvious implications for the chemistry of the nanI sialidase. The first is that nanI can catalyse the formation of sialic acid from DANA through hydration of the C2=C3 double bond, with retention of anomeric configuration. The second is that sialic acid might act as an inhibitor of the enzyme once formed.

It was important to carry out a number of control experiments with nanI to rule out contamination of the DANA stock solution with sialic acid, or another compound. The mass spectrum of a fresh sample of DANA did not show any presence of sialic acid; indeed the hydration of alkene double bonds is not spontaneous at room temperature, usually requiring a metal catalyst such a palladium in the presence of concentrated HCl (McMurray, 1996). Further characterisation of this reaction using 1D ^{1}H NMR provided conclusive evidence

that DANA was being converted to sialic acid by the nanI enzyme in solution at room temperature, 293 K (Chapter 6).

The presence of α-sialic acid in the active site of the co-crystals also suggested another characteristic of this complex, stability. For sialic acid to remain in the active site for the duration of the crystallisation process, which took a minimum of five days at room temperature, the complex must be stable. Sialic acid acts as a poor inhibitor of the influenza viral neuraminidases (K_i ~10^{-3} M), and as such allowed the complex with the enzymes to be determined (Varghese *et al* ., 1992). Trapping of α-sialic acid in a bacterial sialidase however, has not been possible, due to the high turnover number of these enzymes. Given the presence of sialic acid in the nanI complex, it may be possible that the sialic acid also functions as a competitive inhibitor of this enzyme. Further kinetic studies will need to undertaken to investigate this further.

(a)



(b)

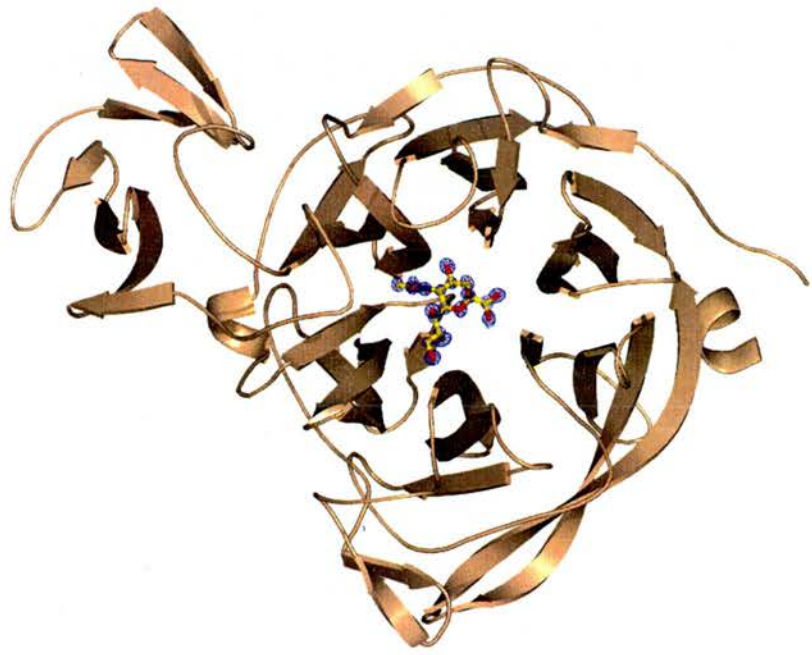
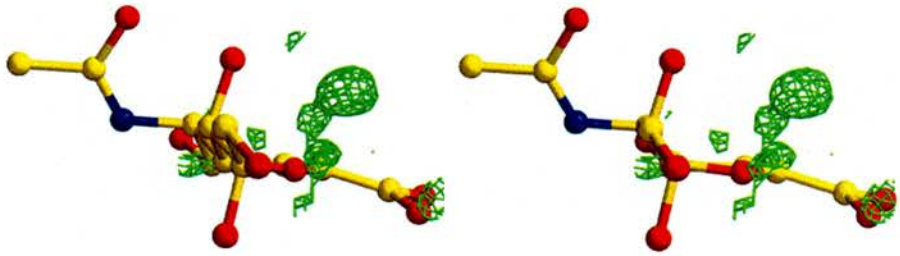


Figure 5.2.2. (a) Schematic representation of *C.perfringens* nanI with the ligand α-Neu5Ac (sialic acid) sitting in the active site. The ligand is shown with the final $2F_O$-F_C electron density map (blue) contoured at 2 σ. (b) view rotated 90°.

(a)



(b)

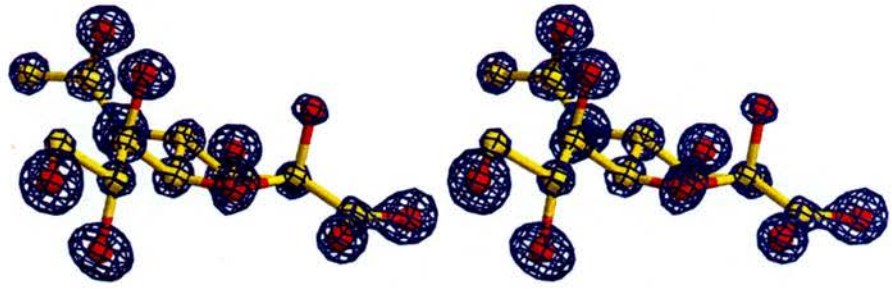


Figure 5.2.3. (a) The ligand Neu5Ac2en (DANA) is shown during the initial refinement, with the F_O-F_C electron density map (green) contoured at 3 σ. (b) The ligand α-Neu5Ac is shown with the final $2F_O$-F_C electron density map (blue) contoured at 2 σ. The sugar ring is observed in a distorted $B_{2,5}$ boat conformation.

5.2.1. Interactions of α-Neu5Ac with the catalytic site.

The presence of α-Neu5Ac (α-sialic acid) in the catalytic site of the nanI sialidase, coupled with the atomic resolution data, provide for the first time a very detailed account of how sialic acid sits in the active site of a bacterial sialidase, and importantly the conformation of the pyranose ring.

The interactions between the protein and sialic acid are very similar to those between Neu5Ac2en (DANA) and the bacterial sialidases from *V.cholerae* (Crennell *et al* ., 1994), *S.typhimurium* (Crennell *et al* ., 1993) and *M.viridifaciens* (Gaskell *et al* ., 1995) and are listed in Table 5.2.1. and shown in Figure 5.2.5.

The carboxylate group of sialic acid interacts with the arginine triad (Arg266, Arg555, Arg615) via strong ionic interactions. The guanidine group of Arg615 faces the carboxylate group directly and forms a very strong interaction with hydrogen bond distances of 2.9 Å. The positive potential generated by the arginine triad will almost certainly play a major role in channelling the substrate into the active site and guiding the negatively charged carboxylate to the correct orientation, as illustrated in Figure 5.2.4.

The O2 hydroxyl sits in the axial orientation on the anomeric carbon and forms two hydrogen bonds, one to the acid/base catalyst, Asp291 and to a water molecule, coordinated also by O7. The reduced density for this hydroxyl compared with the other similar atoms in the ligand suggests that this atom may only have partial occupancy in this structure. Analysis of the B-factors (Table 5.2.2) shows that the O2 hydroxyl has a significantly higher B-value compared with the other ligand atoms, approximately twice the value. It is possible that the active site has a mixture of both α-Neu5Ac and Neu5Ac2en, which could explain the distorted $B_{2,5}$ boat conformation.

Table 5.2.1. B-factor analysis of the α-Neu5Ac (sialic acid) ligand modelled into the active site of the nanI sialidase.

Sialic Acid atom	B-factor (Å^2)
O1A	13.8
O1B	9.5
C1	11.5
C2	11.2
O2	20.71
C3	2.8
C4	9.4
O4	8.4
C5	9.1
C6	8.2
O6	11.3
C7	9.1
O7	10.9
C8	9.3
O8	11.0
C9	12.1
O9	11.2
N5	7.8
C10	7.1
O10	9.8
C11	8.5

The O4 hydroxyl interacts via hydrogen bonds with a conserved arginine, Arg285 and an aspartic acid, Asp328.

The N-acetyl group sits in a hydrophobic pocket made by residues Phe460, Phe353, Phe347, and Ile327. The nitrogen atom of the N-acetyl group also interacts with Asp328, an interaction that has been seen previously in complexes of sialidases with DANA. The formation of this interaction at N5, coupled with the ionic and hydrogen bonding interactions at the carboxylate group, serve to 'lock' the sialic acid into the active site at either end of the molecule. It is thought this characteristic may play a key role in sugar distortion and catalysis (Amaya *et al* ., 2004), and will be discussed further when analysing the covalent sialyl-enzyme intermediate in section 5.4. of this chapter.

The O10 hydroxyl makes a single hydrogen bond to a water molecule sitting above the hydrophobic pocket. The glycerol side chain of sialic acid makes a total of five hydrogen bonds. The O7 hydroxyl makes a single hydrogen bond to the water coordinated also with the O2 hydroxyl, with O8 making two hydrogen bonds to a tyrosine (Tyr485) and a water molecule. The terminal hydroxyl, O9, makes two hydrogen bonds, one to the same water coordinated by O8 and a glutamine side chain (Gln493). Finally, beneath the anomeric carbon is the hydroxyl group of the catalytic nucleophile, Tyr655, which is poised for attack at the anomeric carbon at a distance of 2.96 Å.

The sialic acid ring adopts a distorted $B_{2,5}$ conformation, as shown in Figure 5.2.3. A similar sialic acid ring conformation was seen in the Michaelis complex between an aspartic acid mutant of *T.cruzi* trans-sialidase and α(2,3)-sialyl-lactose (SL) (Amaya *et al* ., 2004). This conformation of the sugar ring results in a pseudo-axial orientation of the glycosidic bond being cleaved, bringing it within 2.5 Å of the acid catalyst (Asp291) for proton transfer, as is seen when the SL is modelled into the nanI structure (Figure 5.4.1). Such a distortion places the anomeric carbon in a suitable position for in-line attack by the nucleophile (Tyr655) with minimal encumbrance from 1,3-diaxial repulsions, and so satisfying the stereochemical requirements for the incipient oxocarbonium ion.

Figure 5.2.4. highlights the prominent position of Trp354 at the mouth of the active site pocket. It is currently unclear if this residue plays any significant role in the chemistry of the active site. Comparisons with other sialidases however, show this feature is unique to the nanI enzyme. Trp354 is unlikely to play a role in substrate selectivity; it is simply too far from where the glycosidic bond would sit in the substrate. It was also known previously that the nanI sialidase has very broad substrate specificity; capable of accepting sialic acid bound either α2-3, α2-6 or α2-8 to the subterminal aglycon, galactose (Roggentin *et al* ., 1995). Its position, sitting above the hydrophobic pocket of the enzyme but extending over the glycerol group of sialic acid, may provide a clue to its function. It may shield this part of the active site from bulk solvent and help contribute to the strength of the hydrogen bonds made to the glycerol group through desolvation effects. This idea will be returned to later when we consider a possible mechanism for the hydration of DANA by nanI in Chapter 6.

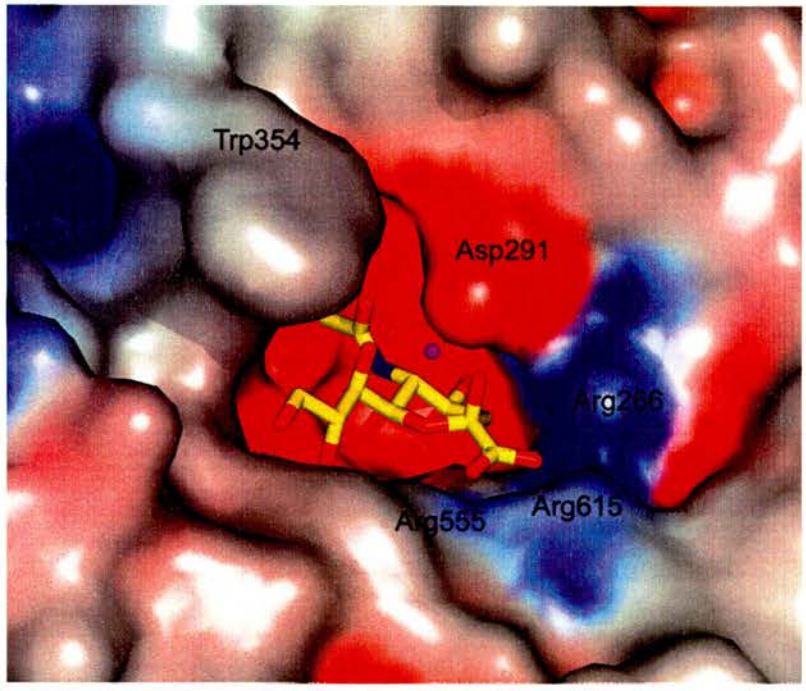


Figure 5.2.4. Electrostatic potential at the active site of the *C.perfringens* nanI with the boud ligand (Neu5Ac). The positive charge created by the arginine triad nicely complements the negative charge of the carboxylate group. This has the effect of 'locking' the ligand in place at C1. The electrostatic potential surface is colour coded ranging from –10 KT (red) to 10 KT (blue).

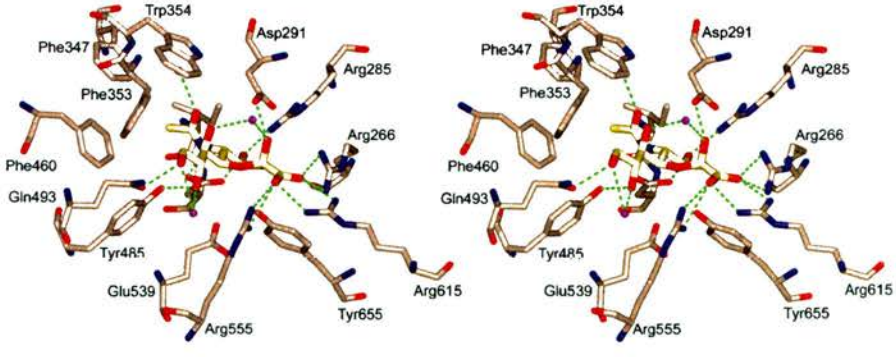


Figure 5.2.5. Stereo view of the active site of nanI in complex with α-Neu5Ac (sialic acid). The hydrogen bonding interactions are drawn as dotted lines; water molecules are represented as magenta spheres.

Table 5.2.2. Comparison of hydrogen bonding interactions between the bound ligands, α-sialic acid and 3-fluoro sialic acid complexed with nanI.

Sialic acid atom	Protein/Water atom	Distance (Å)	
		α-Sialic acid	Covalent intermediate
O1A	Arg615-$N^{\eta 2}$	2.91	2.77
	Arg555-$N^{\eta 1}$	3.15	3.12
	Arg555-$N^{\eta 2}$	3.11	----
	H_2O	----	2.91
	H_2O	----	3.09
O1B	Arg615-$N^{\eta 1}$	2.92	3.00
	Arg266-$N^{\eta 1}$	3.01	2.93
	Arg266-$N^{\eta 2}$	3.02	3.01
	H_2O	----	3.13
O2	Asp291-$O^{\delta 2}$	2.57	----
	H_2O	----	3.13
F1	Arg266-$N^{\eta 1}$	----	2.91
	Asp291-$O^{\delta 1}$	----	2.88
O4	Arg285-$N^{\eta 2}$	2.97	2.91
	Asp328-$O^{\delta 1}$	2.68	2.71
O7	H_2O	3.15	----
	H_2O	----	2.76
O8	H_2O	2.79	2.78
	Tyr485-O	2.82	2.67
O9	H_2O	3.06	3.18
	H_2O	----	2.91
	Gln493-$O^{\varepsilon 1}$	2.73	----
O10	H_2O	2.84	2.91
N5	Asp328-$O^{\delta 2}$	2.86	2.72

Table 5.2.3. Comparison of selected close contacts between the bound ligands, α-sialic acid and 3-fluoro sialic acid complexed with nanI.

Sialic acid atom	Protein atom	Distance (Å)	
		α-Sialic acid	Covalent intermediate
C1	Tyr655-O	2.82	2.43
	Tyr655-C^{z}	3.57	3.04
	Tyr655-$C^{\varepsilon 1}$	3.69	3.12
C2	Tyr655-C^{z}	3.88	2.44
C2	Tyr655-O	2.96	----
C3	Tyr655-O	3.06	2.54
C3	Tyr655- C^{z}	3.62	3.04
C3	Ile267- $C^{\delta 1}$	3.88	3.66
C4	Tyr655-O	3.62	3.12
C4	Ile267-$C^{\delta 1}$	4.00	3.76
C5	Ile327- $C^{\delta 1}$	4.59	4.64
C6	Tyr655-O	3.54	3.08
C9	Phe353-C^{ε}	3.64	3.55
C9	Trp353-C^{z2}	5.66	5.11
C10	Ile327- $C^{\delta 1}$	3.74	3.80
C11	Phe460-$C^{\varepsilon 2}$	3.86	3.71
C11	Ile327- $C^{\delta 2}$	4.23	4.35
C11	Phe347- $C^{\varepsilon 1}$	4.24	4.33

5.3. Structural characterisation of a covalent sialyl-enzyme intermediate.

Until recently it was thought that the mechanism utilised by the bacterial sialidases involved water directly attacking a sole oxocarbonium ion intermediate, via an internal return mechanism. This was a result of kinetic isotope effect and modelling studies on the influenza virus sialidase (Chong *et al* ., 1992a; Taylor and von Itzstein, 1994). The similarity between the active sites of the influenza virus and the bacterial sialidases suggested that both would function with a similar, although not necessarily identical mechanism.

An alternative mechanism was proposed that involved the formation of a covalently bound sialyl-enzyme intermediate (Davies *et al* ., 1998). This latter proposal is analogous to the standard double displacement reactions of retaining glycosidases that hydrolyse acetal linkages (Zechel and Withers, 2001).

Using a fluorinated sialic acid derivative, α-2,3-difluoro sialic acid (**1**), it has been possible to trap this compound covalently bound to the conserved tyrosine residue (Tyr655) and determine the crystal structure to 1.7 Å. This structure, the first to show such an intermediate in a bacterial sialidase, unambiguously identifies tyrosine as the catalytic nucleophile in this enzyme.

Formation of a stable covalent sialyl-enzyme intermediate was achieved by incorporating a fluorine substituent at C3, adjacent to the anomeric center. Flourine then acts to inductively destablise the positively charged oxocarbonium ion transistion state, thereby slowing both the formation and turnover of the covalent intermediate. The presence of an anomeric fluoride group ensures that the intermediate is kinetically accessible, as fluorine is a good leaving group.

The formation of a covalent intermediate with (**1**) was shown prior to undertaking any crystallographic determination. Our collaborators, Professor Stephen Withers and Dr. Andrew Watts at the Department of Chemistry, University of British Columbia, performed labelling studies and peptide mapping experiments using LC-MS/MS (Figure 5.3.1.a).

These studies showed the covalent inactivation of nanI occurring in a time dependent manner, with complete inactivation by (**1**) after only 1 minute, even at

very low concentrations (Figure 5.3.1.b). As a result, a detailed kinetic analysis of inactivation was not possible. However, it became clear from these initial studies that the rates of both inactivation and reactivation of nanI are relatively fast, in the order of seconds. The peptide mapping studies identified the site of attachment as Tyr655, as expected from the previous studies on the TcTS enzyme (Watts *et al* ., 2003). This information proved invaluable in designing successful soaking experiments and trapping the covalent intermediate in the crystal structure.

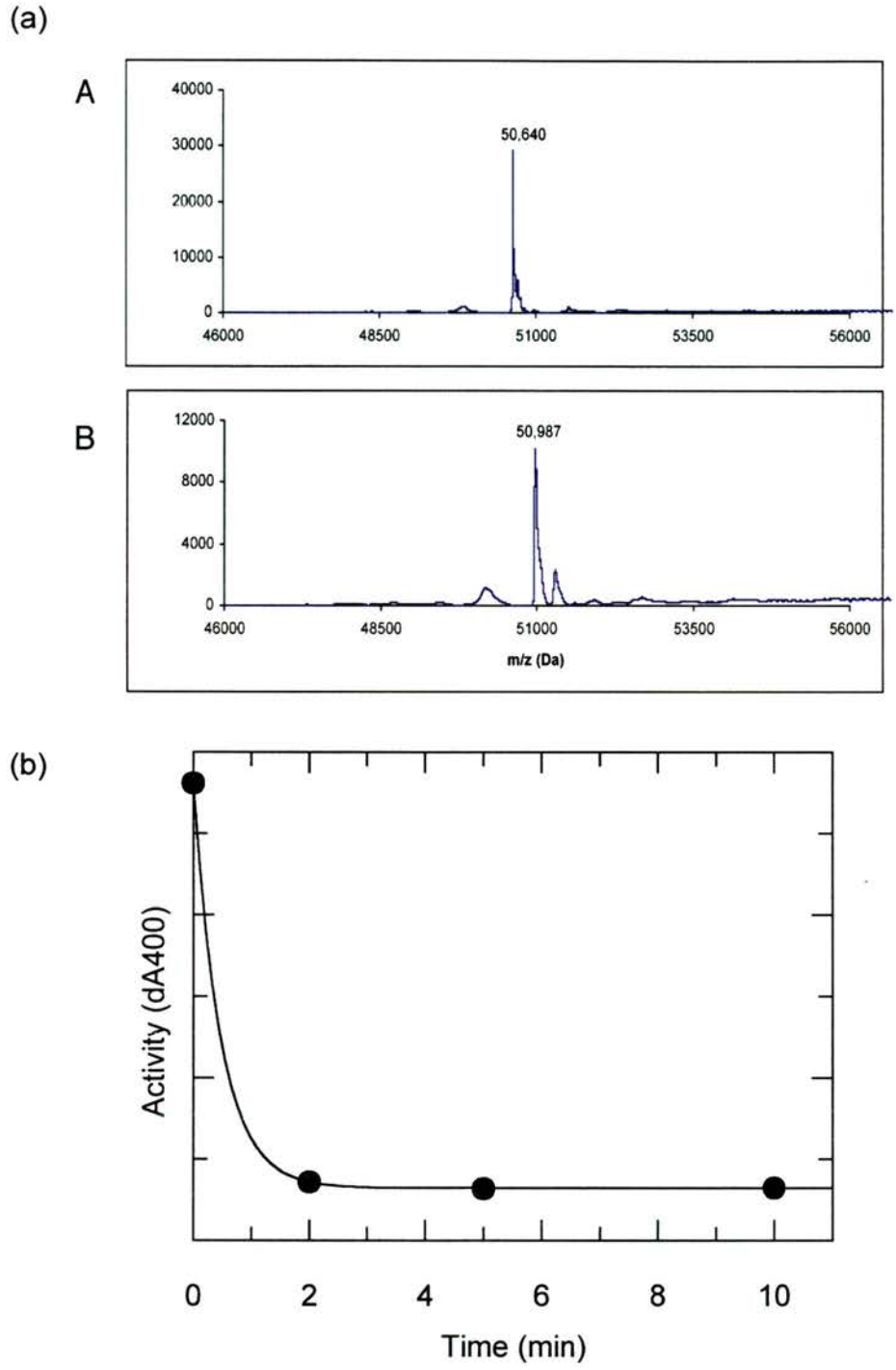


Figure 5.3.1. (a) Mass spectrum of unlabeled nanI sialiadase (A) and of nanI sialidase inclubated with 2 mM (**1**) (B). The mass increase of 347 Da between A and B indicates the covalent attachment of the 3-fluoro sialyl moiety. (b) Time dependent inactivation of nanI with 100 μM (**1**). After a period of ~ 20 minutes the reactivition of nanI was observed (unpublished results) indicating the breakdown of the covalent intermediate.

The $2F_O$-F_C electron density map shows the formation of the covalent complex in the nanI active site (Figure 5.3.2). A covalent bond is clearly visible in the electron density map between the oxygen atom of the Tyr655 side chain and the anomeric carbon of sialic acid. This bond was refined to a distance of 1.44 Å.

Formation of the covalent intermediate and the concomitant release of the aglycon moiety promote changes in the conformation of the sugar ring compared with α-sialic acid. The sugar ring is seen to relax into the $^2C_5$ chair conformation, with a β linkage to the catalytic nucleophile Tyr655. This conformational change in the sugar ring accommodates the change in relative position of Tyr655 and C2 of the substrate that is required to form the covalent bond. The anomeric carbon moves ~1.2 Å in shortening the Tyr655-C2 distance to the covalent bond length of 1.44 Å (Figure 5.3.2).

Table 5.2.1. and 5.2.2 list all the hydrogen bond and close contact interactions made by the covalent complex, respectively. In general these are very similar to those observed for the α-sialic acid complex. However, the conformational change of the sugar ring from the $B_{2,5}$ boat, to the $^2C_5$ chair, does cause a number of subtle differences.

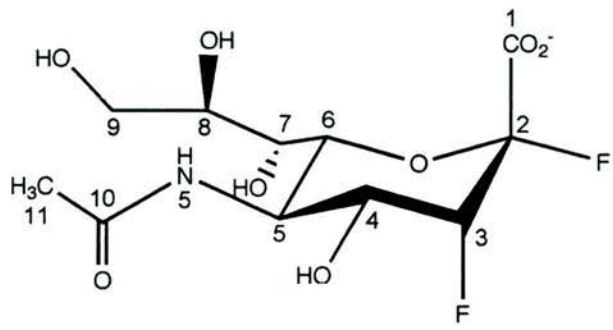
Of particular note are the appearances of seven water molecules, compared with only two in the α-sialic acid complex. The relaxed conformation of the ring places the C1 carboxylate group into the equatorial position, and this now hydrogen bonds to two of these waters.

Several studies have indicated that water rearrangement plays an important role in carbohydrate-protein interactions (Clarke *et al* ., 2001; Ladbury, 1996). The ordering of five extra water molecules in the covalent complex would be expected to decrease the entropy, and subsequently increase the free energy of the system ($\Delta G = \Delta H - T\Delta S$). It has been suggested that each water molecule that is ordered at an interface can contribute –0.5 to –3.0 kcal mol^{-1} to ΔH (Holdgate *et al* ., 1997). This increase in free energy would serve to destabilise the covalent intermediate and may act to drive the reaction forward to product formation.

The electronegative fluorine appears not to have a dramatic effect on the position of the sugar in the active site. It makes two hydrogen bonds to both the acid/base catalyst, Asp291 and Arg266.

The most noticeable difference observed between this complex and the one with α-sialic acid, is the change in hydrogen bonds around the glycerol group of the sugar. The O7 hydroxyl now interacts with two waters molecules, as compared with the one seen previously. The O9 hydroxyl now interacts only with a new water molecule, as compared with Gln493 in the previous structure. Interestingly, this new hydrogen bond points in the opposite direction to that seen with Gln493, suggesting a subtle shift of this group in the active site. There are no changes in the interactions made by the O8 hydroxyl.

(a)



(b)

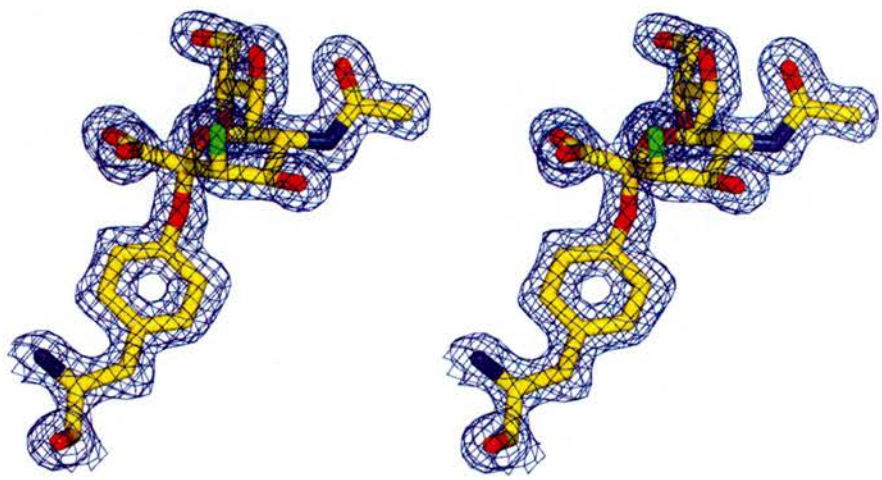


Figure 5.3.2. (a) Chemical structure of (1), 2,3-difluoro sialic acid used for trapping the reaction intermediate. (b) Stereo view of the electron density map ($2F_O$-F_C), contoured at 2σ, for the covalent intermediate bound to Tyr655 in the nanI active site.

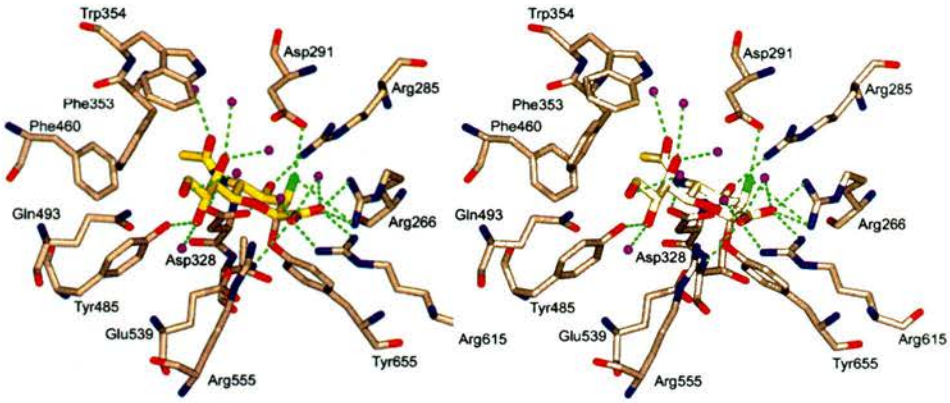


Figure 5.3.3. Stereo view of the catalytic active site with the covalent intermediate. The protein-ligand interactions are depicted as green dotted lines; water molecules are coloured magenta.

5.4. Sugar distortion: Implications for the catalytic mechanism of the nanI sialidase.

The structures of nanI in complex with the product, α-sialic acid, and a covalent sialyl-enzyme intermediate, provide a unique insight into the catalytic mechanism of this enzyme, in particular substrate distortion.

Comparison between the Michealis complex of the *T.cruzi* trans-sialidase with the substrate α2,3-sialyllactose, shows the conformation of the sialic acid sugar rings are very similar, adopting a $B_{2,5}$ boat configuration. The atomic resolution data on the sialic acid complex however, shows the sugar ring in a distorted twisted boat. The twist occurs at the carboxylate end of the sugar, with the O6 atom sitting below the plane of the ring and the C3 atom above (Figure 5.4.1). This figure also illustrates the shallow puckering of the ring at this end of the sugar. In a typical twist boat configuration, both C2 and C5 should be roughly puckered by equal amounts, yet in the sialic acid complex, C2 lies very close to the plane of the sugar ring. This is most likely a consequence of the

very strong ionic interactions between the carboxylate group at C1 and the arginine triad in the active site.

Using the position of sialic acid in the nanI complex, the substrate, α2,3-sialyllactose was modelled into the active site (Figure 5.4.1.a). The PDB coordinates for sialyllactose were taken from the complex with an active site mutant of the *T.cruzi* trans-sialidase (PDB code 1S0I). Although fairly crude, this model does serve to highlight the geometry of the glycosidic bond in relation to the acid base catalyst (Asp291) and nucleophile (Tyr655). In this conformation, the glycosidic bond is positioned in a pseudo-axial orientation, brining it within ~2.5 Å of the acid catalyst for proton transfer. This geometry provides minimal obstruction from 1,3-diaxial repulsions and provides the required stereoelectronic requirements for oxycarbonium ion formation. The catalytic nucleophile is now directly below the anomeric carbon at a distance of ~2.9 Å, and in the perfect position to stabilise the incipient oxocarbonium ion and 'pull' the ring into $^4H_5$ half chair transition state.

The authors of the TcTS study believe that the next stable species along the reaction coordinate is the covalent intermediate, bound to the catalytic nucleophile (Amaya *et al* ., 2004). No direct evidence exists for the formation of this species in the bacterial sialidases with natural substrates. However, the covalent sialyl-enzyme intermediate presented in this chapter provides the first evidence that such an intermediate can exist and be accommodated within the active site.

Formation of the covalent intermediate is achieved via the relaxation of the sugar ring into the $^2C_5$ chair conformation, through electrophilic migration of the anomerc carbon. The anomeric carbon moves 1.2 Å in shortening the distance to the OH atom of Tyr655 between free sialic acid and fluorinated derivative. Figure 5.4.2.b. shows an overlay of the three conformations adopted by the sugar ring in the two complexes. The pseudo half chair adopted by the transition state analogue DANA has also been modelled in. This figure highlights the proposed movement of the anomeric carbon of sialic acid through the reaction mechanism.

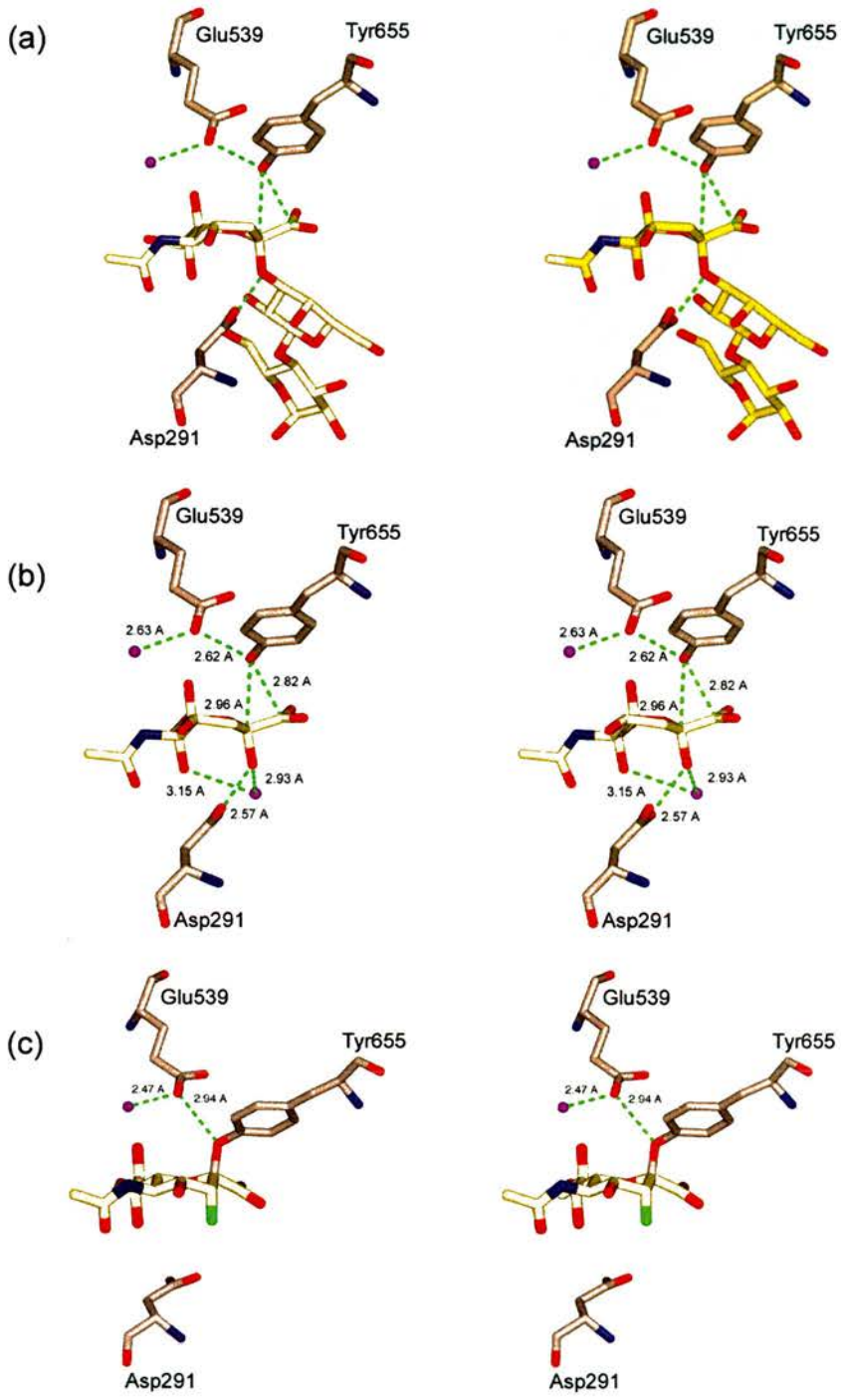


Figure 5.4.1. Stereo views of the relative positions of the catalytic residues in nanI. (a) The modelled structure of the Michaelis complex of nanI with α2,3-sialyllactose. (b) The sialic acid complex. (c) The sialyl-nanI intermediate. Water molecules are shown as magenta spheres, hydrogen bonds and close contacts as dashed green lines.

Also emphasised are the rigid positions of both the N-acetyl group at one end of the sugar and the carboxylate group at the other. These interactions serve to tether the substrate firmly at either end, and may provide the anomeric carbon with more flexibility by removing the need to form tight hydrogen bonds with the other ring constituents. The tight interaction of the carboxylate group with the arginine triad was discussed above as a possible reason for the much reduced puckering at C2 in the sialic acid complex.

An advantage of holding the ring in this distorted twist boat would mean a much reduced energy barrier to ring inversion. The ring conformation in the covalent intermediate is in a relaxed $^{2}C_{5}$ chair, which is one of the low energy states adopted by six membered rings in sugar chemistry. Nucleophilic attack by an incoming water molecule, activated by Asp291, is concomitant with ring inversion back to the twist boat. The conformation of the sugar ring shown in the sialic acid complex, suggests that the C2 would only need to move ~1.2 Å, and would be further stabilised by the subsequent hydrogen bonds created to both the Asp291 and a water molecule (Figure 5.2.4).

In all sialidases studied to date, the carboxylate group of a conserved glutamic acid (Glu539) forms a short hydrogen bond (distance 2.62 Å) with the hydroxyl group of the catalytic tyrosine. What effect this bond has on the reactivity of the tyrosine is unclear. The authors of the *T.cruzi* study have suggested that a deprotonated glutamic acid would act as a base catalyst for the attack of the tyrosine on the anomeric center, therefore acting in a charge relay system (Amaya *et al* ., 2004). This idea has also received support from a recent study on tyrosine mutants of the bacterial sialidase from *M.viridifaciens* (Watson *et al* ., 2003). This study showed that the rate of glycosidic bond cleavage in the wild type enzyme is dependent on the protonation state of the glutamic acid. The length of this hydrogen bond increases with the formation of the covalent intermediate (Figure 5.4.1), from 2.62 to 2.94 Å, indicating a weaker interaction once the intermediate has formed.

Atomic resolution crystallography can provide a number of ways to investigate the protonation state of key catalytic residues. The most obvious is

the appearance of difference density corresponding to hydrogen atoms. Within the structure of the sialic acid complex no such density is observed for the glutamic acid residue. However, the electron density cannot be interpreted in terms of protonation, as many of the ring and hydroxyl hydrogens are also absent in the difference density maps. Atomic resolution data do, however, allow very precise and unbiased positioning of the 'heavier' atoms in the structure, allowing bond lengths to be accurately determined and protonation states inferred. This technique has been used previously to assign protonation states to key catalytic residues in xylanase Xyn10A (Gloster *et al* ., 2004). The bond lengths of the carboxylate group of the glutamic acid are both 1.26 Å from C^{δ} to $O^{\varepsilon 1}$ and from C^{δ} to $O^{\varepsilon 2}$. This value is the same as the mean unrestrained carboxylate C-O bond distance and strongly suggests that the charge on this carboxylate group is delocalised, and that the residue is not protonated.

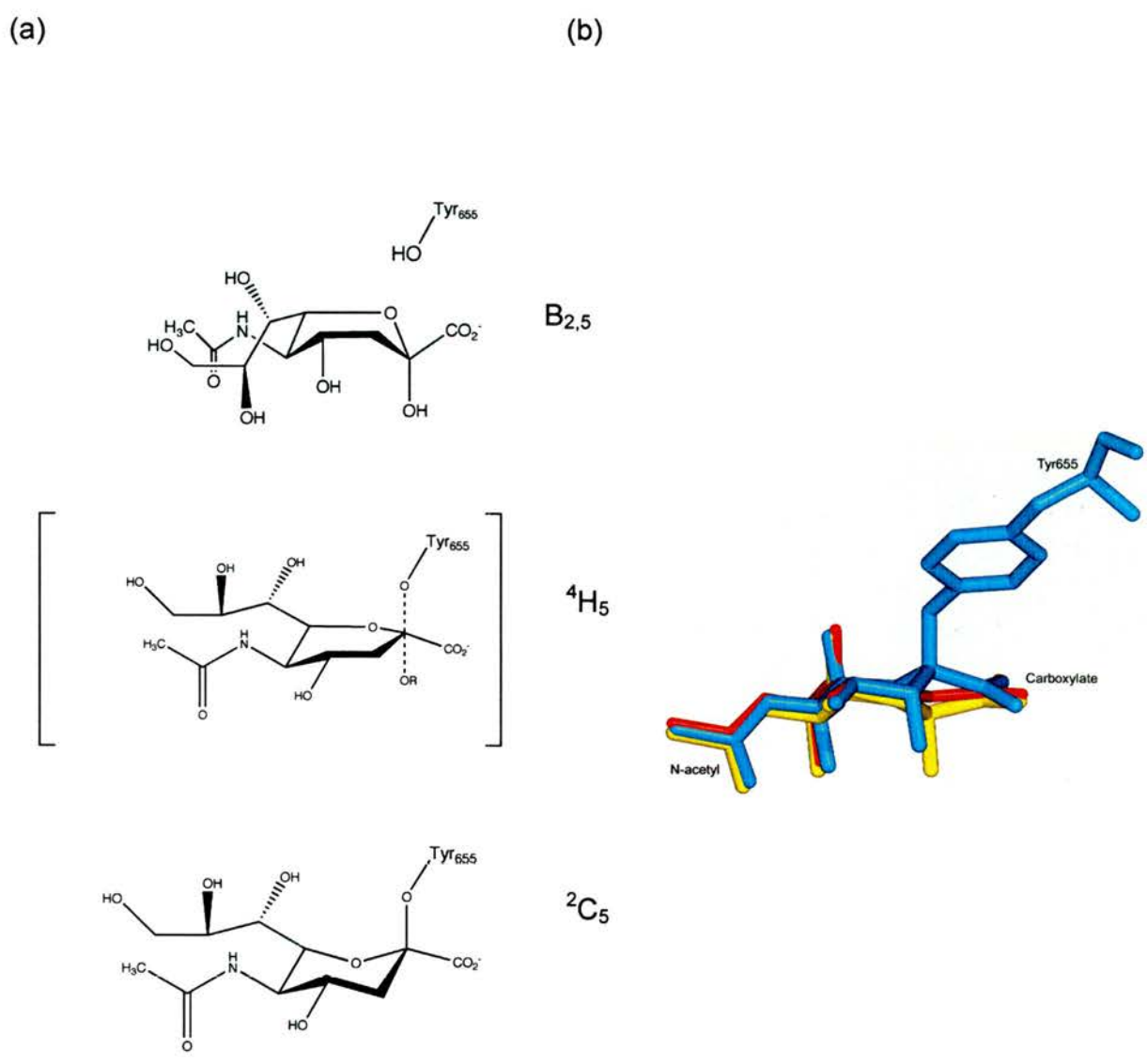


Figure 5.4.2. Sugar distortion in catalytic mechanism of nanI. (a) Sugar distortion observed in the nanI complexes. Top, sialic acid in the $B_{2,5}$ boat; middle, oxocarbonium ion transition state in the $^4H_5$ half chair; bottom, sialyl-nanI intermediate in the relaxed $^2C_5$ chair. (b) Overlay of structures with sialic acid (yellow), DANA (red) modelled in, and the sialyl-nanI intermediate (blue).

The structures of the nanI complexes show that the carboxylate group of Glu539 makes two conserved hydrogen bonds in the active site (Figure 5.4.1). One is to the tyrosine hydroxyl as discussed, the other to a strongly coordinated water molecule present in both complexes. This water molecule is also present in the apo structure, coordinated to a tyrosine residue and a further glutamic acid on the opposite side of the active site pocket, with hydrogen bond distances of 2.60 and 2.78 Å respectively. The presence of such a conserved water molecule might suggest a possible role in modulating the chemistry of the glutamic acid-tyrosine couple in the nanI enzyme. A water molecule in a similar position is seen in the covalent complex with the *T.cruzi* enzyme (Amaya *et al* ., 2004), but this disappears in both the apo and Michaelis complex structures.

Cleavage of a glycosidic bond between two sugar residues requires substantial general acid catalytic assistance, as well as general base catalysis to assist the attack of a nucleophilic water molecule (Zechel and Withers, 2001). The structures presented here clearly point to Asp291 as the acid catalyst for the initial step involving sialyl-enzyme formation, as it is well located to hydrogen bond with the glycosidic oxygen (Figure5.4.1.a). Further, within the sialyl-enzyme intermediate, it is well positioned to function as the general base catalyst assisting nucleophilic attack by the hydroxyl group of a water molecule. The contribution of this aspartic acid to the free energy of hydrolysis for natural substrates was calculated at > 19 kJ/mol for a homologous bacterial sialidase from *M.viridifaciens* (Watson *et al* ., 2004).

In order for the aspartic acid residue to act in this dual role, its pKa must be cycled during the reaction mechanism. As yet, no robust theory has been proposed for how this is achieved. It has previously been argued that this conserved aspartic acid might even be unable to function as an acid catalyst. Its relatively high solvent exposure at the mouth of the active site would suggest it has a pKa that is too low to allow it to function in such a role according to the known pH dependence of the influenza virus neuramindase (Burmeister *et al* ., 1993; Chong *et al* ., 1992b).

The complex of the substrate, α2,3-sialyllactose with the *T.cruzi* trans-sialidase however, shows this solvent exposure is very much reduced upon substrate binding in this enzyme (Amaya *et al* ., 2004). Modelling of α2,3-sialyllactose into the active site of nanI also suggests that solvent exposure would be drastically reduced for this residue. The *T.cruzi* study also points out the relative proximity of the carboxylate group of the sialic acid substrate to this residue (~5 Å), which they argue could substantially raise the pKa of the aspartic acid as a consequence of electrostatic interactions. A similar effect was observed in a ^{13}C-NMR study of a xylanase from *B.circulans* (Joshi *et al* ., 1997; McIntosh *et al* ., 1996). However, no account is taken of the effect the positive potential generated by the arginine triad would have on this proposed electrostatic manipulation of the aspartic acid pKa (Figure 5.2.3). It should be possible to get a more accurate picture of the different protonation states of this residue by trapping the Michaelis complex with either α2,3 or α2,6 sialyllactose and determining the structure at atomic resolution. It should then be relatively straightforward to measure the lengths of the C^{δ}-$O^{\varepsilon 1}$ and C^{δ}-$O^{\varepsilon 2}$ bonds and infer their protonation states. As expected, the lengths of these bonds in the sialic acid complex are characteristic of a deprotonated carboxylate group, with C^{δ}-$O^{\varepsilon 1}$ and C^{δ}-$O^{\varepsilon 2}$ bond lengths of 1.22 and 1.25 Å respectively.

5.5. Comparison of the active site between nanI and the trans-sialidase from *M.decora.*

Given the overall structural similarity between nanI and the leech trans-sialidase, it was of interest to see if the two enzymes were also similar in their reaction mechanism. The leech trans-sialidase has a number of interesting characteristics, the first is the enzyme's strict Neu5Acα2→3Gal linkage specificity, the second an unusual intramolecular trans-sialosyl reaction (Lou *et al* ., 1998a). This second characteristic results in the release of 2,7-anhydro-sialic acid as the final product of glycosidic bond hydrolysis. Complexes of the sialidase with a substrate analogue and with the product resulted in a proposed reaction mechanism (Lou *et al* ., 1998b). This involves the formation of a boat conformation in the sugar ring of the intermediate, which allows the O7 hydroxyl to attack the anomeric carbon of the transition state.

The structure of the leech complexes revealed substantial changes around the glycerol binding pocket compared to previous sialidase structures (Lou *et al* ., 1998a). This results in the substantial movement of a threonine residue into the glycerol binding pocket, compared with the equivalent residue (Thr538) in nanI (Figure 5.5.1). This movement is achieved due to an extra amino acid insertion in this loop in the leech sialidase, which is absent from the nanI structure. These changes were attributed to the resultant axial orientation of the glycerol group and stabilisation of a distorted $B_{2,5}$ boat conformation in both the product and substrate analogue.

A comparison between this active site and the one found in nanI reveals that no such changes exist in the nanI enzyme (Figure 5.5.1). The position of the glycerol group in both of the nanI complexes is equatorial, with well-defined interactions with the glycerol binding pocket. It is unlikely, given the interactions made by both sialic acid and the covalent sialyl-enzyme intermediate in the catalytic site, that nanI could catalyse a similar reaction to the leech enzyme.

The authors of the leech study suggest that the reason behind the unusual chemistry seen in the leech sialidase is a result of the strict linkage specificity

(Lou *et al* ., 1998b). The steric hindrance that restricts the substrate specificity also reduces the solvent accessibility, thus reducing the efficiency of hydrolysis. An intramolecular mechanism through O7 appears to provide an alternative source of nucleophile. No such bulky hydrophobic groups are present in a similar position in the nanI structure, resulting in no obvious steric hinderence or solvent exclusion. It is unlikely therefore, that the nanI sialidase would require such an unusual mechanism.

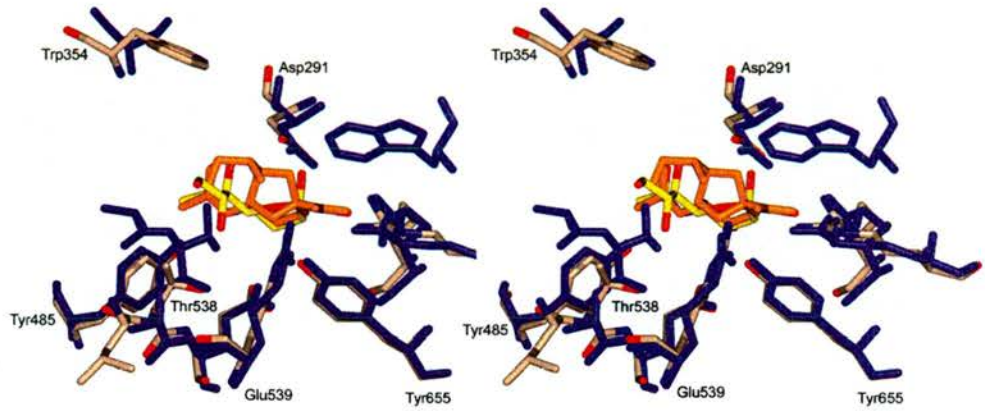


Figure 5.5.1. Stereo view of the superimposed active sites from nanI and the leech trans-sialidase from *M.decora*. Amino acids are numbered for the nanI structure, with the leech residues in blue. The coordinates for the leech trans-sialidase were taken from PDB deposition 2SLI and include the structure of the bound product, 2,7-anhydro sialic acid (orange).

5.6. Discussion.

In general, enzymes achieve their catalytic rate enhancements by lowering transition-state (TS) energies and energetic intermediates and by raising the ground-state (GS) energy (Silverman, 2002). The complexes presented in this Chapter have provided a unique insight into the way the sialidase nanI, from *C.perfringens*, achieves this goal. From these complexes it is now possible to observe the conformation of the reaction product, sialic acid and a covalent sialyl-enzyme intermediate.

Sialidases and trans-sialidases hydrolyse or transfer sialic acids with retention of anomeric configuration (Chong *et al* ., 1992b; Kao *et al* ., 1997). Despite a number of detailed structural studies on the viral, bacterial and eukaryotic sialidases, a number of fundamental questions on their reaction mechanism remain unanswered. While most retaining glycosidases operate via a covalent glycosyl-enzyme intermediate, involving the action of a protein carboxylate nucleophile and acid/base catalyst (Rye and Withers, 2000), sialidases have no suitably positioned carboxylate to function in a similar way. This structural observation, most notably from the influenza virus neuraminidase and subsequently from bacterial sialidases, prompted much of the earlier literature to favour an ionic (rather than covalent) stabilisation of the reaction intermediate (Burmeister *et al* ., 1993; Chong *et al* ., 1992b; Lou *et al* ., 1998b). In addition, these structural studies even questioned the viability of the conserved aspartic acid acting as a general acid/base catalyst (Burmeister *et al* ., 1993; Chong *et al* ., 1992b). However, recent structural studies on the trans-sialidase from the human pathogen *T.cruzi* showed the first structural evidence that sialidase active sites can stabilise a covalent intermediate, with the conserved active site tyrosine acting as the nucleophile (Amaya *et al* ., 2004). Further, the trapping of a Michaelis complex showed the conserved aspartic acid in perfect alignment for proton donation to the glycosidic oxygen, and provided the initial evidence that the pKa of this residue can by modulated through desolvation (Amaya *et al* ., 2004). On the basis of these studies, the authors propose that catalysis by sialidases occurs via a very similar mechanism to that of other

retaining glycosidases, but with interesting differences that have likely evolved in response to the unique chemical properties of sialic acid.

The covalent intermediate trapped in this study would appear to lend support to the idea of covalent catalysis within the bacterial sialidases. However, this intermediate alone only provides evidence that such an intermediate can form, and not that it does with the natural substrate. The electronegative effect of the 3-flouro group on the sialic acid derivative may preferentially favour formation of a covalent bond rather than simply ionic stabilisation from the tyrosine hydroxyl.

The structures of the nanI sialidase in complex with both the reaction product, α-sialic acid and the covalent intermediate, combined with the structure of the apo enzyme (Chapter 4) provide the most detailed picture yet of the reaction mechanism of a bacterial sialidase. The first observation that was made concerning the structure of the nanI sialidase was the rigidity of the active site. Very little difference exists between the apo structure and that of either the covalent intermediate or the sialic acid complex. The original analogy of the lock-and-key hypothesis (Silverman, 2002) of enzyme substrate interactions is a very good description of this system. From the structures presented here, very little if any conformational change occurs in the nanI structure during catalysis. The conformational change observed in the sugar rings of the two complex structures provide a possible answer to how the enzyme achieves stabilisation of the TS and destabilisation of the GS; substrate distortion. In organic chemistry, strain and distortion play an important role in the reactivity of molecules. Most enzymes have at least a portion of the active site that is complementary to the substrate to permit binding on the first collision event. In the active site of nanI this is achieved through the interaction of the carboxylate group at C1 forming strong ionic interactions with the arginine triad and by the interaction of the N-acetyl group at the opposite end of the pocket. These interactions in concert with the active site floor force the ring to adopt a high energy $B_{2,5}$ boat conformation. This would serve to destabilise the substrate and raise the GS energy of the substrate. Close proximity of the conserved acid/base catalyst (Asp291) causes

cleavage of the glycosidic bond through general acid hydrolysis, with subsequent development of a delocalised ionic charge between the anomeric carbon and the enodcyclic oxygen. This oxocarbonium ion is most likely the transition state of the reaction. The nature of this stabilisation is currently being debated in the literature. The structure of the covalent intermediate provides one possible route through the reaction coordinate. The formation of a covalent bond between the anomeric carbon of the sialic acid and the conserved tyrosine (Tyr655) would cause significant stabilisation of the reaction intermediate. The ring is seen to adopt a relaxed $^{2}C_{5}$ chair conformation, and covalent bonds have been known to contribute anywhere from –40 to –110 kcal mol^{-1} to the stability of the intermediate. However, evidence from the nanI complex suggests that this intermediate is not stable. The ordering of five extra water molecules in the covalent complex would be expected to decrease the entropy, and destabilise the energy of the intermediate. Subsequently, the covalent intermediate must breakdown, most likely back to the TS with subsequent formation of product through nucleophilic attack by an activate water molecule.

It has been proposed that strain or distortion of the bound substrate is essential for catalysis (Jencks, 1987). This theory is based on the observation that GS stabilisation of the substrate occurs concomitant with TS stabilisation, and the free energy ($\Delta G^{\ddagger}$) is unchanged from that of the uncatalysed reaction, only displaced downward. To lower the free energy of the catalytic reaction, the enzyme-substrate complex must be destabilised by strain, desolvation or loss of entropy on binding, thereby raising the free energy of both the enzyme-substrate and enzyme-product complexes simultaneously. The observation that little or no conformational movement is observed in the active site of nanI, suggests that much of the conformational change must be occurring in the substrate. This is perfectly illustrated by the structures presented here of nanI (Figure 5.4.1), which shows the conformational changes undertaken by the sugar ring in this study. This is likely to the main mechanism employed by the nanI sialidase in catalysing glycosidic bond cleavage.

An unusual characteristic of the sialidase active site is the choice of tyrosine as the catalytic nucleophile. If we assume that sialidases, like other retaining glycosidases, go through a covalent intermediate, then why was tyrosine chosen instead of the usual carboxylate? Two alternative ideas have been proposed. The first, by the authors of the *T.cruzi* study, believe the tyrosine acts in a charge relay system with the conserved glutamic acid (Amaya *et al* ., 2004). The choice of tyrosine over carboxylic acid was purely an electrostatic one. The tyrosine would minimise electrostatic repulsion from the carboxylate group of the sialic acid, whilst the conserved glutamic acid acts as a base catalyst for the attack of the tyrosine on the anomeric carbon in the transition state. The alternative rationale is one based on transition state stabilisation, where the tyrosine was chosen as a less reactive leaving group to compensate for the inherently higher reactivity of sialidases relative to other glycosidases (Watson *et al* ., 2003).

The rationale for proposing a charge relay system between the conserved glutamic acid and tyrosine residues is plausible, and supported by the observation of a strong hydrogen bond between these two residues in the structures presented here (Figure 5.4.1). However, the argument that the electrostatic repulsion from the substrate carboxylate drove the evolution of tyrosine as the catalytic nucleophile is not straightforward. The effect of the countering positive charge generated by the arginine triad has not been investigated to date. It seems likely, given the proximity of this interaction that any negative charge of the carboxylate would be cancelled out by this interaction. It is currently unclear why this family of glycosidases has chosen a tyrosine-glutamic acid couple to act as the catalytic nucleophile and more mutagenesis studies will need to undertaken to answer this intriguing question.

Chapter 6

Preliminary studies on the reaction mechanism of nanI using NMR

6.1. Summary.

The crystallographic studies on the nanI sialidase highlighted a number of interesting questions concerning the reaction mechanism of this enzyme. These were further investigated using the technique of nuclear magnetic resonance (NMR) spectroscopy.

Subcloning the catalytic domain of nanI raised the possibility that this domain, in isolation from the N-terminal domain, may function differently to the full-length enzyme. It had previously been shown that full-length nanI could hydrolyse Neu5Ac (sialic acid) from the natural substrate analogue α2,3 sialyllactose (Roggentin *et al* ., 1995; Sheu *et al* ., 2002). The hydrolysis of α2,3 sialyllactose by nanI was investigated using 1D ^{1}H NMR and shown to release α-sialic acid as the first product. This is then shown to mutarotate in solution to the thermodynamically more stable β anomer. This result was consistent with previous studies on the bacterial sialidases, which show that the catalytic activity of these enzymes resides solely in the β-propeller domain.

The presence of sialic acid in the active site of crystals grown in the presence of 10 mM Neu5Ac2en (DANA) was discussed in Chapter 5. This surprising observation suggested that the nanI sialidase could hydrate the C2=C3 double bond in this compound, creating α-sialic acid. It was known previously that sialidases could synthesise DANA, a known inhibitor of these enzymes, as a by-product of their catalytic mechanism (Burmeister *et al* ., 1993; Moustafa *et al* ., 2004). Prior to this study, the reverse reaction had not been observed. We sought to investigate this reaction further, again using 1D ^{1}H NMR, to determine whether the initial observation was a crystallographic artefact or not and to understand the mechanism.

Our preliminary results show that DANA is indeed converted to α-sialic acid by nanI in solution at room temperature. In addition to this finding, these

studies have shown that hydration occurs with initial protonation of C3 in the axial position, with subsequent nucleophilic attack by an activated water molecule on C2 to yield α-sialic acid. The source of this initial proton was not identified in this study. However, given the close proximity of the acid/base catalyst Asp291, it is possible that this residue acts to protonate the double bond at C3. An initial mechanism is proposed for the hydration reaction based on the NMR results and the analysis of the nanI ligand complexes discussed in Chapter 5.

6.2. Materials and Methods.

Spectrometer: All the NMR spectra were collected on a 500 Mhz Bruker Avance spectrometer, at 293 K using a quad probe for ^{1}H data acquisition. The samples were transferred to 8 inch Wilmad 5 mm NMR tubes prior to data collection. All data were processed using the software XwinNMR 3.5 on a Windows 2000 workstation. Dr. Thomas Lebl, Chemistry Department, University of St Andrews, carried out all data acquisition and processing.

6.2.1. Hydrolysis of α-2,3 sialyllactose by nanI.

The protein samples used for the NMR experiments were identical to those used in the crystallographic studies. Their preparation is detailed in Chapter 2. All experiments were performed in a final volume of 800 μl, which consisted of 80 μl α-2,3 sialyllactose (200 mM), 50 μl nanI protein (21 mg ml^{-1}), 40 μl of 1 M sodium acetate pH 5.5 and 630 μl D_2O. The sodium acetate was used as the reference peak in the integration calculations. The first spectra were taken 3 minutes after mixing. The delay time between subsequent acquisitions was 400 seconds, with each experiment consisting of 32 scans, taking 3.11 minutes. The control experiment was set up in exactly the same way as above, but without the nanI enzyme.

6.2.2. Hydration of Neu5Ac2en by nanI.

All experiments were performed in a final volume of 800 μl, which consisted of 142 μl of nanI protein (21 mg ml^{-1}), 40 μl Neu5Ac2en (DANA) (200 mM) dissolved in D_2O (sigma), 40 μl of 1 M sodium acetate pH 5.5 and 578 μl D_2O. The sodium acetate was used as the reference peak in the integration calculations. The first spectra were taken 20 minutes after mixing. The delay time between subsequent acquisitions was 30 minutes, with each experiment consisting of 32 scans, taking 3.11 minutes. The control experiment was set up in exactly the same way as above, but without the nanI enzyme.

6.2.3. Investigating the mechanism of hydration catalysed by nanI.

To investigate the mechanism of hydration, a number of experiments were performed initially in ~100 % D_2O with addition of H_2O after 24 hours to monitor the appearance of proton peaks in the reaction product. In these experiments, the nanI protein was buffer exchanged on a spin concentrator (Vivascience) into D_2O. No buffer was used in these experiments. The reaction consisted of 50 μl of nanI protein (21 mg ml^{-1}), 40 μl DANA (200 mM) dissolved in D_2O and 670 μl D_2O. The first spectra were taken 3 minutes after mixing. The delay time between subsequent acquisitions was 400 seconds, with each experiment consisting of 32 scans, taking 3.11 minutes.

6.3. Results.

6.3.1. Hydrolysis of α-2,3 sialyllactose by nanI.

The hydrolysis of α-2,3 sialyllactose (**1**) by nanI was investigated using 1D ^{1}H NMR to determine whether the subcloned catalytic domain remained active independent. Previous studies had shown that the full-length enzyme was capable of hydrolysing the glycosidic bond between sialic acid and lactose in the natural substrate analogue, α-2,3 sialyllactose (Roggentin *et al* ., 1995; Sheu *et al* ., 2002). In effect, this investigation was acting as a control for the structural studies described in Chapters 4 and 5. These results show that the subcloned domain retains its catalytic activity in the absence of the N-terminus, and validating our analysis of the catalytic mechanism from the crystallographic study. A full kinetic characterisation, however, should be carried out and kinetic parameters determined. This would allow the catalytic activity of this domain to be compared quantitatively to the full length enzyme.

The 1D ^{1}H NMR technique allowed the identification of the substrate and product through the differences in the resonances of the axial and equatorial protons on C3 between the α and β anomers of sialic acid (Table 6.1). The stacked spectra in Figure 6.1. represent snapshots of the enzymatic hydrolysis reaction. Scheme 1 shows a chemical representation of the substrate α-2,3-sialyllactose, and products α and β-sialic acid, observed during the reaction.

As the reaction proceeds, the resonances of the H_{3ax} and H_{3eq} protons on C3 of the α-sialic acid (**2**) (α-sialic acid: $\delta[H_{3ax}] = 1.53$, $\delta[H_{3eq}] = 2.63$) decrease in intensity with a concomitant increase in intensity for the resonances of the axial and equatorial protons on C3 of β-sialic acid (**3**) (β-sialic acid: $\delta[H_{3ax}] = 1.71$, $\delta[H_{3eq}] = 2.14$). These results are consistent with the product retaining the anomeric configuration of the substrate, followed by mutarotation in solution to the β form, as shown in previous studies (Friebolin *et al* ., 1980; Moustafa, 2004). The shift in resonances of the N-acetyl group protons from $\delta = 1.93$ in the α-sialic acid to $\delta = 1.96$ in β-sialic acid can also

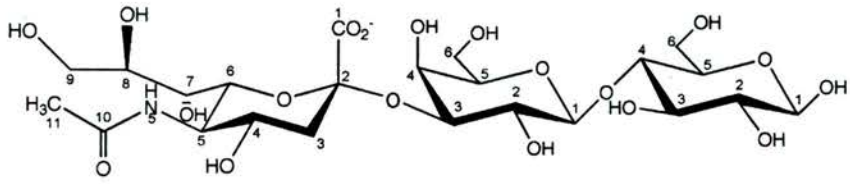
be seen in the spectra. Figure 6.1. shows that by 193 minutes almost all of the substrate has been converted to product. This spectra shows that after 3 minutes all of the α2,3 sialyllactose has been hydrolysed to give free α-sialic acid, which then is seen to mutarotate in solution to the more stable β-sialic acid conformer. It should be possible to calculate the rate that this occurs from these data. A control sample of α-2,3 sialyllactose in the absence of nanI showed no breakdown of the glycosidic bond even after 24 hours at room temperature.

Table 6.1. Chemical shifts of the characteristic signals from the C3 protons of α/β-sialic acid.

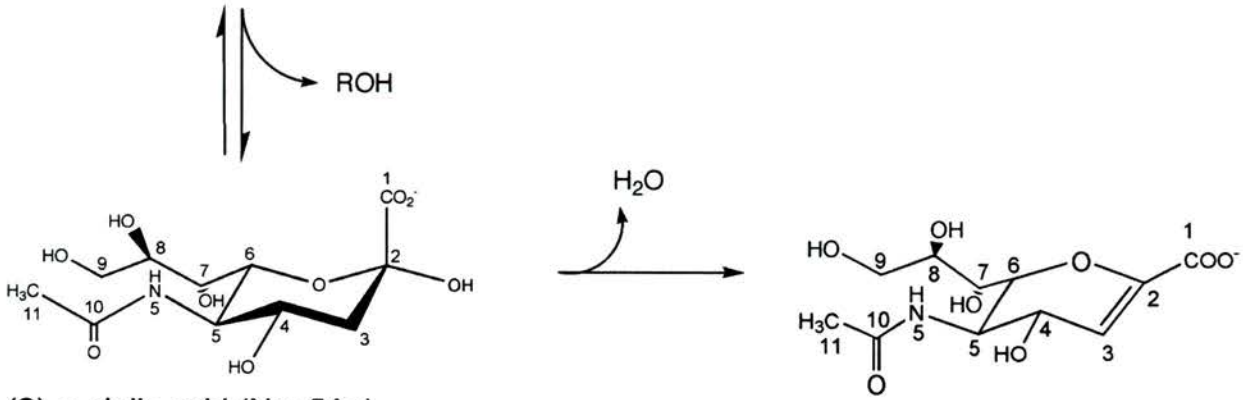
	$H_{3\ axial}$	$H_{3\ equatorial}$	$NCOCH_3$
α-Sialic Acid	1.53	2.63	1.93
β-Sialic Acid	1.71	2.14	1.96
α2,3-Sialyllactose	1.68	2.66	1.97

The study on the influenza virus sialidase (NA) reported by Burmeister and colleagues (Burmeister *et al* ., 1993) showed that this sialidase could synthesise the transition-state analogue Neu5Ac2en as a side reaction of siayl-lactose hydrolysis. They proposed a reaction mechanism whereby the oxocarbonium ion transition state could eliminate the C3 proton and form Neu5Ac2en. They also showed that this sialidase continuously turned over Neu5Ac with the solvent, resulting in continual exchange of the O2 hydroxyl at the anomeric carbon. This results in Neu5Ac being a millimolar inhibitor of this enzyme. The mechanism of proton elimination from C3 of the oxocarbonium ion to form Neu5Ac2en is currently unclear and is not discussed in the Burmeister study. The authors did note that they observed no hydration of Neu5Ac2en with the viral enzyme and concluded that this side reaction was essentially irreversible with the equilibrium lying far over in favour of the Neu5Ac2en formation. The formation of Neu5Ac2en by the bacterial sialidase from *V.cholerae* has also been observed, forming in a similar way the viral enzyme (Moustafa *et al* ., 2004). The observation that

the nanI sialidase can catalyse the formation of Neu5Ac from Neu5Ac2en is the first to show that this reaction can occur in the active site of these enzymes. It would be interesting to investigate if this reaction could be observed in other bacterial sialidases, which perhaps take much longer to convert Neu5Ac2en to Neu5Ac. The *C.perfringens* nanI enzyme may just be unusual in that it performs this reaction on a quicker time scale than the currently studies homologues.

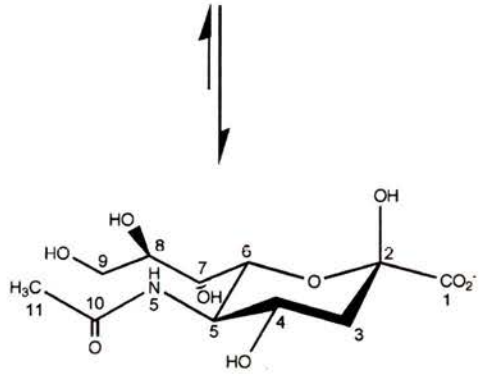


(**1**) α-2,3-sialyllactose



(**2**) α-sialic acid (Neu5Ac)

(**4**) DANA (Neu5Ac2en)



(**3**) β-sialic acid (Neu5Ac)

Scheme 1. The enzyme catalysed hydrolysis of the glycosidic bond between sialic acid and galactose in α-2,3-sialyllactose (**1**) by the action of the nanI sialidase. The first product of hydrolysis is α-sialic acid (**2**), which mutarotates in solution to the more stable β-sialic acid (**3**). Proton elimination at C3 leads to DANA (Neu5Ac2en) (**4**). The overall reaction in the synthesis of DANA leads to water elimination from sialic acid.

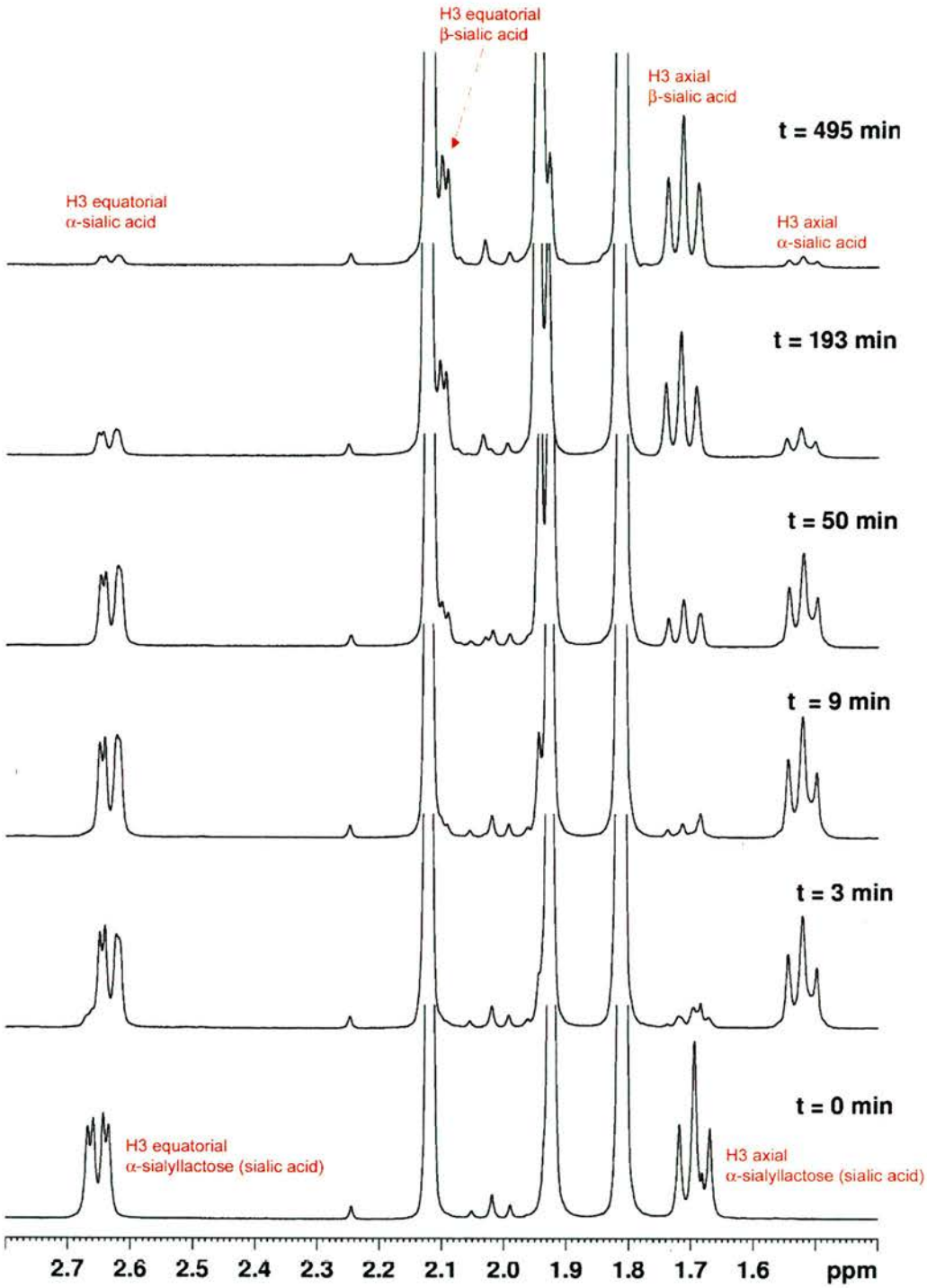
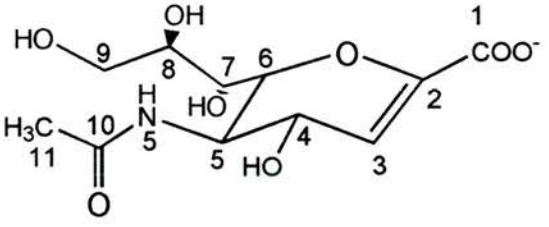


Figure 6.1. Stacked spectra of showing the initial hydrolysis of α2,3-Sialyllactose (10 mM) in the presence of nanI (24 μM), followed by the mutarotation of the release α-sialic acid to the more stable β-sialic acid conformer. The region corresponding to the C3 proton peaks (δ = 1.4 – 2.8 ppm) is shown. The characteristic resonances for α- and β-sialic acid are labelled; the time each spectrum was taken is indicated.

6.3.2. Hydration of Neu5Ac2en by nanI to produce α-Neu5Ac.

The possibility that the nanI sialidase could hydrate the C2=C3 double bond of the Neu5Ac2en (DANA) was investigated using 1D ^{1}H NMR. The proton on the double bonded C3 of DANA (**4**) has a characteristic chemical shift (δ = 5.59 ppm) far removed from those generated in either anomer of sialic acid (Figure 6.1). Using the peak from the acetate buffer (δ = 1.90) as an internal reference, we chose to monitor the turnover of DANA by nanI. The area under the C3 proton peak from DANA was measured at 30 minute intervals during the reaction and plotted against time. These data were compared to a control experiment without the enzyme. Figure 6.2. clearly shows the decrease in amount of DANA over time in the sample containing the enzyme compared without. We believe this experimental evidence indicates that the nanI sialiadase is capable of hydrating the compound DANA to sialic acid. The next most obvious question is by what mechanism?



(4) Neu5Ac2en (DANA).

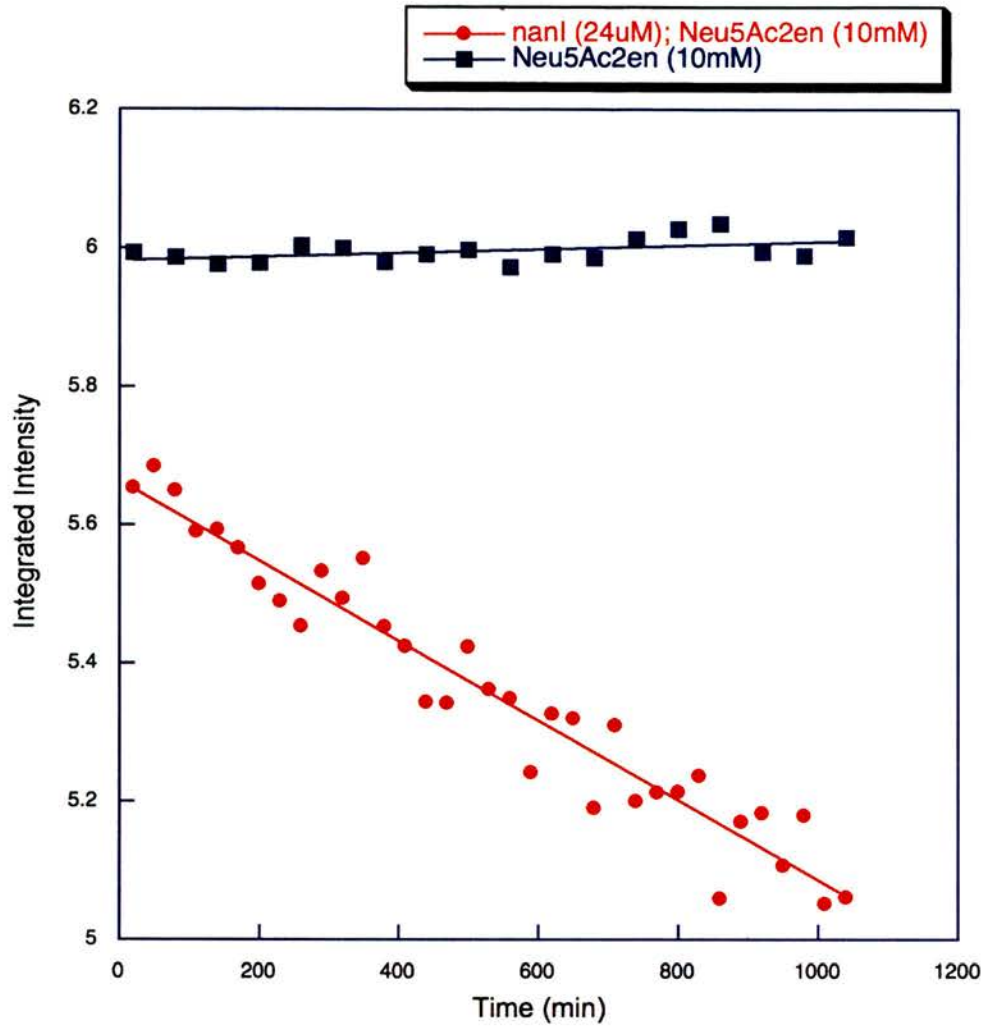


Figure 6.2. Scatter plot of Integrated Intensity of the proton on the double bonded C3 of Neu5Ac2en (DANA) (**4**) compared to a sample of DANA without enzyme.

6.3.3. Preliminary investigation into the mechanism of Neu5Ac2en (DANA) hydration in nanl.

The observation that nanl can catalyse the hydration of the C2=C3 double bond in the compound DANA is the first to our knowledge in the sialidase superfamily. In order to understand the mechanism of hydration in more detail, we chose to undertake a further 1D ^{1}H NMR experiment. The idea was to see if we could identify which face of the pyranose ring of DANA was being attacked in the initial protonation event. The experiment was conducted in 100 % D_2O so the hydration reaction would occur with addition of a deuterium to C3 and a deuterated hydroxyl to C2 to generate sialic acid. It should be possible to measure the chemical shift of the original C3 proton of DANA in the sialic acid product and determine whether it sits in the equatorial or axial position. This would allow us to infer the direction of attack, either from the solvent face or the buried face of DANA. The large volume and solvent exposure of the active site in the nanl structure (Chapter 4) creates the possibility that protons could exchange for deuterium's. We could therefore make no inferences on the identity of the initial proton.

A stack plot of the measured spectra is shown in Figure 6.3. The chemical shifts of only β-sialic acid are visible, although as shown from section 6.3.1. and from the appearance of α-sialic acid in the nanl complex (Chapter 5) the initial product must be α-sialic acid. The first three spectra were taken in 100 % D_2O and show no increase in the axial protons ($\delta[H_{3ax}]$ = 1.71). A small increase in the equatorial protons ($\delta[H_{3eq}]$ = 2.14), however, is observed. Upon addition of 200 μl of H_2O at t = 1440 minutes, the peaks of both axial and equatorial chemical shifts increase, suggesting the addition of a proton to the axial position of C3.

From the spectra obtained, it was possible to calculate a table of integrals for the alkene proton of DANA at C3, the five peaks corresponding to the equatorial protons and approximate triplet axial protons in sialic acid; Table 6.2. Using these data it is possible to calculate the fraction of DANA present in each experiment, f(DANA), and from this the fraction of sialic acid,

f(Sialic acid). This done as follows: f(DANA) = DANA/(DANA + equat + axial), with f(Sialic acid) = 1 – f(DANA).

Analysis of these data indicate that the initial protonation of the alkene double bond of DANA (**4**), occurs from the face of the compound open to the bulk solvent, the face closest to Asp291 in the active site of nanI. The original proton of DANA on C3 is observed in the equatorial position, with the newly acquired proton sitting in the axial orientation. Using this information and the knowledge of the active site gained through the structure of nanI ligand complexes, we can propose two initial mechanisms through which the hydration reaction could be occuring.

The first mechanism, proposed in Figure 6.5a shows the reaction proceeding through a covalent intermediate attached to the conserved tyrosine 655. This mechanism is based upon the reaction proposed by the Alzari group on their TcTS study with the covalent intermediate (Amaya *et al* ., 2004)

The alternative mechanism, proposed in Figure 6.5b, involves going through a single oxocarbonium ion transition state in a similar manner to that proposed for glycosidic bond hydrolysis in the Burmeister paper (Burmeister *et al* ., 1993).

At present it is not possible to distinguish between these two mechanisms. More kinetic and biochemical work is required to investigate this interesting mechanism. The first approach would be to mutate out the conserved tyrosine and see if the reaction can still proceed.

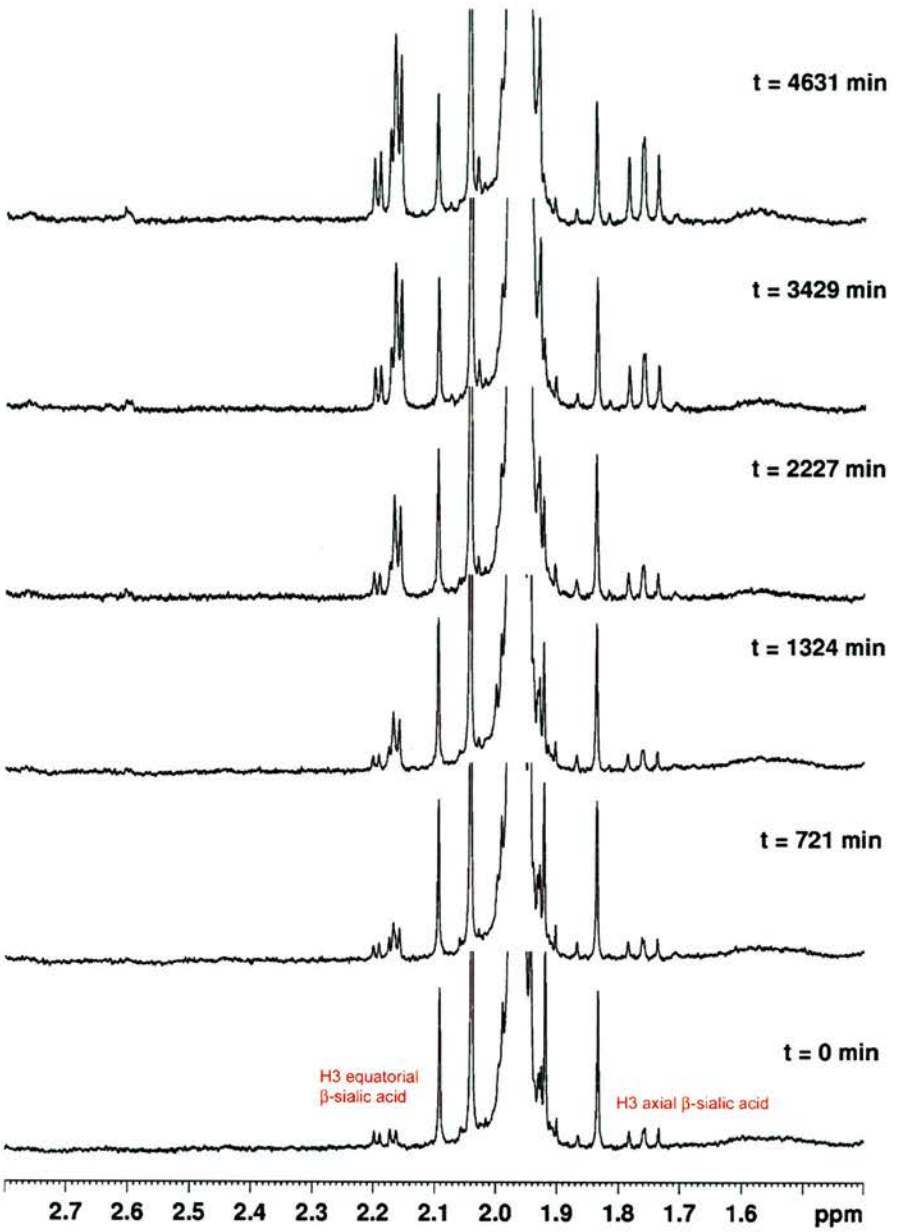


Figure 6.3. Stacked spectra of showing the hydration of Neu5Ac2en (DANA) (10 mM) in the presence of nanI (68 μM) showing the region corresponding to the C3 proton peaks (δ = 1.4 – 2.8 ppm). The first three spectra (t = 0, 721 and 1324) were taken in 100 % D_2O, the subsequent spectra were taken after the addition of 200 μl H_2O to the reaction.

Table 6.2. Table of Integrals for hydration of Neu5Ac2en (DANA) by nanI.

Time (min)	DANA	$H_{3equatorial}$	H_{3axial}	Total	Calc (axial)	Fract (axial)	f(DANA)	f(Sialic acid)
0	38.15	1.07	1.28	40.51	1.28	0.032	0.94	0.06
721	37.51	1.31	1.34	40.15	1.27	0.032	0.93	0.07
1324	39.10	1.87	1.38	42.34	1.34	0.032	0.92	0.08
2227	56.63	2.27	0.72	59.62	1.89	0.032	0.95	0.05
3429	55.64	3.51	1.44	60.59	1.91	0.032	0.92	0.08
4631	65.30	6.06	2.41	73.77	2.34	0.032	0.88	0.11

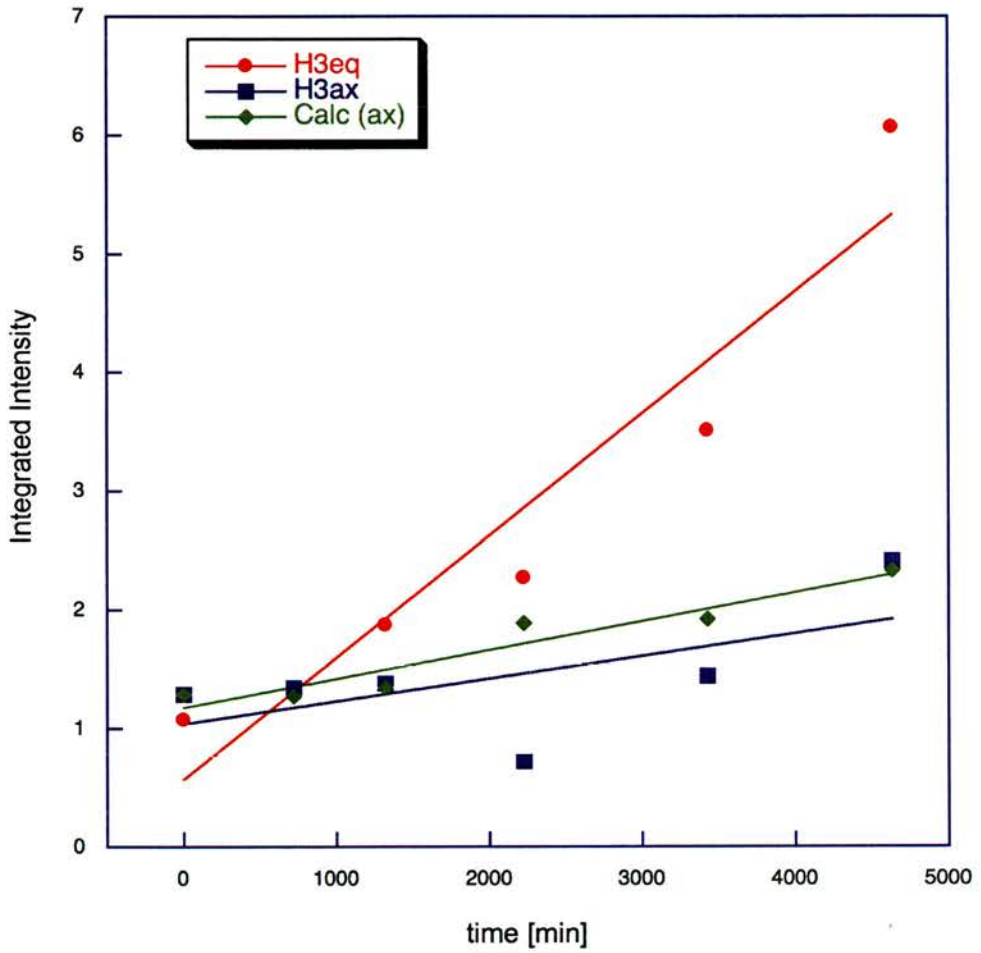


Figure 6.4. Scatter plot showing correlation between the calculated integration of the axial C3 proton (Calc ax) and the measured integration (H_3ax). Notice how only the area under the H_3 equatorial peaks (H_3eq) increases during the first three data points shown. These data indicate that only the H_3 equatorial protons were being measured, and that deuterium appears in the axial position and is not observed in the ^{1}H NMR.

6.4. Proposed mechanism for the conversion of Neu5Ac2en to α-Neu5Ac by nanI.

The structure of free α-sialic acid in the active site of nanI shows a number of interesting features that may help to explain why this enzyme can catalyse the hydration of the alkene double bond in DANA. DANA is a stable compound, due to the presence of the alkene double bond and subsequent planar arrangement around C2. Any effort to destabilise this double bond would require disrupting this planar configuration. The most obvious was to achieve this would be through substrate distortion and ring strain.

One of the most notable differences between the sialic acid complex and the covalent intermediate is the absence of water in the former. The glycerol group of sialic acid makes a number of hydrogen bonds to the residues in the active site that are absent in the covalent intermediate. Most notable among these are the interactions of the O9 hydroxyl to the amino group of Gln493 and O8 to the hydroxyl of Tyr485 (Figure 5.2.3). The interaction to Gln493 is absent in the covalent structure, being replaced by a new hydrogen bond to a water molecule (Figure 5.3.2). This glutamine residue is absent in all other structures of bacterial sialidases, and is not replaced any alternative hydrogen bonding residue

The presence of Trp354 at the mouth of the active site and sitting over the glycerol binding pocket, could contribute to strengthening the hydrogen bonds made to Gln493 and Tyr485 through desolvation. Stronger interactions of the protein with the glycerol group may contribute to ring strain in the DANA compound. Once destabilised, the electrons in the nucleophilic π bond could attack the suitably positioned proton on Asp291 in an electrophilic addition reaction (A; Figure 6.5a). In the complex of sialic acid with nanI, the distance from the anomeric carbon to carboxylate of Asp291 is 3.12 Å.

This would result in a carbocation (oxocarbonium ion) intermediate, with the positive charge delocalised between the anomeric carbon and the endocyclic oxygen (B; Figure 6.5a). The structure of the covalent complex with nanI shows that such an intermediate could be stabilised through a covalent bond with the nucleophilic Tyr655 sitting directly beneath (C; Figure

6.5a). Asp291 could then activate an incoming water molecule to attack the positive charge on the anomeric carbon, creating a protonated alcohol intermediate (D; Figure 6.5a). Loss of a H^+ from this protonated alcohol back to Asp291 gives the product, sialic acid (E; Figure 6.5a).

It is unclear whether the reaction proceeds through a covalent intermediate, as is shown in the proposed mechanism in Figure 6.5a. The alternative mechanism, proposed in Figure 6.5b is conceptually simpler, involving a single oxocarbonium ion intermediate. In this mechanism, the electrophilic addition reaction (A; Figure 6.5b) causes the breaking of the C2=C3 double bond with the addition of the proton on Asp291 forming a new bond to the C3 of DANA. This reaction would generate the oxocarbonium ion (carbocation) transition state, which would be stabilised through resonance with the endocyclic oxygen (B; Figure 6.5b) and through favourable electrostatic interactions with the conserved tyrosine Tyr655. Subsequent nucleophilic attack by an incoming, activated water molecule would create a protonated alcohol intermediate (C; Figure 6.5b). Again, loss of a H^+ from this protonated alcohol back to Asp291 gives the product, sialic acid (E; Figure 6.5b).

6.5. Discussion.

The preliminary investigation of nanI by 1D ^{1}H NMR have shown conclusive evidence that this sialidase can convert the compound Neu5Ac2en (DANA) to α-Neu5Ac (sialic acid). There have been no previous reports of such a reaction in the sialidase superfamily, although a recent study of the sialidase from *V.cholerae* suggests that it to might perform a similar reaction (Moustafa, 2004). It was known that sialidases could synthesise DANA as a by-product during catalysis of natural substrates, having been observed a number of times in structural studies (Burmeister *et al* ., 1993; Buschiazzo *et al* ., 2002). This observation was puzzling, given that DANA acts as a weak inhibitor of most sialidases, and served as the structural template for inhibitor design for the influenza neuraminidase (Colman, 2002). It was generally accepted that this side reaction was unimportant and had no real effect on the enzymes in their natural environment. The results presented here however, could be seen to disagree with this view. Clearly, the active site of the nanI sialidase has evolved to hydrate the alkene double bond of DANA efficiently and immediately upon forming an enzyme-ligand complex, suggesting an evolutionary pressure to do so. Although in principle we cannot rule out the possibility that this reaction is merely coincidental, it seems likely that it has a specific role to play in the biochemistry of the nanI sialidase.

The two most prominent differences between the active site of nanI and those of other bacterial sialidases are Trp354 and Gln493. Both of these residues are in close proximity to the glycerol group of sialic acid and make either indirect or direct interactions with this group respectively. This is the area most likely to yield clues for discovering the nature of the initial destabilising event in the hydration reaction. A number of site-directed mutants at these sites could allow the trapping of DANA in the active site. Combined with the atomic resolution data possible in this system, it should be possible to identify the obviously subtle changes to the interactions with this compound that initiate this reaction.

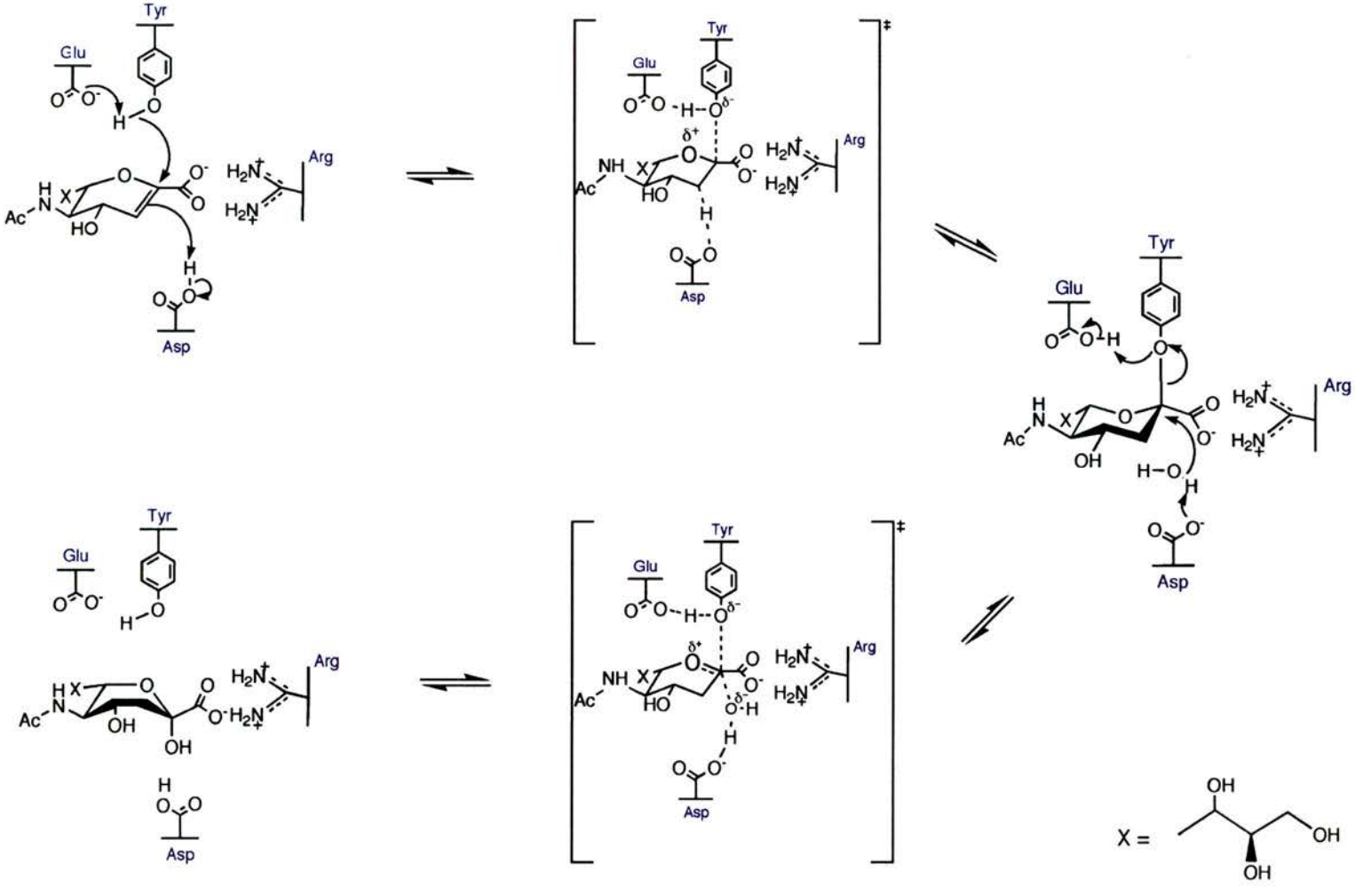


Figure 6.5a. First proposed mechanism of hydration of the C2=C3 double bond of Neu5Ac2en (DANA) catalysed by the nanI sialidase.

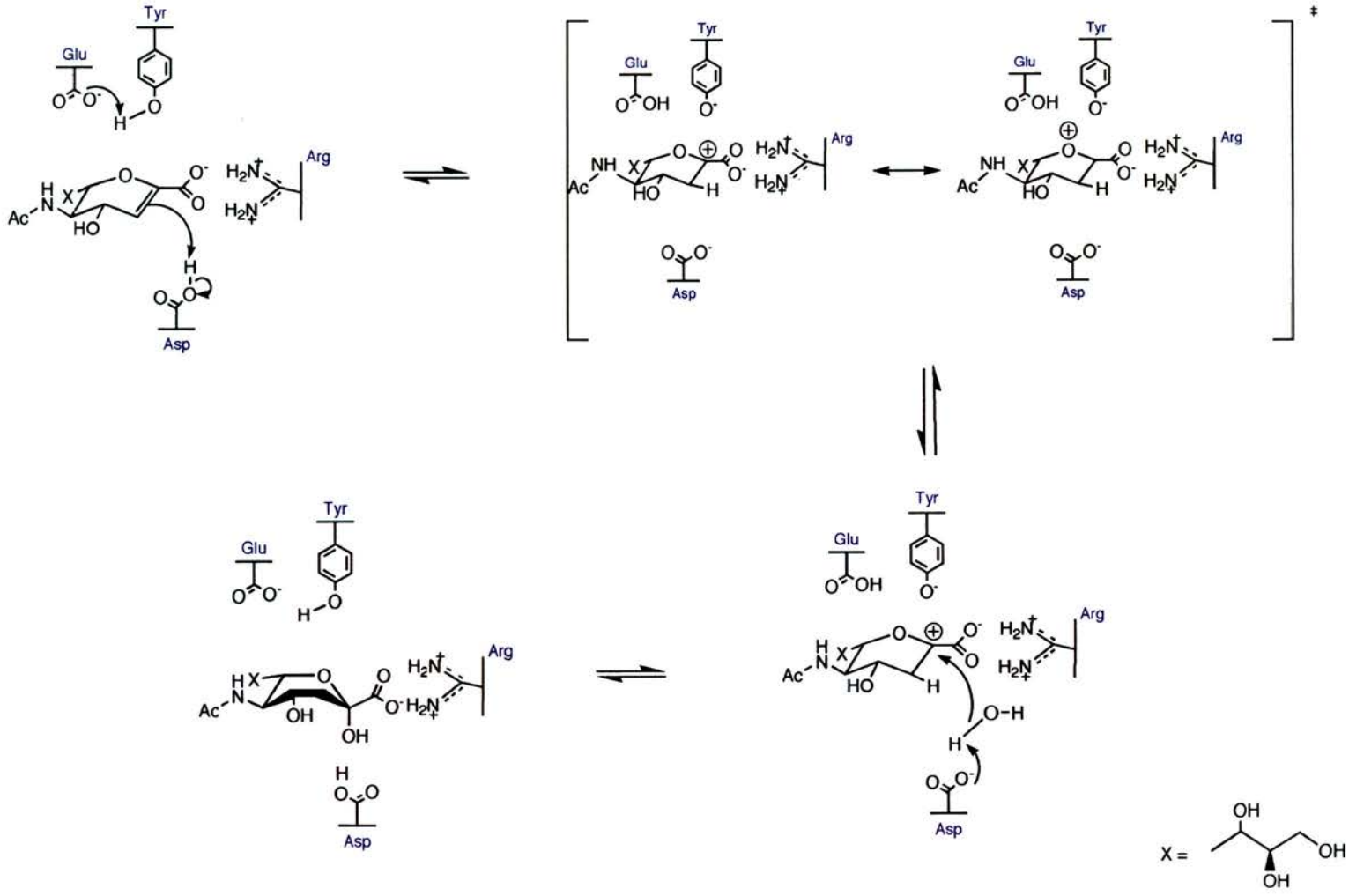


Figure 6.5b. An alternative mechanism of hydration of the C2=C3 double bond of Neu5Ac2en (DANA) catalysed by the nanI sialidase.

Chapter 7

Structural Studies on Active Site Mutants of the 68 kDa sialidase from *Micromonospora viridifaciens.*

Crystallisation, Data collection, Processing, Molecular Replacement and Refinement of the structures.

7.1. Summary.

To understand in greater detail the role played by the conserved amino acids in the active site of sialidases, a structural and biochemical approach was taken. In collaboration with Professor A. Bennet and Dr. J. Watson at the Department of Chemistry, Simon Fraser University, Canada, a range of site-directed mutants were created in the 68 kDa form of the sialidase from the soil bacterium *Micromonospora viridifaciens* (MvNA). Our aim in this study was to provide structural data to compliment the biochemical characterisation. Previous structural studies on this protein had shown that it consists of three domains: a β-propeller, an immunoglobulin module and a galactose-binding jelly roll (Gaskell *et al* ., 1995). A smaller, 41 kDa form of this protein can also be expressed by the bacterium when colominic acid is used as an inducer. This form corresponds to the β-propeller domain only and was not used in this study.

The following point mutants were crystallised and their structures determined by molecular replacement with the wild type structure: Asp92Gly (D92G), Trp370Gly (Y370G) and Trp370Phe (Y370F). Complete x-ray datasets were collected for each mutant, both in house and at the European Synchrotron Radiation Facility (ESRF) on beamlines ID14eh2 and ID14eh3.

Complexes of the mutants with different ligands were studied depending on the nature of the biochemical studies being performed. Structural data on complexes with the proposed transition-state analogue Neu5Ac2en (DANA) were obtained with the D92G and Y370F mutants. The structure of the Y370G mutant was obtained in complex with the β form of the substrate Neu5Ac (sialic acid). The

structure of the covalently modified intermediate was also attempted with the wt structure, using the α2,3-difluoro sialic acid derivative, in a similar manner to that described for the *C.perfringens* nanI sialidase in Chapter 5. However, after refinement of the data the difference density maps (F_o-F_c) showed the presence of the fluorine atom at C3, but not for the covalent bond to the tyrosine (Tyr370). The current data on this complex is unable to resolve this ambiguity, with the most likely explanation being that the α2,3-difluoro sialic acid broke down prior to the soaking experiment. It was decided to model DANA into the difference density and await further follow up studies to resolve this unusual result.

7.2. Crystallisation.

Dr. J. Watson sent the purified 68 kDa MvNA protein flash frozen on dry-ice. The cloning, expression and purification of the sialidase mutants is described elsewhere (Watson *et al* ., 2003). The proteins were expressed and purified without the aid of an affinity tag. All of the mutants were prepared for crystallisation in the same way. The protein was initially thawed upon arrival on ice and the protein concentration determined using the Bicinchroninic Acid (BCA) protein assay (Smith, 1985) (Pierce). The purity of the protein samples were also analysed using SDS-PAGE (Figure 3.2.1). The extra bands seen in the gel correspond to breakdown of the sialidase protein during the heat denaturation step prior to loading the samples into the wells of the gel.

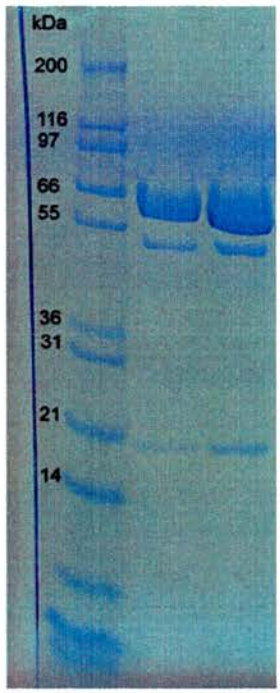


Figure 7.2.1. SDS-PAGE analysis of the Y370G point mutant of the 68 kDa sialidase from *Micromonospora viridifaciens*. L-R, 10 μl of Mark 12 protein standards (Invitrogen), 20 μl fractions of the Y370G protein. The bands at ~50 and ~20 kDa represent breakdown products of the sialidase due to denaturation prior to sample loading.

The protein, once thawed, was dialysed overnight at 4 °C against 10 mM Tris, pH 8.4. using a 50 kDa MWCO dialysis cassette (spectrum). The dialysed protein was again assayed using the BCA assay to determine protein concentration and check that no loss had occurred using the dialysis membrane. The protein solution was then concentrated to 20 mg/ml for crystallisation using a spin concentrator with a 10 kDa MWCO membrane (viva science). All crystallisation experiments were performed using the sitting-drop vapour diffusion method in Douglas Instruments 96-well crystallisation plates (Douglas Instruments) at 21 °C. The crystallisation drops were made up of equal volumes (2 μl) of mother liquor and protein solution.

The protein used in this study was produced using modern recombinant method techniques, in contrast to the original study that used protein purified from the cultured *M.viridifaciens* bacterium (Taylor *et al* ., 1992). It was therefore not surprising that the original crystallisation conditions used in the previous study failed to produce any crystals. New crystallisation conditions for the recombinant MvNA were therefore sought. A range of crystallisation conditions was tested using the sparse matrix and grid screens from Hampton Research, Decode Genetics, SIGMA and Molecular Dimensions.

Initial crystals were obtained in two conditions in the PEG/Ion screen from Hampton Research, condition 1 (20 % PEG 3350, 0.2 M sodium fluoride) and 48 (20 % PEG 3350, 0.2 M di-ammonium hydrogen citrate). Interestingly, only the Y370G mutant crystallised in condition 1. All other mutants crystallised in condition 48. The morphology of the crystals was also different in these two conditions. The Y370G crystals grown in condition 1 formed long rod like crystals belonging to space group $P2_1$, whereas the wt, D92G and Y370F crystals grew as either short rods, square or pyramid shaped crystals. The short rods were typical of the D92G crystals, which belonged to space group $P2_12_12_1$. The square and pyramid shaped crystals, typical of the ones formed by the Y370F mutant sialidase, belonged to space group $P3_221$.

The original crystallisation conditions tended to give a large amount of very small crystals, indicating a great number of nucleation events within the

drop. To slow down the nucleation and attempt to grow large crystals a number of methods can be used. In this instance, lowering the concentration of precipitating agent, PEG 3350, proved successful in achieving this aim (Figure 3.2.2). A range of PEG 3350 concentrations were set up to slow down the rate of nucleation, these were 10, 12, 14, 16, 18, 20 % PEG 3350, plus 0.2 M of either sodium fluoride or di-ammonium hydrogen citrate. In each case, good quality crystals suitable for x-ray diffraction experiments were grown over the course of two to six weeks in the lower concentrations of PEG, typically 10 to 12 % proved optimal.

(a) (b)

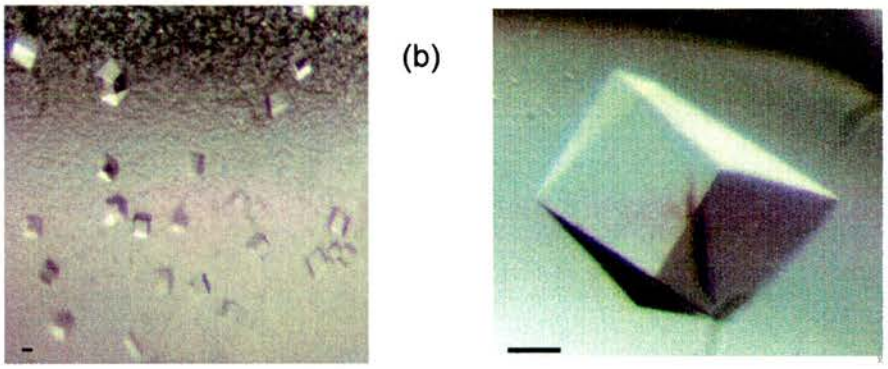


Figure 7.2.2. Optimisation of MvNA crystals for x-ray diffraction experiments. (a) Initial crystals of mutant Y370F grown in 20 % PEG 3550, o.2 M di-Ammonium hydrogen Citrate after 5 days. (b) The same protein solution crystallised with only 10 % PEG 3350 after 3 weeks. Both crystals were grown using the sitting drop technique at 21 °C. The black bar represents 25 μm in each picture.

7.3. Data collection and processing.

7.3.1. Data collection.

The crystals were cryoprotected in 20 % glycerol in mother liquor and flash-frozen in a nitrogen stream at 100 K before data collection. The data collection statistics are given in Table 7.3.1. Diffraction data were collected in house on a Rigaku-MSC MicroMax-007 rotating anode generator and an R-AXIS HTC image plate area detector. At the ESRF, data were collected on beamlines ID14eh2 and eh3 using ADSC Q4R CCD and MARCCD3x3 detector systems, respectively.

Details regarding each x-ray diffraction experiment are given below:

1. **Asp92Gly (D92G) – Neu5Ac2en (DANA) complex.**

   Crystals of the D92G mutant were soaked in 10 mM Neu5Ac2en for 1 hour at room temperature prior to data collection. Data were collected on the in house x-ray generator to a resolution of 2.1 Å. The crystals belonged to space group $P2_12_12_1$. This was a new space group for the 68 kDa form of the protein. 120° of data were collected at 0.5° oscillations with 15 minute exposures.

2. **Asp92Gly (D92G) – Apo structure (lactose).**

   Diffraction data on the apo structure of the D92G mutant was also collected, with subsequent refinement revealing the disaccharide lactose bound in the C-terminal galactose-binding domain. Data for this structure were collected on ID14eh2 to a resolution of 1.7 Å. The crystals were found to index again in space group $P2_12_12_1$ and 120 ° of data were collected at 0.5° oscillations using a 3s exposure time.

3. **Tyr370Phe (Y370F) – Neu5Ac2en (DANA) complex.**

The Y370F mutant crystals were soaked in 10 mM Neu5Ac2en for 30 minutes at room temperature. The crystals diffracted to a resolution of 2.1 Å. Initial images from the crystal indexed in a new trigonal space group, $P3_221$. Data were collected in 0.5° oscillations with 5s exposures. 120° of data were collected.

4. **Tyr370Gly (Y370G) - β-Neu5Ac (Sialic Acid).**

The Y370G mutant crystals were soaked in α-DHP-Neu5Ac for 30 minutes at room temperature. This compound spontaneously broke down in solution to give β-Neu5Ac, which subsequently bound the active site pocket. These crystals grew in a different condition to the others, with 0.2 M sodium fluoride replacing the di-ammonium hydrogen citrate. Diffraction data were collected to a resolution of 1.8 Å. The initial images were incorrectly indexed at the beamline and only 120° degrees of data were collected with 0.5° oscillations and an exposure time of 3s. A more careful examination of the data revealed the space group to be monoclinic, $P2_1$. However, a fortuitous orientation of the crystal on the goniometer produced an overall completeness of 87 %, which we deemed acceptable for refinement.

5. **Wild Type (wt) – α2,3-difluoro sialic acid derivative complex.**

In an attempt to trap a covalently bound sialyl-enzyme intermediate in this sialidase, the α2,3-di-fluoro sialic acid derivative was used. The crystals of the wt enzyme were soaked in the compound for 1 hour at room temperature. Refinement of the data revealed the presence of a mixed population of complexes within the crystal. The majority of the MvNA proteins appeared to have DANA bound, but a small percentage also appeared to have the covalent intermediate. However, the data was insufficient to model and refine the covalent intermediate and the final model was refined with DANA bound. This structure is not discussed in

chapter 8 as work is still ongoing to repeat this experiment and trap the intermediate. The initial images from the diffraction data were indexed in space group $P2_12_12_1$. 171° of data were collected in 0.5° oscillations with 5s exposures. The diffraction data extended to 2.0 Å resolution.

Table 7.3.1. Data collection statistics for the five MvNA complexes.

Complex	D92G – Neu5Ac2en	D92G – lactose	Y370F – Neu5Ac2en	Y370G – βNeu5Ac	Wt – α2,3 di-fluoro Neu5Ac
X-ray source	In house	ID14-2 ESRF	ID14-2 ESRF	ID14-3 ESRF	ID14-2 ESRF
Wavelength (Å)	1.542	0.933	0.933	0.931	0.933
Space group	$P2_12_12_1$	$P2_12_12_1$	$P3_221$	$P2_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 46.6	a = 46.6	a = 143.25	a = 55.44	a = 46.82
	b = 111.61	b = 110.35	b = 143.25	b = 49.54	b = 111.33
	c = 143.86	c = 141.88	c = 160.25	c = 107.04	c = 144.23
Unit-cell angles (°)	α = 90.00	α = 90.00	α = 90.00	α = 90.00	α = 90.00
	β = 90.00	β = 90.00	β = 90.00	β = 101.74	β = 90.00
	γ = 90.00	γ = 90.00	γ = 120.0	γ = 90.00	γ = 90.00
Resolution range (Å)	87 – 2.1	87 – 1.7	124 – 2.1	105 – 1.8	88 – 2.0
	(2.2 – 2.1)	(1.79 – 1.7)	(2.2 – 2.1)	(1.9 – 1.8)	(2.1 – 2.0)
Mosaicity (°)	0.45	0.40	0.55	0.78	0.52
Observed reflections	175 360	288 015	799 157	117 043	415 912
Unique reflections	44 772	82 427	110 930	46 213	51 920
Completeness (%)	89	92	100	87.3	100
	(88)	(72)	(100)	(87.3)	(100)
Redundancy	4.3	4.0	7.2	2.7	6.8
	(3.2)	(3.8)	(7.0)	(2.3)	(6.9)
R_{merge}†	0.068	0.070	0.072	0.090	0.079
	(0.30)	(0.32)	(0.474)	(0.215)	(0.295)
I/σI	6.9	4.7	7.4	5.0	7.5
	(2.3)	(2.1)	(1.5)	(3.0)	(2.3)

Cell dimensions and space group of the wt 68 kDa crystals obtained in the original structural study; PDB code 1EUU:
Space group: $P2_1$ Unit Cell: dim [Å]: a 50.87 b 116.99 c 60.00 angles [°]: α 90.0 β 95.60 γ 90.00

† $R_{merge} = \Sigma | I(k) - \langle I \rangle | / \Sigma I(k)$, where I (k) is the value of the k_{th} measurement of the intensity of a reflection, ‹I› is the mean value of the intensity of that reflection and the summation is over all measurements.

Values in parenthesis are for the highest resolution shell.

7.3.2. Data processing.

For all datasets the integrated intensities output from MOSLFM were scaled and merged using the program SCALA (Evans, 1997) from the CCP4 package (CCP4, 1994). The statistical analysis of the scaling is output to the log file generated by the program. Tables 7.3.2.1 – 7.3.2.4 show the summary of the log file tables from the output of SCALA for the three datasets. Ideally, the intensities of 100 % of the Bragg reflections should be measured and the data should be significant (≥ 2) in terms of the $I/\sigma(I)$ ratio throughout the resolution range (Wilson, 2001). The R_{merge} value is provided for each resolution shell along with the values of $I/\sigma(I)$ and Completeness. These values were used to assess the quality of the data and their useful resolution. All the datasets have a completeness > 85 %, which we deemed acceptable for refinement, along with significant intensities in the outer resolution shell and acceptable merging R-values. The dataset collected on the Y370G-βNeu5Ac complex was very fortuitous, given that the total rotation range collected was only 120°, which falls short of the 180° degrees sometimes required for monoclinic space groups. The orientation of this crystal was clearly adventitious as the completeness of the data was 87 % overall and in the highest resolution shell. However, the redundancy of the measurements was clearly affected and is significantly lower than the other datasets Table 7.3.1. The low $I/\sigma(I)$ value and high R_{merge} for the highest resolution shell in the Y370F complex was considered acceptable due to the low overall merging R-value, high redundancy and high $I/\sigma(I)$ values for this dataset. The log files from the scaling of the D92G-Neu5Ac2en dataset were unfortunately lost due to a computer hardware failure.

It was clear from both the unit cell dimensions and the different space groups obtained, that the crystal packing would be substantially altered compared to that found in the original crystals obtained for the 68 kDa wt protein (Gaskell *et al* ., 1995). Molecular replacement (MR) would therefore be required to locate the positions of the protein in the new crystal forms. Table 7.3.2.5. gives the results of the cell content analysis for each of the five crystals, which would prove invaluable in finding a successful MR solution.

Table 7.3.2.1. Statistics of data processing in the ten resolution shells for the MvNA D92G – lactose dataset, collected on ID14eh2.

Resolution shell (Å)	I/σ(I)	Completeness (%)	R_{merge} (%)
∞ - 5.38	6.5	90.3	7.2
5.38 – 3.80	7.7	95.9	6.6
3.80 – 3.10	10.1	97.7	5.2
3.10 – 2.69	9.6	98.3	5.6
2.69 – 2.40	8.3	98.0	6.9
2.40 – 2.19	6.4	97.7	9.0
2.19 – 2.03	5.0	96.5	11.2
2.03 – 1.90	3.0	93.5	14.6
1.90 – 1.79	3.1	90.6	21.3
1.79 – 1.70	2.1	72.2	32.3

Table 7.3.2.2. Statistics of data processing in the ten resolution shells for the MvNA Y370F – Neu5Ac2en (DANA) dataset, collected on ID14eh2.

Resolution shell (Å)	I/σ(I)	Completeness (%)	R_{merge} (%)
∞ - 6.64	8.8	99.5	3.9
6.64 – 4.70	14.1	99.8	3.9
4.70 – 3.83	14.3	99.9	4.1
3.83 – 3.32	11.6	99.9	5.2
3.32 – 2.97	9.2	100	6.9
2.97 – 2.71	6.9	100	10.0
2.71 – 2.51	4.8	100	14.8
2.51 – 2.35	3.1	100	23.3
2.35 – 2.21	2.2	100	32.6
2.21 – 2.10	1.5	100	47.4

Table 7.3.2.3. Statistics of data processing in the ten resolution shells for the MvNA Y370G – βNeu5Ac (sialic acid) dataset, collected on ID14eh3.

Resolution shell (Å)	I/σ(I)	Completeness (%)	R_{merge} (%)
∞ - 5.69	9.4	78.1	6.2
5.69 – 4.02	8.7	83.5	6.4
4.02 – 3.29	8.2	85.5	6.9
3.29 – 2.85	7.4	86.8	7.9
2.85 – 2.55	6.4	87.9	9.5
2.55 – 2.32	5.7	88.8	11.2
2.32 – 2.15	2.0	89.7	14.3
2.15 – 2.01	3.7	89.1	14.6
2.01 – 1.90	2.7	88.7	17.5
1.90 – 1.80	3.0	85.6	21.5

Table 7.3.2.4. Statistics of data processing in the ten resolution shells for the MvNA wt – α2,3 difluoro sialic acid dataset, collected on ID14eh2.

Resolution shell (Å)	I/σ(I)	Completeness (%)	R_{merge} (%)
∞ - 6.32	14.7	99.1	4.1
6.32 – 4.47	13.5	100	4.5
4.47 – 3.65	12.2	100	5.3
3.65 – 3.16	10.2	100	6.4
3.16 – 2.83	8.7	100	7.7
2.83 – 2.58	7.1	100	9.9
2.58 – 2.39	5.6	100	12.8
2.39 – 2.24	4.4	100	16.5
2.24 – 2.11	3.4	100	21.4
2.11 – 2.00	2.3	100	29.5

Table 7.3.2.5a. Cell content analysis.

Content analysis	D92G – Neu5Ac2en	D92G – lactose	Y370F – Neu5Ac2en
Space group	$P2_12_12_1$	$P2_12_12_1$	$P3_221$
Cell. Vol. (Å^3)	748 256	738 405	2 848 180
V_m[a](Å^3/Da)	2.8	2.7	2.3
Solvent content (%)	54.9	54.3	46.7
No. Mol./Asym. Unit	1	1	3

Table 7.3.2.5b. Cell content analysis.

Content analysis	Y370G – βNeu5Ac	Wt – α2,3-difluoro sialic acid
Space group	$P2_1$	$P2_12_12_1$
Cell. Vol. (Å^3)	287 804	751 894
V_m[a](Å^3/Da)	2.1	2.8
Solvent content (%)	41.4	55.2
No. Mol./Asym. Unit	1	1

V_m[a] is the Mathews' coefficient (Kantardjieff and Rupp, 2003; Mathews, 1968).

7.4. Molecular Replacement.

The different unit cell dimensions and space group symmetry observed for the MvNA crystals in this study necessitated the use of molecular replacement (MR) to locate the positions of the protein(s) in the new crystal forms. The MR program AMoRE (Navaza, 1994; Navaza, 2001), as implemented in the CCP4 suite (v.5.0.1) (CCP4, 1994) was used. This program was chosen over the CNS suite described previously due to the overall speed of the rotation and translation functions, which allows for fast, multiple explorations of crystal configurations. The autoAMoRE function was not used, instead each step of the MR was carried out separately so as to control the output of the program and allow for intelligent user intervention.

The original crystals obtained for the 68 kDa form of MvNA belonged to space group $P2_1$; the unit cell dimensions are given at the bottom of Table 7.3.1 (Gaskell *et al* ., 1995). This space group was observed for the Y370G complex; however, the unit cell dimensions along edges b and c were sufficiently non-isomorphous to suggest that the packing was different in these crystals. The different crystal forms obtained could be split into three different groups defined by their crystal symmetry. The three orthorhombic crystals, the D92G mutants and wt, belonged to space group $P2_12_12_1$ and were isomorphous with one another. The two other mutants, Y370G and Y370F, formed monoclinic or trigonal crystal forms respectively, belonging to space groups $P2_1$ and $P3_221$. For each of these crystal forms a new MR solution was needed to locate the protein in the unit cell. The structure of original wt MvNA protein (PDB code 1EUU) was used as the search model in each case. The solutions obtained are presented below for each crystal form.

7.4.1. Molecular replacement solution for the $P2_12_12_1$ crystal form of MvNA.

The D92G and wt crystals obtained in this study were orthorhombic, belonging to space group $P2_12_12_1$. The crystals were isomorphous, requiring only one solution to be found that could subsequently be used to refine the other datasets against. The cell content analysis suggested the presence of only 1 molecule/asymmetric unit with a solvent content of ~55 % and a Mathews' coefficient of 2.8 $Å^3$ Da^{-1} (Table 7.3.2.5.). Using data in the range 15 – 3.0 Å and rotation search sphere of 38 Å, only one rotation solution was found, with a correlation coefficient of 17.8; see Table 7.4.1.1. This result agreed favourably with the cell content analysis. This solution was then translated within the unit cell to locate the protein. The result of the translation search increased the correlation coefficient to 38.9 and had an R_{factor} of 48.7 %; see Table 7.4.1.2. Rigid body fitting of the search model, using the whole protein as one rigid body, increased the correlation coefficient to 48.8 and dropped the R_{factor} to 46.7 %; see Table 7.4.1.3. The output model built by AMoRE had zero close contacts and analysis of the packing in the graphics program 'O' (Jones *et al* ., 1990) showed good packing of the protein into the crystal.

Table 7.4.1.1. The single solution obtained for the cross-rotation function for the $P2_12_12_1$ crystal form of MvNA.

Solution	α	β	γ	[a]CC_F	[b]RF_F	[c]CC_I	[d]CC_P
1	171.65	56.43	164.37	17.8	54.8	24.0	16.6

Table 7.4.1.2. The result of the translation search applied to the single solution obtained from the rotation search for the $P2_12_12_1$ crystal form of MvNA.

Solution	α	β	γ	Tx	Ty	Tz	CC_F	RF_F	CC_I
1	171.65	56.43	164.37	0.468	0.443	0.107	38.9	48.7	37.4

Table 7.4.1.3. Rigid-body fitting for the translation function solution of the $P2_12_12_1$ crystal form of MvNA.

Solution	α	β	γ	Tx	Ty	Tz	CC_F	RF_F	CC_I
1	171.45	56.51	164.25	0.469	0.443	0.107	48.8	46.7	48.7

7.4.2. Molecular replacement solution for the $P2_1$ crystal form of MvNA.

The Y370G mutant complex crystallised in the same space group as the original crystal structure (PDB code 1EUU). However, as discussed above, the unit cell and hence the crystal packing were substantially different. The cell content analysis suggested, like the $P2_12_12_1$ crystal form, the presence of only 1 molecule/asymmetric unit. The solvent content was slightly lower at 41.4 % and the Mathews' coefficient was 2.1 $Å^3$ Da^{-1} (Table 7.3.2.5.). Using data in the range 30 – 4.0 Å and rotation search sphere of 23 Å, only one rotation solution was found, with a correlation coefficient of 18.5; see Table 7.4.2.1. This result agreed favourably with the cell content analysis. This solution was then translated within the unit cell to locate the protein. The result of the translation search increased the correlation coefficient to 30.8 and had an R_{factor} of 49.8 %; see Table 7.4.2.2. Rigid body fitting of the search model, using the whole protein as one rigid body, increased the correlation coefficient to 50.7 and dropped the R_{factor} to 43.6 %; see Table 7.4.2.3. The output model built by AMoRE had 68 close contacts. Analysis of the packing in the graphics program 'O' showed that these clashes were not serious. Subsequent refinement and model building eliminated them from the model.

Table 7.4.2.1. The single solution obtained for the cross-rotation function for the $P2_1$ crystal form of the Y370G mutant of MvNA.

Solution	α	β	γ	[a]CC_F	[b]RF_F	[c]CC_I	[d]CC_P
1	310.39	59.69	127.11	18.5	53.6	13.7	37.7

Table 7.4.2.2. The result of the translation search applied to the single solution obtained from the rotation search for the $P2_1$ crystal form of the Y370G mutant of MvNA.

Solution	α	β	γ	Tx	Ty	Tz	CC_F	RF_F	CC_I
1	310.39	59.69	127.11	0.383	0.000	0.232	30.8	49.8	19.2

Table 7.4.2.3. Rigid-body fitting for the translation function solution of the $P2_1$ crystal form of the Y370G mutant of MvNA.

Solution	α	β	γ	Tx	Ty	Tz	CC_F	RF_F	CC_I
1	309.55	59.41	129.13	0.381	0.000	0.232	50.7	43.6	41.3

7.4.3. Molecular replacement solution for the $P3_221$ crystal form of MvNA.

The Y370F mutants crystallised in a totally different space group to those observed previously. The crystals were found to be trigonal, belonging to space group $P3_221$. The cell content analysis suggested a whole range of possibilities for the number of molecules per asymmetric unit. We chose to look for three molecules, as the solvent content and Mathews' coefficient for this arrangement was most similar to the other crystal forms, being 46 % and 2.6 $Å^3$ Da^{-1} respectively (Table 7.3.2.5.).

Using data in the range 15 – 3.0 Å and rotation search sphere of 38 Å, a large number of rotation solutions were found with very similar correlation coefficients; see Table 7.4.3.1. The correct solutions were not obvious at this stage in the MR process. However, after a number of different combinations of solutions were tried, the three rotation solutions highlighted (1, 3 & 4) were found

to be correct. The first solution from the rotation search gave the location of the first monomer following translation. Once these coordinates were known, the position of this monomer was fixed and the position of the second monomer was sought. This process was continued until all three monomers had been found. The result of the final translation and rigid-body fitting of the search model on the coordinates are given in Tables 7.4.3.2. and 7.4.3.3. respectively. The final correlation coefficient and R_{factor} were 47.7 and 48.4 %.

The correct solution had only 18 close contacts, compared to >800 for an incorrect solution. The correct choice of searching for three molecules in the asymmetric unit was also justified following the refinement of this model against the dataset, described in the next section. The packing of the three monomers in the unit cell also looked very good when viewed in 'O', with clearly defined solvent channels and close packing of the monomers to the symmetry related molecules.

Table 7.4.3.1. Top 10 solutions for the cross-rotation search function in the MR of the Y370F $P3_221$crystal form of MvNA.

Solution	α	β	γ	[a]CC_F	[b]RF_F	[c]CC_I	[d]CC_P
1	69.18	90.00	153.51	18.2	56.3	26.4	6.9
2	51.51	88.12	334.67	18.0	56.3	25.9	7.3
3	100.92	6.10	221.77	17.9	56.4	26.1	8.4
4	106.58	82.26	63.67	17.8	56.3	26.1	6.8
5	64.64	73.64	152.14	17.6	56.5	25.6	6.0
6	2.02	62.33	134.10	17.5	56.5	25.5	6.6
7	23.25	53.69	115.38	17.5	56.5	25.0	5.9
8	61.65	70.54	150.08	17.4	56.6	25.2	5.5
9	63.03	62.74	151.77	17.3	56.6	25.0	5.7
10	75.62	79.43	127.50	17.3	56.6	24.9	5.7

Table 7.4.3.2. The result of the translation search applied to the three solutions obtained from the rotation search in the MR of the Y370F $P3_221$crystal form of MvNA.

Solution	α	β	γ	Tx	Ty	Tz	CC_F	RF_F	CC_I
1	69.48	90.00	153.51	0.107	0.455	0.449	24.2	54.7	45.6
3	100.92	6.10	221.77	0.289	-0.14	0.249	31.4	52.5	43.8
4	106.58	82.30	63.67	0.186	0.204	0.219	40.1	49.5	40.3

Table 7.4.3.3. Rigid-body fitting for the three solutions to the location of the three monomers in the $P3_221$crystal form of MvNA.

Solution	α	β	γ	Tx	Ty	Tz	CC_F	RF_F	CC_I
1	68.97	90.70	153.19	0.106	0.454	0.448	48.4	47.7	48.8
3	99.03	5.35	222.05	0.291	-0.14	0.251	48.4	47.7	48.8
4	105.99	82.46	63.64	0.185	0.203	0.220	48.4	47.7	48.8

[a]CC_F is the correlation coefficient between the observed amplitudes for the data and the calculated amplitudes for model.
[b]RF_F is the classic R factor between the observed amplitudes for the data and the calculated amplitudes for the model.
[c]CC_I is the correlation coefficient between the observed intensities from the data and the sum of the calculated intensities for all symmetry equivalents of the model.
[d]CC_P is the Patterson correlation coefficient between the data and the model Patterson's evaluated within the defined sphere centred on the Patterson origin.

7.5. Model Refinement.

Following the placement of the wt MvNA structure (PDB 1EUU) into the three different crystal forms, the models required refining and manual modification to more accurately represent the observed diffraction data for each complex. The refinement of all five structures was carried out using the CCP4 program REFMAC5 (v.5.2.0003) (Murshudov, 1997). This program uses the statistical principles of Maximum Likelihood to improve the agreement between the observed and calculated data (Pannu *et al* ., 1998; Pannu and Read, 1996). Agreement is measured by the probability of measuring the data. So in effect, this program tackles the problem of model refinement as 'Given a protein model, what is the probability that the given set of data would be observed?' The model is then adjusted to maximise the probability of the given observations.

The refinement in each case was begun with 20 cycles of rigid body refinement, which is important if the model has come from molecular replacement. Molecular replacement can only place the protein molecules into their correct positions crudely within the asymmetric unit cell. This is due to both the resolution limits of the search (15 – 3 Å in this instance) and the treatment of the protein as one rigid body, which in many cases is an over simplification. This crude positioning is usually reflected in the initial R-factors from the molecular replacement, which are usually around 40 – 45 %. Rigid body refinement was used at the start of the model refinement process to optimise the position of the MvNA monomers within the asymmetric unit of the different crystal forms.

The structure of the MvNA sialidase is well suited to rigid body refinement due to its three-domain architecture. The three domains used corresponded to the β-propeller (residues 47 - 402), the immunoglobulin domain (residues 403 - 502) and the galactose-binding domain (residues 503 - 647); Figure 7.5.1. The initial resolution limits used were 15 – 3.0 Å in the case of the $P2_12_12_1$ and $P2_1$ crystal forms and 15 – 8.0 Å for the $P3_221$ crystal form.

The $P3_221$ crystal form had three molecules in the asymmetric unit and the orientation of the galactose-binding domain relative to the immunoglobulin

and β-propeller domains had changed for one of the monomers. This change was significant when compared to the original wt structure used as the search model. Using the resolution limits of 15 – 3.0 Å, the program was unable to move the domain into the correct orientation. It was decided therefore, to use a very low refinement resolution in an attempt to allow the model more freedom to move within the constraints imposed by the program. This solution worked and the three monomers were correctly positioned and optimised.

Following rigid body refinement, restrained refinement was used to improve the model. For each round of restrained refinement, 10 cycles of maximum likelihood restrained refinement were used. In order to account for the variation in B-factors throughout the model, isotropic B-factors were refined in each round of restrained refinement. For each complex, the weighting term was determined to optimise the restraints on the stereochemistry of the model.

Visual inspection of the electron density maps, $2F_o$-F_c and F_o-F_c, clearly showed peaks of extra density for the ligands in the different complexes, as will be discussed in the following chapter. The ligand descriptions required by REFMAC, which describe the stereochemical parameters for the ligands, were constructed using the CCP4 monomer library sketcher program and corrected using the molecular modelling program Sybyl (Tripos).

Manual inspection of the model and fitting of ligands was carried out using the modelling program 'O' (Jones *et al* ., 1990) with the aid of the $2F_o$-F_c and F_o-F_c electron density maps. A significant missing part of any macromolecular structure at less than 2.5 Å resolution is the ordered water molecules. For each of the MvNA complexes the program ARPwARP (Perrakis *et al* ., 1999) was used to build the solvent atoms into the different models. In some instances several rounds of water divining was needed. The waters were manually inspected to make sure they each sat in an appropriate location and electron density peak. The final refinement statistics for each model are given in Table 7.5.1. Throughout the refinement process the refinement values R and R_{free} were used to monitor how well the current model fitted the diffraction data. Changes to the

model that affected both the R and R_{free} were kept; those that only lowered the R factor were not.

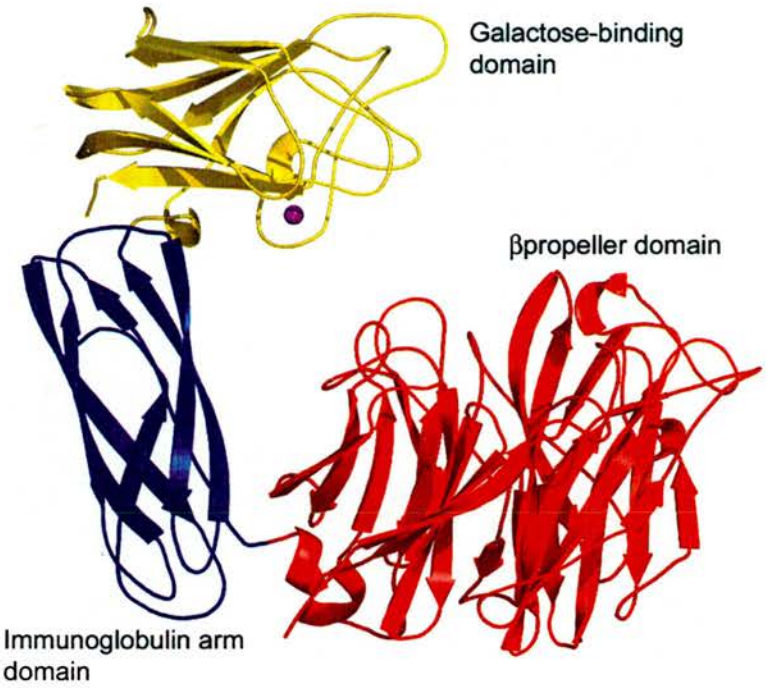


Figure 7.5.1. The structure of the wt *M.viridifaciens* sialidase showing the three rigid body groups used in the refinement of the five complexes. β-propeller domain (residues 47-402) in red, the immunoglobulin arm (residues 403-502) in blue and the galactose-binding domain (residues 503-647) in yellow

Table 7.5.1. Refinement statistics for the five MvNA complexes.

Complex	D92G – Neu5Ac2en	D92G – Lactose	Y370F – Neu5Ac2en	Y370G – βNeu5Ac	Wt – α2,3 di-fluoro Neu5Ac
Resolution range (Å)	87 – 2.1	87 – 1.7	124 – 2.1	105 – 1.8	88 – 2.0
R factor (%)	16.2	16.2	17.5	14.1	16.3
R free (%)	21.7	19.7	23.8	20.1	21.7
Rms deviation in bond length (Å)	0.02	0.02	0.021	0.02	0.02
Rms deviation in bond angles (°)	1.742	1.848	1.835	1.789	1.707
Esd of atom positions† (Å)	0.162 ± 0.038	0.08 ± 0.023	Chain A: 0.198 ± 0.037 Chain B: 0.165 ± 0.026 Chain C: 0.181 ± 0.040	0.103 ± 0.028	0.128 ± 0.04
Number of protein atoms	4555	4568	13 650	4548	4587
Number of water molecules	516	728	1079	849	914
Average B-factor for all atoms (Å^2)	25.2	21.2	Chain A: 31.5 Chain B: 26.4 Chain C: 28.9	15.0	22.7
Main-chain atoms	25.0	19.9	Chain A: 31.0 Chain B: 25.8 Chain C: 28.4	14.1	22.0
Side-chain atoms	26.6	22.6	Chain A: 32.0 Chain B: 26.9 Chain C: 29.5	16.2	23.6
Lactose	-----	44.3	-----	-----	-----
Neu5Ac2en	32.3	-----	Monomer A: 22.9 Monomer B: 27.9 Monomer C: 22.8	-----	-----
β-Neu5Ac	-----	-----	-----	10.7	-----
Neu5Ac2en	-----	-----	-----	-----	14.58

† The standard radial uncertainty of an atom with the average B-factor was estimated using the DPI (Diffraction Precision Indicator) method proposed by Cruickshank {Cruickshank, 1999 #131;Cruickshank, 1999 #132} as implemented in the program ESCET {Schneider, 2002 #127}.

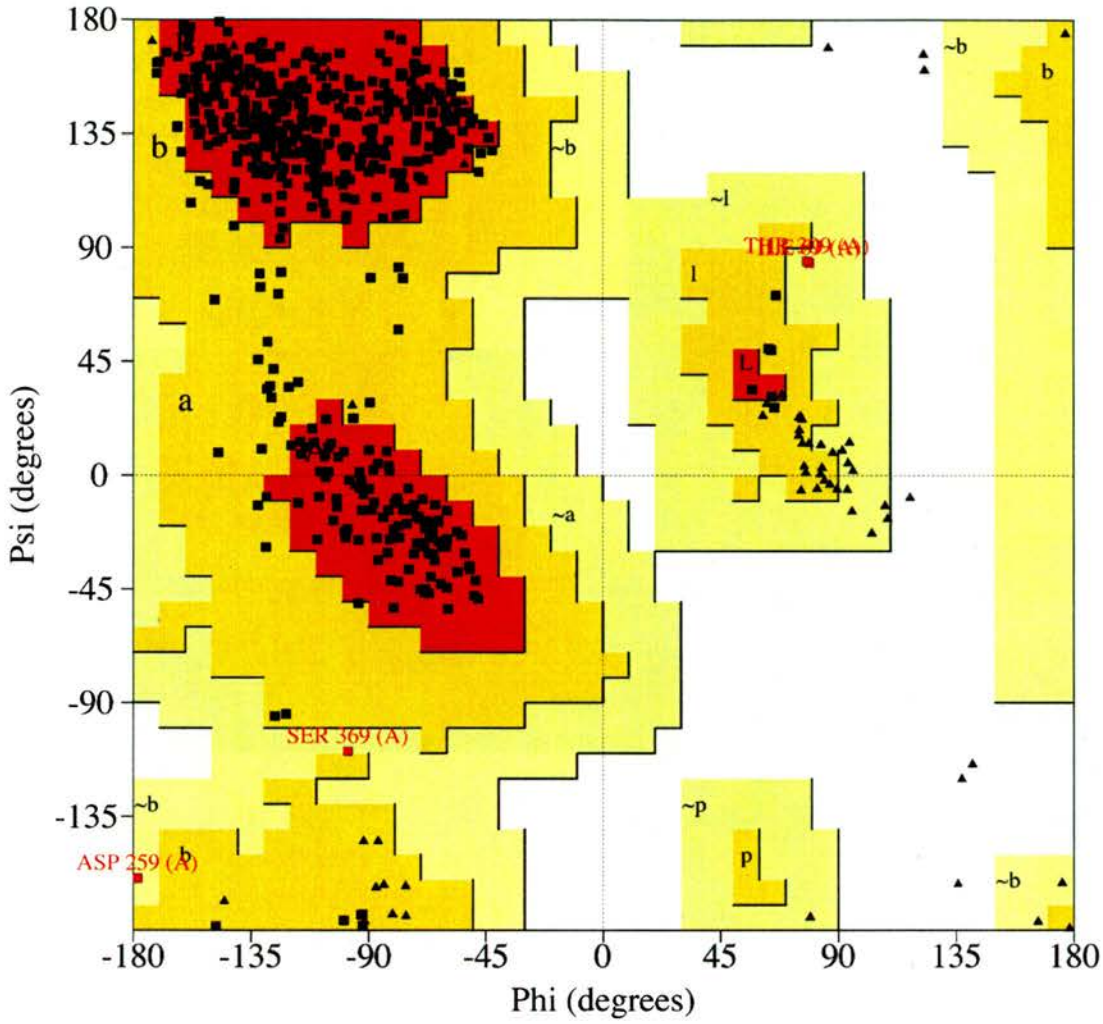
7.6. Structure validation.

The stereochemistry of the refined models of the MvNA complexes was checked using the programs PROCHECK (Laskowski *et al* ., 1993) and WHATIF (Vriend, 1990). Figures 7.6.1. – 7.6.5. show the Ramachandran plots of the main chain torsion angles (ϕ, ψ). There are three plots for the three monomers in the asymmetric unit of the $P3_221$ crystal form. PROCHECK also outputs a range of plots showing how a number of main chain and side chain parameters for each model compare to the average macromolecular structure at the given resolution. These plots have been omitted for brevity; however, all of the structures were classes as either 'inside' or 'better' for each of the parameters.

The Ramachandran plots show that none of the models had main chain torsion angles in the disallowed region of the plot. However, five residues are flagged up as having distorted torsion angle geometries. These are Ile69, Asp131, Asp259, Thr309 and Ser369. The torsion angles of these residues lie in the generously allowed regions of the plot. Some of these residues are flagged in more than one of the models, indicating that these unusual torsion angles are characteristic of the sialidase backbone. On closer investigation all of these residues are in clearly defined electron density and in many cases interact specifically with the ligands bound in the active site. Asp131 and Asp259 interact directly with the N-acetyl and glycerol groups of β sialic acid and 3F-DANA in the Y370G and wt complexes respectively. Ser369 is the residue preceding the conserved catalytic nucleophile Tyr370 and the unusual torsion angles most probably reflect the need to place the tyrosine in the correct orientation. Ile69 and Thr309 do not interact directly with the ligand, however, they appear in all of the models and both fit the electron density very well. These outliers are obviously forced into adopting the unusual torsion angles by the fold of the protein.

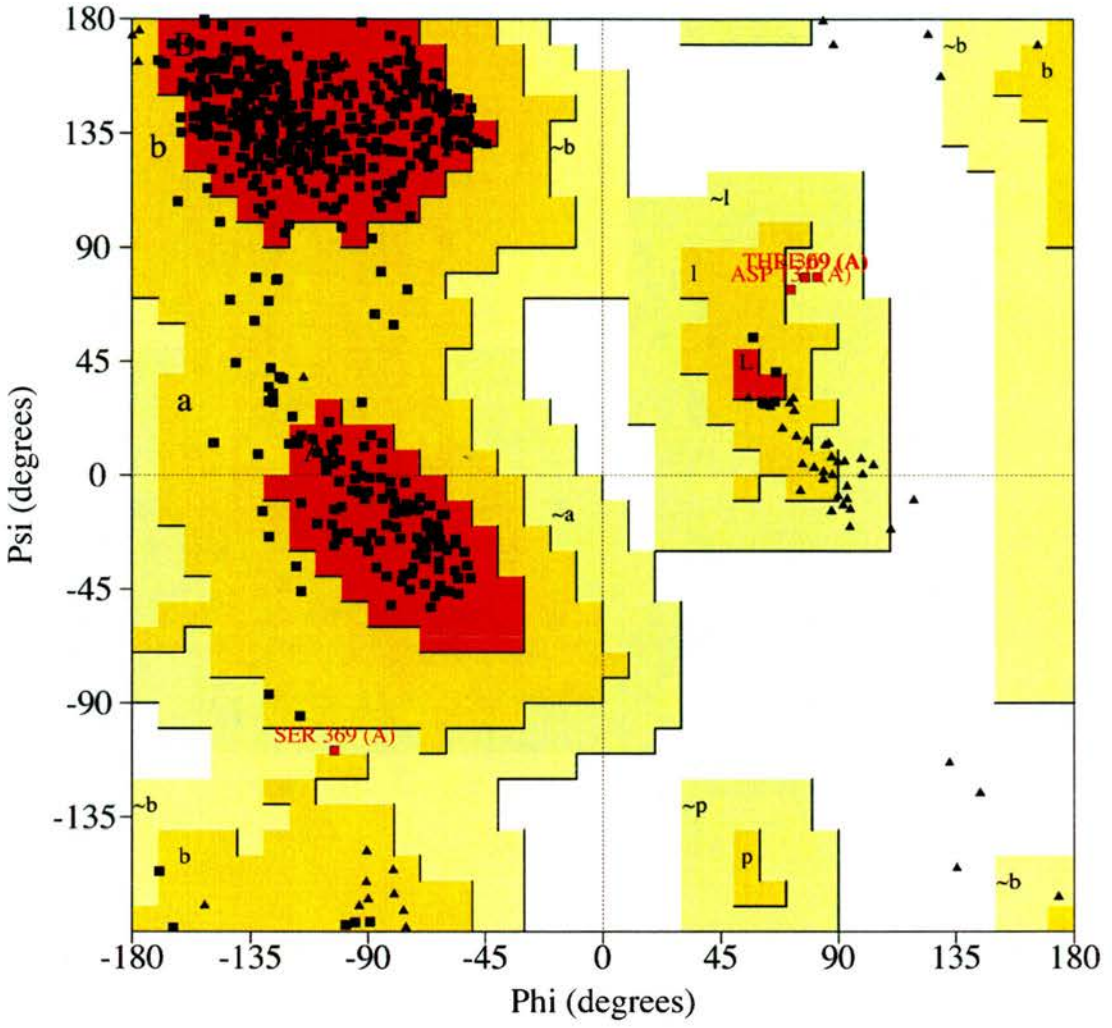
The availability of atomic resolution data (< 1.2 Å) for a protein structure allows the rigorous estimation of standard uncertainties (s.u.s) for individual atomic positions, bond lengths and angles. This was done for the atomic

resolution models of the *C.perfringens* nanI sialidase discussed in the previous chapters. In the model of the MvNA complexes, individual positional errors for the atoms were not available. In such cases it is necessary to derive approximate values. In the past few years Thomas Schneider, who has devised a novel method of calculating estimates for individual coordinate errors, has addressed this problem. Using an approximation based on the recently proposed diffraction precision indicator (DPI; (Cruickshank, 1999a; Cruickshank, 1999b; Cruickshank, 2001)) estimated standard deviations of atomic coordinates are calculated and scaled according to B-factors (Schneider, 2000; Schneider, 2002). This method is utilised in the program ESCET to provide an objective and statistically meaningful comparison of different proteins and conformations. The esd's calculated for each of the MvNA models are shown in Table 7.5.1. This program is used to compare the conformational polymorphism of the MvNA models in chapter 8.



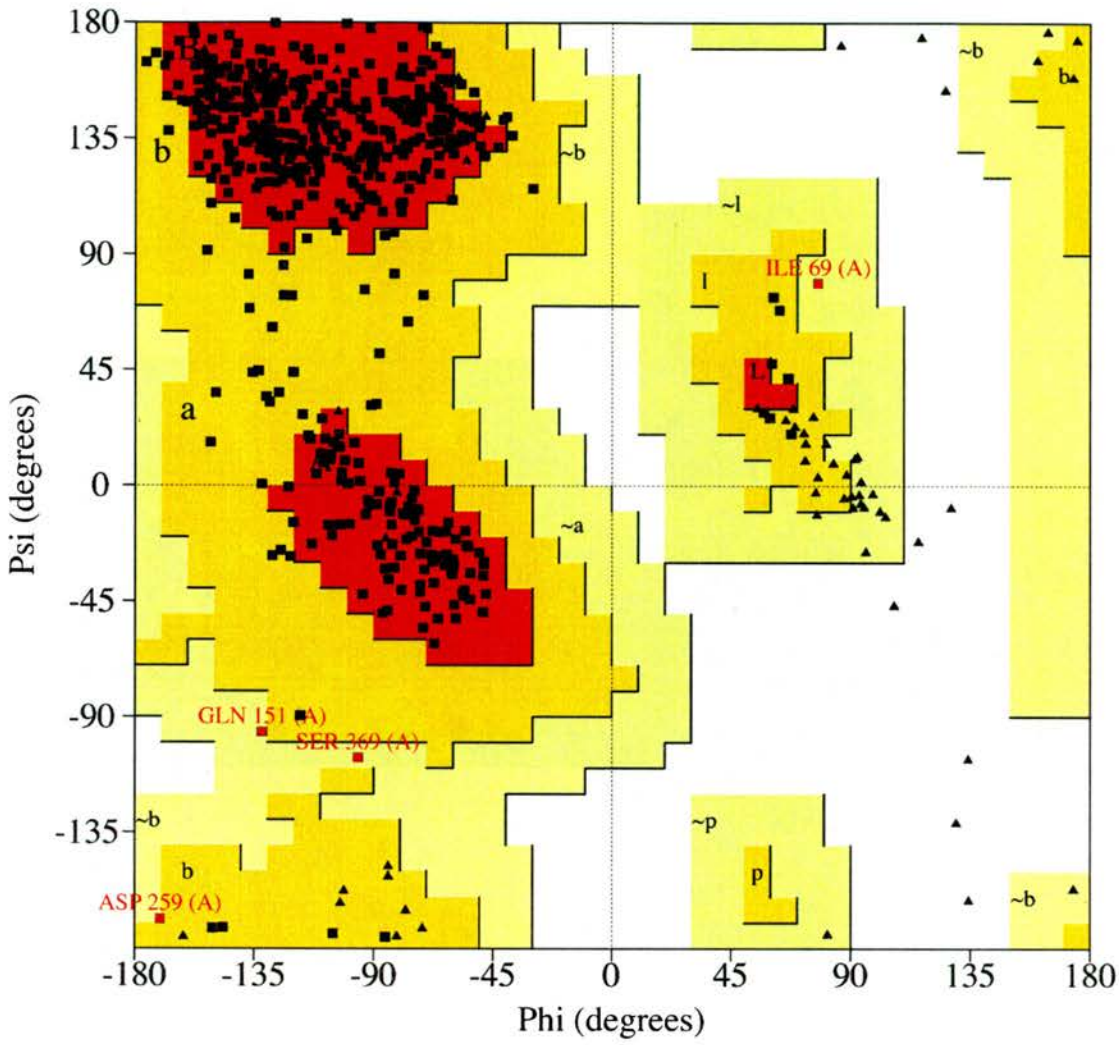
Residues in most favoured regions [A, B, L]	441	89.3 %
Residues in additional allowed regions [a, b, l, p]	49	9.9 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	4	0.8 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	494	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	65	
Number of proline residues	41	
Total number of residues	601	

Figure 7.6.1. Ramachandran plot for the MvNA D92G – Neu5Ac2en (DANA) complex.



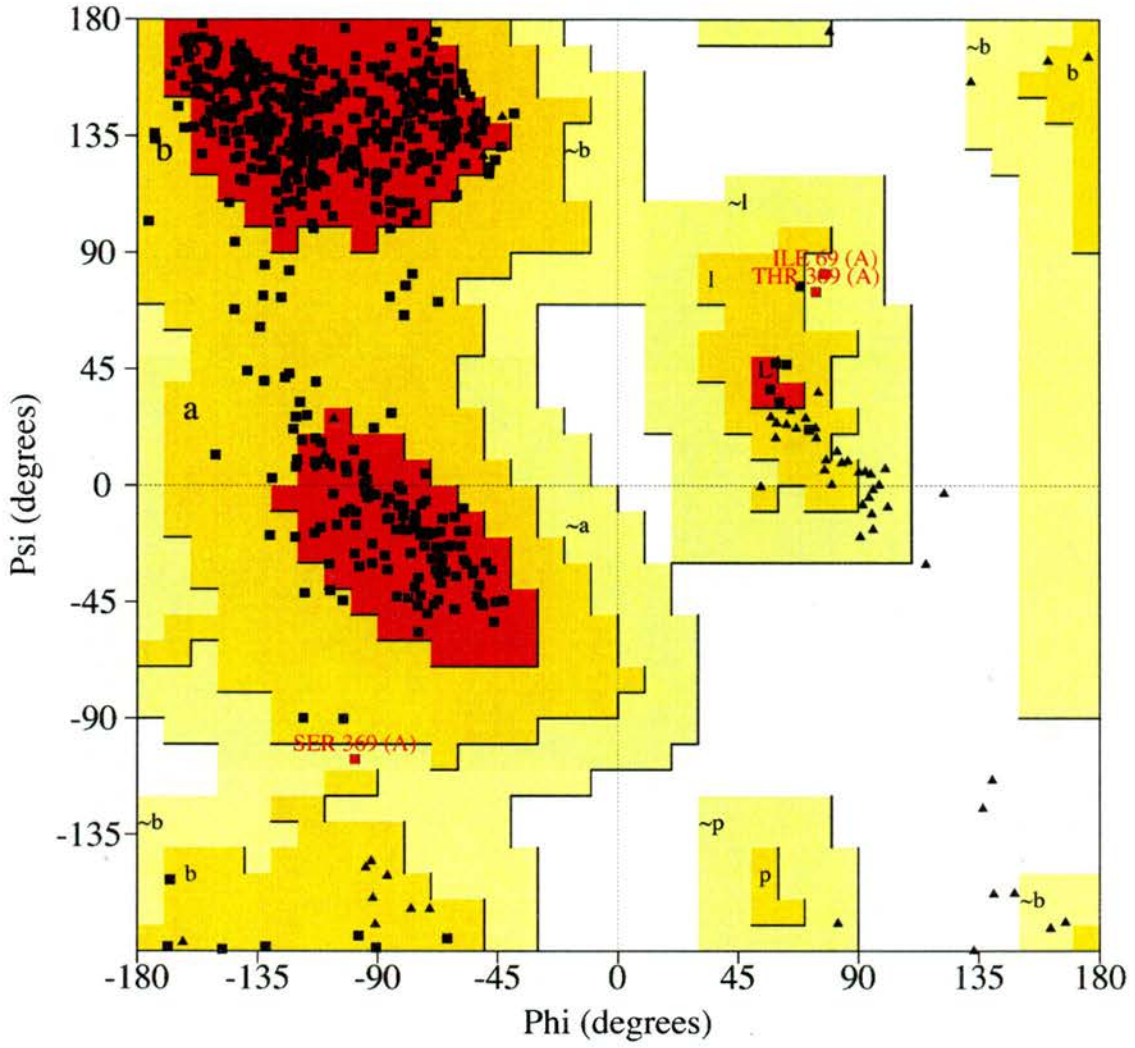
Residues in most favoured regions [A, B, L]	445	90.1 %
Residues in additional allowed regions [a, b, l, p]	45	9.1 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	4	0.8 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	494	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	65	
Number of proline residues	41	
Total number of residues	601	

Figure 7.6.2. Ramachandran plot for the MvNA D92G – Lactose complex.



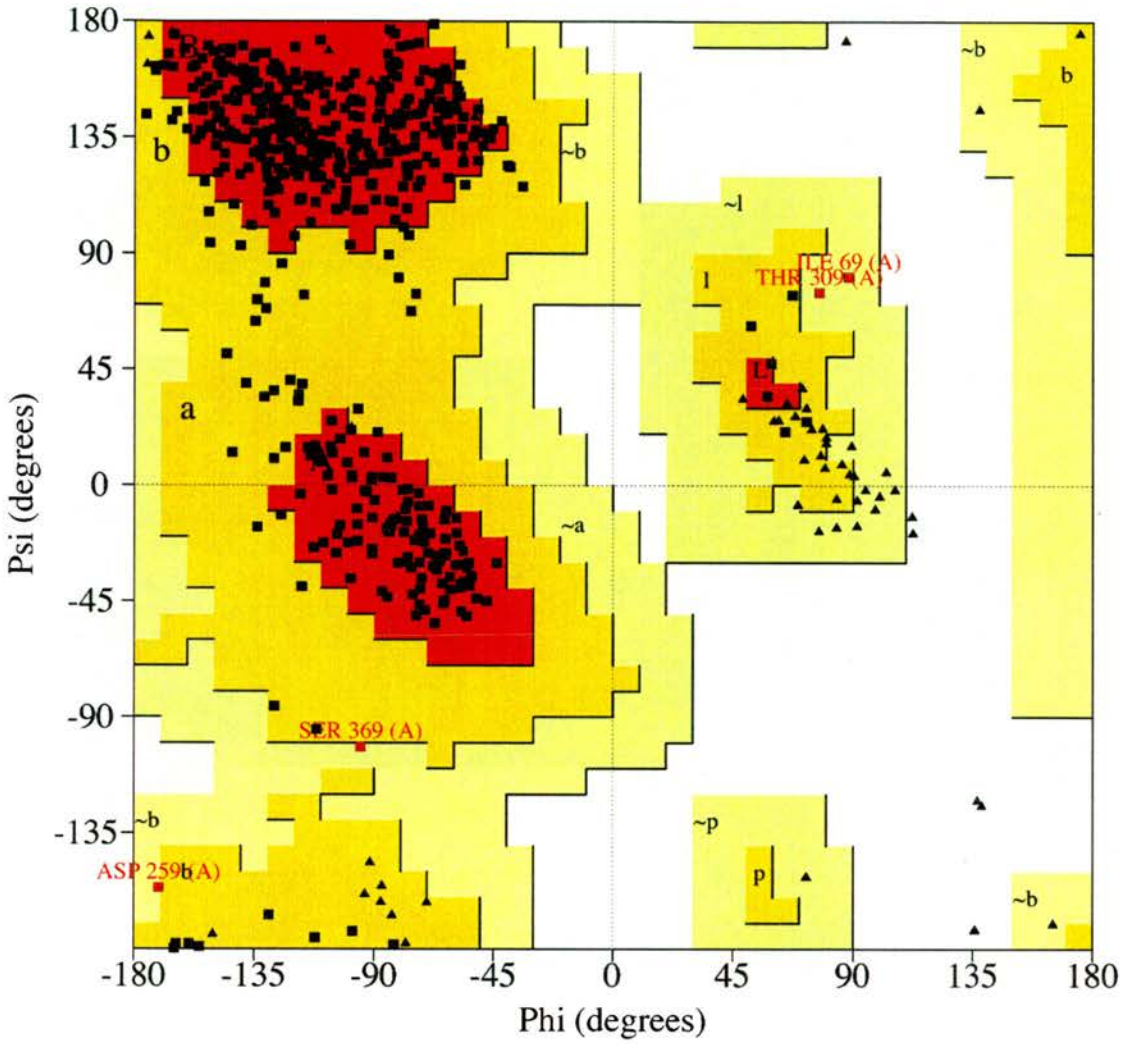
Residues in most favoured regions [A, B, L]	435	87.9 %
Residues in additional allowed regions [a, b, l, p]	56	11.3 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	4	0.8 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	494	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	64	
Number of proline residues	41	
Total number of residues	601	

Figure 7.6.3a. Ramachandran plot for monomer A of the Y370F – Neu5Ac2en (DANA) complex.



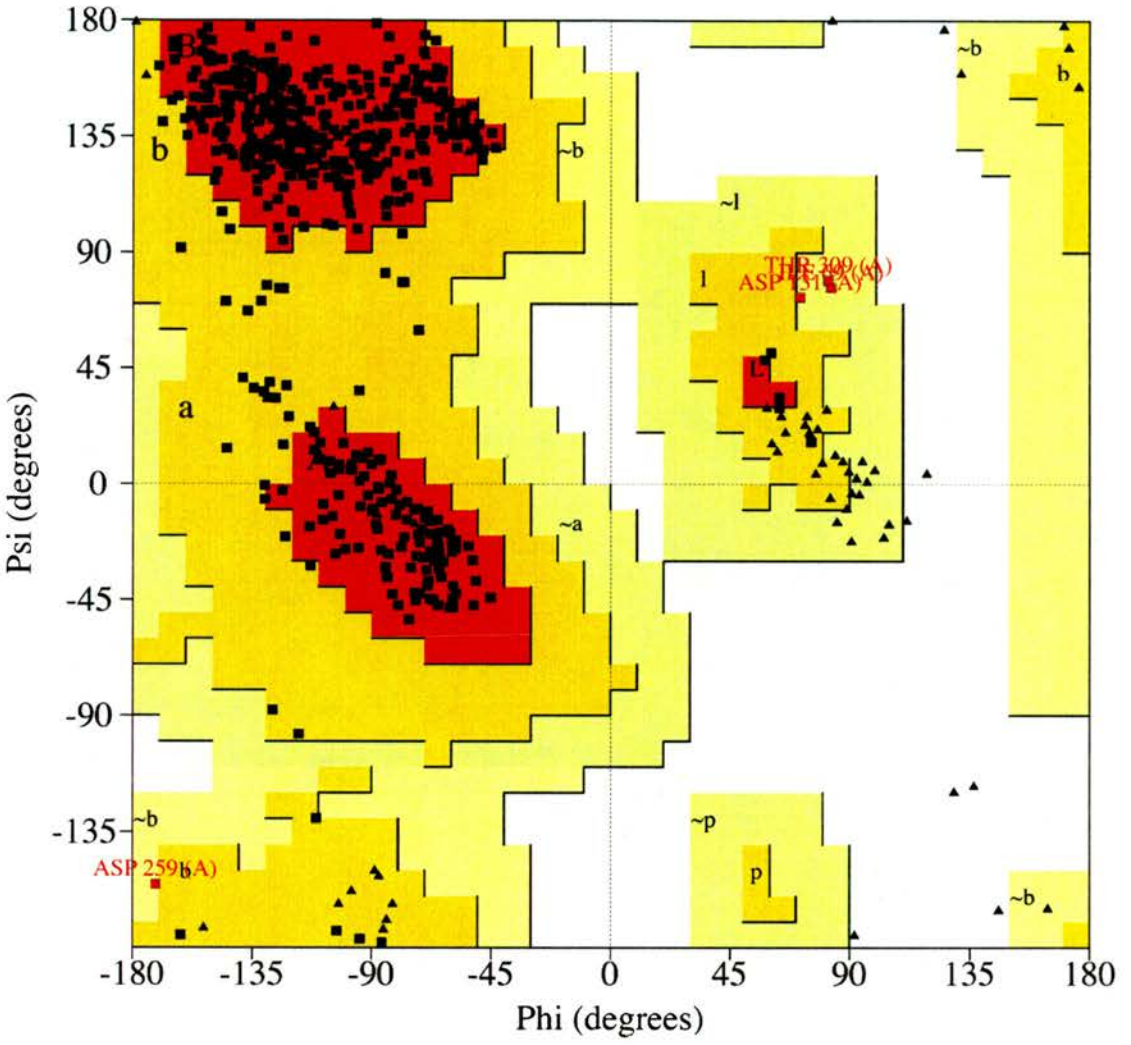
Residues in most favoured regions [A, B, L]	435	87.9 %
Residues in additional allowed regions [a, b, l, p]	57	11.5 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	3	0.6 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	495	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	64	
Number of proline residues	41	
Total number of residues	601	

Figure 7.6.3b. Ramachandran plot for monomer B of the Y370F – Neu5Ac2en (DANA) complex.



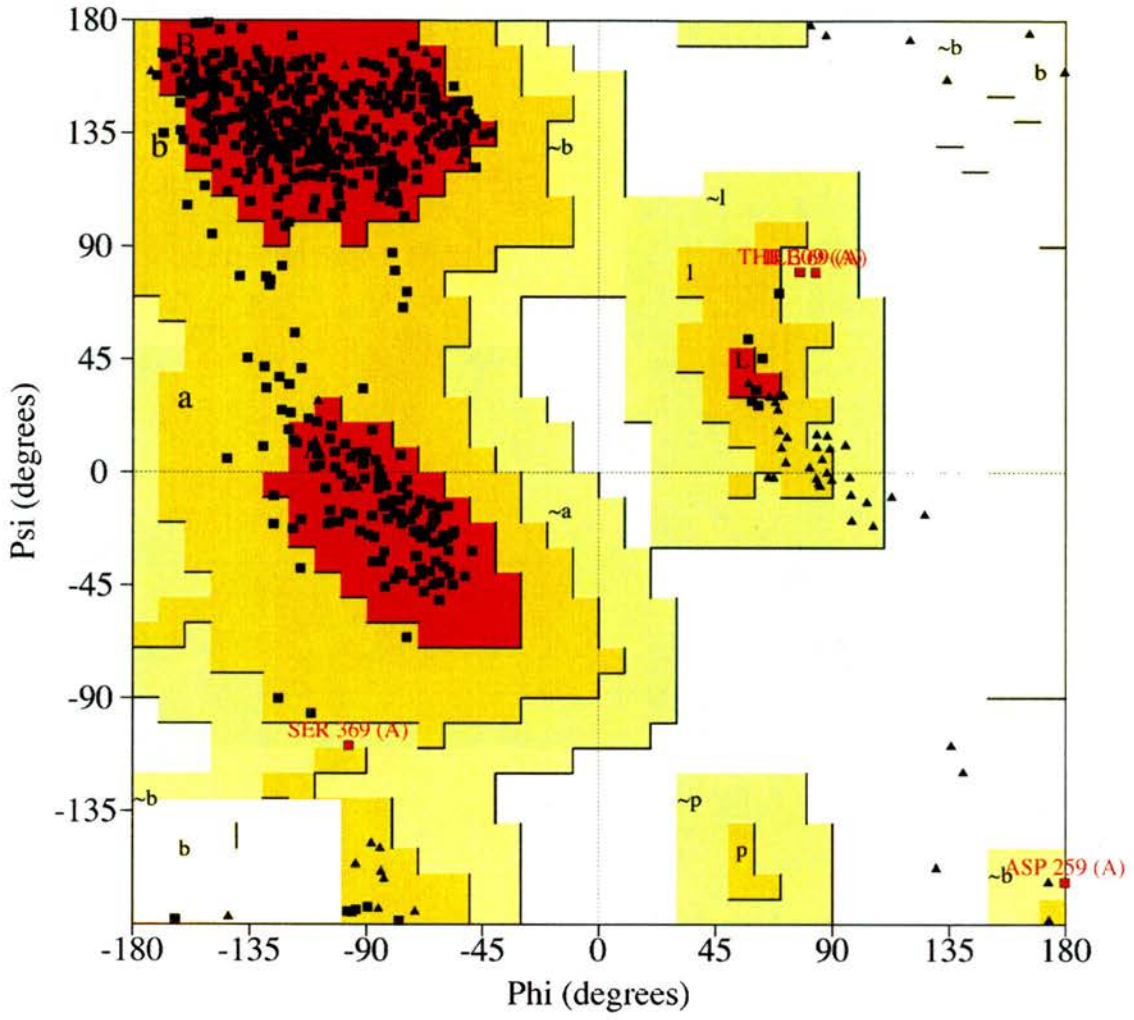
Residues in most favoured regions [A, B, L]	433	87.7 %
Residues in additional allowed regions [a, b, l, p]	57	11.5 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	4	0.6 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	494	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	63	
Number of proline residues	41	
Total number of residues	601	

Figure 7.6.3c. Ramachandran plot for monomer C of the Y370F – Neu5Ac2en (DANA) complex.



Residues in most favoured regions [A, B, L]	443	89.7 %
Residues in additional allowed regions [a, b, l, p]	47	9.5 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	4	0.8 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	494	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	65	
Number of proline residues	41	
Total number of residues	601	

Figure 7.6.4. Ramachandran plot for the MvNA Y370G - βNeu5Ac complex.



Residues in most favoured regions [A, B, L]	444	89.7 %
Residues in additional allowed regions [a, b, l, p]	47	9.5 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	4	0.8 %
Residues in disallowed regions	0	0.0 %
Number of non-glycine and non-proline residues	495	100.0 %
Number of end residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	64	
Number of proline residues	41	
Total number of residues	601	

Figure 7.6.5. Ramachandran plot for the wt MvNA – 3F-Neu5Ac2en (3F DANA) complex.

7.7. Discussion.

The original crystal structure of the *M.viridifaciens* 68 kDa sialidase (MvNA) was determined at 2.5 Å resolution using crystals belonging to space group $P2_1$ (Gaskell *et al* ., 1995). The crystal structure revealed the presence of three clearly identifiable domains, a six-bladed βpropeller at the N-terminus, an immunoglobulin domain and a galactose-binding jellyroll domain at the C-terminus. The function of these last two domains in the biological activity of this enzyme is still not understood. The aim of this structural study was to investigate the catalytic mechanism of this enzyme through targeted mutagenesis of the conserved active site residues followed by kinetic and structural analysis.

The original MvNA protein was purified directly from the cultured bacterium (Taylor *et al* ., 1992); a new recombinant system was established to allow the generation of site directed mutants (Watson *et al* ., 2003). In collaboration with Professor A. Bennet and Dr. J. Watson at Simon Fraser University, Vancouver, Canada, a number of mutants were constructed, purified and provided for crystallisation studies. Of these the following proved successful in the ligand binding and co-crystallisation studies, the D92G, Y370F and Y370G mutants.

The recombinant protein crystallised in a new condition. The original paper describes a crystal grown over two months in 4 % PEG 4000, with an initial protein concentration of 12 mg ml^{-1} in 0.1 M Tris-HCl buffer at pH 8.2 (Gaskell *et al*., 1995). Interestingly, this condition failed to induce the protein to form crystals. Screening the recombinant MvNA found a set of new crystallisation conditions in the PEG/Ion screen from Hampton Research. Two new conditions were found using 10 % PEG 3350 as the precipitant and 0.2 M of either di-ammonium hydrogen citrate or sodium fluoride as the main salt. These new conditions gave three different crystal forms.

All three of these crystal forms diffracted x-rays to higher resolution than were obtained in the original study of 2.5 Å. The highest resolution structure was that of the D92G mutant bound to lactose in the galactose-binding domain, at 1.7 Å. This model has allowed the carbohydrate-interactions made by this domain to

be determined in greater detail, as will be discussed in the following chapter. The crystal conditions presented here provide a reproducible platform from which further structural studies of this protein and its mutants can continue.

Interestingly, the wt, Y370F and D92G proteins crystallised in the di-ammonium hydrogen citrate condition, whilst the Y370G protein only crystallised in the presence of sodium fluoride. The reasons behind this are still unclear. The active site mutations were buried deep within the active site pocket of the β-propeller domain and are not involved in any crystal contacts. The change in crystallisation conditions and crystal forms may have been a result of different purification procedures, (columns, buffers, pH, etc) employed by Dr. J. Watson, who was optimising the purification protocol throughout the study (J. Watson, personal communication). Therefore, each batch of protein may have been exposed to different pH environments that may have subtly altered the charge and arrangement on the surface of the protein.

Another interesting observation from this study was the formation of different crystal forms in the same condition by different mutants. The D92G and wt protein would always form orthorhombic crystals belonging to space group 19, whereas the Y370F crystals was always form trigonal crystals of space group 154. Again, the reasons behind this are unclear, but the effects on the downstream data collection, processing and refinement are extreme. The orthorhombic crystals diffracted much further and had lower mosaicity compared to the trigonal crystal form (Table 7.3.1.) and subsequently had much better refinement parameters (Table 7.5.1.) that resulted in more accurate models.

The appearance of these different crystal forms was an unexpected, but very welcome result from this study. The unusual domain structure of this enzyme poses a number of significant questions with ramifications in protein engineering and biotechnology. It is clear from the structure of this enzyme that the three domains are very separate and one could be forgiven in believing that this enzyme was manufactured by simply splicing these three domains together. The three domains appear to have very distinct functions, the β-propeller domain houses the active site pocket and functions independently of the other two

(Gaskell *et al* ., 1995). The immunoglobulin domain appears to act as an arm, connecting the catalytic domain to the galactose-binding C-terminal domain. The flexibility of these domains relative to one another is one question that has so far been unanswered in the literature. The different packing arrangements observed for MvNA protein in the crystal forms could be used to investigate this question. The results from the molecular replacement, especially in the trigonal form, suggest that this three-domain protein exhibits significant conformational polymorphism and this is explored further in the next chapter.

Refinement of the wt MvNA crystals soaked in the α2,3-di-fluoro sialic acid derivative was found to contain a mixed population of both DANA and covalent intermediate complexes. The ratio clearly favoured the DANA complex, as this was easily refined against the data with the sugar ring planar around C3-C2-O6. However, a positive peak (4 σ) in the difference density map was observed in sitting above C3 of the DANA compound, in a position one would expect fluorine to sit in a covalent complex. A fluorine could be refined in this position with half occupancy, although refinement with full occupancy resulted in a large negative peak (6 σ). No peaks were observed for the covalent bond from C2 of the sugar to the tyrosine (Tyr370) and any attempt to refine this bond resulted in the appearance of strong negative peak (7 σ) in the difference maps. In the final model therefore, DANA was refined in the active site, although the evidence suggests the presence of the covalent intermediate in a small population of the protein molecules. This experiment need repeating to completely trap the covalent intermediate in this sialidase.

Chapter 8

Structural Studies on Active Site Mutants of the 68 kDa Sialidase from *Micromonospora viridifaciens*.

Structural Analysis of the Complexes.

8.1. Summary.

Structural studies were carried out on three active site mutants of the 68 kDa sialidase from the soil bacterium, *M. viridifaciens* (MvNA). The principle aim of this project was to understand the effects these mutations had on both the structural integrity and interactions to substrate and products within the active site pocket. The three mutations described are those of the conserved catalytic nucleophile Tyr370 to phenylalanine (Y370F) and glycine (Y370G), and the conserved acid/base catalyst Asp92 to glycine (D92G). The structures of these mutants, in complex with either the transition state analogue Neu5Ac2en (DANA) or β-Neu5Ac (sialic acid), have proved informative in understanding the results from kinetic and product studies carried out by our collaborators (Watson *et al* ., 2003; Watson *et al* ., 2004). The structures of these complexes are described in the context of these biochemical studies and the information they provide about the catalytic mechanism of this sialidase.

The two extra domains attached to the C-terminus of the β-propeller fold make this sialidase one of the most structurally diverse members of the sialidase superfamily. The immunoglobulin 'arm' domain connects the catalytic β-propeller to a carbohydrate-binding module (CBM) at the C-terminus. This domain has now been classified as a type C small sugar binding CBM (Coutinho and Henrissat, 1999a; Coutinho and Henrissat, 1999b). This unique class of CBM has the lectin-like property of binding optimally to mono-, di- or tri-saccharides and has distinct structural differences to the type A and B classes (Boraston *et al* ., 2004). The previous structural study identified the monosaccharide galactose as the possible target ligand for this domain (Gaskell *et al* ., 1995). However, the

data on this complex was the result of a simple 'quick and dirty' soaking experiment carried out on the fly at the synchrotron (G. Taylor pers. Comm.). The result was a dataset only 62 % complete extending to 2.5 Å resolution, allowing only crude placement of the sugar into the binding pocket (G. Taylor, pers. comm.). During the course of this study, the disaccharide lactose was observed bound to this domain in one of the D92G mutant structures and refined to a resolution of 1.7 Å. The higher resolution of this structure allows a more detailed analysis of the protein-carbohydrate binding interactions in this domain. This structure has important implications for the role of the type C CBM in the function of this enzyme. Superimposition of the original wt structure (PDB 1EUU) and the lactose complex presented here, show that the location of the galactose modelled in the original wt structure was possibly incorrect.

Protein models derived from different crystal forms provide a unique opportunity to investigate the conformational variability within a structure. Three different crystal forms of the protein were observed during data collection on these mutants: monoclinic, orthorhombic and trigonal, providing seven independently refined models. An objective comparison between the seven MvNA protein structures was made using the program ESCET (Schneider, 2002). This analysis has provided valuable insight into the flexibility of this unique multi-domain sialidase, highlighting the regions of the protein that are conformationally invariant and those regions that are flexible. These data shed new light on the relationship between the catalytic β-propeller domain and the C-terminal type C CBM.

8.2. Objective Comparison of the MvNA protein structures.

Understanding the relationship between macromolecular structure and function often relies on the comparison of different structural models of a molecule. The structural study of the MvNA sialidase mutants provided a popuLacion of crystal structures from different crystal forms and molecules reLaced by non-crystallographic symmetry, with which to carry out such a comparison. Although a single crystal structure is thought to represent a rather static model of a protein molecule, the comparison of different conformers of a given molecule, as obtained from different structure determinations, can provide valuable insight into its flexibility. The availability of these models provided the perfect opportunity to analyse the conformational variability within the unique three-domain structure of this sialidase.

The original structure was thought to be quite flexible, with the three domains able to move as independent rigid-bodies. This hypothesis fitted with the idea of the galactose-binding domain as the molecular anchor, attaching to carbohydrate surfaces likely to contain sialic acid. The β-propeller domain would then be able to cleave off the sialic acid for the bacterium to use as a carbon or energy source. The original structure of the MvNA protein also seemed to suggest a flexible structure, with only a di-glycine linker, at residues Gly402 and Gly403, between the immunoglobulin 'arm' and the β-propeller domain (Gaskell *et al* ., 1995). The only other direct link between these two domains is a disulphide bond between Cys351 on the β-propeller and Cys405 on the immunoglobulin 'arm'. The aim of this analysis was to compare the seven independent structures of the MvNA sialidase and determine the extent to which the three domains are flexible relative to each other.

In such a comparative analysis, the identification of the part of the molecule that is conformationally invariant with respect to a set of conformers is a critical step, as the corresponding subset of atoms constitutes the reference for subsequent analysis, for example by least-squares superposition. The application of error-scaled difference distance matrices and their simultaneous

analysis using a genetic algorithm has been implemented in the program ESCET for the purpose of objectively comparing protein structures (Schneider, 2002).

When two conformations of a molecule are compared, the central question is whether the relative positions of atoms are different. As all structural models contain errors, different in this context means significantly different with respect to the precision of the atomic coordinates in the structural models being compared. The ESCET program calcuLaces for each pair of monomers an error-scaled difference distance matrix (Schneider, 2000), that takes into account the different levels of coordinate precision, both for different models and individual atoms, via a modified form of Cruickshank's DPI (Cruickshank, 1999a; Cruickshank, 1999b). This modified form of the DPI is used to place the errors in different structures onto a common scale by exploiting the experimentally observed correLacion between standard uncertainties (esd's) and B values (Schneider, 2000). Once on a common scale the significance of the movement can be determined and judgements made about the possible biological implications this may have.

For the seven models from the structural study (Chapter 7), 21 error-scaled C_α-C_α difference distance matrices with 179 101 independent elements each were calculated and analysed by the ESCET program. The uncertainties for the differences in C_α-C_α distances ranged from 0.037 Å for atom pairs with low B-factors, to 0.409 Å for atom pairs with high B-factors, with a mean of 0.201 Å. Pairwise comparison of the seven models (Table 8.2.1.) shows that the D92G-Dana complex is identical to the wt complex, within experimental error. Also, monomers A and B in the asymmetric unit of the Y370F $P3_221$ crystal form can also be considered structurally identical. The corresponding error-scaled difference distance matrix for the remaining five structurally dissimilar MvNA models (Figure 8.2.1.), very clearly shows that the major differences between the molecules are the conformations of the immunoglobulin 'arm' and the galactose-binding domain, from residues 400 to 647.

	1	2	3	4	5	6	7
	Y370G	Y370F monomer A	Y370F monomer B	Y370F monomer C	D92G Lactose	D92G Dana	Wt
<esd> [Å]	0.11	0.21	0.18	0.20	0.08	0.16	0.13
Y370G		72.5	72.7	74.4	76.5	92.6	92.2
Y370F monomer A			100	91.1	77.4	76.6	75.3
Y370F monomer B				88.1	78.6	77.6	76.3
Y370F monomer C					75.3	75.3	73.9
D92G Lactose						92.5	881
D92G Dana							100
Wt							

Table 8.2.1. The following table contains for every pair of models the percentage of elements in the error-scaled difference distance matrix that are similar than the threshold value of low limit = 5 σ. If for a pair of models this value is larger than 98.0 %, the two models can be considered identical. The first line gives the mean coordinate uncertainties in Angstroms.

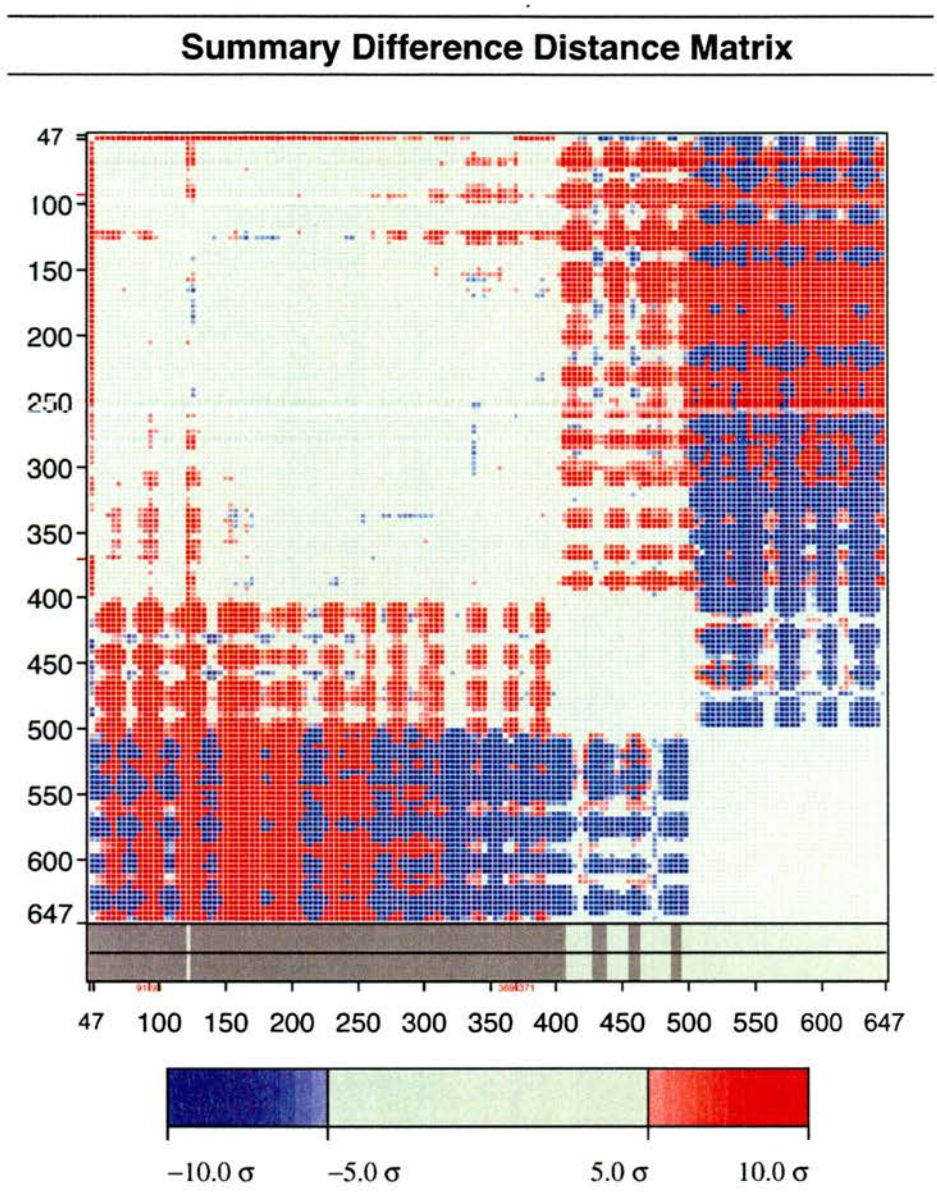


Figure 8.2.1. Summary of the 21 error-scaled difference distance matrices generated for the five structurally dissimilar models of the MvNA sialidase. All changes in C_α-C_α distances smaller than a threshold 5σ are shown in grey; differences between this lower limit and an upper limit of 10σ are shown using a colour ramp where red stands for expansion and blue for contraction. All differences larger than the upper limit are shown in full blue and full red respectively. The two bars at the bottom of the matrix show the parts of the protein that were identified to be conformationally invariant (dark grey) using a tolerance of 5σ.

The analysis of the 21 error-scaled difference distance matrices by the program ESCET identified the following residues belonging to a conformationally invariant region: A48-A91, A93-A121, A125-A369, A371-A407, A427-A438, A486-A494 A454-A643. Mapping this analysis onto the structure of the MvNA sialidase provides a graphical representation of the conformationally invariant and flexible regions of the structure (Figure 8.2.2.).

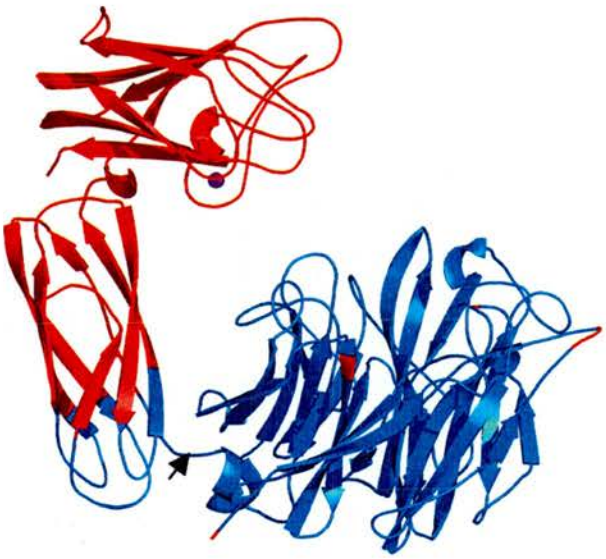


Figure 8.2.2. ESCET analysis showing in blue the conformationally invariant region of the MvNA sialidase between the five models analysed. The region of the protein shown in red identifies the flexible domains, which are the immunoglobulin and galactose-binding domains, respectively. The di-glycine linker is shown using a black arrow. The sodium ion coordinated within the galctose-binding domain is coloured magenta.

It is interesting to see that the flexibility of the immunoglobulin arm only occurs approximately a quarter of the way up, with the bottom quarter closest to the β-propeller domain remaining rigid. Intuitively one would have thought that the whole immunoglobulin domain would be flexible given the di-glycine link to the β-propeller domain. The disulphide bond formed between Cys351 and Cys405 is only ~8 Å from the di-glycine linker and clearly has a strong effect in holding this part of the immunoglobulin arm domain in place. It would be interesting to see the effect of removing this disulphide had on the flexibility of this region of the structure.

The newly identified conformationally invariant region of the structure was used to perform a least-squares superposition of the five different conformers of the MvNA sialidase (Figure 8.2.3.). The superposition showed that the majority of the flexibility lay in the galactose-binding domain, which moved as a rigid-body either towards or away from the active site pocket of the β-propeller domain. The maximum movement of this domain was ~8 Å between the Y370G conformer (orange; Figure 8.2.3.) and the Y370F monomer A conformer (yellow; Figure 8.2.3.), respectively.

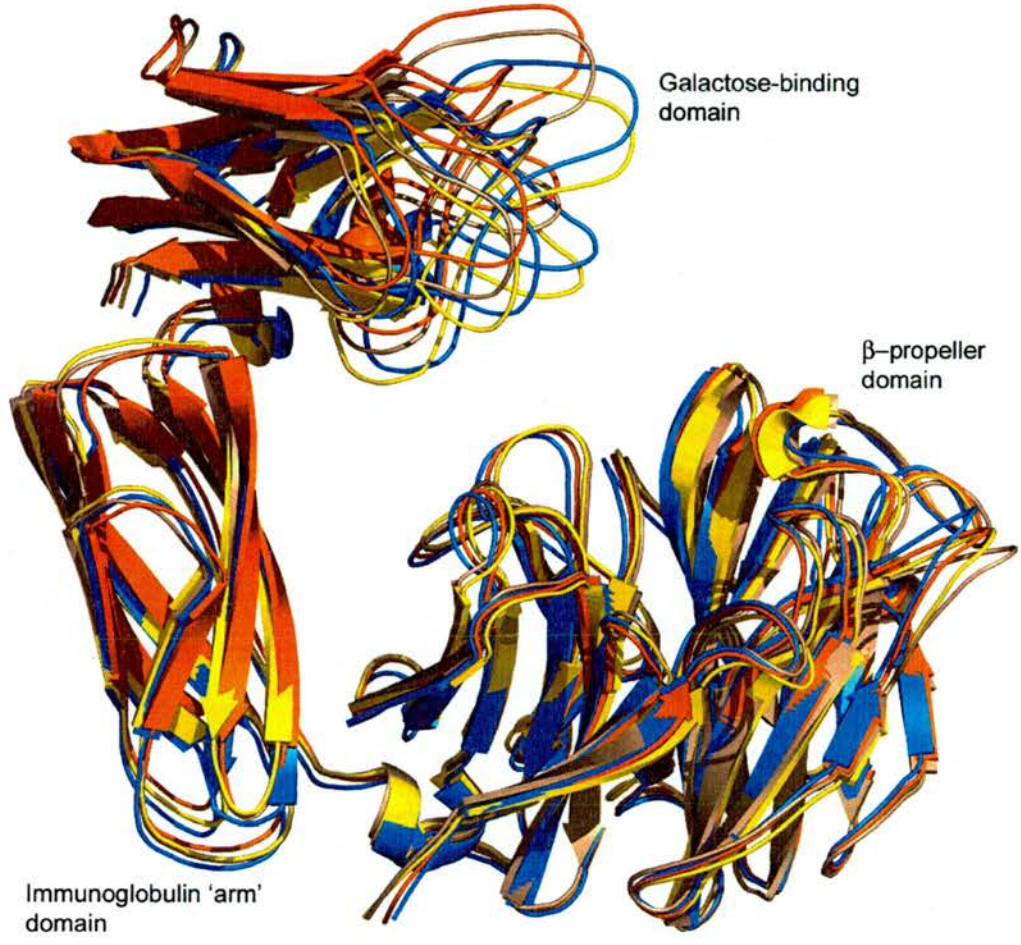
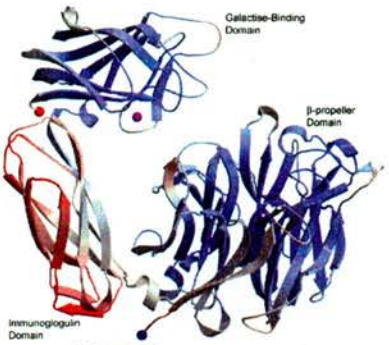


Figure 8.2.3. Least-squares superimposition of four different conformers of MvNA. The four conformers were superimposed using the conformationally invariant β-propeller domain in the program LSQKAB (Kabsch *et al* ., 1976). The four conformers are: Y370G (orange), D92G-Lactose (wheat), Y370F-monomer A (yellow) and Y370F-monomer C (blue). The D92G-Dana structure was omitted for clarity. The wt conformer is identical to the D92G-Dana structure, and the Y370F-monomer B conformer is identical to the Y370F-monomer A structure, as determined using ESCET.

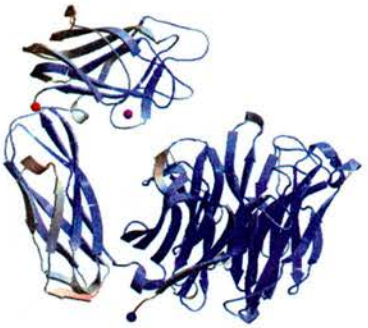
Further evidence that the galactose-binding domain moves either towards or away from the β-propeller domain as a rigid-body comes from analysis of the refined B-factors. B-factors represent the distribution of positions occupied by atoms over time or in different unit cells of the crystal and are reLaced to the atomic displacement parameter u, by the equation $B = 8 \pi^2 \langle u^2 \rangle$. Crystallographers often use B-factors to identify parts of the model that have large coordinate errors and therefore likely to be flexible within the crystal structure. Figures 8.2.4. shows the seven MvNA models coloured according to B-factor. The figures show that on average the B-factors for the galactose-binding domain are lower than for the immunoglobulin domain, indicating that the atoms of the galactose-binding domain have lower coordinate errors and their positions more accurately determined within the crystal structures.

The B-factor analysis correlates well with the results from the ESCET study, which showed that the conformationally invariant region between the different models is the β-propeller domain. This domain also has the lowest coordinate error in each model, indicated by the low B-factors.

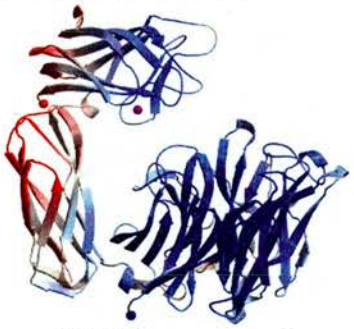
The implications of this result for the function of the galactose-binding domain in the biology of this sialidase are discussed in section 8.5. The following section focuses on the interactions made by the disaccharide lactose to this 'galactose-binding' CBM domain.



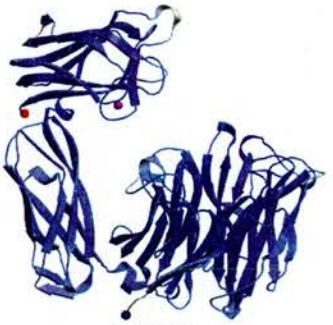
Y370F monomer A
B-factor range (Å^2) 20 - 60



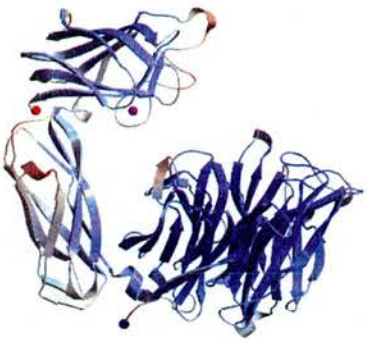
Y370F monomer B
B-factor range (Å^2) 20 - 60



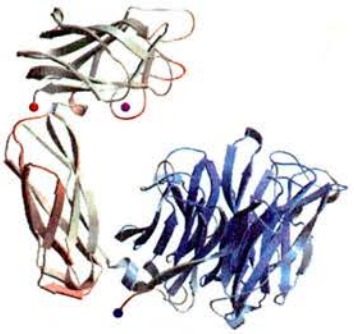
Y370F monomer C
B-factor range (Å^2) 20 - 60



Y370G
B-factor range (Å^2) 5 - 50



D92G-Lactose
B-factor range (Å^2) 5 – 50



D92G-Dana
B-factor range (Å^2) 5 - 50

Wt - α2,3 di-fluoro sialic acid
B-factor range (Å^2) 10 - 60

Figure 8.2.4. The seven MvNA crystal structures are shown coloured according to B-factor, from dark blue to bright red. The B-factor ranges are given under each figure.

8.3. The molecular recognition of lactose by the 'Galactose-Binding' CBM Domain.

The modular construction of the MvNA sialidase, composed of a non-catalytic carbohydrate binding module (CBM) linked to a catalytic glycosidase domain, is a feature seen in other glycoside hydrolases (Boraston *et al* ., 2004). Similar to the catalytic modules of glycoside hydrolases, CBMs are divided into families based on amino acid sequence similarity. There are currently 39 defined families of CBMs (Coutinho and Henrissat, 1999a; Coutinho and Henrissat, 1999b), which display substantial variation in ligand specificity. The galactose-binding domain of MvNA has been assigned to CBM family 32. A further level of classification has been applied to the CBM families based on structural and functional similarity. Using this criteria the CBM families are grouped into three types: 'surface-binding' CBMs (Type A), 'glycan-chain binding' (Type B), and 'small-sugar binding' CBMs (Type C). Clearly the MvNA CBM32 falls into this latter class of Type C modules.

The role of the CBMs is mainly to target the enzymes to specific structural polysaccharides and enhance the catalytic efficiency by increasing the effective concentration of the enzyme on the surface of the substrates (Bolam *et al* ., 1998; Gill *et al* ., 1999). The bacterial sialidases from *C.perfringens* and *V.cholerae*, like the MvNA sialidase, contain a catalytic domain together with one or more CBMs. A recent kinetic analysis showed the ability of these enzymes to hydrolyse polyvalent substrates with much greater efficiency than their monovalent counterparts (Thobhani *et al* ., 2002). Confirming a similar role in both substrate targeting and catalytic efficiency for the CBM domains in the sialidase superfamily.

The crystal structure interpretation of the C-terminal domain of the MvNA sialidase (residues 503-647) was originally aided by using the N-terminal domain of a fungal galactose oxidase from *Dactylium dendroides*, with which it shares 35.4 % sequence identity (PDB 1GOF) (Gaskell *et al* ., 1995; Ito *et al* ., 1994). The two domains were found to be very similar in structure and superimpose with

an r.m.s.d. of 1.16 Å for 139 C_α atoms. The topology is that of a β-sandwich with a five-stranded antiparallel β-sheet packed on top of a three-stranded antiparallel β-sheet, and can be described as a jelly-roll (Figure 8.3.1). The structure of the C-terminal domain of MvNA was also found to bind the monosaccharide galactose (Gal) through a soaking experiment, with the similar residues involved in binding the Gal in both domains. The diffraction data for the MvNA-Gal complex however, was relatively weak, being only 62 % complete and extending to only 2.5 Å resolution. Gal was modelled into the difference density, but a full analysis of the binding interactions could not be made due to absence of solvent modelling in the refinement and the ambiguity in the exact location of the hydroxyl groups in the sugar.

The serendipitous trapping of the disaccharide lactose (Lac) in one of the D92G mutants (D92G-Lactose; PDB 1W8O) at 1.7 Å resolution has now allowed the protein-carbohydrate interactions mediated by this domain to be analysed in detail. The structure of the D92G-Neu5Ac2en complex (PDB 1W8N), refined at 2.1 Å resolution, also had a peak of difference density in this domain, into which Gal was modelled. However, the refined average B-factor for the Gal was very high, 89 $Å^2$, indicating a high degree of disorder for this ligand. The refined average B-factor for the Lac in comparison was 44 $Å^2$. Also, the difference density was insufficient to accurately determine the positions of the hydroxyl groups of the Gal in the D92G-Neu5Ac2en complex. However, refinement of the Gal modelled in the same orientation as the Gal unit of the Lac complex, resulted in no negative peaks in the difference density maps. Refinement of the Gal in other orientations did produce negative peaks for the hydroxyls, indicating that this orientation was likely to be correct. The high coordinate uncertainty associated with this ligand however, introduces too much error to make any analysis meaningful.

The high B-factors for these ligands indicate weak binding of the two sugars to the C-terminal domain. This assumption is borne out by the analysis of the protein-carbohydrate interactions made in the Lac complex. The analysis of the Lac complex is presented in section 8.3.1.

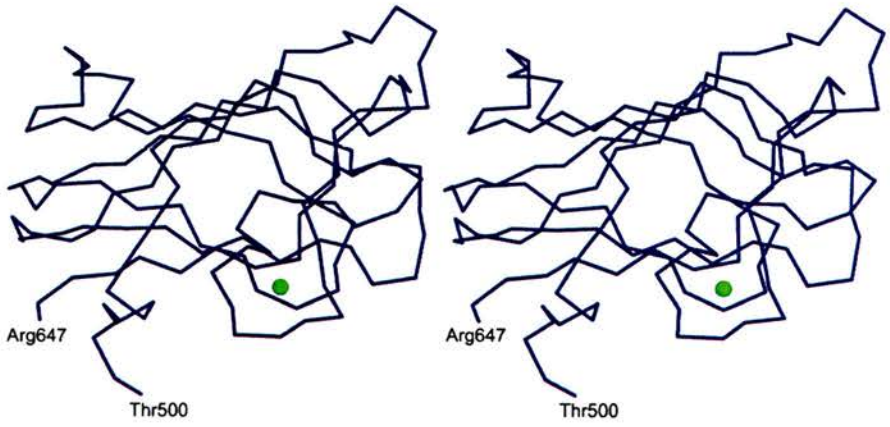
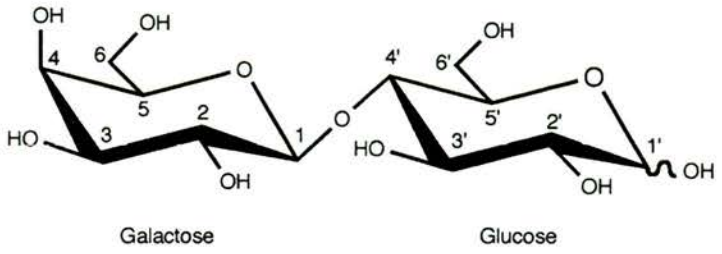


Figure 8.3.1. Stereo view C_α trace of the C-terminal galactose-binding domain (residues 500-647) of the MvNA sialiadase. The metal ion (presumed sodium) is shown as a green sphere, stabilising the loop region in which it is coordinated.

8.3.1. Structural characterisation of lactose binding to the galactose-binding domain of MvNA.

The structure of the D92G mutant of MvNA in complex with lactose was determined to 1.7 Å resolution (chapter 7). Analysis of the electron density maps of the D92G-Apo complex resulted in the identification of a large positive peak in the difference density map (F_o-F_c) at the site originally identified as binding Gal. The difference density was sufficiently unambiguous to allow the accurate placement of the disaccharide α-lactose (**1**) (Lac; α-D-galactopyranosyl-(1→4)-β-D-glucopyranose) into the binding site of the C-terminal galactose-binding domain (Figure 8.3.2.). The direction of the sugar was determined from the large positive peak for the O4 hydroxyl of galactose in the axial position. This results in the reducing end of the sugar pointing out into the solvent. The hydrogen bond interactions formed between the protein residues and the hydroxyl groups of the sugar units are summarised in Table 8.3.1. and Figure 8.3.3.

(a)



(b)

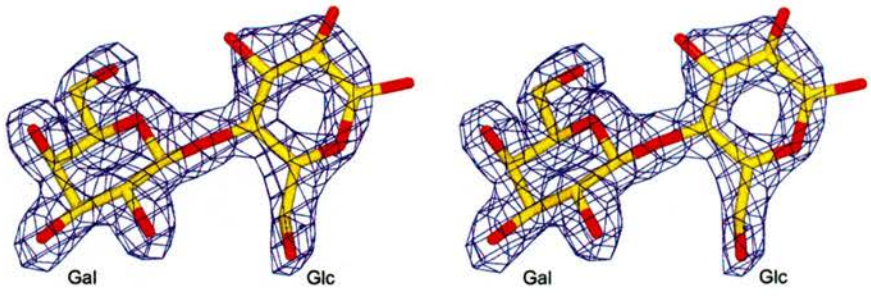


Figure 8.3.2. (a) Chemical structure of (**1**), α-Lactose (Lac; α-D-galactopyranosyl-(1→4)-β-D-glucopyranose). (b) Stereo view of the electron density map ($2F_o$-F_c), contoured at 2 σ, for α-Lactose bound to the C-terminal galactose-bininding CBM of MvNA.

The sugar makes only three direct hydrogen bond interactions with the protein and these are all at the non-reducing end of the sugar. The O4 hydroxyl makes a strong hydrogen bond (2.66 Å) with the NE2 atom of His539. The other two protein interactions are made by the O3 hydroxyl, to both the carboxylate group of Glu522 and the amide group of Arg572. The remaining hydrogen bond interactions are made to a constellation of nine water molecules located within the binding site. Figure 8.3.2. illustrates the intricate hydrogen bond network mediated by these water molecules. Ser575 in particular plays a crucial role in linking this hydrogen bond network to the glucose unit of the disaccharide.

The other structural feature of this binding site is the location of Trp542 sitting above the galactose sugar ring. The aromatic ring of Trp542 appears to form a stacking interaction with the sugar ring and is located ~ 3.5 Å from the C4 atom. This type of stacking interaction is ubiquitous in CBM protein-carbohydrate

interactions, where the hydrophobic pLacform can be either planar, twisted or form a sandwich (Boraston *et al* ., 2004).

The binding site topography and ligand conformation strongly suggest that this CBM is specific for galactose. The binding pocket appears to only accommodate the axial position of O4, with the strong hydrogen bond to His539. Modelling of the O4 hydroxyl to the equatorial position, which would create glucose, shortens the distance to the NE2 atom of His539 to ~2.06 Å. This would create an unfavourable electrostatic repulsion and is unlikely to be accommodated in the binding site. The conformation of the bound Lac also appears to mirror the solution conformation of the sugar, having the effect of minimising any energetic penalty paid upon binding.

At first glance is appears that the binding of Lac to this domain should be fairly weak, being mostly mediated by water molecules, which carry a large entropic penalty for ordering. A recent biochemical investigation carried out on this domain calculated the dissociation constant (K_D) for Lac as 1.3 mM using isothermal titration calorimetry (Boraston *et al* ., 2003). Comparison with the uncomplexed structure shows that only five of the water molecules are ordered upon Lac binding, with the other waters already present within the binding site (Table 8.3.1). This still leaves the entropic penalty paid for ordering the five water molecules clustered beneath the glucose sugar unit in Figure 8.3.2. Several studies have indicated that water rearrangement plays an important role in the thermodynamics of carbohydrate-protein interactions (Clarke *et al* ., 2001; Ladbury and Chowdhry, 1996). Clearly, the entropic penalty paid for ordering these five water molecules are offset by the enthalpy gain on forming a greater number of hydrogen bonds in the complex than could have been achieved without the water molecules. It has been proposed that the hydrogen bonds in these types of carbohydrate-protein interactions have higher occupancies than in bulk solution, which could contribute to stabilising the complex (Clarke *et al* ., 2001).

The comparison with the uncomplexed CBM also shows that no significant rearrangement of the binding site takes place upon Lac binding, the r.m.s.d. was 0.56 Å for the 147 equivalent C_α positions.

The individually measured K_D value for the binding of both galactose and lactose to the CBM domain of the MvNA sialidase may far underestimate the true strength of substrate binding in *vivo*. In Nature weak interactions between protein and carbohydrates are often compensated for by the phenomenon of avidity resulting from multivalent interactions (Boraston *et al* ., 2004; Liang *et al* ., 1997). In these cases, multiple clustered carbohydrate-binding sites interact simultaneously with the carbohydrate ligands, which present multiple recognition elements. This type of interaction results in increased association constants relative to any one of the isoLaced carbohydrate-binding sites (Kiessling *et al* ., 2000). A recent study investigating the role of multivalency in the mode of action of bacterial sialidases concluded that the extra CBM domains associated with the catalytic β-propeller have dramatic effects on the K_m values for polyvalent substrates compared with their monovalent counterparts (Thobhani *et al* ., 2002). It appears that the extra CBM domains, associated with bacterial sialidases have increased the catalytic efficiency of these enzymes by enabling them to bind tighter and more specifically to polyvalent sialic acid containing glycoconjugates.

Table 8.3.1. Table of hydrogen bonding interactions between α-Lactose and the C-terminal CBM of MvNA.

Sugar atom	Protein/Water atom	Distance (Å) Lactose
O2	H_2O^*	2.61
	H_2O^*	2.74
O3	Glu522-$O^{\varepsilon 2}$	2.49
	Arg572-$N^{\eta 1}$	2.97
	H_2O^*	3.09
O4	His539-$N^{\varepsilon 2}$	2.73
	H_2O^*	2.66
	H_2O	2.60
O3'	H_2O	2.75
O5'	H_2O	3.06
	H_2O	3.01
O6'	H_2O	2.37

* Denotes water molecules already present in the uncomplexed binding site of C-terminal CBM in MvNA.

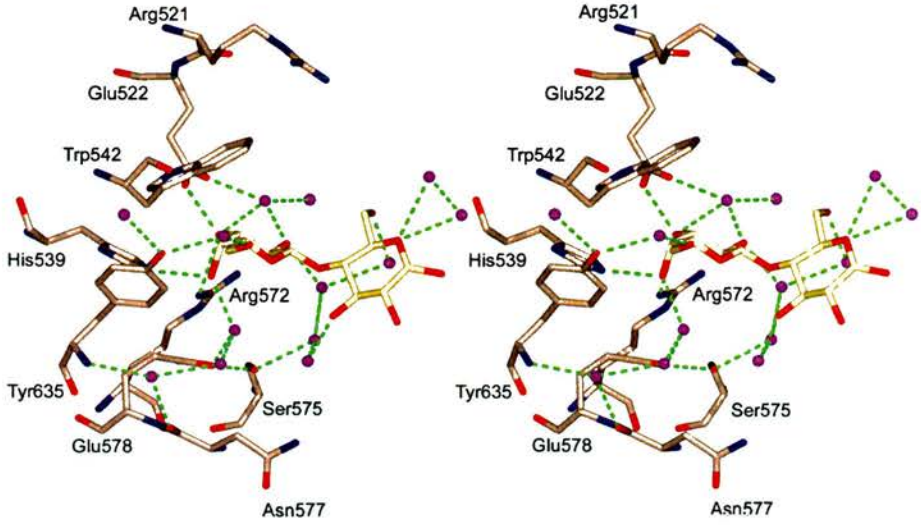


Figure 8.3.3. Stereo view of the C-terminal CBM domain of MvNA in complex with α-Lactose. The hydrogen bonding interactions are drawn as dotted lines; water molecules are represented as magenta spheres.

8.3.2. Comparison between the binding of lactose and galactose to the C-terminal CBM of MvNA.

The availability of two crystal structures of the MvNA C-terminal CBM in complex with both the monosaccharide galactose (Gal) and the disaccharide lactose (Lac) allows a useful comparison to be made. The orginal wt structure (PDB code 1EUU) of the MvNA sialidase was refined with Gal bound to the C-terminal CBM, from which it derived its name as a 'galactose-binding domain'. However, the low completeness of the data allowed only a crude model of this interaction to be made. A superposition of this structure with the 1.7 Å resolution Lac complex presented here, shows that the position of the Gal in the wt structure is significantly different to the Gal unit of Lac (Figure 8.3.4). This could indicate one of two possibilities. The first is that Gal and Lac bind differently to the CBM, the second, that the position of the Gal in the wt structure was modelled incorrectly.

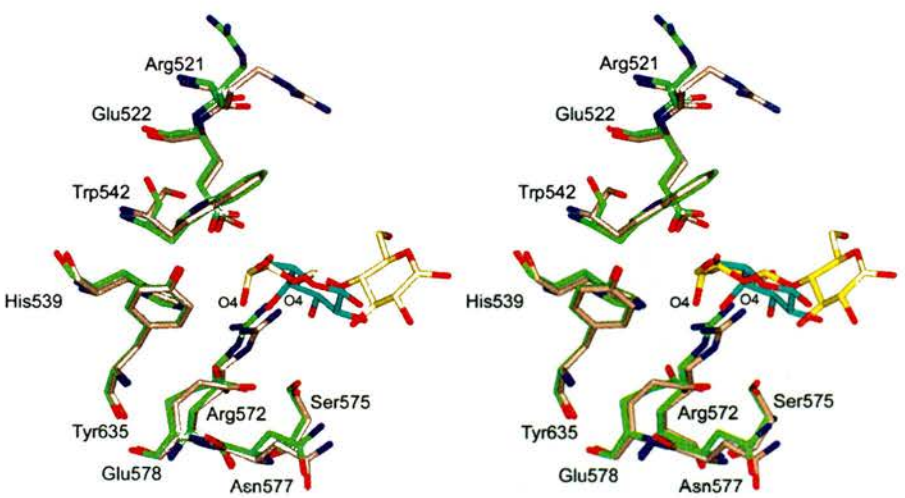


Figure 8.3.4. Superimposition of the C-terminal CBM of MvNA from the wt (PDB code 1EUU) structure (green) and the D92G structure (PDB code 1W8O) structure (wheat). The Gal, present in the wt structure is shown in cyan. The Lac, present in the D92G structure is shown in yellow. The O4 hydroxyls of both ligands are labelled for clarity.

Table 8.3.2. Comparison of hydrogen bonding interactions between the bound ligands, Galactose and Lactose complexed with MvNA.

Sugar atom	Protein/Water atom	Distance (Å)	
		Galactose	Lactose
O2	H_2O^*	----	2.61
	H_2O^*	----	2.74
O3	Glu522-$O^{\varepsilon 2}$	----	2.49
	Arg572-$N^{\eta 1}$	----	2.97
	Arg572-$N^{\eta 2}$	3.18	----
	Ser575-O^{γ}	3.17	----
	H_2O^*	----	3.09
O4	His539-$N^{\varepsilon 2}$	3.88	2.73
	Ser575-O^{γ}	4.16	----
	Arg572-$N^{\eta 1}$	3.40	----
	Glu578-$O^{\varepsilon 1}$	3.70	----
	H_2O^*	----	2.66
	H_2O	----	2.60
O6	H_2O^*	----	----
O3'	H_2O	----	2.75
O5'	H_2O	----	3.06
	H_2O	----	3.01
O6'	H_2O	----	2.37

* Denotes water molecules already present in the uncomplexed binding site of C-terminal CBM in MvNA. No waters were modelled in the wt-Gal complex (PDB code 1EUU).

The major interactions between the galactose and the wt CBM binding site are: hydrogen bonding of the O3 and O4 hydroxyls of galactose to Arg572; hydrogen bonding of O4 to Oε1 of Glu578, and a weak interaction between O4 and Nε2 of His539; a weak interaction with Oγ of Ser575; and finally, van der Waals interactions between C6 and Trp542. These interactions are significantly different to those observed for Gal unit in the lactose complex, listed in Table 8.3.2.

The major differences between the two complexes are the alternate positioning of Glu578 and the position of Gal further away from His539. In the wt-Gal complex the carboxyl group of Glu578 makes hydrogen bonding interactions to both the O4 hydroxyl of Gal and the $O\gamma$ of Ser575. In the D92G-Lac complex, Glu578 is held tightly by the $O\gamma$ of Ser575 and makes no direct interactions with the sugar (see Table 8.3.2 and Figure 8.3.4).

The exact nature of the ligand for the CBM of MvNA is currently unknown. In comparing the two complexes, it becomes clear that the D92G-Lac complex suggests the target ligand for this domain is the non-reducing end of galactose, bound as either the monosaccharide or as part of a larger sugar. The location of the non-reducing end of the Gal unit so close to the $N\varepsilon 2$ atom of His539, makes the accommodation of any configuration for the O4 hydroxyl other than axial impossible. The position of the Gal sugar in the wt-Gal complex by contrast could accommodate either configuration at C4. The dissociation constants (K_D) have been published for both the Gal and Lac sugars to the CBM of MvNA. The published values were 0.9 (± 0.1) mM and 1.0 (± 0.0) mM respectively by UV difference spectroscopy, and 0.5 (± 0.1) mM and 1.3 (± 0.0) mM respectively by ITC (Boraston *et al* ., 2003). The K_D values determined by the UV difference method suggest that the binding of both Gal and Lac are of equal strength. Whereas, the values determined by ITC suggest that the binding of Gal is twice as strong as that of Lac. If the values for the K_D were similar, it could be argued that the binding would also be similar and the interactions observed in the more accurate D92G-Lac complex would represent the true binding interactions made to both sugars. If the values for the K_D are different however, it could also be argued that this suggests a different mode of binding for both ligands.

The structural data presented here illustrates how the disaccharide Lac interacts with this type C CBM, and provides evidence for the specificity of this domain for Gal, with O4 in the axial position. It is interesting that no direct hydrogen bonds are made from the protein to the glucose sugar of Lac. The only interaction with the CBM is via the non-reducing end of the Gal sugar. A higher resolution structure with the monosaccharide galactose is needed however, to

determine if the interactions made to these sugars by the CBM are similar or different.

8.4. Structural analysis of the catalytic mutants of MvNA, the bacterial sialidase from *M.viridifaciens*.

Sialidases and trans-sialidases contain seven strictly conserved amino acids, namely: an arginine triad that binds the carboxylate group of the substrate via electrostatic interactions; a distal glutamate residue that forms a salt bridge with one of the conserved arginine residues; and three residues that are important for catalysis, a tyrosine-glutamic acid dyad and an aspartic acid. These residues were discussed in detail for the *C.perfringens nanI* catalytic domain in chapters 4 and 5. The effects of mutating these residues on the structural, kinetic and biochemical characteristics of bacterial sialidases have not been investigated previously in the literature. The aim of this collaboration with Prof. A. Bennet's group at Simon Fraser University in Vancouver, Canada, was to undertake such a study using the 68 kDa form of the sialidase from the soil bacterium *M.viridifaciens* (MvNA), for which the wt structure was already determined (Gaskell *et al* ., 1995).

The three mutations for which crystal structures were obtained are those of the conserved catalytic nucleophile Tyr370 to phenylalanine (Y370F) and glycine (Y370G), and the conserved acid/base catalyst Asp92 to glycine (D92G). All known natural sialidases catalyse the hydrolysis of substrates with retention of anomeric configuration (Davies *et al* ., 1998), an observation that is compatible with this family of enzymes operating via a standard double-displacement mechanism (Amaya *et al* ., 2004; Davies *et al* ., 1998; Vasella *et al* ., 2002). Recently, our collaborators showed that three separate tyrosine mutants of MvNA are inverting sialidases (Watson *et al* ., 2003). Such an observation is evidence for the conserved tyrosine residue being the catalytic nucleophile in bacterial sialidases and that water acts as the nucleophile in the inverting mutant enzymes by occupying the 'hole' created when smaller amino acid side chains

replace the tyrosine. Two tyrosine mutants, Y370G and Y370F, were investigated to study the effects these mutations had on the structural integrity of the active site and on the interactions made to both substrates and products. The structures of these mutants in complex with the product β-Neu5Ac (sialic acid) and Neu5Ac2en (DANA) are presented below in sections 8.4.2 and 8.4.3 for the Y370G and Y370F mutants, respectively.

The proposed catalytic role of the conserved aspartic acid, D92, is as a general acid / base catalyst in the reaction mechanism (Varghese and Colman, 1991), based on studies of the influenza virus sialidase. The results from previous mutagenic studies on this aspartic acid residue led to the conclusion that it is important for catalysis, but its specific energetic contribution to the catalysis was not determined (Chien *et al* ., 1996; Ghate and Air, 1998; Kleineidam *et al* ., 2001). A series of kinetic and product studies were performed on the D92G mutant of MvNA to answer this question. In addition, in order to understand any structural implications this mutation might have on the integrity of the active site, the crystal structure of the D92G mutant in the apo form and complexed with the inhibitor Neu5Ac2en (DANA), were studied (Watson *et al* ., 2004).

8.4.1. Contribution of the active site aspartic acid to catalysis in MvNA.

Crystal structures of the D92G mutant sialidase from *M.viridifaciens* (MvNA) were determined in the apo form and complexed with the transition state analogue, Neu5Ac2en (DANA), at 1.7 and 2.1 Å resolution, respectively; see chapter 7. Kinetic and product studies revealed that the replacement of the conserved aspartic acid with glycine results in a catalytically competent retaining sialidase, that possess significant activity against activated substrates (Watson *et al* ., 2004). The conclusion from these results was that the D92G mutant sialidase was operating via a double displacement mechanism, in a similar manner to that already discussed for the *C.perfringens nanI* sialidase (chapters 5 and 6). The striking finding from this study was that the D92G mutant was

remarkably active, with a k_{cat}/K_m (the specificity constant (Silverman, 2002)) value 82 % of that measured for the wt enzyme using the fluorogenic substrate MU-αNeu5Ac.

The energetic contribution of the aspartic acid residue to catalysis clearly needed further investigation. The data, summarised in Table 8.4.1, shows that a pronounced change in relative catalytic activity occurs between the wt and D92G mutant sialidases on altering the leaving group from 4-Methylumbelliferone or 4-Nitrophenol, in the activated substrates, to lactose one of the natural aglycons.

Table 8.4.1. Relative catalytic activity of the wild-type and D92G mutant sialidases. Taken from Watson *et al* . (Watson *et al* ., 2004).

Leaving group	Relative k_{cat}[a]	Relative k_{cat}/K_m[b]
4-Nitrophenol	7.8	1.5
4-Methylumbelliferone	4.4	1.2
3'-Lactose	1.5×10^3	1.6×10^3
6'-Lactose	2.3×10^3	2.6×10^3

[a]Equals k_{cat}(wild-type)/k_{cat}(D92G).
[b]Equals k_{cat}/K_m(wild-type)/k_{cat}/K_m(D92G).

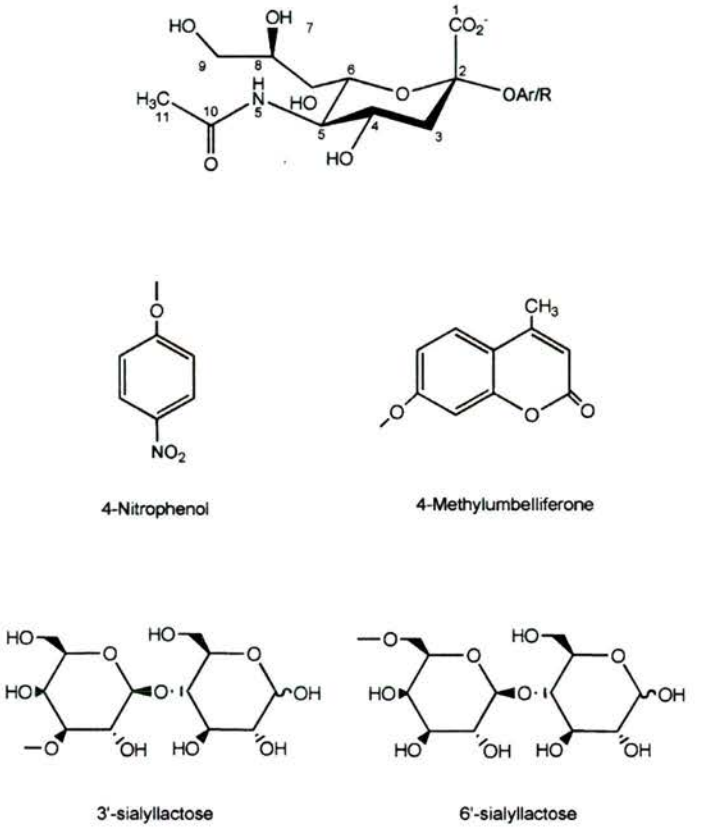


Figure 8.4.1. Chemical structures of the substrates used in the kinetic analysis of the MvNA mutants.

The effect of leaving group ability on k_{cat} (the catalytic rate constant (Silverman, 2002)) and k_{cat}/K_m for the D92G mutant were analysed using Brønsted plots (Watson *et al* ., 2004). A Brønsted plot is a plot of the log of k_{cat} (or K_2) against the pK_a of the leaving group (Fersht, 1985). The important parameter that is derived from the Brønsted plot is the β_{lg} value. The sign and magnitude of this value are an indication of the charge developed in the transition state. The calculated β_{lg} were –0.37 ± 0.02 and –0.72 ± 0.03 for the kinetic terms k_{cat} and k_{cat}/K_m, respectively for the D92G mutant (Watson *et al* ., 2004). For the wt sialidase catalysed reaction, k_{cat} was found to be limited by a step, or steps, which occur after cleavage of the glycosidic bond (Watson *et al* ., 2003). A similar finding has been reported for the type A influenza virus sialidase (Chong *et al* ., 1992). The identity of this rate-limiting step(s) in the wt catalysed reaction

could be the formation of a covalent sialyl-enzyme intermediate, similar to that seen in both the trans-sialidase from *T.cruzi* (Amaya *et al* ., 2004) and the large sialidase *nanI* from *C.perfringens,* with the di-fluro sialic acid derivative (chapter 5). The results of the Brønsted analysis show that substitution of the conserved aspartic acid by a glycine residue results in glycosidic bond cleavage becoming partially rate-limiting for k_{cat}. In other words, the first-order rate constants for the reaction of the Michaelis complex (k_{cat}) are influenced by the leaving group ability (β_{lg} = –0.37). As a result, it was concluded that the transition state for the D92G mutant catalysed cleavage of the glycosidic bond in the natural substrate analogues, 3'-siayllactose and 6'-siayllactose, is at least 19 kJ/mol higher in energy ($\Delta\Delta G^{\ddagger}$ at 37 °C) than the corresponding transition state for the wt sialidase reaction. A similar conclusion was also made for the kinetic parameter k_{cat}/K_m, where the β_{lg} value increases from –0.30 in the wt to –0.72 in the D92G mutant. The conserved aspartic acid residue therefore displays a greater catalytic effect when the substrate possesses a 'bad' leaving group, such as lactose. It is expected that a similar effect would be observed with all natural sialosides, including glycolipids and glycoproteins, given that the terminal residue in the aglycon is always a carbohydrate unit.

The introduction of a glycine residue in place a normally larger amino acid, such as aspartic acid, could have the effect of introducing flexibility into the protein structure that could have implications for analysing the subsequent kinetic analysis. In order to investigate if this were the case with the D92G mutant, the crystal structure was determined (chapter 7). The structure shows that the substitution caused no significant perturbation of the active site (Figure 8.4.2). The r.m.s.d. between the wt structure (PDB code 1EUS) and the D92G mutant (PDB code 1W8O) was 0.43 Å for C_α 47 – 402, which make up the catalytic β-propeller domain. In particular, no movement in the loop that contains the D92G mutation is observed. This confirmed that the kinetic effects of the mutation were a result of only losing this chemical group from the active site and not the result of some gross conformation rearrangement.

Interestingly, the MvNA sialidase is one of only two bacterial sialidases where this aspartic acid forms a direct hydrogen bond interaction to the O4 of the ligand (2.7 Å), the other one being the sialidase from *V.cholerae*, whose equivalent aspartic acid is Asp62 (PDB code 1KIT). In the wt MvNA structure, Asp92 also interacts with NH1 and NH2 of Arg87 (2.7 and 2.8 Å). In the D92G-Neu5Ac2en complex (PDB code 1W8N), there is no change in the position of the ligand in comparison with the wt catalytic domain structure (PDB code 1EUS). However, as can be seen in Figure 8.4.2. in the D92G structure, a water molecule sits in the equivalent position of the carboxylate group in the wt structure, making a hydrogen bonding interaction to the O4 hydroxyl of DANA. This observation suggests that such a water molecule could still act in the catalytic mechanism of this mutant, but would remain inactivated due to the absence of the aspartic acid. This would explain the shift in the rate-determining step that is observed in this mutant (Watson *et al* ., 2004).

The structural similarity between the D92G mutant and the wt enzyme, combined with the kinetic data from our collaborators suggest that the active site of bacterial sialidases is very rigid. A similar observation was made in the analysis of the *C.perfringens nanI* structure and fits with the observation that nearly all glycosidases have a rigid catalytic scaffold, analogous to the 'lock' in the lock and key metaphor of enzyme catalysis. This study also lends evidence to the idea that the rate-limiting step for the wt sialidase, which is presumed to involve a conformational change, occurs in the substrate rather than in the catalytic site. This idea also correlates well with the observed substrate distortion seen in the *nanI* structure and in the trans-sialidase from *T.cruzi* (Amaya *et al* ., 2004).

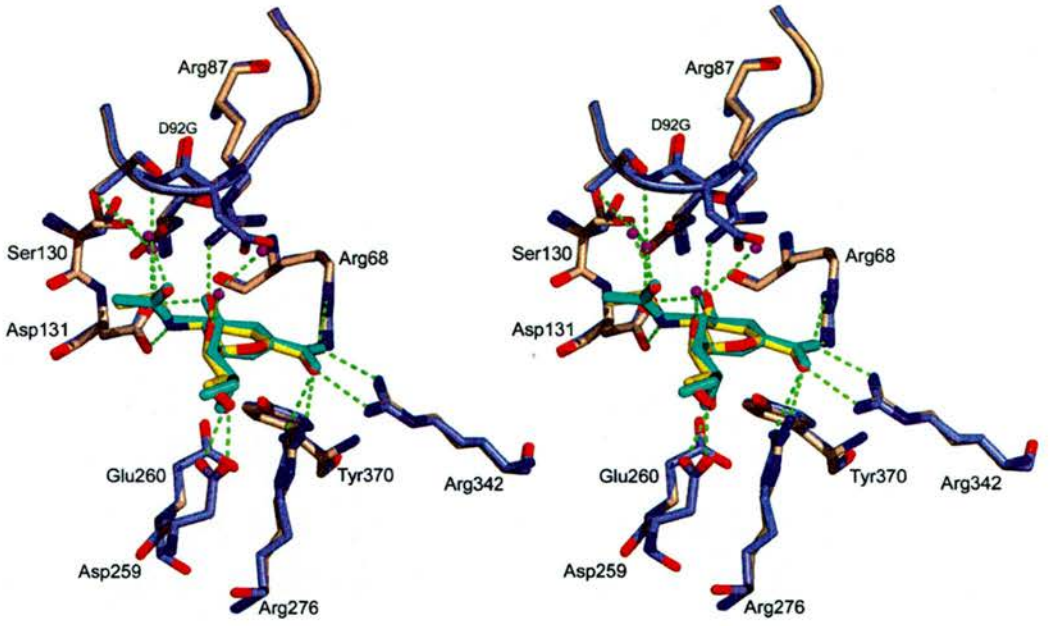


Figure 8.4.2. Stereoview of the superimposed active site residues of the wild-type catalytic domain with Neu5Ac2en (PDB code 1EUS) and the D92G Neu5Ac2en complex (PDB code 1W8N). Residues of the wild-type structure are shown in slate and the mutant in wheat. The hydrogen bonding interactions are drawn as green dotted lines. The ligand, Neu5Ac2en is shown in yellow from the mutant complex. The ligand from the wild-type complex is shown in cyan. The loop containing the D92G mutation is shown as a C_α trace from residues 89-91, with the mutation labelled.

8.4.2. Mutagenesis of the conserved active-site tyrosine changes a retaining sialidase into an inverting sialidase.

In a series of complementary kinetic and product studies on three tyrosine mutants of MvNA, our collaborators found that they could change the mechanism of this enzyme from one of retention of anomeric configuration to one of inversion. In this new mechanism, water efficiently functions as the nucleophile (Watson *et al* ., 2003).

Three mutants, Y370A, Y370D and Y370G, were produced recombinantly in *E.coli*, and all were found to be catalytically active against the activated substrate 4-methylumbelliferone α-D-N-acetylneuraminide (MU-αNeu5Ac; Figure

8.4.1) Table 8.4.2. The Y370D mutant was also shown to catalyse the hydrolysis of natural substrate analogues, such as 3'-sialyllactose (Watson *et al* ., 2003). In order to probe the structural details of this new active site, the structure of the Y370G mutant was solved in complex with the reaction product β-Neu5Ac (sialic acid). The mutant crystallised in the same space group as the original wt structure, $P2_1$, but with different unit cell dimensions and diffraction data collected to 1.8 Å resolution (chapter 7). The difference density was unambiguous for the β-sialic acid ligand, Figure 8.4.3.

Table 8.4.2. Kinetic parameters of wild-type MvNA sialidase and Y370 variants using MU-αNeu5Ac at pH 5.25 and 37 °C. Taken from (Watson *et al* ., 2003).

	Relative k_{cat}[a]	Relative k_{cat}/K_m[b]
Wild-type	100 ± 5	100 ± 19
Y370A	2.9 ± 0.2	0.6 ± 0.16
Y370D	28 ± 2	1.4 ± 0.3
Y370G	86 ± 6	18 ± 4

[a] Absolute k_{cat} value for the wild-type is 51.5 ± 2.7 s^{-1}.
[b] Absolute k_{cat}/K_m for wild-type is (7.2 ± 1.4) × 10^6 M^{-1} s^{-1}.

The overall structure of the Y370G active site is very similar to the previously reported wt structure (PDB code 1EUS), with an r.m.s.d. of only 0.36 Å for the 359 C_α that make up the β-propeller domain. However, removal of the tyrosine residue creates two very significant changes to the active site. The first is the creation of a large cavity directly beneath the O2 atom of the β-sialic acid, in which three water molecules are seen forming hydrogen-bond interactions with each other and with protein residues. The second is the movement of the side chain of Asn310 to make a strong hydrogen bond to one of these waters (2.70 Å). Figures 8.4.4 and 8.4.5. illustrate the 'hole' created by the Y370G mutation and the hydrogen bonding interactions in this area of the active site, respectively.

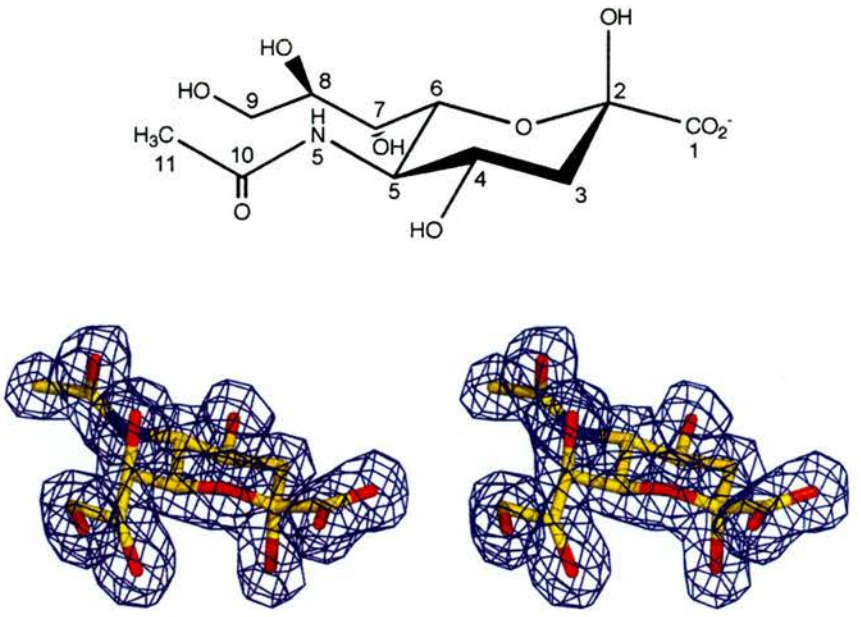


Figure 8.4.3. (a) Chemical structure of β-Neu5Ac (sialic acid). (b) Stereo view showing the refined $2F_o$-F_c electron density map for the β-sialic acid ligand, contoured at 2 σ.

Our collaborators concluded from the results of their biochemical analysis on the Y370D mutant, that hydrolysis occurs via a dissociative mechanism (S_N1) to give a transient oxocarbonium ion transition state (TS), that is rapidly trapped by a water molecule from the top face of the sugar ring. This is in contrast to the wt enzyme, which was shown to act via a concerted mechanism (Watson *et al* ., 2003). The structure of the Y370G complex neatly demonstrates the source of the attacking water from the 'hole' generated by the absence of the phenol ring of Tyr370 (Figure 8.4.4). Furthermore, analysis of the hydrogen bond interactions, (Table 8.4.3) show the location of the conserved active site general base (Glu260) only 2.71 Å from O2 and 2.7 Å from one of these water molecules. Thus, it appears as though the oxocarbonium ion could be stabilised through an interaction with the carboxylate group of this residue. By forming this strong

hydrogen bond to the water molecule, this residue could also serve to hold a prospective water molecule in a good position to attack the C2 atom of the TS to give the inverted product.

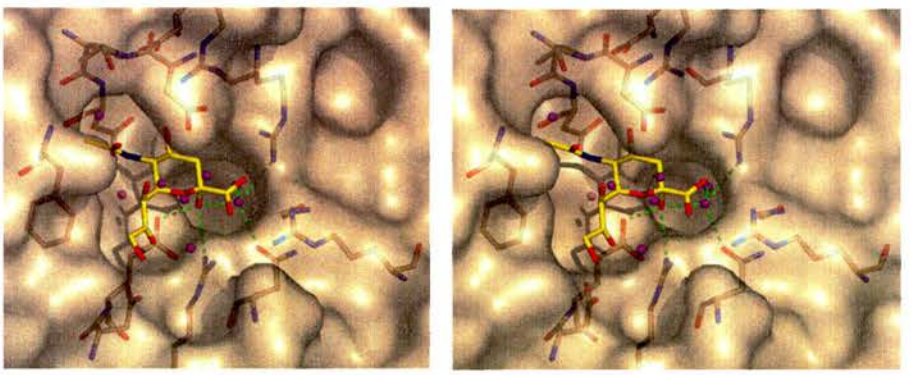


Figure 8.4.4. Stereo view of the active site of the MvNA Y370G mutant. The view is looking down onto the ligand, β-Neu5Ac, shown in yellow. The molecular surface is shown in half transparency. The 'hole' generated by the mutation can be seen directly beneath the O2 atom of the ligand. The water molecules are shown as magenta coloured spheres, with selected hydrogen-bonds as dashed green lines.

Table 8.4.3. Hydrogen bonding interactions between the bound ligand, β-sialic acid and the active site of the Y370G MvNA mutant.

β-Sialic acid	Protein/Water atom	Distance (Å)
O1A	Arg276-$N^{\eta 1}$	3.20
	Arg276-$N^{\eta 2}$	2.99
	Arg342-$N^{\eta 2}$	2.94
	H_2O	3.11
O1B	Arg342-$N^{\eta 1}$	3.07
	Arg68-$N^{\eta 1}$	2.98
	Arg266-$N^{\eta 2}$	3.02
O2	Arg276-$N^{\eta 1}$	3.05
	H_2O	2.70
	Glu260- $O^{\varepsilon 2}$	2.71
O4	Arg87-$N^{\eta 2}$	2.77
	Asp131-$O^{\delta 1}$	2.71
O6	H_2O	2.72
O7	H_2O	2.66
	H_2O	3.17
O8	Asp259-$O^{\delta 1}$	2.66
O9	H_2O	2.86
	Asp259-$O^{\delta 2}$	2.53
O10	H_2O	2.74
N5	Asp131-$O^{\delta 2}$	2.76

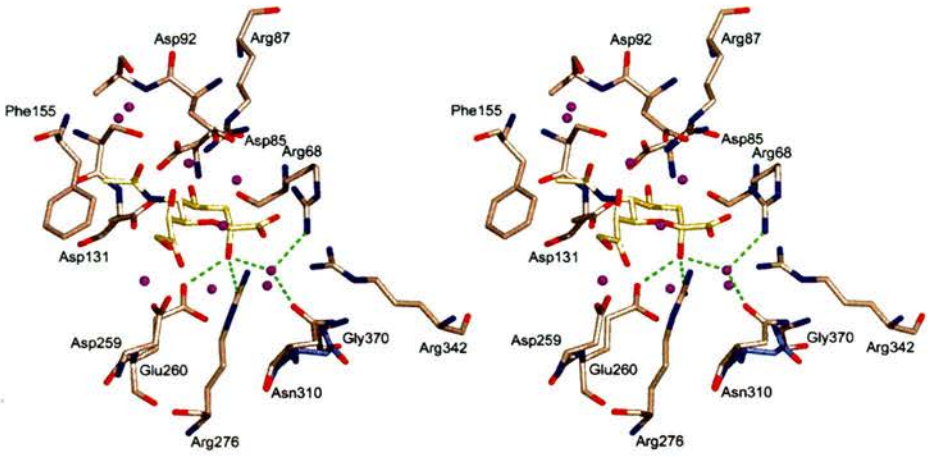


Figure 8.4.5. Stereo view of the active site of the MvNA Y370G mutant. The mutant active site is shown in wheat, the position of Asn310 in the wt structure (PDB code 1EUS) is shown in slate. The selected hydrogen bonds to the O2 atom of the ligand, β-Neu5Ac, are shown as dashed green lines. Three water molecules (magenta) can be seen sitting in the 'hole' generated by the absence of Tyr370. The movement of Asn310 in the mutant structure can also clearly be seen making a new hydrogen bond to one of these water molecules.

In order to probe what size of 'hole' is required to accommodate a bound water to act as a nucleophile, more extensive substitutions of the catalytic tyrosine have been performed (Watson *et al* , in press). The hydrolytic activity displayed by two additional mutants Y370E and Y370F were 1 and 4 % that of the wt enzyme when MU α-Neu5Ac was used as the substrate. This level of activity, although low, was sufficient however, to allow the identification of the initial hydrolysis product using standard ^{1}H NMR spectroscopy. The results from these investigations showed that in contrast to the Y370G/A/D mutants, the Y370E/F mutants operated with retention of configuration. It was concluded therefore, that the addition of an additional CH_2 unit between the Y370D and Y370E mutants, was sufficient to close off the 'hole' and switch the mechanism to one of retention.

The structural study is still being carried out on these mutants. However, a complex was obtained of the Y370F mutant with the inhibitor DANA (chapter 7). This is an interesting mutant, as the active site has no obvious nucleophile to

stabilise the TS and/or form a covalent intermediate. Given that the nucleophilic oxygen has been replaced in the Y370F mutant, possible mechanisms of hydrolysis are: (1) generation of an oxocarbonium ion intermediate that has a lifetime sufficient to allow aglycon departure prior to capture by water from the same face, i.e., internal return; (2) generation of an oxocarbonium ion intermediate that is trapped by a proximal residue, possibly Glu260, and this sialyl-enzyme intermediate is then hydroylsed (S_N1); or (3) direct double displacement of the leaving group by an alternate nucleophile, such as Glu260 (S_N2).

The Y370F mutant crystallised in the trigonal space group, $P3_221$, with three molecules in the asymmetric unit. Non-crystallographic symmetry was not used in the refinement to constrain the three models, and each monomer was therefore refined independently. Superimposing the wt structure (PDB code 1EUS) onto the three monomers shows that the active site makes no significant changes to accommodate the phenylalanine side chain. The r.m.s.d. for the 357 equivalent C_α atoms was 0.34 Å, 0.36 Å and 0.31 Å, respectively for monomers A-C. Figure 8.4.6 shows the active site of the Y370F-DANA monomer A complex with the Glu260 and Tyr370 side chains of the wt structure superimposed. The phenol ring of Phe370 sits in the same plane as the tyrosine ring in the wt structure.

As DANA is considered to be a TS analogue of the reaction (Burmeister *et al* ., 1993), we analysed the distance from the carboxylate group of Glu260 to the C2 atom of DANA in each monomer (Table 8.4.4). For completeness, the remaining hydrogen bond interactions to DANA in monomer A are given in Table 8.4.5. The equivalent distances in monomers B and C are not significantly different.

Table 8.4.4. Distances between the carboxylate group of Glu260 in each monomer of Y370F to the C2 atom of Neu5Ac2en (DANA) compared to the wt structure (PDB code 1EUS).

	DANA C2 – E260-$O^{\varepsilon 1}$ distance (Å)	DANA C2 – E260-$O^{\varepsilon 2}$ distance (Å)
Wt	5.46	4.56
Y370F monomer A	4.78	4.35
Y370F monomer B	5.15	4.19
Y370F monomer C	5.21	4.40

Table 8.4.5. Hydrogen bonding interactions between the bound ligand, Neu5Ac2en (DANA) and the active site of monomer A of the Y370F MvNA mutant.

DANA	Protein/Water atom	Distance (Å)
O1A	Arg68-$N^{\eta 1}$	2.87
	Arg68-$N^{\eta 2}$	2.89
	Arg342-$N^{\eta 1}$	2.84
O1B	Arg342-$N^{\eta 2}$	2.95
	Arg276-$N^{\eta 1}$	3.00
	Arg276-$N^{\eta 2}$	2.99
	H_2O	3.06
O4	Arg87-$N^{\eta 2}$	3.00
	Asp131-$O^{\delta 1}$	2.61
	Asp131-$O^{\delta 2}$	2.99
O7	H_2O	3.07
O8	Asp259-$O^{\delta 1}$	2.68
O9	H_2O	2.83
	Asp259-$O^{\delta 2}$	2.48
O10	H_2O	2.84
N5	Asp131-$O^{\delta 2}$	2.55

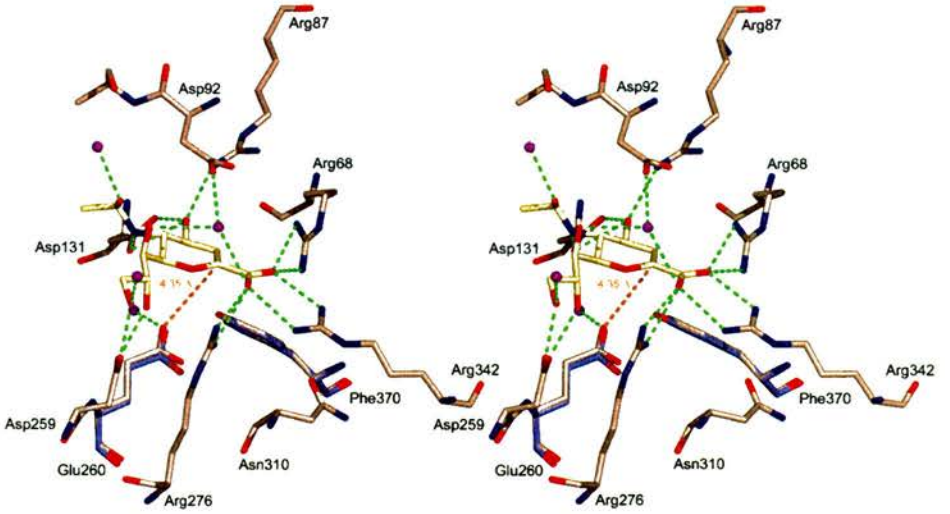


Figure 8.4.6. Stereo view of the active site of the Y370F-Neu5Ac2en (DANA) monomer A complex. The Y370F residues are shown in wheat, the superimposed residues from the wt structure (Glu260 and Tyr370) are shown in sLace. Selected hydrogen bond interactions are shown as green dashed lines. The distance between the OE2 atom of Glu260 and the C2 atom of DANA is shown as an orange dashed line with the distance, in Ångstroms, indicated.

In monomer A only, the Glu260 side chain moves 0.44 Å closer to the C2 of DANA (Figure 8.4.6) compared with the wt structure. The esd (of atomic positions) for this monomer is 0.217 ± 0.049 Å (Table 7.5.1; chapter 7), making this a shift of 2 σ. It is unclear if this movement is significant however, as this shift is not observed in the other two monomers.

To investigate the mechanism used by the Y370F mutant, a series of kinetic experiments were carried out similar to those described for the D92G and Y370G mutants in sections 8.4.1 and 8.4.2 of this chapter. Listed in Table 8.4.5. are the derived kinetic parameters for the wt, Y370D, Y370E and Y370F mutants using an activated and a natural substrate analogue (3' sialyllactose), respectively. By analysing at the effects of leaving group ability on the activity of the mutant sialidase, our collaborators concluded that this mutant was operating via a dissociative (S_N1) mechanism (Watson *et al* , in press). This conclusion was based on the assumption that in S_N1 reactions a larger drop in reactivity is observed on decreasing the leaving group ability than seen for a comparable S_N2

reaction (Lowry and Richardson, 1987); as is seen for the Y370F mutant in Table 8.4.6.

Table 8.4.6. Kinetic parameters of wild-type MvNA sialidase and Y370 variants using MU-αNeu5Ac and 3' sialyllactose at pH 5.25 and 37 °C.
Taken from (Watson *et al* , in press)

	MU Neu5Ac		3' sialyllactose	
	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
Wild-type[a]	51 ± 0.1	$7.23 \pm 0.04 \times 10^6$	138 ± 5	$3.55 \pm 0.6 \times 10^5$
Y370D[a]	14.6 ± 0.6	$1.0 \pm 0.1 \times 10^5$	0.017[b]	8.5
Y370E	0.51 ± 0.01	$4.09 \pm 0.4 \times 10^3$	0.0027[b]	24.1
Y370F	2.22 ± 0.13	$6.21 \pm 1.5 \times 10^4$	0.00036[b]	0.683

[a] Data taken from reference (Watson *et al* ., 2003).
[b] No error is associated with these values because they are estimates based on a single rate measurement at high concentration of substrate and enzyme.

Given that the lifetime in aqueous solution for the sialosyl oxocarbonium ion is around 3×10^{-11} seconds (Horenstein and Bruner, 1996), it is likely that the cationic intermediate is captured by a proximal enzyme nucleophile. The structural data presented here show that only Glu260 is close enough to act in this capacity, being located ~4.35 Å from the C2 of the bound DANA (Figure 8.4.6). A Y370F/E260G double mutant has been produced to investigate this hypothesis.

8.5. Discussion.

The observation that the monosaccharide galactose is capable of being recognised by the C-terminal domain of the MvNA sialidases led to the naming of this domain as a 'galactose-binding' module (Gaskell *et al* ., 1995). Subsequent sequence and structural analyses of this domain have classified it as a Type C 'small sugar binding' carbohydrate binding module (CBM), belonging to CBM family 32 (Boraston *et al* ., 2004). The functional relationship between the 'galactose-binding' CBM and the catalytic β-propeller domain of the MvNA sialidase, however, is not fully understood. The structural determination of the different MvNA mutants for this study, combined with the fortuitous tapping of the disaccharide lactose in the CBM, have provided new structural information with which to investigate this relationship.

The ESCET analysis showed that the 'galactose-binding' CBM, along with the immunoglobulin domain are conformationally flexible with respect to the β-propeller domain (section 8.2). The maximum movement of the CBM domain relative to the β-propeller was ~8 Å between the two most structurally divergent conformers. This movement was less than expected based on a modelling study performed on the wt structure when it was first solved (Taylor *et al* ., 1999). This showed that the di-glycine linker connecting the β-propeller to the immunoglobulin 'arm' would have afforded much greater flexibility than that seen in these structures. This study has shown that although the three-domain sialidase does display some conformational flexibility, the overall orientation of the CBM sitting ~ 20 Å over the active site of the enzyme is invariant. This observation suggests that this spatial arrangement of the three domains has some functional significance for the enzyme. In order to understand the significance of this spatial arrangement it is necessary to identify the nature of the target ligand. Once this information is known, it should be possible to place the unusual structure of this sialidase in the context of its substrate.

A recent study into the structure of a family 6 CBM also investigated the binding of a number of carbohydrate ligands to the CBM of the MvNA sialidase

(Boraston *et al* ., 2003). This study was the first to quantify, using ITC and UV difference spectroscopy, the binding of both galactose and lactose to the CBM of MvNA. The authors also suggest that sialic acid itself could also be a target ligand, although no experimental evidence was given. This study showed that the K_D of binding was in the millimolar range for both sugars, and that the binding of Gal is tighther than to Lac, with a K_D value of 0.5 mM compared to 1.3 mM respectively (Boraston *et al* ., 2003). The data presented in section 8.2 of this chapter provide a structural understanding of this carbohydrate-protein interaction, showing clearly how lactose binds into the CBM domain.

The binding site looks very similar to those observed previously in lectins, being mediated predominantly via hydrogen bonds to water molecules. This observation supports the suggestion by Boraston and colleagues (Boraston *et al* ., 2003), that the distinction between lectins and TypeC CBMs is blurring as more structural information is obtained on the CBM domains. The binding site of the MvNA CBM shows three direct protein-carbohydrate hydrogen bond interactions, and these seem to dictate the specificity of this CBM for galactose, with the O4 hydroxyl accommodated in the axial position only. This model, combined with the quantitative study by Boraston *et al* (Boraston *et al* ., 2003), provides the strongest evidence yet that the target ligand is galactose. The overall position of the CBM above the β-propeller domain and the binding interactions to lactose, appear to make this domain well suited to binding cell-surface glycans, oligosaccharides or polysaccharides that cannot be completely enveloped by the CBM.

These observations fit very neatly with the proposed biological role of this sialidase. The enzyme is secreted by the bacterium, presumably to scavenge sialic acid sugars from the surrounding soil and vegetation to use as a carbon and energy source. Sialic acid residues are found most commonly in the terminal positions of the oligosaccharide chains of glyconconjugates, linked either α 2 → 3 or α 2 → 6 to D-Gal (Brooks *et al* ., 2002). As such, the cleavage of the terminal sialic acid sugar by the β-propeller domain would expose the subterminal galactose. The CBM domain, sitting above the catalytic domain ~23

± 8 Å away, would be in a perfect position to attach to the exposed galactose and anchor the enzyme onto the glyconconjugate, Figure 8.5.1. Such a role for the CBM fits the proposed function of these domains in other glycosidases. In this role, the CBM further potentiates catalytic activity by mediating prolonged and intimate association between the sialidase and the substrate. In a recent kinetic study it was shown that bacterial sialidases possessing CBM domains have increased catalytic activity toward polyvalent substrates due to tighter binding and lowering of the K_m (Thobhani *et al* ., 2002).

Although the data presented here cannot prove it, the invariable spatial arrangement of the three-domains in the MvNA sialidase may play a role in substrate disruption. It has long been known that sialic acids have the ability to influence the three-dimensional stability of glycoproteins and gangliosides and form a multitude of branched glyco-structures (Brooks *et al* ., 2002). The positioning of the CBM of MvNA above the catalytic domain may have a role in increasing the access to sialic acid present in these branched configurations, in a similar way to the disruption of crystalline cellulose observed in other glycosidases (Din *et al* ., 1994).

Figure 8.5.1. Cartoon representation of the MvNA sialidase showing the location of the bound ligand lactose sitting in the CBM of the D92G-lactose complex. The distance from the O1 atom of the glucose sugar unit to the active site in the β-propeller is ~23 ± 8 Å, as calculated from the ESCET analysis of the MvNA structures.

The structural study of the MvNA active site mutants was initiated initially to underpin and complement a series of detailed kinetic and product studies carried out by our collaborators (Watson *et al* ., 2003; Watson *et al* ., 2004). The aims of these kinetic studies were to investigate the role of the conserved active site residues, principally the catalytic nucleophile Tyr370 and acid/base catalyst Asp92. The role of the conserved tyrosine as the catalytic nucleophile in sialidase catalysed hydrolysis has now been established (see Chapter 5 and (Amaya *et al* ., 2004; Watts *et al* ., 2003). However, the debate as to why nature uses a tyrosine nucleophile in sialidases and trans-sialidases is still unresolved. Withers and co-workers proposed that because sialic acids have a carboxylate group attached to the anomeric centre, it would be electrostatically unfavourable for a carboxylate residue to displace the leaving group nucleophilically (Watts *et al* ., 2003). The structure of the covalent intermediate with the trans-sialidase from *T.cruzi* was used to support this hypothesis (Amaya *et al* ., 2004).

However, the negative charge on the substrates carboxylate group, upon binding to this family of enzymes, should be more than counter-balanced by the three conserved arginines to which it forms strong ionic interactions. This point was not addressed in the *T.cruzi* paper (Amaya *et al* ., 2004).

The results of the kinetic and product studies performed on the MvNA mutants discussed in this chapter have led to an alternative proposal (Watson *et al* ., 2003). That the natural selection of tyrosine, rather than the usual carboxylate group seen in other glycosidases (Davies *et al* ., 1998; Zechel and Withers, 2001), could originate from differences in intrinsic reactivity of the covalently bound intermediate. This hypothesis is based on the observation that ketal-based sialic acids are inherently more reactive than acetal-based glycosides, and that nature has selected to stabilise the sialyl-enzyme intermediate by making the leaving group (tyrosine) worse. Such stabiliation would allow time for aglycon departure and incoming of water in the bacterial sialidases or an acceptor sugar in the case of the trans-sialidase.

The structures of the MvNA mutants, in complex with the different ligands described, have helped to underpin these kinetic and product studies. The analysis of the active sites have shown that the sialic acid binding pocket is extremely rigid, with little or no significant movement observed upon mutation of key catalytic residues, even to glycine. This observation is significant in itself, as it reinforces the ideas forwarded earlier in chapter 5 on the *C.perfringens* structure, that substrate distortion plays a significant role in the catalytic mechanism of bacterial sialidases, by forcing the sugar to 'fit' the active site. A range of non-hydrolysable substrates has now been developed to aid in the mechanistic and structural investigation of this bacterial sialidase.

Future Work and Discussion

The primary aim of this investigation was the structural determination of the 77 kDa nanI isoenzyme from *Clostridium perfringens*. The persistant breakdown of the full length protein into two domains led to the subcloning of a 50 kDa fragment, which retained the full catalytic activity of the enzyme. The subsequent crystallisation and structure solution of this domain has shown that the fragment folds into the canonical six-bladed β-propeller structure, with the addition of a small β-barrel inserted into the second blade. Clearly the function of this domain should be investigated. Studies from other sialidases have indicated that the extra domains, where present, are usually involved in binding to carbohydrate structures on the sialylated glycoconjugates substrates (Taylor et al., 1999). These Carbohydrate Binding Modules (CBMs) have been shown to greatly enhance the kinetic properties of the sialidases (Thobhani et al., 2002). The ability of this domain to recognise various carbohydrates could be tested using Isothermal Titration Calorimetry (ITC). Such instruments are used to measure the thermodynamics of protein-ligand interactions, and a range of small carbohydrates could be added to a solution of the 50 kDa domain to see if any interactions can be measured.

The structure of the 50 KDa catalytic domain of nanI has also given us the ability to predict what the nanJ structure should look like in this region. The sequence alignment presented in chapter 1 shows that both proteins have very similar sequences for the corresponding region of the nanJ protein. This implies that the β-barrel domain seen in nanI has some beneficial function to the enzyme, as it likely appears in both structures. The structure of the nanI catalytic domain also has some bearing of the origins of the leech intra-molecular trans-sialidase (Lou et al., 1998), which was used as a molecular replacement model for the structure solution (chapter 3). It was proposed that this sialidase was isolated from a contaminating bacterium in the leech, and was not actually encoded in the genome. This suggestion was made on bioinformatics evidence that showed the primary structure of the intra-molecular trans-sialidase was more

closely related to the microbial sialidases than those found in higher eukaryotes (E.Vimr, personal communication). The results presented here for the structure of the nanI sialidase from *C.perfringens* clearly supports this hypothesis, with both sialidases having very similar structures, especially in the β-barrel domain. It will be interesting to see the results of sequencing the leech genome, if it is attempted, to confirm or disprove this.

The structure of the N-terminal domain of the *C.perfringens* sialidase was not solved during the course of this study. Sequence alignment with the intra molecular trans-sialidase suggests that it forms a similar fold as seen in this enzyme, namely a lectin jelly roll. This domain was sequenced from the full length construct provided for this study, and was shown to express well with a C-terminal six-histidine affinity tag in *E.coli*. Crystallisation of this domain is currently underway to determine its structure. It should also be possible to investigate the target ligand for this domain using the ITC method described above. It would prove insightful to know which carbohydrates this domain targets in understanding the role of this sialidase in the pathogenesis of this bacterium.

The observation that *C.perfringens* has three sialidase isoenzymes was surprising, given that the sialidases of this bacterium have been studied extensively over the past 20 years (Roggentin et al., 1989; Roggentin et al., 1995; Schauer et al., 1995). Both nanI and nanJ have signal sequences at their N-termini and the nanI sialidase is secreted by the bacterium (Roggentin et al., 1995). However, the nanJ sialidase was not detected in any previous biochemical studies on this bacterium. The gene encoding this sialidase was PCR amplified from the genome of *C.perfringens* strain A99 (a kind gift from Dr. P. Roggentin), and like the nanI isoenzyme was found to degrade easily during purification. A sub-cloning strategy has been designed to split the enzyme into different domains based on the structure of the nanI sialidase and on the domain boundaries identified using BLAST. It would be fascinating to see how these domains were arranged around the β-propeller structure and work is currently underway in our laboratory to crystallise the full-length enzyme.

The complex of the nanI catalytic domain with the substrate α-sialic acid (Neu5Ac) is the first such structure in a non-viral sialidase. The atomic resolution diffraction collected from these crystals allowed the very best possible structural model to be built, and has allowed the interactions of this carbohydrate with the active site to be analysed in great detail. The structure of the covalent sialyl-enzyme intermediate was only collected at 1.6 Å, as attempts to collect data on this soak at the ESRF proved unsuccessful. The aim in the next few months will be to try and get higher resolution data for this complex, which will allow the measurement of the tyrosine – C2 bond length to be made. The high-resolution data for this complex, however, has proved very interesting. We now have structural data on two different members of the sialidase superfamily with the covalently bound intermediate, the structure of nanI and the *T.cruzi* (Amaya et al., 2004) structure discussed in chapter 5. Active site mutations of the nanI enzyme and complexes with this compound should prove useful to determining the role of the conserved glutamic acid in stabilising the covalent bond.

In summary, the successful crystallisation of sub-cloned 50 kDa catalytic domain of the nanI sialidase from *C.perfringens*, and the ability to collect atomic resolution data from the crystals has provided a new and valuable tool in sialidase research. The ideas forwarded by Alzari and colleagues from results of the *T.cruzi* structures can now be tested on a strict hydrolyse, and comparisons made. The structure should also prove useful to the wider glycosidase community for investigating the novel hydrolysis mechanism utilised by the sialidase superfamily.

Bibliography:

Achyuthan, K. E., and Achyuthan, A. M. (2001). Comparative enzymology, biochemistry and pathophysiology of human exo-alpha-sialidases (neuraminidases). Comp Biochem Physiol B Biochem Mol Biol *129*, 29-64.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Amaya, M. F., Buschiazzo, A., Nguyen, T., and Alzari, P. M. (2003). The high resolution structures of free and inhibitor-bound Trypanosoma rangeli sialidase and its comparison with T. cruzi trans-sialidase. J Mol Biol *325*, 773-784.

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 (Camb) *12*, 775-784.

Anderson, D. H., Weiss, M. S., and Eisenberg, D. (1997). Charges, hydrogen bonds, and correlated motions in the 1 A resolution refined structure of the mating pheromone Er-1 from Euplotes raikovi. J Mol Biol *273*, 479-500.

Baumgartner, S., Hofmann, K., Chiquet-Ehrismann, R., and Bucher, P. (1998). The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. Protein Sci *7*, 1626-1631.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nat Struct Biol *10*, 980.

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 Res *28*, 235-242.

Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P., and Gilbert, H. J. (1998).

Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem J *331 (Pt 3)*, 775-781.

Bond, C. S. (2003). Topdraw: A program for drawing protein topology diagrams. Bioinformatics *19*, 311-312.

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

Boraston, A. B., Notenboom, V., Warren, R. A., Kilburn, D. G., Rose, D. R., and Davies, G. (2003). Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and "galactose-binding" domains. J Mol Biol *327*, 659-669.

Brooks, S. A., Dwek, M. V., and Schumacher, U. (2002). Sialic acids. In Functional & Molecular Glycobiology. (BIOS Scientific Publishers Ltd.), pp. 203-220.

Brünger, A. T. (1997). Free R Value: Cross-Validation in Crystallography. In Methods Enzymology, R. M. Sweet, ed. (Academic Press), pp. 366-396.

Brünger, A. T., Adams, P. D., and al., e. (1998a). Crystallography and NMR system: A new software suite for macromolecular structure determination. Acta Crystallographica Section D *54*, 905-921.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998b). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 905-921.

Burmeister, W. P., Cusack, C., and Ruigrok, R. W. H. (1994). Calcium is needed for the thermostability of influenza B virus neuraminidase. J Gen Virology *75*, 381-388.

Burmeister, W. P., Henrissat, B., Bosso, C., Cusack, C., and Ruigrok, R. W. H. (1993). Influenza B virus neuraminidase can synthesise its own inhibitor. Structure (Camb) *1*, 19-26.

Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C., and Alzari, P. M. (2002). The Crystal Structure and Mode of Action of Trans-Sialidase, a Key Enzyme in Trypanosoma cruzi Pathogenesis. Molecular Cell *10*, 757-768.

Buschiazzo, A., Tavares, G. A., Campetella, O., Spinelli, S., Cremona, M. L., Paris, G., Amaya, M. F., Frasch, A. C., and Alzari, P. M. (2000). Structural basis of sialyltransferase activity in trypanosomal sialidases. Embo J *19*, 16-24.

Canard, B., and Cole, S. T. (1990). Lysogenic phages of Clostridium perfringens: mapping of the chromosomal attachment sites. FEMS Microbiol Lett *54*, 323-326.

CCP4 (1994). The CCP4 suite: Programs for Protein Crystallography, Number 4. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Chavas, L. M., Tringali, C., Fusi, P., Venerando, B., Tettamanti, G., Kato, R., Monti, E., and Wakatsuki, S. (2004). Crystal structure of the human cytosolic Sialidase Neu2: Evidence for the dynamic nature of substrate recognition. J Biol Chem.

Chien, C. H., Shann, Y. J., and Sheu, S. Y. (1996). Site-directed mutations of the catalytic and conserved amino acids of the neuraminidase gene, nanH, of Clostridium perfringens ATCC 10543. Enzyme Microb Technol *19*, 267-276.

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

Chong, A. K. J., Pegg, M. S., Taylor, N. R., and von Itzstein, M. (1992b). Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. Eur J Biochem *207*, 335-343.

Chong, A. K. J., Pegg, M. S., and von Itzstein, M. (1991). Influenza virus sialidase: effect of calcium on steady-state kinetic parameters. Biochem Biophys Acta *1077*, 65-71.

Clarke, C., Woods, R. J., Gluska, J., Cooper, A., Nutley, M. A., and Boons, G. J. (2001). Involvement of water in carbohydrate-protein binding. J Am Chem Soc *123*, 12238-12247.

Colman, P. M. (2002). Neuraminidase inhibitors as antivirals. Vaccine *20 Suppl 2*, S55-58.

Corfield, A. P. (1992). Bacterial sialidases - roles in pathogenicity and nutrition. Glycobiology *2*, 509-521.

Corfield, A. P., Veh, R. W., Wember, M., Michalski, J. C., and Schauer, R. (1981). The release of N-acetyl- and N-glycolloyl-neuraminic acid from soluble complex carbohydrates and erythrocytes by bacterial, viral and mammalian sialidases. Biochem J *197*, 293-299.

Coutinho, P. M., and Henrissat, B. (1999a). Carbohydrate-Active Enzymes server at URL: http://afmb.cnrs-mrs.fr/CAZY/.

Coutinho, P. M., and Henrissat, B. (1999b). Carbohydrate-active enzymes: an integrated database approach. In Recent Advances in Carbohydrate Bioengineering, H. J. Gilbert, Davies, G. J., Henrissat, B., Svensson, B., ed. (Cambridge, Royal Society of Chemistry), pp. 3-12.

Crennell, S. J., Garman, E., Laver, G., Vimr, E. R., and Taylor, G. (1994). Crystal structure of Vibrio cholerae neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. Structure (Camb) *2*, 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. Proc Natl Acad Sci U S A *90*, 9852-9856.

Crennell, S. J., Takimoto, T., Portner, A., and Taylor, G. (2000). Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. Nat Struct Biol *7*, 1068-1074.

Cruickshank, D. W. (1999a). Remarks about protein structure precision. Acta Crystallogr D Biol Crystallogr *55 (Pt 3)*, 583-601.

Cruickshank, D. W. (1999b). Remarks about protein structure precision. erratum. Acta Crystallogr D Biol Crystallogr *55*, 1108.

Cruickshank, D. W. J. (2001). Coordinate Uncertainty. In International Tables for Crystallography., E. Arnold, ed. (Dordrecht/Boston/London, Kluwer Academic), pp. 403-414.

Dauter, Z., Lamzin, V., and Wilson, K. (1997). The benefits of atomic resolution. Curr Opin Struc Biol *7*, 681-688.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure *3*, 853-859.

Davies, G., Sinnott, M. L., and Withers, S. G. (1998a). Comprehensive Biological Catalysis. In, M. L. Sinnott, ed. (San Diego., Academic Press.).

Davies, G., Sinnott, M. L., and Withers, S. G. (1998b). Glycosyl Transfer. In Comprehensive Biological Catalysis, M. L. Sinnott, ed. (Academic Press, San Diego.), pp. 119-209.

Davies, G. J., Ducros, V. M., Varrot, A., and Zechel, D. L. (2003). Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography. Biochem Soc Trans *31*, 523-527.

Demuth, D. R., Golub, E. E., and Malamud, D. (1990). Streptococcal-host interactions. Structural and functional analysis of a Streptococcus sanguis receptor for a human salivary glycoprotein. J Biol Chem *265*, 7120-7126.

Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A., and Kilburn, D. G. (1994). C1-Cx revisited: intramolecular synergism in a cellulase. Proc Natl Acad Sci U S A *91*, 11383-11387.

Dirita, V. J. (2001). Molecular basis of Vibrio Cholerae Pathogenesis. In Principles of Bacterial Pathogenesis, E. A. Groisman, ed. (Academic Press), pp. 457-508.

Engh, R. A., and Huber, R. (1991). Accurate Bond length and angle parameters for X-ray protein-structure refinement. Acta Cryst A *47*, 392-400.

Evans, P. R. (1997). Joint CCP4 ans ESF-EACBM Newsletter 33, pp. 22-24.

Ferrero-Garcia, M. A., Trombetta, S. E., Sanchez, D. O., Reglero, A., Frasch, A. C., and Parodi, A. J. (1993). The action of Trypanosoma cruzi trans-sialidase on glycolipids and glycoproteins. Eur J Biochem *213*, 765-771.

Fersht, A. (1985). Chemical catalysis. In Enzyme Structure and Mechanism. (New York., W. H. Freeman and comp.), pp. 47-98.

Friebolin, H., Brossmer, R., Keilich, G., Ziegler, D., and Supp, M. (1980). [1H-NMR-spectroscopic evidence for the release of N-acetyl-alpha-D-neuraminic acid as the first product of neuraminidase action (author's transl)]. Hoppe Seylers Z Physiol Chem *361*, 697-702.

Garman, E. (1999). Cool data: quantity AND quality. Acta Crystallogr D Biol Crystallogr *55 (Pt 10)*, 1641-1653.

Gaskell, A., Crennell, S. J., and Taylor, G. (1995). The Three Domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure (Camb) *3*, 1197-1205.

Ghate, A. A., and Air, G. M. (1998). Site-directed mutagenesis of catalytic residues of influenza virus neuraminidase as an aid to drug design. Eur J Biochem *258*, 320-331.

Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S., Simpson, P. J., Williamson, M. P., Hazlewood, G. P., and Gilbert, H. J. (1999). The type II and X cellulose-binding domains of Pseudomonas xylanase A potentiate catalytic activity against complex substrates by a common mechanism. Biochem J *342 (Pt 2)*, 473-480.

Gloster, T. M., Williams, S. J., Roberts, S., Tarling, C. A., Wicki, J., Withers, S. G., and Davies, G. J. (2004). Atomic resolution analyses of the binding of xylobiose-derived deoxynojirimycin and isofagomine to xylanase Xyn10A. Chem Commun (Camb), 1794-1795.

Holdgate, G. A., Tunnicliffe, A., Ward, W. H., Weston, S. A., Rosenbrock, G., Barth, P. T., Taylor, I. W., Pauptit, R. A., and Timms, D. (1997). The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. Biochemistry *36*, 9663-9673.

Holm, L., and Sander, C. (1998). Touring protein fold space with Dali/FSSP. Nucleic Acids Res *26*, 316-319.

Horenstein, B. A., and Bruner, M. (1996). Acid-catalyzed solvolysis of CMP-N-acetyl neuraminate: Evidence for a sialyl cation with a finite lifetime. J Am Chem Soc *118*, 10371-10379.

Hoyer, L. L., Roggentin, P., Schauer, R., and Vimr, E. R. (1991). Purificaton and properties of cloned Salmonella typhimurium LT2 sialidase with virus-typical kinetic preference for siayl alpha2 -> 3 linkages. J Biochem *110*, 462-467.

Hutchinson, E. G., and Thornton, J. M. (1996). PROMOTIF - A program to identify and analyse structural motifs in proteins. Protein Science *5*, 212-220.

Ito, N., Phillips, S. E., Yadav, K. D., and Knowles, P. F. (1994). Crystal structure of a free radical enzyme, galactose oxidase. J Mol Biol *238*, 794-814.

Jancik, J., and Schauer, R. (1974). Sialic acid--a determinant of the life-time of rabbit erythrocytes. Hoppe Seylers Z Physiol Chem *355*, 395-400.

Jencks, W. P. (1987). Economics of enzyme catalysis. Cold Spring Harb Symp Quant Biol *52*, 65-73.

Jermyn, W. S., and Boyd, E. F. (2002). Characterization of a novel Vibrio pathogenicity island (VPI-2) encoding neuraminidase (nanH) among toxigenic Vibrio cholerae isolates. Microbiology *148*, 3681-3693.

Johnson, N. P., and Mueller, J. (2002). Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. Bull Hist Med *76*, 105-115.

Jones, T. A., Bergdoll, M., and Kjeldgaard, M. (1990). O: A macromolecular modeling environment. In Crystallographic and Modeling Methods in Molecular Design., S. Ealick, ed. (Springer-Verlag), pp. 189-195.

Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for the building of protein models in electron density maps and the location of errors in these models. Acta Cryst A *47*, 110-119.

Joshi, M. D., Hedberg, A., and McIntosh, L. P. (1997). Complete measurement of the pKa values of the carboxyl and imidazole groups in Bacillus circulans xylanase. Protein Sci *6*, 2667-2670.

Kabsch, W., Kabsch, H., and Eisenberg, D. (1976). Packing in a new crystalline form of glutamine synthetase from Escherichia coli. J Mol Biol *100*, 283-291.

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

Kao, Y. H., Lerner, L., and Warner, T. G. (1997). Stereoselectivity of the Chinese hamster ovary cell sialidase: sialoside hydrolysis with overall retention of configuration. Glycobiology *7*, 559-563.

Katz, A. K., Glusker, J. P., Beebe, S. A., and Bock, C. W. (1996). Calcium ion coordination: A comparison with that of Beryllium, Magnesium and Zinc. J AM CHEM SOC *118*, 5752-5763.

Kiessling, L. L., Gestwicki, J. E., and Strong, L. E. (2000). Synthetic multivalent ligands in the exploration of cell-surface interactions. Curr Opin Chem Biol *4*, 696-703.

Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. Acta Crystallogr D Biol Crystallogr *55 (Pt 2)*, 484-491.

Kissinger, C. R., Gehlhaar, D. K., Smith, B. A., and Bouzida, D. (2001). Molecular replacement by evolutionary search. Acta Crystallogr D Biol Crystallogr *57*, 1474-1479.

Klein, P. J., Bulla, M., Newman, R. A., Muller, P., Uhlenbruck, G., Schaefer, H. E., Kruger, G., and Fisher, R. (1977). Thomsen-Friedenreich antigen in haemolytic-uraemic syndrome. Lancet *2*, 1024-1025.

Kleineidam, R. G., Kruse, S., Roggentin, P., and Schauer, R. (2001). Elucidation of the role of functional amino acid residues of the small sialidase from Clostridium perfringens by site-directed mutagenesis. Biol Chem *382*, 313-319.

Kleywegt, G. J. (2001). Structure Validation. In International Tables for Crystallography, M. G. R. Arnold, ed. (Dordrecht/Boston/London, Kluwer Academic), pp. 497-506.

Kleywegt, G. J., and Brünger, A. T. (1996). Checking your imagination: applications of the free R value. Structure (Camb) *4*, 897-904.

Kleywegt, G. J., and Jones, T. A. (1997). Detecting folding motifs and similarities in protein structures. Methods Enzymology *277*, 525-545.

Kobasa, D., Takada, A., Shinya, K., Hatta, M., Halfmann, P., Theriault, S., Suzuki, H., Nishimura, H., Mitamura, K., Sugaya, N., *et al.* (2004). Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. Nature *431*, 703-707.

Kruse, S., Kleineidam, R. G., Roggentin, P., and Schauer, R. (1996). Expression and purification of a recombinant "small" sialidase from Clostridium perfringens A99. Protein Expr Purif *7*, 415-422.

Ladbury, J. E. (1996). Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. Chem Biol *3*, 973-980.

Ladbury, J. E., and Chowdhry, B. Z. (1996). Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. Chem Biol *3*, 791-801.

Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. J Mol Biol *231*, 1049-1067.

Lawrence, M. C., Borg, N. A., Streltsov, V. A., Pilling, P. A., Epa, V. C., Varghese, J. N., McKimm-Breschkin, J. L., and Colman, P. M. (2004). Structure of the haemagglutinin-neuraminidase from human parainfluenza virus type III. J Mol Biol *335*, 1343-1357.

Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 & ESF-EAMCB Newsletter on protein crystallography, No 26.

Liang, R., Loebach, J., Horan, N., Ge, M., Thompson, C., Yan, L., and Kahne, D. (1997). Polyvalent binding to carbohydrates immobilized on an insoluble resin. Proc Natl Acad Sci U S A *94*, 10554-10559.

Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A., and Cambillau, C. (1997). Atomic resolution (1.0 A) crystal structure of Fusarium solani cutinase: stereochemical analysis. J Mol Biol *268*, 779-799.

Lou, Y., Li, S.-C., Chou, M.-Y., Li, Y.-T., and Lou, M. (1998a). The crystal structure of an intramolecular trans-sialidase with a NeuAcalpha2->3Gal specificity. Structure (Camb) *6*, 521-530.

Lou, Y., Li, S. C., Li, Y. T., and Lou, M. (1998b). The 1.8 A structures of Leech intramolecular trans-sialidase complexes: Evidence of its enzymatic mechanism. JMB *285*, 323-332.

Lowry, T. H., and Richardson, K. S. (1987). Mechanism and theory in organic chemistry, 3rd edn (New York, Harper & Row).

Maina, C. V., Riggs, P. D., Grandea, A. G., 3rd, Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A., and Guan, C. D. (1988). An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. Gene *74*, 365-373.

Mathews, B. W. (1968). Solvent content of protein crystals. J Mol Biol *33*, 491-497.

McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996). The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a 13C-NMR study of bacillus circulans xylanase. Biochemistry *35*, 9958-9966.

McMurray, J. (1996). Organic Chemistry, Brooks/Cole Publishing Company).

Miller, W., Wang, P., and Flashner, M. (1978). Mechanism of Arthrobacter sialophilus neuraminidase: the binding of substrates and transition-state analogues. Biophys Biochem Res Commun *83*, 1479-1487.

Milligan, T. W., Straus, D. C., and Mattingly, S. J. (1977). Extracellular neuraminidase production by group B streptococci. Infect Immun *18*, 189-195.

Miyagi, T., Wada, T., Yamaguchi, K., and Hata, K. (2004). Sialidase and malignancy: A minireview. Glycoconjugate J *20*, 189-198.

Moustafa, I., Connaris, H., Taylor, M., Zaitsev, V., Wilson, J. C., Kiefel, M. J., von Itzstein, M., and Taylor, G. (2004a). Sialic acid recognition by Vibrio cholerae neuraminidase. J Biol Chem *279*, 40819-40826.

Moustafa, I. M. (2004) Structural studies on Vibrio cholerae neuraminidase, University of St Andrews, St Andrews. PhD Thesis.

Moustafa, I. M., Connaris, H., Taylor, M., Zaitsez, V., Wilson, J. C., Kiefel, M. J., von Itzstein, M., and Taylor, G. (2004b). Sialic Acid Recognition by Vibrio cholerae Neuraminidase. J Biol Chem *279*, 40819-40826.

Munoz Barroso, I., Moralejo, F. J., and Villar, E. (1994). Ionic dependence of the sialidase activity of hemagglutinin-neuraminidase glycoprotein in Newcastle disease virus membrane. Biochem Soc Trans *22*, 366S.

Murshudov, G. N. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr *53*, 240-255.

Murshudov, G. N., and Dodson, E. J. (1997). Simplified error estimation a la Cruickshank in macromolecular crystallography. CCP4 Newsletter *January*.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Cryst A *50*, 157-163.

Navaza, J. (2001). Implementation of molecular replacement in AMoRe. Acta Crystallogr D Biol Crystallogr *57*, 1367-1372.

Nees, S. R., Veh, R., and Schauer, R. (1975). Purification and Characterisation of neuraminidases from Clostridium perfringens. Hoppe-Seyler's Z Phys Chem *356*, 1027-1042.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281-296.

Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.

Pannu, N. S., Murshudov, G. N., Dodson, E. J., and Read, R. J. (1998). Incorporation of prior phase information strengthens maximum-likelihood structure refinement. Acta Crystallogr D Biol Crystallogr *54*, 1285-1294.

Pannu, N. S., and Read, R. J. (1996). Improved structure refinement through maximum likelihood. Acta Cryst A *52*, 659-668.

Paulson, J. C., Weinstein, J., and Schauer, A. (1989). Tissue-specific expression of sialyltransferases. J Biol Chem *264*, 10931-10934.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A *85*, 2444-2448.

Perrakis, A., Morris, R., and Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. Nat Struct Biol *6*, 458-463.

Puchelle, E., Girard, F., Houdret, N., and Bailleul, V. (1975). [Action of neuraminidases isolated from Diplococcus pneumoniae and Clostridium perfringens on the viscosity and elastic properties of bronchial secretions]. Biorheology *12*, 219-224.

Rhodes, G. (2000). Crystallography made crystal clear: A guide for users of Macromolecular model., Second edn (London, Academic Press).

Roggentin, P., Kleineidam, R. G., and Schauer, R. (1989). Conserved sequences in bacterial and viral sialidases. Glycoconjugate J *6*, 349-353.

Roggentin, P., Kleineidam, R. G., and Schauer, R. (1995). Diversity in the Properties of Two Sialidase Isoenzymes Produced by Clostridium perfringens spp. Bio Chem Hoppe-Seyler *376*, 569-575.

Roggentin, P., and Schauer, R. (1987). Purification and characterisation of sialidase from Clostridium sordelli G12. Glycoconj J *4*, 349-359.

Roggentin, P., Schauer, R., Hoyer, L. L., and Vimr, E. R. (1993). The sialidase superfamily and its spread by horizontal gene transfer. Mol Microbiol *9*, 915-921.

Rood, J. I. (1998). Virulence genes of Clostridium perfringens. Annu Rev Microbiol *52*, 333-360.

Rosenberg, A., ed. (1995). Biology of Sialic Acids (New York, Plenum Press).

Rosenberg, A., and Schengrund, C. L. (1976). Sialidases. In Biological Roles of Sialic Acid., C. L. Schengrund, ed. (New York, Plenum Press), pp. 295-359.

Rossmann, M. G., and Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst *15*, 24-31.

Rothe, B., Rothe, B., Roggentin, P., and Schauer, R. (1991). The sialidase gene from Clostridium septicum: cloning, sequencing, expression in Escherichia coli and identification of conserved sequences in sialidases and other proteins. Mol Gen Genet *226*, 190-197.

Rye, C. S., and Withers, S. G. (2000). Glycosidase mechanisms. Curr Opin Chem Biol *4*, 573-580.

Schauer, R. (1982). Chemistry, metabolism, and biological functions of sialic acids. Adv Carbohydr Chem Biochem *40*, 131-234.

Schauer, R. (1985). Sialic acids and their roles as biological masks. Trends Biochem Sci *10*, 357-360.

Schauer, R. (2000). Achievements and challenges of sialic acid research. Glycoconj J *17*, 485-499.

Schauer, R., Kelm, S., Reuter, G., Roggentin, P., and Shaw, L. (1995). Biochemistry and Role of Sialic Acids. In Biology of Sialic Acids, A. Rosenberg, ed. (New York, Plenum Press.), pp. 7-67.

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*, 1117-1125.

Schneider, T. R. (2000). Objective comparison of protein structures: error-scaled difference distance matrices. Acta Crystallogr D Biol Crystallogr *56 (Pt 6)*, 714-721.

Schneider, T. R. (2002). A genetic algorithm for the identification of conformationally invariant regions in protein molecules. Acta Crystallogr D Biol Crystallogr *58*, 195-208.

Schneider, T. R. (2004). Reliability of Cruickshank's DPI for medium to high resolution protein models., S. L. Newstead, ed.

Schwarzenbach, D., Abrahams, S., Flack, H., E, P., and Wilson, A. J. C. (1995). Statistical descriptors in crystallography. II. Report of a working group on expression of uncertainty in measurement. Acta Crystallographica Section A *51*, 565-569.

Sheldrick, G. M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. Acta Crystallogr A Found of Cryst *46*, 467-473.

Sheldrick, G. M., and Schneider, T. R. (1997). SHELXL: high-resolution refinement. In Methods Enzymology.

Sheu, S.-Y., Tseng, H.-j., Huang, S.-p., and Chien, C.-h. (2002). Cloning, expression, and deletion analysis of large *nanH* of *Clostridium perfringens* ATCC 10543. Enzyme and Microbial Technology *31*, 794-803.

Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S., and Hayashi, H. (2002). Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater. Proc Natl Acad Sci U S A *99*, 996-1001.

Siaito, M., and Yu, R. K. (1995). Biochemistry and Function of Sialidases. In Biology of the Sialic Acids, A. Rosenberg, ed. (New York, Plenum Press), pp. 261-313.

Silverman, R. B. (2002). The Organic Chemistry of Enzyme Catalysed Reactions., Academic Press).

Smith, P. K. (1985). Measuremnet of protein using bicinchroninic acid. Analytical Biochemistry *150*, 76-85.

Taylor, G. (1996). Sialidases: structures, biological significance and therapeutic potential. Curr Opin Struct Biol *6*, 830-837.

Taylor, G., Crennell, S. J., Thompson, C., and Chuenkova, M. (1999). The Sialidase superfamily: Structural Theme and Variations. In Sialobiology and Other Novel Forms of Glycosylation., F. A. Troy, ed. (Osaka, Gakushin Publishing Company), pp. 187-195.

Taylor, G., Dineley, L., Glowka, M., and Laver, G. (1992). Crystallization and preliminary crystallographic study of neuraminidase from Micromonospora viridifaciens. J Mol Biol *225*, 1135-1136.

Taylor, G. L. (2003). The Phase Problem. Acta Crystallogr D Biol Crystallogr *59*, 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. J Med Chem *37*, 616-624.

Thobhani, S., Ember, B., siriwardena, A., and Boons, G.-j. (2002). Multivalency and Mode of Action of Bacterial Sialidases. J AM CHEM SOC *125*, 7154-7155.

Traving, C., Schauer, R., and Roggentin, P. (1994). Gene structure of the 'large' sialidase isoenzyme from Clostridium perfringens A99 and its relationship with other clostridial nanH proteins. Glycoconj J *11*, 141-151.

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. J Mol Biol *221*, 473-486.

Varghese, J. N., Laver, W. G., and Colman, P. M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. Nature *303*, 35-40.

Varghese, J. N., McKimm-Brechkin, J. L., Caldwell, J. B., Kortt, A. A., and Colmann, P. M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. Proteins *14*, 327-332.

Varghese, J. N., Smith, P. W., Sollis, S. L., Blick, T. J., Sahasrabudhe, A., McKimm-Breschkin, J. L., and Colman, P. M. (1998). Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. Structure *6*, 735-746.

Varki, A. (1992). Diversity in the sialic acids. Glycobiology *2*, 25-40.

Vasella, A., Davies, G. J., and Bohm, M. (2002). Glycosidase mechanisms. Curr Opin Chem Biol *6*, 619-629.

Vimr, E. R., and Troy, F. A. (1985). Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. J Bacteriol *164*, 845-853.

von Itzstein, M., Wu, W. Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Van Phan, T., Smythe, M. L., White, H. F., and Oliver, S. W. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature *363*, 418-423.

Vriend, G. (1990). WHAT IF: A molecular modeling and drug design program. J Mol Graph *8*, 52-56.

Wang, Z., Luecke, H., Yao, N., and Quiocho, F. A. (1997). A low energy short hydrogen bond in very high resolution structures of protein receptor--phosphate complexes. Nat Struct Biol *4*, 519-522.

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*, 12682-12690.

Watson, J. N., Newstead, S. L., Dookhun, V., Taylor, G., and Bennet, A. J. (2004). Contribution of the active site aspartic acid to catalysis in the bacterial neuraminidase from Micromonospora viridifaciens. FEBS Letters. 2004 Nov 5;577(1-2):265-9.

Watts, A. G., Damager, I., Amaya, M. F., Buschiazzo, A., Alzari, P. M., Frasch, A. C., and Withers, S. G. (2003). Trypanosoma cruzi Trans-sialiadase Operates through a Covalent Sialyl-enzyme Intermediate: Tyrosine is the Catalytic Nucleophile. J AM CHEM SOC *125*, 7523-7533.

Weaver, R., and Hedrick, P. (1997). Genetics, 3 edn, Wm. C. Brown Publishers).

Wicki, J., Rose, D. R., and Withers, S. G. (2002). Trapping Covalent Intermediates on beta-Glycosidases. Methods in Enzymology *354*, 84-105.

Wilson, K. (2001). Monochromatic data collection: Principles of monochromatic data collection. In International Tables for Crystallography of Biological Macromolecules, M. G. R. Arnold, ed. (Dordrecht/Boston/London, Kluwer Academic Publishers), pp. 177 - 195.

Wolfenden, R., Lu, X., and Young, G. (1998). Spontaneous hydrolysis of glycosides. J Am Chem Soc *120*, 6814-6815.

Zechel, D. L., and Withers, S. G. (2001). Dissection of nucleophilic and acid-base catalysis in glycosidases. Curr Opin Chem Biol *5*, 643-649.