

# University of St Andrews



Full metadata for this thesis is available in  
St Andrews Research Repository  
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

**Structural studies on**  
***Vibrio cholerae* neuraminidase**

**Ibrahim M. Moustafa**

**A thesis submitted for the degree of Doctor of Philosophy**

**University of St-Andrews**

**January 2004**



Th E590

To Mum, Dad, Raghda and Heidi

## Declaration

(i) I, Ibrahim M. Moustafa, hereby certify that this thesis, which is approximately 55,000 words in length has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date 30/1/2004

Signature of candidate

Ibrahim

(ii) I was admitted as a research student in January 2000 and as a candidate for the degree of Ph.D. in January 2000; the higher study for which this is a record was carried out in the University of St-Andrews between 2000 and 2004.

Date 30/1/2004

Signature of candidate

(iii) I hereby 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 30/1/2004

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 regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and 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 30/1/2004

Signature of candidate

## Abstract

*Vibrio cholerae* neuraminidase (VCNA) plays a significant role in the cholera disease, though its role is not yet fully understood. *Vibrio cholerae* secretes VCNA, as a part of a mucinase complex, which cleaves the terminal sialic acid from the higher gangliosides to produce the GM1 ganglioside, the receptor of cholera toxin (CT). It also increases the fluidity of the mucin enhancing the access of *V. cholera* to the enterocytes. The 3D structure of the native enzyme has been described previously; it is composed of a central  $\beta$ -propeller domain flanked by two lectin-like domains with unknown functions. For a deeper understanding of the enzymatic action of VCNA, a crystallographic study has been undertaken to solve the structure of VCNA in complex with a number of the enzyme inhibitors as well as with its substrate. In addition, the functions of the lectin domains have been investigated. NMR, and isothermal titration calorimetry (ITC) have been used in addition to crystallography to investigate VCNA.

The structures of VCNA with substrates  $\alpha$ -2,3-sialyllactose (refined to 1.9 Å resolution) and  $\alpha$ -2,6-sialyllactose (1.6 Å), and complexes with four inhibitors: DANA (1.9 Å), FANA (2.1 Å), GANA (2.6 Å), and CANA (2.8 Å) have been obtained, where FANA, GANA, and CANA are derivatives of DANA (2-deoxy-2,3-didehydro-5-N-acetylneuraminic acid) modified at the 5-N-acetyl group.

The DANA-complex structure, in addition to exhibiting the inhibitor bound in the active site of the catalytic domain, surprisingly, displayed  $\alpha$ -sialic acid bound at the main groove of the N-terminal lectin domain, thereby revealing the function associated with this domain. Subsequently, the sialic acid-binding site in the N-terminal lectin domain has been confirmed in the two sialyllactose-complexes. Similarly, the complex structure of FANA (2-deoxy-2,3-didehydro-5-N-trifluoroacetylneuraminic acid) revealed the

hydrolysed product (N-trifluoroacetylneuraminic acid, or FNANA) bound at the sialic acid-binding site. Another surprise has been obtained from the structure of VCNA in complex with  $\alpha$ -2,3-sialyllactose: the refined structure revealed the inhibitor DANA bound at the neuraminidase catalytic site, indicating that the bacterial enzyme is able to synthesize its own inhibitor as shown before for the neuraminidases from influenza virus and *Trypanosoma rangeli*. Further evidence of DANA formation during the action of VCNA on its substrate has been obtained from  $^1\text{H}$  NMR described in the thesis. The affinity of the new sialic acid-binding site has been determined using the two thiosialosides (Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me and Neu5Ac-2-S- $\alpha$ -(2,6)-Glc $\beta$ Me) as ligands to give a  $K_d \approx 30 \mu\text{M}$ . Moreover, the interactions of the thiosialoside (Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me) with the lectin domain has been mapped using STD NMR (saturation transfer difference NMR spectroscopy). As the crystallographic structures revealed no clue for the ligand recognized by the C-terminal lectin domain, we carried out an NMR screening of a small carbohydrate library. Unfortunately, no evidence for any specific binding could be observed for any of the screened carbohydrates, so the nature of the ligand recognized by this domain remains a mystery.

A comparison is made of the catalytic site of VCNA with other neuraminidases whose structures have been determined, as well as an analysis of the recognition pattern of sialic acid by the N-terminal lectin domain and other sialic acid-binding lectins.

The information obtained from the crystal structures of VCNA with the various ligands deepens our understanding of how the cholera enzyme elicits its action, and forms the basis for the design of more potent inhibitors. Finally, a model describing the dual function of VCNA *in vivo* is discussed.

## Acknowledgement

I would like to thank Prof. Garry Taylor, my supervisor, for his patience and continued support throughout my Ph.D. and during my time in St-Andrews. Especial gratitude is due to Margaret Taylor and Dr. Helen Connaris for their help in obtaining the pure protein. I also wish to acknowledge the invaluable help, teaching and friendship of Dr. Viatcheslav Zaitsev who has an immense knowledge of crystallography. It has been my pleasure to work with everybody in Garry's group, many thanks for their friendship and stimulating conversation.

I wish to thank Prof. Mark von Itzstein for giving me the opportunity to carry out the NMR study using the NMR facility in his medicinal chemistry lab. in Institute of Glycomics, Griffith University, Australia. Especial thanks should be given to Dr. Jennifer Wilson for her great help in NMR, and to Dr. Milton Kiefel and Dr. Robin Thomson in Mark's group for providing me with the compounds used in the NMR work. I also wish to thank Dr. Malcolm White for the informative discussions and for using the ITC facility in his lab. in St-Andrews.

Thanks should also be given to the Egyptian Government and NODCAR, for supporting and financial assistance to be given the opportunity to study for my Ph.D. in the United Kingdom. My special thanks to Prof. Abdulla Molokhia, and Prof. Mohamed Badawy in NODCAR, and to Prof. Rifaat Hilal in Cairo University, without whose help none of this could have happened. Thanks should also be given to my colleagues in NODCAR.

Last, but by no means least, my thanks must also be given to family, especially my wife Raghda, without whose patience, support and encouragement I would have found it very difficult to complete this Ph.D.

## Abbreviations

$^1\text{H}$ NMR	Proton NMR
3D	Three dimensions
aa	amino acids
AAA	<i>Aleuria aurantia</i> agglutinin
AMP	Adenosine monophosphate
<i>C. perfringens</i>	<i>Colistridium perfringens</i>
cAMP	Cyclic AMP
CANA	2-deoxy-2,3-didehydro-5-N-chloroacetylneuraminic acid
CCD	Charged-coupled devices
CHA	<i>Capaea hortensis</i>
COSY	Correlation Spectroscopy
CT	cholera toxin
CTX	genetic element coding CT
CTX $\phi$	Bacteriophage coding CTX genes
cv	Cross validation
DANA	2-deoxy-2,3-didehydro-5-N-acetylneuraminic acid
FANA	2-deoxy-2,3-didehydro-5-N-trifluoroacetylneuraminic acid
FID	Free induction decay
FNANA	N-trifluoroacetylneuraminic acid
FT	Fourier transform
Gal	Galactose
GANANA	2-deoxy-2,3-didehydro-5-N-glycolylneuraminic acid
Glc	Glucose
GlcNAc	N-acetylglucosamine

GMI	Ganglioside
Gs $\alpha$	heterotrimeric G-protein
HA	Haemagglutinin
HN	haemagglutinin-neuraminidase
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
kDa	kilo Dalton
LFA	<i>Limax flavus</i>
M	Molar concentration
<i>M. decora</i>	<i>Macrobdeella decora</i>
<i>M. viridifaciens</i>	<i>Micromonospora viridifaciens</i>
MAL	<i>Maackia amurensis</i> leukoagglutinin
MHC	Major histocompatibility complex
MLF	Maximum-likelihood amplitude function
MR	Molecular replacement
mtf	Molecular topology file
NANA	N-acetylneuraminic acid
NDV	Newcastle disease virus
NMR	nuclear magnetic resonance
NOE	Nuclear Overhauser effect
O.D	Optical density
PAGE	Polyacrylamide gel
PEG	Polyethylene glycol
ppm	Part per million
PT	Pertussis toxin

rmsd	Root mean square deviation
rpm	Revolution per minute
<i>S. typhimorium</i>	<i>Salmonella typhimorium</i>
SDS	Sodium dodecyl sulfate
STD NMR	Saturation transfer difference NMR spectroscopy
<i>T. cruzi</i>	<i>Trypanosoma cruzi</i>
<i>T. rangeli</i>	<i>Trypanosoma rangeli</i>
TCP	toxin coregulated pilus
TeNT	Tetanus neurotoxin
TFA	<i>Tritrichomonas foetus</i>
Tris	Tris[hydroxymethyl]aminomethane
VCNA	<i>Vibrio cholerae</i> neuraminidase
VPI	vibrio pathogenicity island
VPI $\phi$	Bacteriophage coding VPI genes
VPI-2	vibrio pathogenicity island-2
WGA	Wheat germ agglutinin
WHO	World Health Organization

## Symbols

$\epsilon$	Absorption coefficient
$\delta$	Chemical shift
$\nu$	Frequency
$\lambda$	Wavelength
$(\phi, \psi)$	Main-chain dihedral angles

$\alpha(hkl)$	phase factor of reflection ( $hkl$ )
$\sigma(I)$	Error in intensity
$\rho(x, y, z)$	Electron density at point ( $x, y, z$ ) in the unit cell
$\alpha, \beta, \gamma$	Angles of unit cell
$\alpha, \beta, \gamma$	Eulerian angles
$\Delta G$	Gibbs free energy change
$\Delta H$	Enthalpy change
$\phi_n$	Phase factor
$\Delta S$	Entropy change
$(u, v, w), \mathbf{u}$	Interatomic vector
$ F_{hkl} $	Structure factor amplitude
$\langle \mu^2 \rangle$	Mean-square displacement of atomic vibrations
$\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$	Unit cell vectors in reciprocal space
$a, b, c$	Unit cell dimensions
$\mathbf{a}, \mathbf{b}, \mathbf{c}$	Unit cell vectors in real space
$B$	Temperature factor
$d_{hkl}$	Spacing between ( $hkl$ ) set of planes in real lattice
$F_c$	Calculated structure factor
$f_e$	Electronic scattering factor
$F_{hkl}$	Structure factor of $hkl$ reflection
$f_j$	Atomic scattering factor
$F_o$	Observed structure factor
$h, k, l$	Miller indices
$I(hkl)$	Intensity of ( $hkl$ ) reflection
$k$	Rate constant

$K_a$	Association constant
$K_d$	Dissociation constant
$K_i$	Inhibition constant
L	Probability likelihood
Q, $q$	heat released or absorbed in ITC
$R(\mathbf{R})$	Rotation function in MR
S	Wave vector
$T(\mathbf{t})$	Translation function in MR
$V_m$	Mathews' coefficient
$2\theta$	Scattering angle

### **Amino acids list (triple letter codes)**

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

<b>Contents</b>	<b>Page</b>
<b>Abstract</b>	i
<b>Acknowledgement</b>	iii
<b>Abbreviations</b>	iv
<b>Chapter 1 Introduction</b>	
1.1 Overview	1
1.2 Cholera: a brief history	2
1.2.1 Treatment	3
1.2.2 Virulence factors	4
1.2.3 Molecular basis of <i>V. cholerae</i> pathogenesis	4
1.3 Sialic acids	5
1.4 Neuraminidases	7
1.5 The enzyme of interest: VCNA	12
1.5.1 Historical perspective	12
1.5.2 Enzyme activity	12
1.5.3 Role of VCNA in cholera	13
1.5.4 Targeting VCNA	15
1.5.5 Native structure of VCNA	16
1.6 Structure-based drug design	19
1.7 Objective of the project study	20
<b>Chapter 2 X-ray Crystallography: Theory</b>	
2.1 Introduction	21
2.2 Diffraction theory	21
2.2.1 Crystal lattice and symmetry	21

2.2.2 Scattering of X-rays by a distribution of electrons	23
2.2.3 Diffraction pattern of a crystal	25
2.2.4 calculating the electron density	30
2.2.5 Reciprocal space and the Ewald sphere	31
2.2.6 Patterson function	33
2.2.7 Molecular replacement	34
2.2.8 Structure refinement	36
2.3 Crystallization	39
2.4 X-ray radiation sources	40
2.5 X-ray detectors	42

### **Chapter 3 Protein expression and purification**

3.1 Introduction	44
3.2 Materials and Methods	45
3.2.1 Original protocol	45
3.2.2 New expression system	47
3.2.3 Neuraminidase activity assay	48
3.2.4 Bradford assay	49
3.3 Results	50
3.3.1 Protein expression	50
3.3.2 Protein purification	51
3.4 Discussion	55

### **Chapter 4 Crystallization, Data Collection, Data processing,**

#### **Molecular Replacement and Refinement**

4.1 Introduction	56
4.2 Materials and methods	57

4.2.1 Crystallization	57
4.2.2 Data collection	59
4.2.2.1 Synchrotron data	60
4.2.2.2 In-house data	60
4.2.3 Data processing	61
4.2.4 Molecular Replacement (MR)	64
4.2.5 Model refinement	66
4.2.6 Model quality	68
4.3 Results	68
4.3.1 crystallization	68
4.3.2 Data collection and data processing	70
4.3.2.1 Space group and unit cell	71
4.3.2.2 Scaled data	74
4.3.3 Molecular Replacement	77
4.3.4 Model refinement	85
4.3.5 Structure validation	88
4.4 Discussion	107

## **Chapter 5 Structures of the VCNA-Complexes**

5.1 Introduction	110
5.2 Comparison between free and bound enzyme	111
5.2.1 Overall structure	111
5.2.2 Crystal packing	113
5.2.2.1 Packing pattern	113
5.2.2.2 Analysis of contacts within crystals	118
5.2.2.3 Nature of contact residues	121

5.3 VCNA-DANA complex	126
5.3.1 Catalytic site	129
5.3.2 Sialic acid-binding site	134
5.3.2.1 Interactions at the sialic acid-binding site	136
5.3.2.2 Specificity of the sialic acid binding	140
5.3.2.3 What about the C-terminal lectin domain?	143
5.3.3 Calcium ion and its role	146
5.3.4 Surface charge: implication for sialic acid recognition	149
5.4 VCNA-3'-sialyllactose complex	152
5.5 VCNA-6'-sialyllactose complex	155
5.6 VCNA complexes with DANA-derivatives modified at C5	160
5.6.1 VCNA-FANA complex	160
5.6.1.1 Binding of FANA at the catalytic site	162
5.6.1.2 Binding at the sialic acid-binding site	166
5.6.2 VCNA-GANA complex	169
5.6.3 VCNA-CANA complex	172
5.7 Implications of structures for the mechanism	176
5.8 Model of the VCNA action: the multifunctional enzyme	179
5.9 Discussion	181

## **Chapter 6 Comparison of VCNA with other neuraminidases**

### **And lectins recognizing sialic acids**

6.1 Introduction	183
6.2 Comparison of VCNA-DANA binding with other neuraminidases	184
6.2.1 Interactions of carboxylate group	185
6.2.2 Interactions of O4	196

6.2.3 Interactions of N-acetyl group	198
6.2.4 Interactions of glycerol group	199
6.3 Comparison with other sialic acid-binding lectins	201
6.3.1 Interactions of carboxylate group	202
6.3.2 Interactions of the anomeric hydroxyl and the ring oxygen	214
6.3.3 Interactions of O4	214
6.3.4 Interactions of N-acetyl group	214
6.3.5 Interactions of glycerol group and sugar ring	215
6.4 Discussion	219

## **Chapter 7 Isothermal Titration Calorimetric study on VCNA**

7.1 Introduction	224
7.2 Isothermal Titration Calorimetry: Overview	226
7.3 Materials and methods	229
7.3.1 ITC experiment	229
7.3.2 Competitive inhibition of VCNA	231
7.4 Results	232
7.4.1 Binding affinity of the thiosialosides towards VCNA	232
7.4.2 Competitive inhibition of thiosialosides <b>1</b> and <b>2</b>	244
7.4.3 Implications of inhibition results on ITC interpretation	247
7.5 Discussion	248

## **Chapter 8 NMR studies on VCNA**

8.1 Introduction	251
8.2 Background	251
8.2.1 Nuclear magnetic resonance (NMR): Overview	251
8.2.2 Protein-ligand interaction and STD NMR	253

8.3 Materials and methods	255
8.3.1 Materials	255
8.3.2 Methods	256
8.3.2.1 General	256
8.3.2.2 Hydrolysis of $\alpha$ -2,3-sialyllactose by VCNA	256
8.3.2.3 STD NMR of the Neu5,9Ac <sub>2</sub> -2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me	257
8.3.2.4 Screening a small library of sugars	258
8.4 Results	259
8.4.1 Hydrolysis of $\alpha$ -2,3-sialyllactose by VCNA	259
8.4.2 STD NMR of the Neu5,9Ac <sub>2</sub> -2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me	269
8.4.2.1 1D <sup>1</sup> H NMR spectrum of the ligand	269
8.4.2.2 The STD NMR spectrum	273
8.4.3 Screening of carbohydrate library by STD NMR	277
8.5 Discussion	288
<b>Appendix 1</b> Determination of the absorption coefficient ( $\epsilon_{280}$ ) of VCNA	292
<b>References</b>	293

# Chapter 1

## Introduction

### 1.1 Overview

Cholera is an ancient epidemic and pandemic worldwide disease caused by the intestinal infection with the bacterium *Vibrio cholerae*, characterized by watery diarrhoea leading to severe dehydration and death, if untreated properly. Seven pandemics have been recorded since 1817. The current seventh pandemic started in 1962 and involved almost the whole world. At the molecular level, the pathogenesis of cholera is a multifunctional process and involves several virulence factors secreted by the bacterium. One of these virulence factors is the *V. cholerae* neuraminidase (VCNA) enzyme, the subject of this thesis. Several studies have shown that VCNA augments the pathogenesis of *V. cholerae* by acting synergistically with cholera toxin (CT), the principal virulent factor in cholera. VCNA acts to cleave the terminal sialic acid from higher order gangliosides producing the GM1 ganglioside, the CT receptor, thereby enhancing the binding of the toxin to the enterocytes. Additionally, removal of the negatively charged sialic acid from the cell surface reduces the viscosity of the gastrointestinal mucus facilitating the access of the bacterium to the target epithelial cells. Therefore, interfering with the neuraminidase function of VCNA is thought to be an effective way to protect from cholera. Since its discovery, the cholera enzyme has been the target of extensive inhibitor studies. The study presented in this thesis was initiated in order to gain a deeper insight into the enzymatic action of VCNA at the molecular level using mainly the X-ray crystallographic technique. In addition, the NMR and isothermal titration calorimetric

techniques were applied in our study. A greater understanding of VCNA could help the development of better inhibitors that might form part of a treatment of the cholera disease, or at least provide specific tools to investigate the role of VCNA in pathogenesis.

A brief account of cholera, the pathogenesis of the disease and the current treatment, is presented in this chapter. Moreover, a detailed account of the enzyme under study and its role in the pathogenesis of cholera is provided. An overview of the neuraminidase family, and its diverse sialic acid substrate that is incorporated in so many biological functions is given in this introductory chapter. Finally, the objective of this study and how the thesis is arranged are described at the end of this chapter.

## **1.2 Cholera: a brief history**

Cholera was always much feared because it regularly occurred in epidemics with high mortality rates (Sack et al., 2004, and references therein). The first recorded cholera pandemic began in 1817 with spread of the disease outside the Indian subcontinent along trade routes to the west as far as southern Russia. A second pandemic started in 1826 and reached the major European cities by the early 1830s, continuing up to 1925, four other pandemics were recorded and involved Africa, Australia, Europe, and all the Americas. The epidemiological study of John Snow in 1854 in London associated the disease with contaminated drinking water. The causative agent *V. cholerae* was not identified until 1884 during the fifth pandemic (Koch, 1894), Koch postulated that the bacteria produced a toxin, which stimulates the massive loss of fluid from the intestine. The seventh pandemic began in Indonesia in 1961; the causative agent El Tor O1 biotype of *V. cholerae* (first isolated in 1905 in the El Tor village in Egypt) was identified. In 1992, a newly described non-O1 *V. cholerae* serogroup, designated O139 Bengal, caused unusual

outbreaks of cholera in India and Bangladesh (Sack et al., 2004). According to the WHO report in year 2000, nearly 120 countries have reported cases of cholera to WHO since 1991 (World Health Organization, 2000). Today, cholera still remains a burden mainly for underdeveloped countries, which cannot afford to establish or to maintain necessary hygienic and medical facilities (Reidl & Klose, 2002).

### **1.2.1 Treatment**

The main treatment of patients infected with *V. cholerae* is by replacing the amount of lost fluid and ions to prevent further dehydration, in most instances, Oral Rehydration Solution (ORS) is given to the patient. Without treatment the case-fatality rate for severe cholera is about 50% (Sack et al., 2004). Patients with severe cholera usually receive a course of antibiotics, however, antimicrobial resistance has limited the effectiveness of antimicrobial drugs (Sack et al., 2001), and increased the risk of *V. cholerae*.

In addition, vaccination against cholera was developed since the early days after its discovery; the whole cell-injectable vaccine was used during this time. Nonetheless, the vaccine was short-lived (6 months) and is no longer recommended for any purpose (Sack et al., 2004). However, developing a new bacterial vaccine containing a few selected antigens of the pathogen is an ongoing research area. Very recently, Stewart-Tull and co-workers demonstrated that vaccines containing *V. cholerae* mucinase complex (neuraminidase and proteinases) protected the animals from a live *V. cholerae* challenge (Stewart-Tull et al., 2004). Without doubt developing a safe chemotherapeutic agent against *V. cholerae* would be of great importance to protect from cholera and to lessen the fear from the eighth pandemic.

### **1.2.2 Virulence factors**

In *V. cholerae*, the major virulence genes required for pathogenesis are in clusters and can apparently propagate laterally and disperse among different strains. Three important genetic elements that distinguish a pathogenic from non-pathogenic strains, these are: (1) the CTX genetic element carrying the genes encoding cholera toxin (Waldor & Mekalanos, 1996), (2) the vibrio pathogenicity island (VPI), which carries genes for the pilus colonisation factor TCP (toxin coregulated Pilus) (Taylor et al., 1987), (3) the vibrio pathogenicity island-2 (VPI-2) encodes several gene clusters including genes required for the utilization of amino sugars as well as neuraminidase (Jermyn & Boyd, 2002). The first two elements are encoded by bacteriophages: CTX is the genome of a lysogenic bacteriophage designated CTX $\phi$  (Waldor & Mekalanos, 1996), and the VPI element is the genome of a filamentous bacteriophage designated VPI $\phi$  (Karaolis et al., 1999). It should be mentioned that the TCP element encoded by VPI $\phi$  functions as both a colonisation factor and a receptor for CTX $\phi$  (Karaolis et al., 1999). It has been shown that non-toxigenic *V. cholerae* strains can be converted by phage transduction with CTX $\phi$ , this could conceivably take place in the gastrointestinal environment yielding a new toxigenic strains (Reidl & Klose, 2002).

### **1.2.3 Molecular basis of *V. cholerae* pathogenesis**

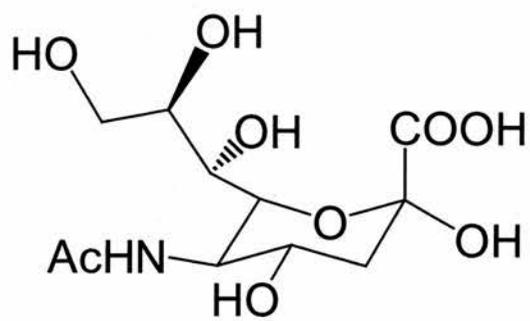
The pathogenesis of cholera is due almost totally to the action of CT, which acts by causing constitutive activation of adenylate cyclase leading to elevated cyclic AMP (cAMP) levels in intestinal epithelial cells (Dirita, 2001). Cyclic AMP activates protein kinase A, which causes the opening of ion channels in the membrane to trigger secretion

of the chloride and bicarbonate ions by intestinal crypt cells resulting in disruption in absorption by villus cells.

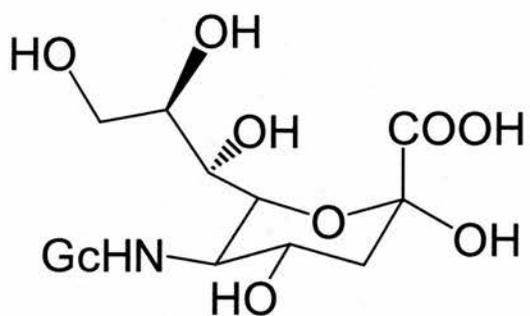
Modification of cAMP production by CT is a well-understood mechanism (Dirita, 2001). Cholera toxin is an AB<sub>5</sub> heterodimer composed of five identical 11 kDa B-subunits associated to form a pentameric ring-like structure, and the A-subunit (A1- and A2-chains linked via a disulfide bond and extensive non-covalent interactions) assembles non-covalently with the B-subunits (Zhang et al., 1995). The C terminus of the A2-chain protrudes through the central pore of the ring and tethers the A- and B-subunits together. The B-subunit is a lectin that binds to the ganglioside GM1 at the cell surface. The A1-chain enters the cytosol and acts as ADP-ribosyltransferase that modifies the heterotrimeric G-protein, Gs<sub>α</sub>, to activate adenylyl cyclase, which in turn induces the intestinal ion secretion seen in cholera. The number of CT-receptors (GM1) on the surface of the host cells can be increased by processing the higher gangliosides, which have a terminal sialic acid linked to the GM1 core, using the secreted *V. cholerae* neuraminidase enzyme. Degrading the higher gangliosides by removing the terminal sialic acids is an example of the implementation of sialic acid in the pathogenesis of microbes (Corfield, 1992). Sialic acids and the enzymes specialized in processing these ubiquitous carbohydrates are discussed in the following two sections.

### **1.3 Sialic acids**

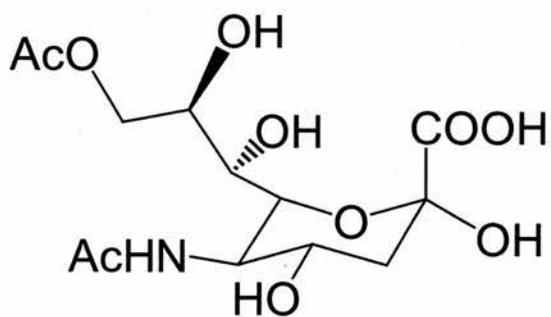
The term 'sialic acid' is a generic term comprising a wide family of compounds, more than 40 members, arising from different substitutions at the backbone of the nine carbon-



1 Neu5Ac, Ac = COCH<sub>3</sub>



2 Neu5Gc, Gc = COCH<sub>2</sub>OH



3 Neu5,9Ac<sub>2</sub>

**Scheme 1.1** The most common natural sialic acids

carbohydrate neuraminic acid or Neu5Ac **1**, (Scheme 1.1) (Angata & Varki, 2002; Schauer, 1995). Neuraminic acid is the most ubiquitous sialic acid and is the biosynthetic precursor for all other sialic acids (Varki, 1992). Many of these sialic acids are found specifically in a given tissue, or subcellular compartment (Schauer, 1995), and usually found as  $\alpha$ -glycoside occupying the non-reductive terminal (or side chain position) of hetero-oligosaccharides in glycoproteins and glycolipids (Maru et al., 2002). Among the members of the sialic acid family, three members are the most common in nature: N-acetylneuraminic acid (Neu5Ac **1**), N-glycolylneuraminic acid (Neu5Gc **2**), and acetylated neuraminic acid (Neu5,9Ac<sub>2</sub> **3**) (Schauer, 1995; Varki, 1992).

The terminal position of sialic acids means that they are appropriately presented for recognition and implication in many biological processes. Because of its negative charge ( $pK=2$ ), sialic acid is incorporated in binding and transport of positively charged compounds, and in aggregation (via  $Ca^{+2}$  bridges) or disaggregation (due to repulsive forces) of cells. They affect the immune system by acting as a mask or as a recognition site on molecules and cells. Various normal cellular processes are accompanied by changes (qualitative or quantitative) in the sialic acid profile of cells, such as activation, differentiation, maturation, cell growth, and hormonal regulation (Reuter & Gabuis, 1996; Pilatte et al., 1993; Schauer, 1985). Moreover, sialic acids are essential components of receptors for certain pathogens producing neuraminidase enzymes (Corfield, 1992), as discussed below.

#### **1.4 Neuraminidases**

Neuraminidases or sialidases (EC.3.2.1.18), both terms are used interchangeably, are a family of glycohydrolyase enzymes that catalyse the cleavage of the  $\alpha$ -ketosidically

linked non-reducing sialic acids from various polysaccharide, glycolipids, and glycoproteins (Yu, 1995). Neuraminidases have a wide distribution among viruses, bacteria, fungi, mycoplasma, protozoa and mammalian species. Indeed, neuraminidases are considered to form a superfamily with a common origin in higher animals (Roggentin et al., 1993). The idea of neuraminidase superfamily receives support from the fact that neuraminidases share a common catalytic domain of  $\beta$ -propeller structure first observed in influenza enzyme (Taylor, 1996). Among the diverse neuraminidases, those from viruses and bacteria are the most studied enzymes; several mammalian neuraminidases have been cloned and sequenced, for example from plasma membrane (Miyagi et al., 1999), lysosomes (Bonten et al., 1996), and the major histocompatibility complex (MHC) (Milner et al., 1997), although those latter two turn out to be the same gene that encodes a lysosomal enzyme.

As neuraminidases are being associated with the removal of terminal sialic acids decorating the cell surface in several tissues, they are incorporated in various biological functions in animals (Schauer, 1985). Regulating the sialic acid profile of cells is required for various functions, such as the immune system (Pilatte et al., 1993), dictating the half-lives of circulating cells (Bratosin et al., 1995), and apoptosis (Malisan & Testi, 2002). For non-mammalian enzymes, neuraminidases have been shown implicated as virulence factors in a range of pathogens, for example: influenza virus (the cause of influenza), Newcastle disease virus (respiratory disease of chicken), parainfluenza viruses (respiratory disease of humans), *V. cholerae* (cholera), and *Clostridia* (gas gangrene, peritonitis), for a more complete list of diseases implementing neuraminidases as virulence factors see the review (Taylor, 1996). More recently Rosenberg (Rosenberg, 2004) proposed a new direction in neuro-AIDS research, therein the author put an assumption that desialylation of the virion surface and the target cell surface by a

microbial neuraminidase each, additively, greatly enhance HIV-1 mediated syncytium formation and virus binding and entry.

Additionally, many pathogenic and non-pathogenic neuraminidase-producing bacteria can use sialic acid as a carbon and energy source and possess both permeases to transport the sugar inside the cell, and a cascade of enzymes for its catabolism (Corfield, 1992).

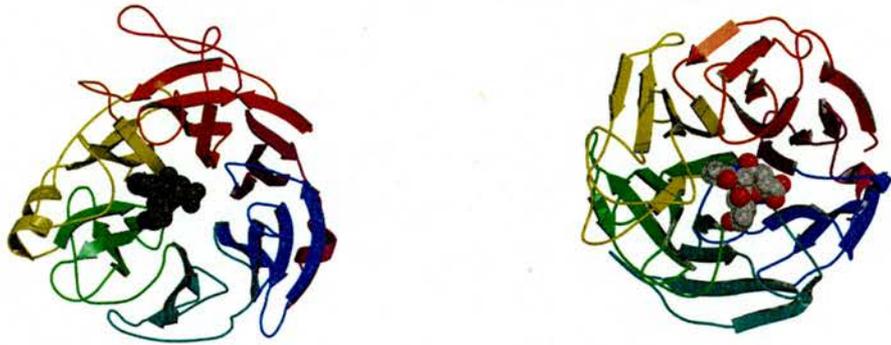
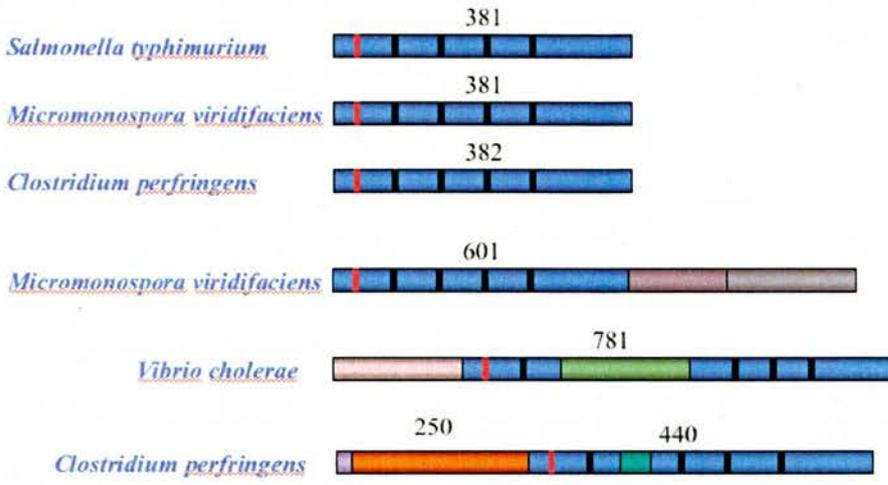
Beside neuraminidases, which process the terminal non-reducing sialic acids, there are trans-neuraminidases produced by certain parasites that catalyse the transfer of carbohydrate-linked sialic acid to other carbohydrates. Examples for these enzymes are trans-neuraminidases from: *Macrobodella decora*, *Trypanosoma cruzi* (Taylor, 1996).

Some members of the neuraminidase family are associated with dual functions; of these are the haemagglutinin-neuraminidases (HNs) of paramyxoviruses. These enzymes are presented on the outer membrane of the virion; beside their catalytic cleavage of the terminal sialic acid (neuraminidase function) they induce the attachment of the virus to the cell surface via sialic acid recognition (haemagglutination function). These dual functions are hypothesized to be associated with a single site within the  $\beta$ -propeller fold of these enzymes (Crennell et al., 2000; Lawrence et al., 2004; Zaitsev et al., 2004). Additionally, neuraminidases from *V. cholerae* (Crennell et al., 1994), *Micromonospora viridifaciens* (Gaskell et al., 1995), *Macrobodella decora* (Luo et al., 1998) and *Trypanosoma rangeli* (Amaya et al., 2003) contain lectin domains, which are associated with carbohydrate recognition. We have elucidated the function associated with one of the two-lectin domains in *V. cholerae* in our study discussed in chapter 5.

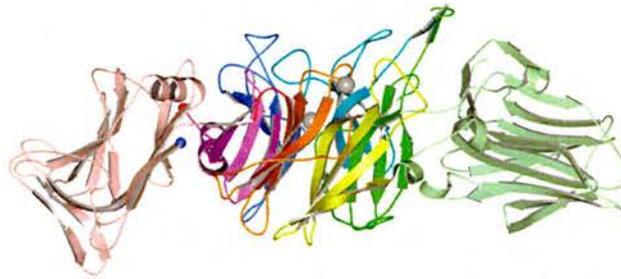
Non-viral neuraminidases show a large range of sizes, from small neuraminidases as those from *S. typhimurium* and *M. viridifaciens* (~40 kDa) up to over 100 kDa as in *C. perfringens*. Also, neuraminidases exhibit variations in their activity and substrate

specificity. The turnover of bacterial neuraminidases is typically 1000 times that of the viral enzymes (Taylor, 1996). The nature of the glycosidic linkage and the structure of the penultimate sugar of the sialoglycoconjugate dictate the specificity of the enzymes. Such variations can be demonstrated in the following examples: the neuraminidase from *V. cholerae* cleaves sialic acids with both  $\alpha$ -2,3/6 linkage with more preference for  $\alpha$ -2,3 over  $\alpha$ -2,6 (Corfield et al., 1983), the neuraminidase from influenza and *S. typhimurium* showed greater specificity towards sialic acids with the  $\alpha$ -2,3 linkage (Hoyer et al., 1991), whereas the neuraminidase from *Micromonospora viridifaciens* cleaves all sialic acid linkages ( $\alpha$ -2,3/6/8) with more preference towards the  $\alpha$ -2,6 linkage (Sakurada et al., 1992), however the enzyme from *Arthobacter ureafaciens* has a unique property that it cleaves internal sialic acids besides the terminal one (Iwamori et al., 1997).

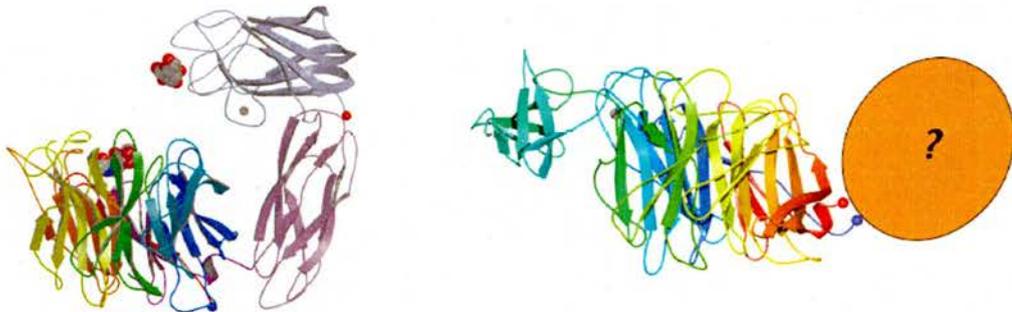
The 3D structures of the neuraminidases from the following sources have been determined (see figure 1.1): *V. cholerae* (Crennell et al., 1994), *Salmonella typhimurium* (Crennell et al., 1996), *Micromonospora viridifaciens* (Gaskell et al., 1995), *Trypanosoma rangeli* (Amaya et al., 2003), influenza virus (Burmeister et al., 1993), Newcastle disease virus haemagglutinin-neuraminidase, or HN-NDV, (Crennell et al., 2000), *Macrobodella decora* (Luo et al., 1998), *Trypanosoma cruzi* (Buschiazzo et al., 2002), and HN from parainfluenza virus type III (Lawrence et al., 2004); figure 1.1 displays examples of the structures of the above neuraminidases. The 3D structures of these enzymes revealed that they share the same topology of the catalytic  $\beta$ -propeller domain and well-conserved residues required for the catalytic functions as will be discussed in chapter 6. In addition, the non-viral neuraminidases share a recurring motif known as 'Asp-box' and a RIP/RLP motif; the latter include one of the catalytic arginine triad, characteristic for all neuraminidases.



*S. typhimurium* neuraminidase, 2SIM    *M. viridifaciens* neuraminidase (small), 1EUS

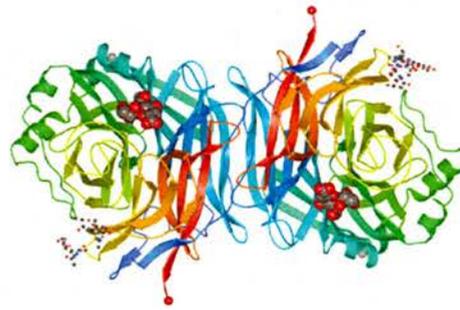
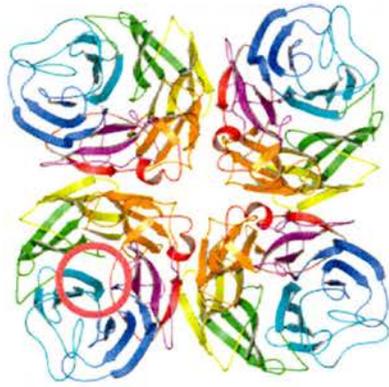


*V. cholerae* neuraminidase, 1KIT



*M. viridifaciens* neuraminidase (large), 1EUU    *C. perfringens* neuraminidase (unpublished)

**Figure 1.1** (continued)



Influenza virus neuraminidase, 1NSD    Haemagglutinin-neuraminidase NDV, 1E8V

**Figure 1.1.** The top displays the schematic representation of the amino acids sequences of non-viral neuraminidases: small neuraminidases from *S. typhimurium*, *C. perfringens*, *M. viridifaciens*, and large neuraminidases from *C. perfringens*, *M. viridifaciens* and *V. cholerae*. The conserved motifs Asp-boxes and RIP/RLP are shown as black and red bars in the schematic diagram, the bars in the diagram are colour coded to match the colours of corresponding domains in the secondary structures representations. The 3D structures of some members of the neuraminidase family are shown in the figure, the bound inhibitor DANA is shown in some of the structures as space-filling model. The  $\beta$ -sheets of the structures are coloured differently. The structure of the large sialidase from *C. perfringens* is incomplete, as the structure of a whole domain is still to be determined (personal communication, Newstead, S., St-Andrews University). The pdb-codes of the drawn structures are given in the figure.

## 1.5 The enzyme of interest: VCNA

### 1.5.1 Historical perspective

Early in 1940s, George Hirst noticed that allantoic fluid from chick embryos infected with influenza virus agglutinate red blood cells in the cold; when warmed to 37°C the cells dispersed as the virus eluted and they could not be re-agglutinated with fresh virus (Hirst, 1941). This suggested that the virus possessed an enzyme whose substrate was the receptor for the virus on the surface of red cells. The ability of culture fluids of *V. cholerae* to render red blood cells inagglutinable by influenza virus was first noticed by Burnet & Stone (Burnet & Stone, 1947), the active principle was described as the receptor destroying enzyme (RDE). Gottschalk & Lind in 1949 provided early evidence about the chemical activity of RDE, but the enzyme was not characterized as a glycosidase until

1956 (Gottschalk & Lind, 1949; Heimer & Meyer, 1956). The enzyme has been termed neuraminidase or sialidase (EC.3.2.1.18), and its action defined as 'the hydrolytic cleavage of the glycosidic bond joining the keto group of N-acetylneuraminic and D-galactosamine' (Gottschalk, 1957). By this time other neuraminidases, beside influenza enzyme, from viral origins: Newcastle disease, fowl plague and mumps were known. To clarify the function of neuraminidase for the virus-host cell relationship and because of the limited amount of viral material at that time, attempts were made to isolate the enzyme from the culture fluids of *V. cholerae* (Schramm & Mohr, 1959). The interest in the enzyme continued and several studies followed to characterize the physical and kinetic properties of the cholera enzyme, VCNA showed maximum activity at pH 5.6 and a value of 1.2 mM for the Michaelis constant ( $K_m$ ) towards the sialyllactose substrate (Pye & Curtain, 1961; Ada et al., 1961). The secreted enzyme has been shown to be a part of a mucinase complex, which also contains endo- $\beta$ -N-acetylhexosaminidase and proteinase activity (Stewart-Tull et al., 1986).

### 1.5.2 Enzyme activity

Studies on the cholera enzyme showed that its activity is stimulated by divalent cations ( $\text{Ca}^{+2}$ ,  $\text{Mn}^{+2}$ ,  $\text{Co}^{+2}$ , and  $\text{Mg}^{+2}$ ), calcium ions demonstrated the maximum stimulation at 1 mM  $\text{CaCl}_2$ , and the enzyme activity is inhibited by EDTA at 10 mM or greater concentrations (Ada, et al., 1961; Holmquist, 1975). An early study also showed that sialic acid itself, when used at high concentration, had some inhibition activity against VCNA (Mohr, 1960). Interesting phenomena were noted for the activity of VCNA. In 1974, Sedlacek & Seiler (Sedlacek & Seiler, 1974) observed that VCNA remains attached to the membrane of neuraminidase-treated cells even after repeated four

times washing. Noticeably, the rate of the enzymatic action on its aggregated substrate, gangliosides in the micelles form, was observed to be increased by up to 30-fold compared to the free substrate (Venerando et al., 1982). Nonetheless, the available data at that time were insufficient to explain these phenomena.

### **1.5.3 Role of VCNA in cholera**

The interest in the cholera enzyme increased as several studies showed that VCNA enhances the pathogenesis of *V. cholerae* and acts as a virulence factor. Stark and Wiegand (Stark & Wiegand, 1974) observed that when canine intestinal loops were pre-treated with VCNA there was an increase in the output fluid after an incubation with CT, they suggested that VCNA released by the bacteria attacks intestinal glycoproteins and gangliosides unmasking receptor sites for CT. Later on in their work Gascoyne *et al.* supported the same idea (Gascoyne & van Heyningen, 1979). Schneider and Parker (Schneider & Parker, 1982) postulated the role of mucinase complex enzymes (VCNA is secreted as a part of this complex) in degrading the intestinal mucin to increase vibrio penetration of the mucus and adherence to the small intestine. Therein, the authors proposed that sialic acids released from mucin by VCNA can be used as nutrients for the bacteria, which possess N-acetylneuraminidase to produce pyruvic acid for energy. In their work, Snyder & Walker (Snyder & Walker, 1987) have presented evidence suggesting that mucin in the small intestine has toxin binding sites involving sialic acids which mimic GM1 moieties; such competitive binding sites should effectively be destroyed by VCNA. An exclusive evidence to clarify the role of VCNA as a virulence factor in the pathology of cholera came from genetic studies by cloning isogenic strains of the *nanH* gene, encoding VCNA, which differ only in the ability to synthesise this

particular enzyme (Vimr et al., 1988). It was concluded that VCNA plays a subtle but significant role in the binding and uptake of CT by susceptible cells (Galen et al., 1992). Moreover, as mentioned before (section 1.2.2) in a recent genetic study it has been shown that the *nanH* gene is present in all toxigenic strains and is absent from all non-toxigenic strains of *V. cholerae* (Jermyn & Boyd, 2002).

#### 1.5.4 Targeting VCNA

The hypothesis that VCNA works as a virulence factor encouraged the synthesis of small molecule inhibitors for the enzyme (Chan et al., 1997; Zbiral, et al., 1989; Holzer et al., 1993; Khorlin et al., 1970; Meindl et al., 1974; Meindl & Tuppy, 1969; Nohle et al., 1985; Schreiner et al., 1991a; Schreiner et al., 1991b; Vorwerk & Vasella, 1998; Wilson et al., 2000). Most of these tested inhibitors were derived from the inhibitor 2,3-didehydro-2-deoxy-5-N-acetylneuraminic acid (Neu5Ac2en, also known as DANA). In the beginning, the structural similarity of DANA to sialic acids (they differ only by one double bond between C2-C3) was the motivation to find a specific neuraminidase inhibitor based on the chemistry of DANA. Thereafter, Chong *et al.* (Chong et al., 1992), in their study on the influenza neuraminidase, provided evidence that the hydrolysis of substrate by neuraminidases proceeds via formation of an endocyclic sialosyl cation transition-state intermediate close to the structure of DANA. Also, the mechanistic study of VCNA showed that the enzyme cleaves the terminal  $\alpha$ -ketosidically linked sialic acid with full retention of configuration and formation of an oxocarbenium ion in the transition state (Guo & Sinnott, 1993). This generated a considerable interest in designing neuraminidase inhibitors based around the transition-state-analogue DANA. The best hit

against VCNA tested so far is 2,3-didehydro-2-deoxy-5-N-trifluoroacetylneuraminic (or FANA) with a  $K_i$  value of 1.8  $\mu\text{M}$ .

### 1.5.5 Native structure of VCNA

Non-viral neuraminidases including VCNA show little overall sequence similarity, apart from the recurring motif composed of an eight amino acids known as an 'Asp-box' (Ser/Thr-X-Asp-X-Gly-X-Thr-Trp/Phe), which repeats three to five times along the sequences, and the RIP/RLP motif (Roggentin et al., 1989). The cholera enzyme has been characterized as a large neuraminidase (runs from amino acids 25 to 781, 83 kDa; amino acids 1 to 24 work as a signal sequence) sharing 16-24% sequence identities with other neuraminidases and possesses four consensus Asp-boxes. Figure 1.2 displays the compared repeated conserved amino-acid sequences in the neuraminidases of *Clostridium sordelli* G12, *C. perfringens* A99, *Salmonella typhimurium* LT-2 and *V. cholerae* 395 as shown by Roggentin et al (Roggentin et al., 1989). Indeed, the Asp-box motif has been identified in other protein families rather than neuraminidases, but yet the function associated with these motifs is uncertain (Copley et al., 2001).

Solving the structure of the native VCNA using X-ray crystallography was a major step forward in the study of this enzyme. The structural data in conjunction with the outlined accumulative facts about VCNA was very informative. The structure (figure 1.3) was solved to a resolution of 2.4 Å at room temperature (Crennell et al., 1994), and showed that the enzyme, 757 amino acid residues (aa 25-781), has three domains: a central  $\beta$ -propeller domain, the canonical form in all neuraminidases and two flanking lectin-like domains with identical topologies though they share only 23 % sequence identity. The first lectin-like domain (wing1) runs from the amino terminus at residue 25

G12	+89	Ala	Lys	Ser	Thr	Asp	Asn	Gly	Gln	Thr	Trp	Asp	Tyr
A99	+71	Ala	Arg	Ser	Thr	Asp	Asn	Gly	Lys	Thr	Trp	Ser	Tyr
LT-2	+71	Ala	Arg	Ser	Thr	Asp	Gly	Lys	Lys	Thr	Trp	Asn	Lys
395	+263	Arg	Thr	Ser	Arg	Asp	Gly	Gly	Ile	Thr	Trp	Asp	Thr
G12	+158	Val	Tyr	Ser	Asp	Asp	Asn	Gly	Glu	Thr	Trp	Ser	Asp
A99	+140	Ile	Tyr	Ser	Asp	Asp	Asn	Gly	Leu	Thr	Trp	Ser	Asn
LT-2	+145	Tyr	Lys	Ser	Thr	Asp	Asp	Gly	Val	Thr	Phe	Ser	Lys
395	+585	Ile	Tyr	Ser	Asp	Asp	Gly	Gly	Ser	Asn	Trp	Gln	Thr
G12	+226	Ile	Tyr	Ser	Lys	Asp	Asn	Gly	Glu	Thr	Trp	Thr	Met
A99	+208	Ile	Tyr	Ser	Lys	Asp	Asn	Gly	Glu	Thr	Trp	Thr	Met
LT-2	+210	Ile	Tyr	Ser	Thr	Asp	---	Gly	Ile	Thr	Trp	Ser	Leu
395	+653	Phe	Leu	Ser	Lys	Asp	Gly	Gly	Ile	Thr	Trp	Ser	Leu
G12	+273	Tyr	Ile	Ser	Tyr	Asp	Met	Gly	Ser	Thr	Trp	Glu	Val
A99	+255	Tyr	Ile	Ser	His	Asp	Leu	Gly	Thr	Thr	Trp	Glu	Ile
LT-2	+254	Phe	Glu	Thr	Lys	Asp	Phe	Gly	Lys	Thr	Trp	Thr	Glu
395	+718	Trp	Phe	Ser	Phe	Asp	Glu	Gly	Val	Thr	Trp	Lys	Gly

**Figure 1.2.** Comparison of the conserved, repeated-amino-acid sequences in the neuraminidases of: *Clostridium sordelli* G12, *C. perfringens* A99, *Salmonella typhimurium* LT-2 and *V. cholerae* 395. The five conserved amino acids in the Asp-box (Ser/Thr-X-Asp-X-Gly-X-Thr-Trp/Phe) are coloured differently (Roggentin et al., 1989).

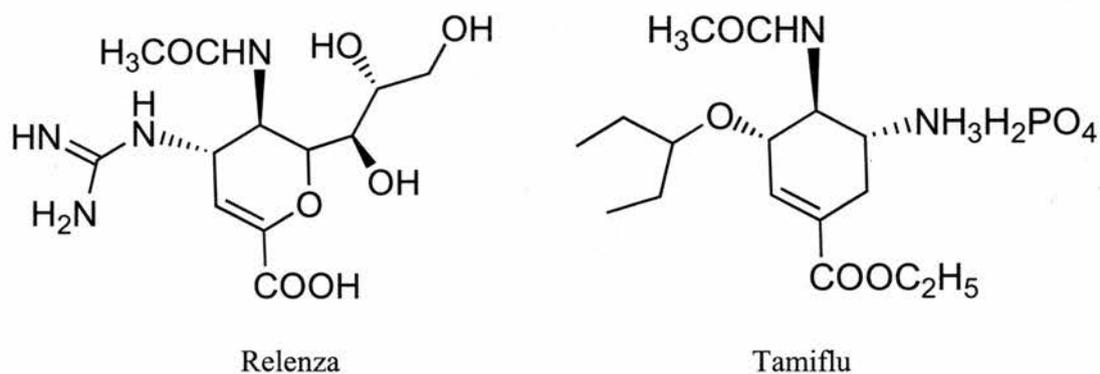


**Figure 1.3.** Ribbon representation of the 2.4 Å resolution structure of the native VCNA (pdb, 1KIT) (Crennell et al., 1994). The calcium ions are drawn as spheres coloured in magenta, the C $\alpha$  of the C- and N-termini are shown as red and blue spheres, respectively.

to residue 216. The chain departs from the central catalytic domain to form the second lectin-like domain (wing2), which runs from residues 347-543 inserted between the second and third  $\beta$ -sheets of the  $\beta$ -propeller domain. These lectin domains suggested carbohydrate recognition, but their exact function(s) and the true substrates for these lectins were to be established. The reported structure demonstrated the role of calcium ions, which stimulates the enzyme activity (section 1.5.2), in that they support structural elements carrying important catalytic residues. In the original reported structure of VCNA, a complex with DANA was only obtained at 4.5 Å resolution data collected from crystals soaked in DANA. Comparison with the high-resolution structure of *S. typhimurium* neuraminidase complexed with DANA (Crennell et al., 1993), allowed the active site interactions to be tentatively assigned. However, solving the complex structure of VCNA with DANA would be of great importance for a deeper understanding of the action of this enzyme.

## 1.6 Structure-based drug design

The ultimate aim in developing a new drug is to have a bioactive compound that interacts selectively and specifically with a certain receptor. Knowing the 3D structure of such a receptor, either free or more preferably in complex with its ligand, deepens our understanding for the system being investigated. Dissecting the interaction between the receptor and its ligand enabled researchers to approach the problem of how to interfere with the function of the targeted receptor more rationally. Such a rational approach has been demonstrated successfully for the influenza neuraminidase and resulted in the development of highly potent inhibitors for the enzyme ( $K_i$  approaching subnanomolar range) (Colman, 2002). The outcome of this approach was the marketed two drugs: Zanamivir (or Relenza), and Oseltamivir (or Tamiflu) (scheme 1.2). Relenza was developed based on the chemistry of the sialic acid analogue DANA (von Itzstein, et al., 1993), and Oseltamivir was based on the chemistry of the carbocyclic DANA analogue (Kim, et al., 1997).



**Scheme 1.2** Chemical structures of Relenza & Tamiflu drugs, which target the influenza neuraminidase.

## 1.7 Objective of the project study

Following the well-studied example of influenza neuraminidase, we undertook a crystallographic study of the *V. cholerae* enzyme, with a view to developing a potent inhibitor that could be used as a potential drug for cholera. Herein, we aimed to solve the crystal structure of VCNA in complex with the substrates  $\alpha$ -2,3- and  $\alpha$ -2,6-sialyllactose in addition to other four enzyme inhibitors (DANA, FANA, GANA, and CANA). Meanwhile, we aimed to elucidate the function(s) of the two lectin domains of the cholera enzyme. Whilst pursuing this study, we applied other techniques, nuclear magnetic resonance (NMR) and isothermal titration calorimetry (ITC), beside the crystallographic technique.

The thesis has been arranged mainly based on the technique being discussed in the different chapters. Because of the size of the crystallographic part, it has been distributed in different chapters (2, 3, 4, and 5). In the next chapter, we present the required theoretical background for protein crystallography, the main technique applied in our study. Chapter 3 presents the work carried out to make the protein required for the different experiments. In Chapter 4, we present the crystallographic data, which are discussed in Chapter 5. We devoted Chapter 6 to compare the main results obtained for the cholera enzymes with other related enzymes (neuraminidases and sialic acid-binding lectins). Chapter 7 presents and discusses the ITC experiment, and finally, Chapter 8 describes the NMR results.

## Chapter 2

### X-ray Crystallography: Theory

#### 2.1 Introduction

X-rays are a form of electromagnetic radiation of wavelength of the order (1-10 Å), which is comparable to the intermolecular spacing in a crystal. When an X-ray beam is incident on a crystal, diffraction is observed. The regularity of the three-dimensional structure acts to reinforce the scattering caused by any one molecule, so that a detectable diffraction effect can be recorded and analysed. From the nature of the diffraction, both the crystal and the molecular structure can be derived (Sherwood, 1976).

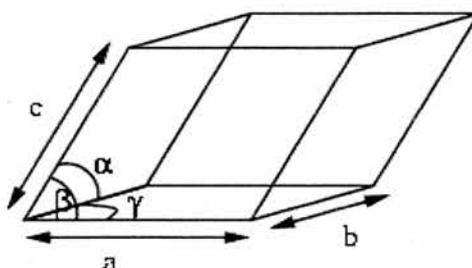
Disclosing the 3D-structural information of a crystallized molecule in an X-ray experiment requires: (1) a single crystal diffracting to high quality resolution, (2) a source that generates a monochromatic beam of X-rays, (3) a detector to record the diffraction pattern (raw data) of the diffracted rays, (4) data analysis and subsequent manipulation of the data employing standard computing programs to extract the structural information hidden in the diffraction pattern. This chapter presents the theoretical background of: the X-ray diffraction theory, growing single crystals, generation and detection of the X-rays.

#### 2.2 Diffraction theory

##### 2.2.1 Crystal lattice and symmetry

Crystals are found to have a repeating unit, the unit cell, with basis vectors **a**, **b**, and **c**, and angles  $\alpha$ ,  $\beta$ , and  $\gamma$  between them, see figure 2.1. The enormous number of unit cells in a crystal are stacked in three dimensions with the origins of the unit cells forming a grid

or lattice. Lattice planes are designated by Miller indices; a set of planes with  $(hkl)$  indices are dividing the vectors  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$  of the unit cell into  $h$ ,  $k$ , and  $l$  equal parts, in the same order (Drenth, 2001).



**Figure 2.1.** Unit cell with axes  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$ ; the angles between the axes are  $\alpha$ ,  $\beta$ , and  $\gamma$  as defined in the figure.

The symmetry operations can be identified as an operation which, when applied, results in a structure indistinguishable from the original structure. The symmetry elements in crystals include: rotation axes, mirror planes, and centre of inversion. Adding translation operation results in the symmetry elements of screw axes and glide planes. Not all combinations of symmetry elements are allowed, and there are 230 combinations (space groups). Space groups with mirror planes and/or inversion centres are not allowed for protein crystals, which crystallize only in 65 enantiomorphic space groups. Some of the space groups supply more favourable packing conditions for proteins; the most favoured space group is  $P2_12_12_1$ . By regarding only the symmetry elements of: rotation axes, mirror planes, and inversion centres; there are 32 different combinations of symmetry operations (point groups) passing through one point that not affected by symmetry operations (point group symmetry). Only 11 enantiomorphic point groups are allowed for protein crystals. The point groups are subdivided into seven crystal systems: cubic, hexagonal, trigonal, tetragonal, orthorhombic, monoclinic, and triclinic based on a

chosen set of axes displaying the external symmetry. The unit cell can be either primitive (*P*) or centred (*C, I*), leading to 14 types of different unit cells called Bravais lattices.

### 2.2.2 Scattering of X-rays by a distribution of electrons

The scattering of an X-ray beam by a crystal results from interaction between the electric component of the beam and the electrons in the crystal. When a monochromatic beam hits an electron, the electron starts to oscillate in the direction of the electric vector of the beam. This oscillating electron acts as the aerial of a transmitter and radiates X-rays with the same frequency (elastic or Thomson scattering) or lower frequency (inelastic or Compton scattering) as the incident beam. Compton scattering contributes to the background in a diffraction experiment, and Thomson scattering is responsible for the interference effect in the diffraction (Drenth, 2001).

The electromagnetic theory showed that the average value of the amplitude of the electric component of the scattered wave  $E_{scat}$  is related to that of the incident beam  $E_{in}$  as:

$$\frac{E_{scat}}{E_{in}} = \frac{e^2}{4\pi\epsilon_0 r m c^2} \frac{(1 + \cos^2 2\theta)}{2} \quad (2.1)$$

where

$\epsilon_0$  = permittivity constant of magnitude  $8.854 \times 10^{-12} \text{ Fm}^{-1}$

$c$  = the velocity of electromagnetic radiation

$e$  = electron charge

$m$  = mass of the electron

$r$  = the distance from the electron at which the scattered wave is sampled

$2\theta$  = the angle through which the incident X-rays are scattered

Although Thompson model of scattering assumed that the electron is perfectly free and responds passively to the incident beam, it can be applied to the electrons bound to atoms given that the frequency of the incident beam is much greater than the characteristic frequency of the electron around its nucleus. Defining the quantity  $f_e$  as:

$$f_e = \frac{e^2}{4\pi\epsilon_0 r m c^2} \quad (2.2)$$

we have,

$$\frac{E_{scat}}{E_{in}} = f_e P(2\theta) \quad (2.3)$$

where  $P(2\theta)$  is some function of the scattering angle  $2\theta$  only, assuming that the calculation is carried out for an incident wave in a defined state of polarization. The function  $P(2\theta)$  is called the polarization factor, and if we disregard this, Equation 2.3 can be rewritten as:

$$E_{scat} = f_e E_{in} \quad (2.4)$$

bearing in mind that in any real case the polarization factor must be inserted as appropriate. The quantity  $f_e$  is known as the electronic scattering factor which represents an intrinsic property of the electron.

Equation 2.4 can be extended to give the total scattered wave amplitude from an array of electrons as

$$\frac{E_{scat}}{E_{in}} = \sum_n f_e e^{i\phi_n} \quad (2.5)$$

where  $\phi_n$  is the appropriate phase factor for each electron.

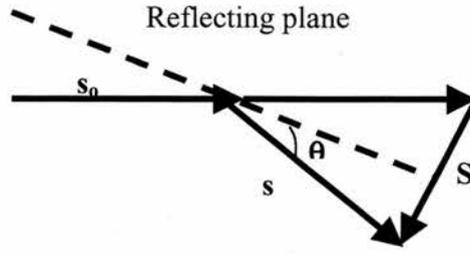
### 2.2.3 Diffraction pattern of a crystal

Information on the crystal structure is contained in the manner in which the crystal perturbs the incident waves into diffracted waves—the diffraction pattern. The geometric locations of the spots with respect to the origin allows accurate determination of the unit-cell parameters and the crystal orientation, and the set of intensities of reflections provides information about the structure of the content of unit cell in the crystal (Sherwood, 1976).

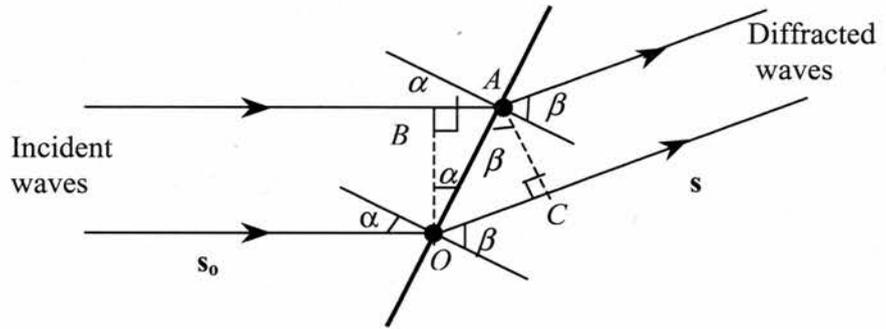
To extract the hidden information from the diffraction pattern, a proper model that relates the diffraction pattern to the unit cell content is required. This model can be expressed mathematically as:

$$F(\mathbf{S}) = \int_{\mathbf{r}} f(\mathbf{r}) e^{2\pi i \mathbf{S} \cdot \mathbf{r}} d\mathbf{r} \quad (2.6)$$

where the function  $f(\mathbf{r})$  describes the contents of the unit cell, and the exponential function contains the phase information carried by the waves in the diffraction event. In the above integral, the function  $F(\mathbf{S})$  is called the Fourier transform of the function  $f(\mathbf{r})$ ; the inverse Fourier transform of the function  $F(\mathbf{S})$  results in the function  $f(\mathbf{r})$ . The space defined by the variable  $\mathbf{r}$  is known as the Real space, and that defined by the variable  $\mathbf{S}$  is referred to as the Reciprocal or Fourier space. The parameter  $\mathbf{S}$  is the scattering vector, which describes the change of the wave direction upon diffraction ( $\mathbf{S} = \mathbf{s} - \mathbf{s}_0$ ) as shown in figure 2.2. The vectors  $\mathbf{s}_0$  and  $\mathbf{s}$  are the wave vectors—vectors in the direction of wave propagation—of the incident and diffracted waves, respectively. Both of vectors  $\mathbf{s}_0$  and  $\mathbf{s}$  are of magnitude  $1/\lambda$ , which is chosen arbitrarily. From the geometry of figure 2.2, the vector  $\mathbf{S}$  has a magnitude of  $(2\sin\theta/\lambda)$ .



**Figure 2.2.** The scattering vector  $\mathbf{S}$  is the vector difference  $\mathbf{s} - \mathbf{s}_0$ , angle  $\theta$  is half of the scattering angle. A plane that makes equal angles with  $\mathbf{s}_0$  and  $\mathbf{s}$  can be regarded as a mirror reflecting the incident beam.



**Figure 2.3.** The geometrical arrangement for two incident waves hitting the obstacle at two points  $O$  and  $A$  at some arbitrary angle  $\alpha$  and diffracted at some arbitrary angle  $\beta$ , to show the path difference between the incident and the diffracted waves. From the depicted geometry, the phase difference  $\Delta\phi$  is related to the path difference  $d$  as:

$$\Delta\phi = \frac{2\pi d}{\lambda}, \quad \text{and } d = OC - BA$$

$$\mathbf{s}_0 \cdot \mathbf{r} = \frac{1}{\lambda} r \sin \alpha, \quad \mathbf{s} \cdot \mathbf{r} = \frac{1}{\lambda} r \sin \beta$$

$$\therefore \Delta\phi = \frac{2\pi}{\lambda} r (\sin \beta - \sin \alpha) = 2\pi(\mathbf{s} - \mathbf{s}_0) \cdot \mathbf{r} = 2\pi\mathbf{S} \cdot \mathbf{r}$$

The vector product,  $2\pi\mathbf{S} \cdot \mathbf{r}$ , in the exponent of Equation (2.6) represents the phase difference between the waves passing through the origin and a point at a relative distance  $r$ ; this can be shown from the geometrical arrangement in figure 2.3.

Equations (2.6) and (2.5) describe the same event:

- In Equation (2.6) an obstacle is described by a continuous amplitude function  $f(\mathbf{r})$ , whereas in Equation (2.5) we have a discrete system of electrons characterised by the electronic scattering factor  $f_e$ .

- Also, the term  $2\pi\mathbf{S}\cdot\mathbf{r}$  in (2.6) is equivalent to the phase factor  $\phi_h$  in (2.5).

The average number of electrons within a volume element  $d\mathbf{r}$  is  $\rho(\mathbf{r})d\mathbf{r}$ , where  $\rho(\mathbf{r})$  is the electron density function. So, the scattering of electrons in a volume element  $d\mathbf{r}$  will be  $f_e\rho(\mathbf{r})d\mathbf{r}$ .

Followed from the above discussion, the mathematical model describing the diffraction can be expressed as

$$F(\mathbf{S}) = \int_{\mathbf{r}} f_e \rho(\mathbf{r}) e^{2\pi i \mathbf{S} \cdot \mathbf{r}} d\mathbf{r} \quad (2.7)$$

where the integral represents the amplitude of the diffraction pattern resulting from the scattering of X-rays by a distribution of electrons  $\rho(\mathbf{r})$ , and it is over the entire volume we are interested in. Since we are interested in the content of the unit cell and the integration is over the variable  $\mathbf{r}$ ; Equation (2.7) can be rewritten as:

$$F(\mathbf{S}) = f_e \int_{\text{unit cell}} \rho(\mathbf{r}) e^{2\pi i \mathbf{S} \cdot \mathbf{r}} d\mathbf{r} \quad (2.8)$$

The above equation can be expressed in coordinate system referenced to the unit cell which is defined by three crystallographic vectors  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$ .

Any point  $\mathbf{r}$  within the unit cell can be expressed as the relative positions  $x$ ,  $y$ , and  $z$  with respect to the dimensions of the unit cell as

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

$$\therefore d\mathbf{r} = V dx dy dz$$

where  $0 \leq x \leq 1$      $0 \leq y \leq 1$      $0 \leq z \leq 1$ , and  $V$  is the unit cell volume

$$\text{Consequently, } F(\mathbf{S})/f_e = V \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) e^{2\pi i \mathbf{S} \cdot (x\mathbf{a} + y\mathbf{b} + z\mathbf{c})} dx dy dz$$

$$\text{or } F(\mathbf{S}) = V \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) e^{2\pi i \mathbf{S} \cdot (x\mathbf{a} + y\mathbf{b} + z\mathbf{c})} dx dy dz \quad (2.9)$$

bearing in mind that the diffraction pattern amplitude  $F(\mathbf{S})$  represents the relative scattering ability of the contents of the unit cell as compared to a single electron.

Diffraction from a crystal can be considered as if the incident X-rays are reflected by a set of parallel planes acting as a mirror. Planes are designated by the Miller indices  $h$ ,  $k$ , and  $l$ . And points of the reciprocal lattice (known as reflections) are defined by the set of planes  $(hkl)$ , which are responsible for a particular reflection of the incident waves.

To obtain a diffraction pattern from a crystal, certain conditions should be met: (1) The Bragg's law, given in Equation (2.10) should be fulfilled

$$\lambda = 2d_{hkl} \sin \theta \quad (2.10)$$

where  $d_{hkl}$  is the spacing between the  $(hkl)$  set of planes of the real lattice, and  $\theta$  is the Bragg angle equal to one-half of the scattering angle.

(2) And Laue equations (2.11) have to be satisfied.

$$\mathbf{a} \cdot \mathbf{S} = h, \quad \mathbf{b} \cdot \mathbf{S} = k \quad \text{and} \quad \mathbf{c} \cdot \mathbf{S} = l \quad (2.11)$$

where  $h$ ,  $k$ , and  $l$  are whole numbers, either positive, negative, or zero.

The scattering vector  $\mathbf{S}$ , which is perpendicular to the reflecting plane, has a length of " $1/d_{hkl}$ ".

Now, back to the Equation (2.9) and applying the Laue conditions to the vector product of the exponential term, we obtain the Equation:

$$F(hkl) = V \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) e^{2\pi i (hx + ky + lz)} dx dy dz \quad (2.12)$$

where  $F(hkl)$  is called the structure factor; the name refers to its dependence on the structure.

An alternative way of representing the distribution of electrons within the unit cell is to consider the electrons as associated with atoms, and then investigate the distribution of

atoms within the unit cell. This leads to a more practical way to calculate the structure factor using the summation:

$$F_{hkl} = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad (2.13)$$

in which  $(x_j, y_j, z_j)$  defines the positions of the nucleus of  $j$ -atom. The term  $f_j$  is defined as the ‘atomic scattering factor’. At zero scattering angle the value of  $f_j$  is the atomic number  $Z_j$ ; as the Bragg angle increases  $f_j$  falls off slowly— because of the dependence of  $|S|$  on  $\theta$ . Also, the value of  $f_j$  falls off more sharply due to the fact that the nucleus and its associated electrons are oscillating about an average position; this results in a change in the interaction of the incident X-ray beam with the electrons and a decrease in the scattering power of the atom. Therefore, a correction for the atomic scattering factor needs to be accounted for as follows:

$$f_j = f_j^o e^{-B_j(\sin^2 \theta / \lambda^2)} \quad (2.14)$$

where  $f_j^o$  is calculated assuming that the nucleus of the atom is stationary, and  $B_j$  is known as the ‘Temperature factor’. The parameter  $B_j$  is related to the mean-square displacement of the atomic vibrations,  $\langle \mu^2 \rangle$  through the relation:

$$B_j = 8\pi^2 \langle \mu_j^2 \rangle \quad (2.15)$$

The  $B$  values obtained for most proteins are too large to be seen as reflecting purely thermal motion and must certainly reflect disorder (Rhodes, 2000). Values of  $f_j^o$  for all atoms as a function of  $\sin \theta / \lambda$ , based on different models of the atomic electron density function, are to be found in Volume C of the *International Tables for X-ray Crystallography* (Wilson & Prince, 1999).

In X-ray diffraction experiment, we have the intensities of reflections  $I(hkl)$  which is related to the structure factor as:

$$I(hkl) = |F(hkl)|^2 \quad (2.16)$$

Values of the structure factor in the derived Equations are on the absolute scale, in terms of the scattering by one electron, whereas the experimental values of  $F(hkl)$  are normally on an arbitrary scale. The average value of the scattered intensity on an absolute scale,  $\langle I_S \rangle_{abs}$  — S denotes the use of the vector notation — can be estimated as follows:

$$\begin{aligned} \text{We have,} \quad I_S &= |F(\mathbf{S})|^2 = F(\mathbf{S}) \cdot F^*(\mathbf{S}) \\ &= \sum_i \sum_j f_i f_j e^{2\pi i(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{S}} \end{aligned} \quad (2.17)$$

where  $F^*(\mathbf{S})$  is the complex conjugate of  $F(\mathbf{S})$ . For a large number of reflections,  $\mathbf{S}$  varies considerably; assuming that the angles  $2\pi(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{S}$  are evenly distributed over the range 0 to  $2\pi$  for  $i \neq j$ , the average value for the terms with  $i \neq j$  will be zero and only the terms with  $i=j$  remain, giving

$$\langle I_S \rangle_{abs} = |F(\mathbf{S})|^2 = K e^{-2B \sin^2 \theta / \lambda^2} \sum_j (f_j^o)^2 \quad (2.18)$$

where the thermal parameter B is averaged over all atoms and K is the scaling factor that is used to transform the intensities and structure factors to the absolute scale.

## 2.2.4 Calculating the electron density

To obtain the electron density, which is the ultimate aim of an X-ray diffraction experiment, we apply the inverse Fourier transform to the Equation (2.12). Considering that the integration will be carried out in the Fourier space of the diffraction pattern which

is discrete not continuous, the integration degenerates to a summation over all reciprocal lattice points indexed by  $h$ ,  $k$  and  $l$  as in the following Equation:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)} \quad (2.19)$$

where  $\alpha(hkl)$  is the phase factor of the structure factor which needs both values of magnitude and phase to be defined. The derivation of the correct values of structure factors  $F_{hkl}$ , in magnitude and phase, from the relative intensities  $|F_{hkl}|^2$  constitutes the 'Phase Problem'. Calculating the electron density of the contents of the unit cell using the above summation in Equation (2.19) is called the Fourier synthesis.

## 2.2.5 Reciprocal space and the Ewald sphere

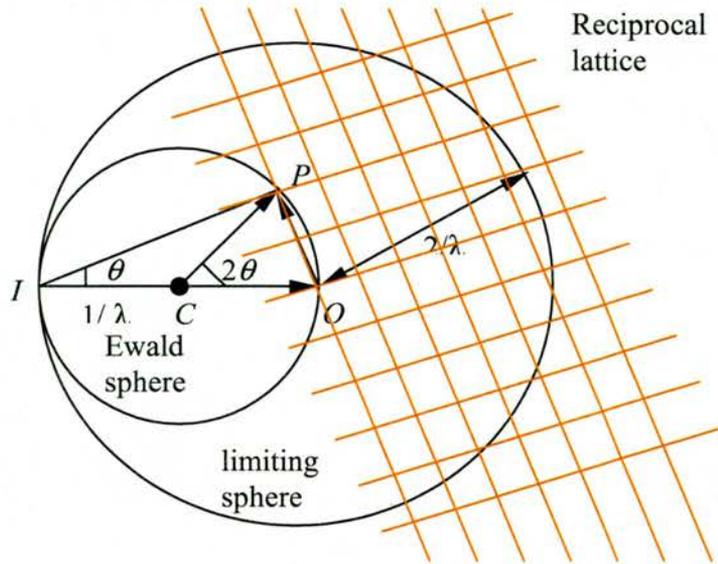
The reciprocal lattice is an array of lattice points in the reciprocal or Fourier space; and unlike real space, reciprocal space is imaginary. The reciprocal-lattice concept is most useful in constructing the directions of diffraction. The reciprocal lattice unit vectors  $\mathbf{a}^*$ ,  $\mathbf{b}^*$ , and  $\mathbf{c}^*$  can be used to define a point in the reciprocal space. The scattering vector  $\mathbf{S}$  in the reciprocal space represents a vector from the origin to a point  $hkl$  of the reciprocal lattice corresponding to diffraction maximum. The vector  $\mathbf{S}$  can be written as:

$$\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad (2.20)$$

The definitions of the vectors  $\mathbf{a}^*$ ,  $\mathbf{b}^*$ , and  $\mathbf{c}^*$  of the reciprocal space in terms of those of the real space  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$  are expressed by equations:

$$\mathbf{a}^* = \frac{\mathbf{b} \wedge \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \wedge \mathbf{c}}, \quad \mathbf{b}^* = \frac{\mathbf{c} \wedge \mathbf{a}}{\mathbf{a} \cdot \mathbf{b} \wedge \mathbf{c}} \quad \text{and} \quad \mathbf{c}^* = \frac{\mathbf{a} \wedge \mathbf{b}}{\mathbf{a} \cdot \mathbf{b} \wedge \mathbf{c}} \quad (2.21)$$

The reciprocal lattice may be thought of as being rigidly associated with the crystal, although it is not defined to be within the physical limits of the crystal as the real lattice



**Figure 2.4.** The reciprocal lattice with the origin  $O$  is superimposed on the Ewald sphere of radius  $1/\lambda$ . The incident wave in the direction  $IO$  hit the crystal at  $C$  and the emerging wave diffracted in the direction  $CP$ ; vector  $\vec{OP}$  represents the scattering vector  $\mathbf{S}$ . The points  $O$  and  $P$  in the reciprocal lattice are in diffraction positions when touching the Ewald sphere. The outer sphere centred at  $O$  and with radius  $2/\lambda$  is the limiting sphere. It can be shown that Bragg's law is followed in the reciprocal space from the geometrical construct (the spheres are drawn as 2D circles for simplicity):

$$OP = |\mathbf{S}| = \frac{2}{\lambda} \sin \theta \quad \text{but} \quad |\mathbf{S}| = \frac{1}{d_{hkl}}$$

$$\therefore \frac{1}{d_{hkl}} = \frac{2}{\lambda} \sin \theta \Rightarrow \lambda = 2d \sin \theta \quad (\text{Bragg's law})$$

does. The interpretation of the diffraction pattern using the reciprocal lattice concept can be understood through the construction of 'Ewald sphere', shown in figure 2.4, as follows:

- 1- Imagine a crystal sitting at point  $C$ ; draw a vector  $\mathbf{s}_0$  representing the direction of the incident wave of wavelength  $\lambda$ .
- 2- The diffracted wave described by a vector  $\mathbf{s}$  emerging from the crystal will have the same length as  $\mathbf{s}_0$  but with a different direction. The end of vector  $\mathbf{s}$  may lie anywhere on a sphere of radius  $1/\lambda$  (Ewald sphere). The points  $O$  and  $P$  at the circumference of the

sphere are the end points of the vectors  $\mathbf{s}_o$  and  $\mathbf{s}$ , respectively. The reciprocal lattice based on vectors  $\mathbf{a}^*$ ,  $\mathbf{b}^*$  and  $\mathbf{c}^*$  can be plotted using  $O$  as the origin and superimposed on the Ewald sphere. The vector  $\overline{OP}$  represents the scattering vector  $\mathbf{S} = \mathbf{s} - \mathbf{s}_o$ . As the crystal rotates about an axis through  $C$ , the associated reciprocal lattice rotates as well around the origin point  $O$ . For various rotations, different reciprocal lattice points will come to lie on the Ewald sphere and give rise to diffraction maxima. Only those reciprocal lattice points lie within the 'limiting sphere' of radius  $2/\lambda$  will come into diffraction positions.

### 2.2.6 Patterson function

Patterson (Patterson, 1934) presented a method for locating the atomic positions in simple molecules without knowledge of the phase angles required for the Fourier synthesis in Equation (2.19). The method involves the calculation of the Patterson function,  $P(\mathbf{u}) = P(u, v, w)$ , where  $\mathbf{u}$  is the interatomic vectors:

$$P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 e^{-2\pi i(hu + kv + lw)} \quad (2.22)$$

Equation (2.22) is the recipe for deriving the Patterson function from the relative intensity data  $|F_{hkl}|^2$ . The relevance of the Patterson function to the crystal structure is better understood when writing the function  $P(\mathbf{u})$  in a different form as shown below.

Equation (2.22) can be written symbolically as

$$P(u, v, w) = \frac{1}{V} \mathfrak{F}(|F_{hkl}|^2) = \frac{1}{V} \mathfrak{F}(F_{hkl} \cdot F_{hkl}^*)$$

which is read as  $P(u, v, w)$  is the Fourier transform ( $\mathfrak{F}$ ) of  $|F_{hkl}|^2$ .

The convolution theorem states that the Fourier transform of a product of two functions is the convolution of the individual Fourier transforms. Where the convolution of the two functions  $f(r)$  and  $g(u-r)$ , written as  $f(r) * g(u-r)$ , is defined by the integral:

$$c(u) = f(r) * g(u-r) = \int_r f(r)g(u-r)dr \quad (2.23)$$

in which, the variable  $u$  determines the position of the function  $g(u-r)$  with respect to the origin; the integral is evaluated for different  $u$  values and is non-zero only when the two functions  $f(r)$  and  $g(u-r)$  overlap. The convolution integral may be thought of as we slide the function  $g(u-r)$  over the function  $f(r)$  for different values of  $u$ .

Now, back to Patterson function, we have

$$P(u, v, w) = \frac{1}{V} [\mathfrak{S}(F_{hkl}) * \mathfrak{S}(F_{hkl}^*)]$$

$$\text{but, } \frac{1}{V} \mathfrak{S}(F_{hkl}) = \rho(x, y, z) = \rho(r) \quad \text{and, } \frac{1}{V} \mathfrak{S}(F_{hkl}^*) = \rho(-x, -y, z) = \rho(-r)$$

$$\text{Therefore, } P(\mathbf{u}) = V[\rho(r) * \rho(-r)] = V \int_{\text{unit cell}} \rho(r)\rho(u+r)dr \quad (2.24)$$

where,  $f(r) = \rho(r)$ , and  $g(r) = \rho(-r)$  in the definition of the convolution integral (2.23). The Patterson function  $P(\mathbf{u})$  is the convolution of the electron density function  $\rho(r)$  and its centrosymmetric image  $\rho(-r)$ , scaled by  $V$ , the volume of the unit cell.

### 2.2.7 Molecular replacement

The term molecular replacement (MR) refers to the method of using a known model structure to solve an unknown structure, given that the two structures are homologous. The idea of MR discussed below was first introduced by Rossmann & Blow in their

pioneering work (Rossmann & Blow, 1962) to detect the non-crystallographic symmetry (NCS), which relates only to a localized volume in the asymmetric unit. Now the term MR is limited to the case where an unknown structure is to be solved with a known search model.

MR has grown to be one of the most powerful tools of the macromolecular crystallographer (Read, 2001). To explain the idea of MR, let us consider two structures of the same macromolecule in different crystal forms, for example the native and the ligand-bound forms of the same protein. Assume that the structure of the native form is known and will be used to solve the structure of the ligand-bound form (unknown). To solve the new structure using the phases of the known structure, the phasing model needs to be superimposed on the unknown structure in its unit cell, and to have the same location and orientation. In theory, it should be possible to conduct a computer search of all orientations and positions of the model in the new unit cell. However, the number of trial orientations and positions of the phasing model is enormous. By working with Patterson function, applying the rotation and translation search, within certain volume  $v$ , on the Patterson vectors of the phasing model will superimpose the two sets of Patterson vectors. If the structure rotates, its Patterson rotates in the same way. The rotation and translation that gives maximum overlap between the two sets of Patterson vectors will define the position of the phasing model in the unit cell of the unknown structure.

Specifying the orientation and location of a molecule in three dimensions requires six variables: three Eulerian angles for rotation and three coordinates for translation. The rotation and translation search is conducted separately rather than simultaneously. The rotation function can be expressed as:

$$R(\mathbf{R}) = \frac{1}{v} \int_{\Omega} P_t(\mathbf{u}) \times P_s(\mathbf{R}^{-1}\mathbf{u}) d\mathbf{u} \quad (2.25)$$

where the rotation matrix  $\mathbf{R}$  that orients the search model in the same way as the molecule, whose structure is to be determined, in the target crystal, is defined by calculating the overlap within a conveniently chosen region  $\Omega$  of volume  $v$  of the observed Patterson function (the target function  $P_t$ ) and a rotated version of the Patterson function corresponding to the search model  $P_s$ .

Having determined the orientation of the model in the target cell, the next step is to find the correct position by running the translation function which can be expressed as:

$$T(\mathbf{t}) = \int_{\Omega} P_t(\mathbf{u}) \times P_s(\mathbf{u}_r + \mathbf{t}) du \quad (2.26)$$

where the overlap between the Patterson of the target structure  $P_t$  and the Patterson of the translated model  $P_s$  (now with the correct orientation) is calculated by the above integral for different translation vectors  $\mathbf{t}$ .

It should be mentioned that both rotation and translation functions are expanded into the Fourier series, not shown, to perform the calculations. Applying the rotation and translation functions to compare the Patterson functions of different crystals is called 'the cross-rotation and cross-translation functions'. When applied within the same crystal, it is designated as 'self-rotation and self-translation functions'.

### 2.2.8 Structure refinement

The ultimate aim of refinement is to optimise simultaneously the agreement of an atomic model with observed structure data and with *a priori* chemical information. The structure factors are calculated for the model structure and we seek the parameters that minimize the difference between the observed structure factor  $F_o$  and the calculated structure factor  $F_c$ . The crystallographic refinement parameters includes: atomic positions

$(x_j, y_j, z_j)$ , B-factors which can be refined as either isotropic (one parameter per atom) or anisotropic in the case of atomic resolution data (six parameters per atom), and occupancy  $n_j$  since the molecules in the crystal occasionally have different conformations. The number of observations needs to exceed the number of refinement parameters.

Assuming that the average geometry observed in small molecules is preserved in the chemical moieties of the protein, we can apply constraints and restraints on the model during refinement. In crystallographic refinement we search for the global minimum of the target function, Equation (2.27), as a function of the refinement parameters (Hendrickson, 1985).

$$E = E_{chem} + (wE)_{X-ray} \quad (2.27)$$

$E_{chem}$  describes covalent interactions (bond lengths, bond angles, torsion angles, chiral centres and planarity of aromatic rings), and non-covalent interactions (intramolecular as well as intermolecular and symmetry-related).  $(wE)_{X-ray}$  is related to the difference between the observed and calculated data, where  $w$  is a weight chosen to balance the gradients (with respect to atomic parameters) arising from the two terms (Brunger et al., 2001).

The traditional form of  $E_{X-ray}$  in Equation (2.27) can be written as:

$$E_{X-ray} = \sum_{hkl \in \text{working set}} (|F_o| - k|F_c|)^2 \quad (2.28)$$

where  $hkl$  are the indices of the reciprocal-lattice points of the crystal and  $k$  is a relative scale factor. The diffraction data are divided into two sets: a large working set (usually comprising 90% of the data) and a complementary test set (comprising the remaining 10%) for what is called 'cross validation' (Brunger, 1992). The above summation is carried over the working set.

$E_{chem}$  can be expressed as follows:

$$\begin{aligned}
E_{chem} = & \frac{1}{2} \sum_{bonds} k_b (b - b_o)^2 + \frac{1}{2} \sum_{bond\ angles} k_\theta (\theta - \theta_o)^2 \\
& + \frac{1}{2} \sum_{dihedral\ angles} k_\phi [1 + \cos(n\phi - \delta)] + \sum_{non-bonding} \left( \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right)
\end{aligned} \tag{2.29}$$

where the first two terms are harmonic potentials represent displacements from the ideal bond length ( $b_o$ ) and bond angle ( $\theta_o$ );  $k_b$  and  $k_\theta$  are the force constants which determine the flexibility of the bond length and bond angle, respectively. The third term is a periodic potential function for the torsion angle ( $\phi$ );  $n$  is the multiplicity, ( $\delta$ ) is the phase factor, and  $k_\phi$  gives a qualitative indication of the relative barriers to rotation. The non-bonded van der Waals interaction is given by the fourth term (Lennard-Jones potential), where ( $r_{ij}$ ) is the distance between two atoms  $i$  and  $j$ . The coefficients A (repulsive term) and B (attractive term) are of a more empirical nature. Atom pairs that are involved in a bond, a bond angle, or a torsion angle do not contribute to the non-bonded interaction energy.

Minimization of the target function in Equation (2.27) can be performed by two methods (Eyke & Watenpaugh, 2001): 1) least-squares method which poses the problem as ‘Given these data, what are the parameters of the model that give the minimum variance of the observations?’. 2) Maximum likelihood method in which the problem is posed as ‘Given this model, what is the probability that the given data set would be observed?’. The likelihood of a model given a set of observations is the product of the probabilities of all of the observations given the model. If  $P_a(\mathbf{F}_i; \mathbf{F}_{i,c})$  is the conditional probability distribution of the structure factor  $F_i$  given the model structure factor  $F_{i,c}$  then the likelihood of the model is:

$$L = \prod_i P_a(F_i; F_{i,c})$$

Instead of maximizing the likelihood, it is more common to minimize the negative logarithm of the likelihood,

$$E_{X\text{-ray}} = -\log L = -\sum_i \log P_a(F_i; F_{i,c}) \quad (2.30)$$

The progress of the refinement is monitored by comparing the measured structure factor  $|F_o|$  with the calculated structure factor of the model  $|F_c|$  for the working set (R-factor) and for the test set (free R-factor) using the following equations (Drenth, 1999):

$$R = \frac{\sum_{hkl \in \text{working set}} \left| |F_o| - |F_c| \right|}{\sum_{hkl \in \text{working set}} |F_o|} \quad (2.31)$$

and

$$R_{\text{free}} = \frac{\sum_{hkl \in \text{test set}} \left| |F_o| - |F_c| \right|}{\sum_{hkl \in \text{test set}} |F_o|} \quad (2.32)$$

### 2.3. Crystallization

Obtaining suitable protein crystals diffracting at high resolution often remains the bottleneck in a crystallographic study. To grow crystals of any compound, molecules have to be brought to a supersaturated, thermodynamically unstable state, which may develop in a crystalline or amorphous phase when it returns to equilibrium. Supersaturation can be achieved by slow evaporation of the solvent or by varying parameters such as concentration, temperature, pH and ionic strength (Ducruix & Giege, 1999). Crystallization starts by a nucleation phase (i.e. the formation of the first ordered aggregates), which is followed by a growth phase. Nucleation is very sensitive to

supersaturation, crystallization rates increase when supersaturation increases. In growth phase, several factors can cause cessation of crystal growth (Ducruix & Giege, 1999): growth defects, poisoning of the faces with impurities, aging of the molecules. That is why a highly purified protein sample is required to grow crystals.

Several techniques are used to grow crystals of macromolecules including dialysis methods, vapour diffusion methods, batch methods, interface methods, and crystallization in gels and related methods. Most crystallographers prefer vapour phase diffusion, which provides an easy way to practice crystallization (Mcperson & Giege, 2001). It should be mentioned that nucleation and growth phases in crystallization are dependent on the method used. Figure 2.5 shows the principle of hanging drop diffusion method. Where a droplet (down to 2  $\mu$ l or less) containing the protein, buffer, crystallising agent, and additives is equilibrated against a reservoir containing a solution of crystallizing agent at higher concentration than the droplet. Equilibration proceeds by diffusion of the volatile species (water) until vapour pressure in the droplet equals that in the reservoir; consequently, the concentration of all constituents in the drop will increase.



**Figure 2.5.** Schematic representation of hanging drop diffusion method. The drop containing protein and additives is placed on a siliconized-cover slip, which is inverted over a reservoir with the precipitating agent at higher concentration than that in the drop.

## 2.4. X-ray radiation sources

X-rays can be derived principally from two effects (Sherwood, 1976): (a) If an electron within an atom is ejected by some means, then it is possible for another electron

in an orbital of higher energy within the same atom to fall into vacant orbital emitting radiation in the x-ray range; this principle is operating in the rotating anode tubes of the in-house X-ray generators. (b) If fast moving charged particles suffer violent accelerations, electromagnetic radiation is emitted. If the accelerations are particularly rapid, the wavelength of this radiation is in the x-ray range; and this is how x-ray radiation is generated at synchrotron.

In the rotating anode tubes, X-rays are obtained by bombarding a rapidly water-cooled rotating metal disk (e.g., Cu) with electrons. The X-rays produced emerge through windows of beryllium. The beam directed through a collimator, a narrow metal tube that selects and reflects the X-rays into parallel paths, producing a narrow beam (Rhodes, 2000).

In a synchrotron source, the charged particles (electrons or positrons) are made to move in close trajectories, often circular or elliptical, and the radiation is generated in cones tangent to the path followed by the particles. Radiofrequency electromagnetic radiation produced from a set of magnets around the synchrotron is used to accelerate the particles; as the energy of the accelerated particles increase the strength of the accelerating field need to be changed to keep the particles moving in the same ring. The change in the steering field must be carefully synchronized to the change in energy or the beam will be lost, hence the name `synchrotron`.

Although synchrotron sources emit a spatially confined narrow beam, the wavelength of this radiation spans a wide continuous spectrum. For easily interpreted data from x-ray experiment, a monochromatic intense focused beam is necessary. A nearly monochromatic beam can be obtained by passing the generated beam through filters, which could be absorption filter or crystal monochromators. Also, collimators and focussing mirrors are used to produce a narrow parallel beam.

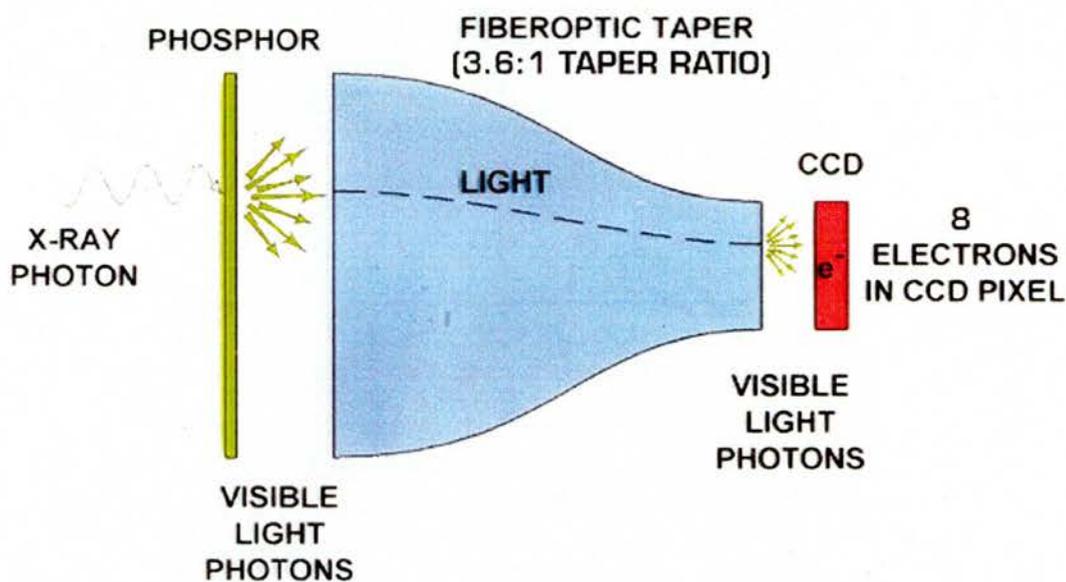
X-rays generated at synchrotron have several advantages over the in-house source:

- i) Due to the shorter wavelengths of the synchrotron radiation, absorption and radiation damages are reduced.
- ii) The wavelength of the synchrotron radiation can be tuned to a selected wavelength from the wide spectrum; this is of use in many applications.
- iii) Since the beam at synchrotron is narrow, higher resolution data can be collected for small crystals (as in our case).
- iv) Because of the quality of the synchrotron beam, shorter exposure time (typically 1-10 s) is used allowing rapid data collection.

## **2.5. X-ray detectors**

The simplest X-ray detector, and for years the workhorse of detectors, is x-ray-sensitive film, which has been almost replaced by image plates and CCD detectors. Image plates are plastic sheets with a coating of small crystals of a phosphor, such as  $\text{BaF:Eu}^{2+}$ . X-rays can stimulate the crystals into a stable excited state in which  $\text{Eu}^{2+}$  loses an electron to the  $\text{F}^-$  layer that contains electron vacancies introduced by the manufacturing process. Further stimulation by visible light causes the electrons to drop back to the Eu layer, producing visible light in proportion to the intensity of the absorbed X-rays. After X-ray-exposure, data are read from the plate by a scanner in which a fine laser beam induces luminescence from a very small area of the plate, and a photocell records the intensity of the emitted light. The intensities are fed into a computer and an image of the diffraction pattern is reconstructed. Image plates can be erased by exposure to bright visible light and reused (Rhodes, 2000).

CCDs (charged-coupled devices) are the latest design in area detector. CCD detectors are photon counters, solid-state devices that accumulate charge in direct proportion to the amount of light that strikes them. CCDs are coated with phosphors that emit visible light in response to X-rays. A typical CCD is a 2.5 cm square array of 25- $\mu\text{m}$  pixels, each of which accumulates charge during data collection. Fiber-optic taper is used to demagnify the light image down to the size of the CCD chip that detects the light image as an electric charge image. The electric charge image is read out of the CCD chip and digitised then fed into a computer. After geometric and intensity corrections are applied, the resulting data are produced. The diagram in figure 2.6 shows the basic principles of CCD detector.



**Figure 2.6.** A diagram showing the basic principles of CCD detector. The picture adapted from the web site of ADSC (Area Detector Systems Corporation), <http://www.adsc-xray.com>

## Chapter 3

### Protein expression and purification

#### 3.1 Introduction

To pursue the crystallographic and the biophysical studies on VCNA, a significant amount of pure homogenous protein sample was required. In earlier work on crystallization of VCNA, a protocol for the enzyme expression and purification was reported (Taylor et al., 1992; Vimr et al., 1988), where the expressed enzyme was retained in the periplasm of the hosting bacteria, and extracted using the Osmotic Shock Fluid method (Neu & Heppel, 1965). We expected it would be straightforward to obtain the required amount of the enzyme by applying the same protocol; however, we could not get enough amount of the pure protein, necessary for our studies. Fortunately, in our laboratory, a molecular biologist (Dr. Helen Connaris) managed to insert the cDNA of the enzyme to a new expression system, which gave higher yield of the required protein—paper to be published.

This chapter is concerned with the experimental work of the expression and purification of VCNA. The two protocols are described in the following sections.

## 3.2 Materials and methods

### 3.2.1 Original protocol

In this protocol, *E. coli* HB101 transformed with the expression vector pCVD364— a gift from Prof. E. Vimr— was used to produce VCNA. The enzyme, retained in the periplasm, was extracted by Osmotic shock Fluid method (Neu & Heppel, 1965).

In detail: a glycerol stock of cells was prepared and cells from that stock were grown at 37 °C on a rapid rotary shaker in Luria-Bertani (LB) medium supplemented with 100 µg/ml of ampicillin. Cells were harvested during the late exponential— stationary phase by centrifugation and subjected to osmotic shock standard procedures as follows:

- 1) The harvested cells were suspended in 20% sucrose with 30 mM Tris-HCl, pH 8.0 at room temperature (80 ml per 1.0 g wet cells).
- 2) The cell suspension was centrifuged for 20 min. at 8,000 x g at 4 °C and the cell pellet was rapidly mixed with a volume of cold water equal to the original volume of the suspension.
- 3) The suspension was incubated for 10 min. on an ice bath before centrifuged at 8,000 x g for 20 min. at 4 °C.
- 4) The supernatant (osmotic shock fluid) that contained the enzyme was kept for further manipulation to purify the enzyme.

To purify the enzyme, the osmotic shock fluid was supersaturated with 50% (w/v) ammonium sulfate to get rid of some of the contaminants. The supersaturated solution was then centrifuged for 45 min. at 20,000 x g at 4 °C; the ammonium sulfate pellet was resuspended in the least volume of 20 mM Tris-HCl buffer, pH 7.6. Undissolved materials were removed by centrifugation and the clean solution was dialysed against the same buffer containing 100 mM NaCl. The dialysed sample solution was subjected to

HPLC chromatography (BioCad) using anion exchange column POROS HQ20 (Pharmacia). Ion exchange chromatography (IEX) separates molecules based on differences between the overall charges of the proteins (Scopes, 1994).

Binding took place under low ionic conditions; elution was achieved by increasing the ionic strength of the buffer system of 20 mM Tris-HCl pH 7.6 containing 250 mM NaCl, to break up the ionic interaction. Fractions were checked for neuraminidase activity (NA) applying a modified fluorometric assay using substrate 2`-(4-methylumbelliferyl)- $\alpha$ -D-N-acetylneuraminic acid sodium salt (Sigma Chemicals Co.), as will be described later. Those with NA<sup>+</sup> were pooled together, concentrated down to 1 ml by ultrafiltration, using an Amicon microfilter with 10 KDa cutoff, and passed over the size exclusion (Gel filtration) chromatography Superdex 200 Hi-Load 16/60 column (Amersham-Pharmacia Biotech) equilibrated in 20 mM Tris-HCl pH 7.6, 100 mM NaCl, using an HPLC system (BioCad), after dialysis against the same buffer.

Gel filtration chromatography separates proteins on the basis of size (Scopes, 1994). Molecules move through a bed of porous beads, diffusing into the beads to greater or lesser degree. Small molecules diffuse further into the pores of the beads, and therefore move through the bed more slowly, while larger molecules enter less or not at all, and thus move through the bed more quickly. Both molecular weight and three-dimensional shape determine the degree of retention.

The purified protein was checked by SDS-PAGE gel. The pure sample, after exchanging its buffer with a system composed of 20 mM Tris-HCl buffer pH 7.6, 150 mM NaCl, 10 mM CaCl<sub>2</sub> and 0.025% NaN<sub>3</sub> (to prevent bacterial growth), was assayed for the total protein using Bradford assay as will be described below. The protein sample in

its final buffer was concentrated using Amicon microfilter with a 10 kDa cutoff to a final protein concentration of 6.0 mg/ml ready for the crystallization experiment.

### 3.2.2 New expression system

The vector pET-30b(+) (Novagen) containing the cDNA construct of the enzyme VCNA, which was lacking the signal sequence that would have directed the enzyme to the periplasm as in the first protocol, was developed and used to transform *E. coli* BL21(DE3) cells. This expression system produced the enzyme in the cytoplasm (data will be published).

In detail: for the transformation, 1  $\mu$ l of the cDNA of the enzyme was added to 50  $\mu$ l of the competent cells, *E. coli* BL21(DE3). The mixture was left in an ice bath for 30 min. without shaking, then subjected to heat shock for 45 seconds at 42 °C and placed back in the ice bath for 2 min. Pre-warmed LB (450  $\mu$ l) was added to the heat shocked cells and shaken for 1 hour at 37 °C. The grown cells were used to plate 50 and 100  $\mu$ l onto two LB/Kanamycin plates and incubated overnight at 37 °C. The plates were checked for cultured colonies and one colony was used to prepare a small culture of cells by inoculating 10 ml of LB medium supplemented with 30  $\mu$ g/ml Kanamycin. The cells were grown at 37 °C with shaking at 200 rpm till the O.D. (optical density at  $\lambda_{600}$ ) reached (0.5-0.6). The small culture cells were used to inoculate 1L LB medium, supplemented with 30  $\mu$ g/ml Kanamycin, at 37 °C till (O.D. 0.5-0.6). The cells then subjected to a heat shock for 20 min. at 42 °C followed by 10 min. at 25 °C. The cells were then induced with 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) and continued to grow overnight at 25 °C

with shaking at 180 rpm. The cells were harvested by centrifugation at 8,000 x g for 20 min. at 4 °C.

The cell pellet was resuspended in a small volume ( $\approx$ 20 ml) of 20 mM Tris-HCl buffer pH 7.6, and 1 tablet of protease inhibitors cocktail was dissolved in the cell suspension. Cells were disrupted by sonication 4 x 30 s. In order to digest the DNA, nuclease deoxyribonuclease I (Dnase I) was added to the disrupted cells to a final concentration of 20  $\mu$ g/ml and left for 30 min. at room temperature. The treated cell suspension was centrifuged for 30 min. at 20,000 x g at 4 °C, and the supernatant containing the desired protein was kept for purification.

Purification of the enzyme was run in the same way employed to purify the osmotic shock fluid in the previous protocol, with one exception: two anion exchange chromatography columns were used in the HPLC step without the need to use the Gel filtration column. Before loading onto the second column, the sample was diluted (two times) to bring the salt concentration down to ( $\approx$ 120-150 mM) using 40 mM Tris-HCl buffer, pH 7.6. The purified protein was checked for purity using SDS-PAGE and NA activity of the enzyme was checked. The enzyme in its final buffer system containing 20 mM Tris-HCl pH 7.6, 150 mM NaCl, 10 mM CaCl<sub>2</sub> and 0.025% NaN<sub>3</sub> was concentrated down to 10.0 mg/ml ready for use.

### **3.2.3 Neuraminidase activity assay**

The assay of the NA activity was carried out using the modified fluorometric assay (Myers et al., 1980; Potier et al., 1979) using the substrate 2`-(4-methylumbelliferyl)- $\alpha$ -D-N-acetylneuraminic acid sodium salt (Sigma Chemicals Co.). Upon the action of the

enzyme, the substrate is cleaved into free sialic acid and free 4-methylumbelliferone. The fluorescence of the free 4-methylumbelliferone was determined in a TD-700 Laboratory Fluorometer (Turner Designs) spectrofluorometer using excitation light at 365 nm and fluorescence emission at 450 nm. Unhydrolysed substrate had an excitation maximum at 315 nm and fluorescence maximum at 374 nm, and thus only slightly interfered with 4-methylumbelliferone determination. The spontaneous hydrolysis of the substrate at pH 5.5 was corrected using a blank solution.

To perform the neuraminidase assay, 1  $\mu$ l of the enzyme sample was added to 90  $\mu$ l sodium acetate buffer pH 5.5 containing 10 mM  $\text{CaCl}_2$  and the hydrolysis reaction was initiated by adding 9  $\mu$ l of the substrate and incubation at 37 °C for 5 min. Adding 3 ml of 250 mM solution of glycine-NaOH, pH10.5, quenched the reaction; and the fluorescence of the solutions was then measured.

### **3.2.4 Bradford assay**

This assay involves the binding of Coomassie Brilliant Blue G-250 to protein. The binding of the dye to protein causes a shift in the absorption maximum of the dye from 465 to 595 nm, and it is the increase in absorption at 595 nm that is monitored (Bradford, 1976).

A series of BSA protein solutions containing 2,4,6,8 and 10  $\mu$ g protein in a volume up to 0.5 ml water was prepared, and 0.5 ml Bradford reagent (Sigma Chemicals Co.) was added to each solution in a disposable cuvet. The contents mixed and the absorbance was measured at  $\lambda_{595}$  after 2 min. and before 1 hour against a reagent blank prepared from 0.5 ml water and 0.5 ml Bradford reagent. The weight of BSA was plotted against the corresponding absorbance resulting in a standard curve used to determine the required

unknown concentration of VCNA. The enzyme sample was prepared by making the appropriate dilution of few microliters enzyme into total volume 0.5 ml with water, and 0.5 ml of the reagent was added to measure the absorbance in the same way.

### **3.3 Results**

#### **3.3.1 Protein expression**

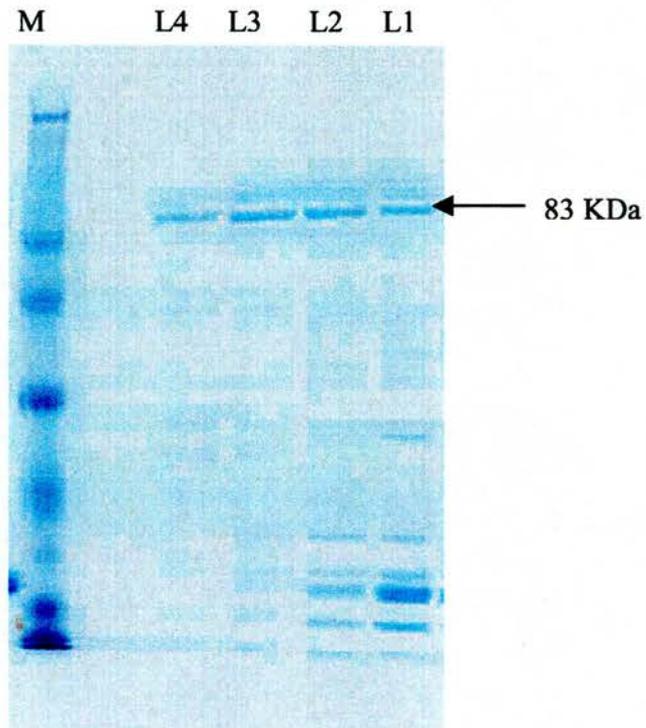
Expressing the enzyme applying the first protocol, in which the expressed enzyme was retained in the periplasm of the bacterial cells and extracted using the osmotic shock fluid method, resulted in a poor expression yield. Barely, 0.1 mg was obtained from 1-liter culture cells. Figure 3.1 displays the SDS-PAGE of the protein sample at different purification stages; the poor expression is clearly demonstrated by the faint band corresponding to VCNA of molecular weight ( $\approx 83$  kDa).

The newly developed protocol for expression, in which the enzyme was produced in the cytoplasm of the hosting cells and the expression was induced using IPTG, gave much higher yield as can be seen from the SDS-PAGE in figure 3.2. About 75 milligrams of pure enzyme were obtained from 1L of cultured LB-medium. The enzyme expressed in the cytoplasm possessed the same activity as the enzyme having the signal sequence and retained in the periplasm.

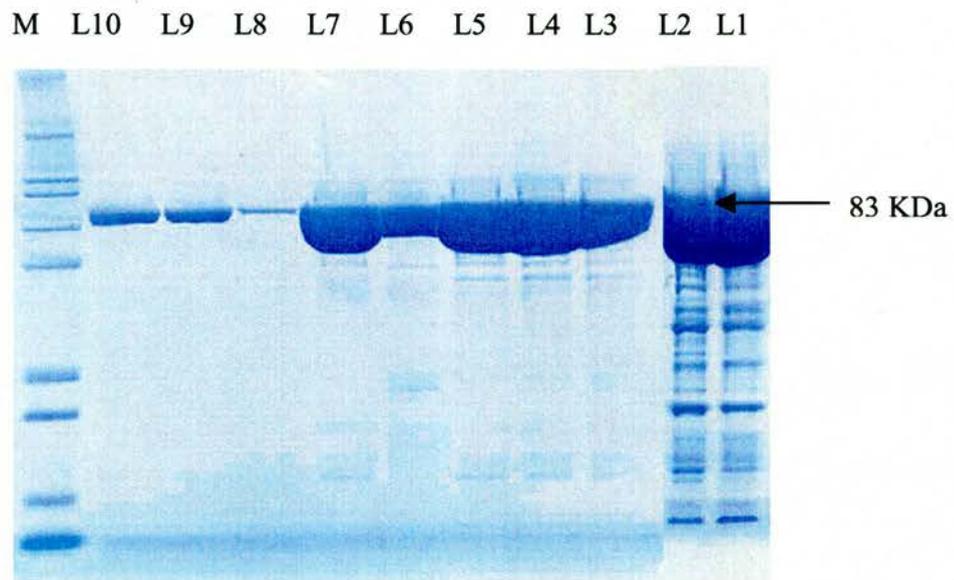
### **3.3.2 Protein purification**

Enzyme samples with high degree of purity, required for the proposed study, were obtained employing HPLC chromatography as described in the material and methods section. VCNA is negatively charged and therefore bound to the positively charged groups of the anion exchange chromatographic column. Figure 3.1 shows the SDA-PAGE of the enzyme samples produced applying the first protocol, at different purification stages. A purification step using the anion exchange column was necessary before passing the sample to the gel filtration column; however, in the published work (Taylor et al., 1992), only gel filtration column was used for purification of the osmotic shock fluid.

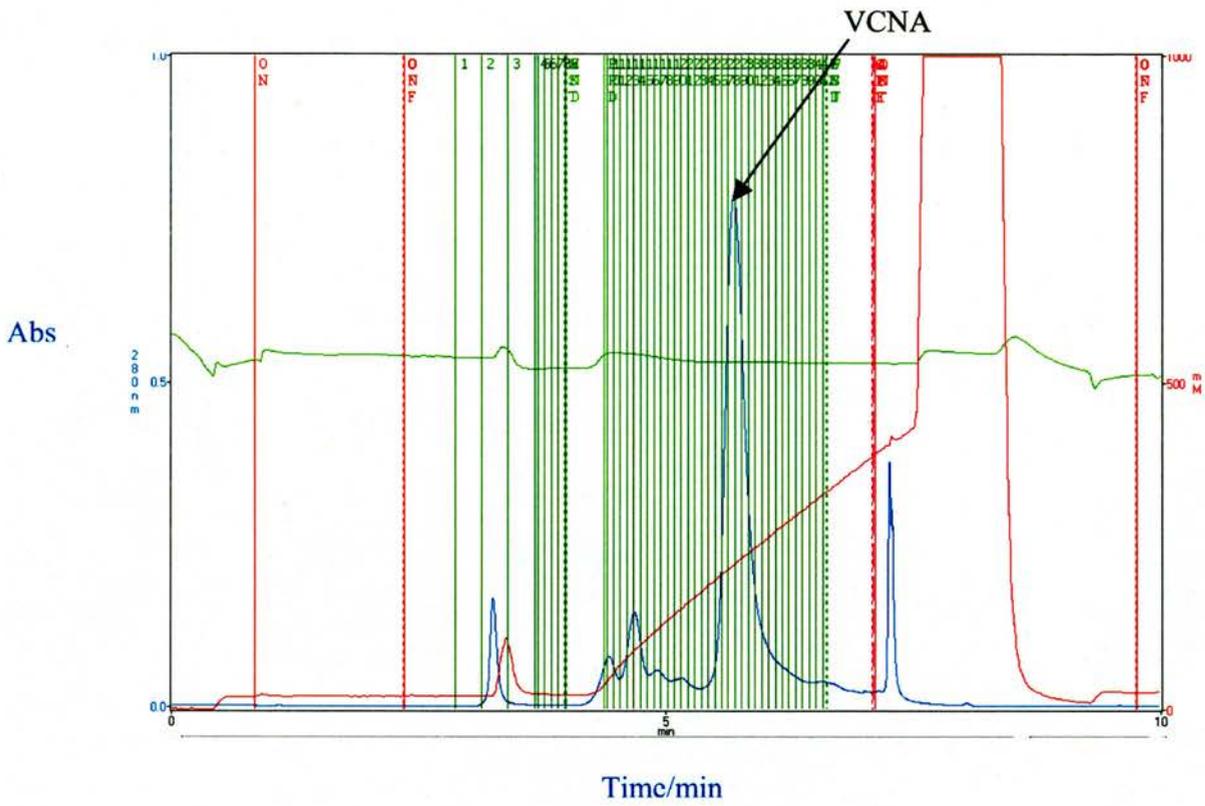
For the enzyme produced using the newly developed protocol, two anion exchange columns were enough to obtain enzyme samples with satisfying purity. Sample fractions after the first and the second anion columns are shown in figure 3.2. Figure 3.3 demonstrated the chromatogram of the anion-exchange chromatography of the enzyme obtained from the new protocol.



**Figure 3.1.** SDS-PAGE of protein sample produced using the first protocol that shows fractions at different purification stages. Lane M is the molecular weight marker; L1 represents the osmotic shock fluid sample that applied to the anion-exchange column. L2 & L3 correspond to different fractions eluted after the anion exchange column. L4 corresponds to fraction eluted from the gel filtration column.



**Figure 3.2.** SDS-PAGE of protein sample produced using the new expression system showing fractions at different purification stages. Lane M is the molecular weight marker; lanes 1 & 2 represent the cell extract that was applied to the chromatographic column. Lanes 3-6 correspond to fractions after the first anion-exchange column. Lanes 7-10 represent fractions eluted from the second column with different sample concentrations.



**Figure 3.3.** The chromatogram of the anion-exchange chromatographic separation of VCNA eluted from the second anion column. The mean peak represents the VCNA protein separated from 5.0 ml sample loaded onto the column and eluted using 20 mM Tris pH 7.6 and 250 mM NaCl.

### 3.4 Discussion

Obtaining a sufficient amount of the purified VCNA enzyme was a critical step for continuing the proposed study. When starting the project study we thought it would not be a problem to produce the required enzyme guided by the published work, which claimed that 1 mg of pure enzyme could be obtained from growing 1-litre medium (Taylor et al., 1992). However, a very small amount of the VCNA enzyme was obtained using the osmotic shock fluid method to extract the enzyme from the periplasm of the bacterial cells.

Developing a new expression system in our laboratory enabled us to obtain enough amount of the protein. In the new expression system, the enzyme was produced in the cytoplasm, and the system was designed to use the advantage of IPTG-induction. Clearly, the high expression yield obtained from the new expression system could be attributed to two features: the IPTG-induction, and the use of the whole cell extract, rather than the periplasm only as shown in the first case.

In the published work (Taylor et al., 1992), only a gel filtration column was used to purify the enzyme from the osmotic shock fluid. However, in our case, the sample of the osmotic shock fluid was not clean enough to apply directly onto the gel filtration column. So, we decided to use an anion exchange chromatographic column before passing the sample through the gel filtration column. In the protocol of the new expression system, two anion exchange columns were enough to obtain the protein sample of a high degree of purity. The accumulating high negative charge on the surface of the VCNA enzyme, net charge (-27), allowed efficient binding to the positive functional group of the anion column, which resulted in a good separation.

## Chapter 4

# Crystallization, Data Collection, Data Processing, Molecular Replacement and Refinement

### 4.1 Introduction

The primary concern of this chapter is to describe the different steps that were involved in our crystallographic study, from growing crystals to finding the 3D-model that fits to the X-ray data. There are steps, which are constant in any structural crystallographic study: firstly, growing suitable crystals that can diffract to a high resolution. Secondly, collecting the crystallographic data by shooting the crystals with the X-ray beam. Thirdly, processing the collected data and extracting the required information that describe the diffracting object; and most importantly the structure factor is obtained. As mentioned in chapter two, both the magnitude and the phase of the structure factor are required in order to generate the electron density map of the diffracting molecule in the unit cell of the crystal. Since only the intensity can be measured directly in an X-ray experiment, from which the magnitude of structure factor is calculated, a fourth step is necessary to obtain the phase information. In our case the molecular replacement (MR) technique was used to obtain the phase information of the structure factor employing the suitable phasing models. Finally, in a fifth step the model solutions are refined and the quality of the models are assessed and validated.

The three-dimensional structures of six complexes of VCNA with the following ligands were obtained: 2-deoxy-2,3-didehydro-5-N-acetylneuraminic acid (Neu5Ac2en, or DANA **1**),  $\alpha$ -2,3-sialyllactose (or 3'-sialyllactose **2**),  $\alpha$ -2,6-sialyllactose (or 6'-sialyllactose **3**), 2-deoxy-2,3-didehydro-5-N-trifluoroacetylneuraminic acid, (or FANA **4**), 2-deoxy-2,3-didehydro-5-N-chloroacetylneuraminic acid (or CANA **5**), and 2-deoxy-2,3-

didehydro-5-N-glycolylneuraminic acid (or GANA), where FANA, CANA, and GANA are derivatives of DANA with a modified N-acetyl group at C5. The chemical structures of the ligands **1-6** are drawn in scheme 4.1.

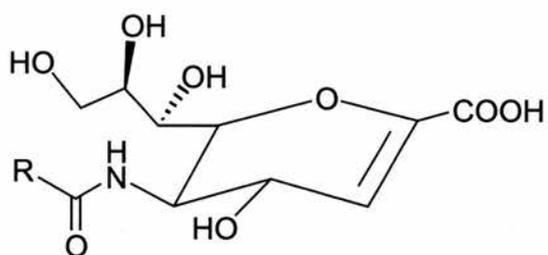
Only the five steps mentioned above are discussed in this chapter; the thorough analyses of the protein-ligand interactions in the 3D-structures of the complexes **1-6** are the subject of the next chapter.

## **4.2 Materials and methods**

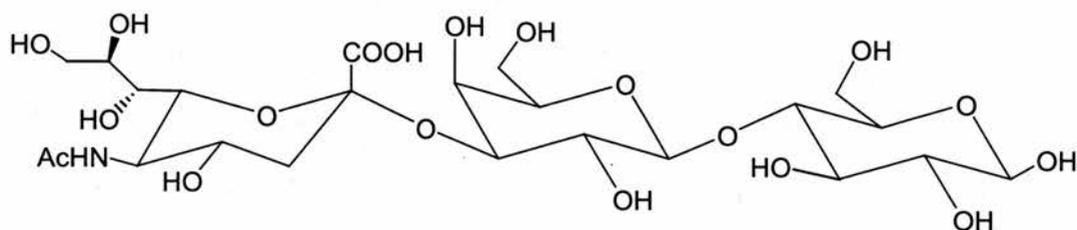
### **4.2.1 Crystallization**

Crystals of the six complexes were obtained using the hanging-drop vapour diffusion technique. Ligands **1 - 3** were purchased from Sigma Chemicals Co., and ligands **4 - 6** were gifts from Prof. Mark von Itzstein (Institute for Glycomics, Griffith University, Queensland, Australia). The protein sample used in crystallization of DANA complex and 3'-sialyllactose complex were produced using the Osmotic Shock Fluid protocol, while crystals of the other complexes were obtained using protein sample produced from the new expression system, see chapter 3.

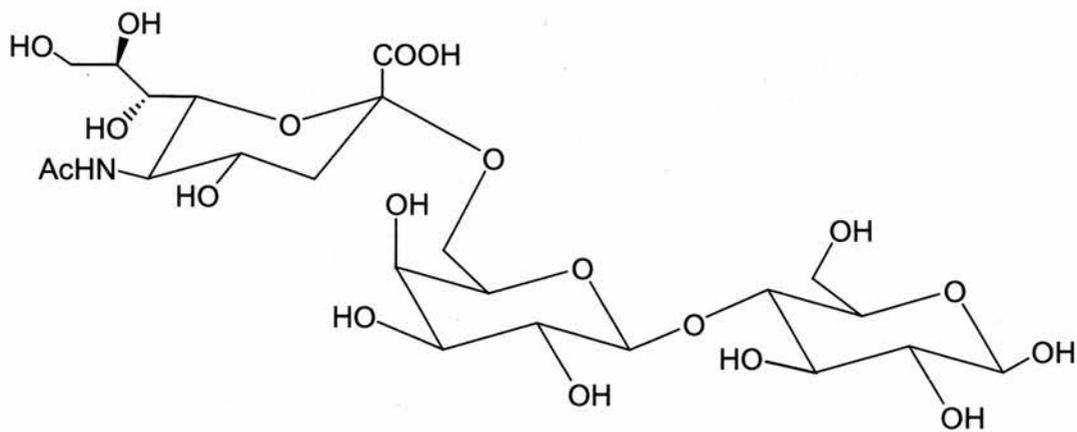
The cocrystals of DANA-complex were obtained from a drop containing: 2  $\mu$ l enzyme of concentration 6.0 mg/ml in (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 0.025% NaN<sub>3</sub>), 1  $\mu$ l 30 mM DANA solution (prepared in 20 mM Tris-HCl pH 7.6), and 1  $\mu$ l precipitating agent composed of (10% PEG 3.35 K plus 40 mM NaCl). The reservoir solution contained (20% PEG 3.35 K, and 150 mM NaCl) in 20 mM Tris-HCl pH 7.6. It took one month for the crystals to grow at 18 °C. The cocrystals of 3'-sialyllactose-complex were obtained in the same way as for the DANA-complex, using 15 mM ligand



- |   |                        |      |
|---|------------------------|------|
| 1 | R= -CH <sub>3</sub>    | DANA |
| 4 | R= -CF <sub>3</sub>    | FANA |
| 5 | R= -CH <sub>2</sub> Cl | CANA |
| 6 | R= -CH <sub>2</sub> OH | GANA |



2 3'-Sialyllactose



3 6'-Sialyllactose

**Scheme 4.1** Compounds used in the crystallization experiments.

solution (prepared in 20 mM Tris-HCl pH 7.6) instead of DANA. It took only few days for the crystals of the 3`-sialyllactose-complex to grow.

The cocrystals of 6`-sialyllactose-complex were obtained using 15 mM 6`-sialyllactose solution and protein concentration of 4.0 mg/ml in the same way as that of the 3`-sialyllactose-complex, with one change that a cocktail of three disaccharides: sucrose, lactose, and maltose (10 mM each) was added to the solution of the precipitating agent.

The cocrystals of the complexes: FANA, CANA, and GANA were obtained as follows: first, the native protein crystals were grown by adding 1  $\mu$ l of the precipitating agent (contained the sugar cocktail of sucrose, lactose & maltose of 10 mM concentration) to 1  $\mu$ l protein sample of concentration 4.0 mg/ml. It took a few days for the native crystals to appear; following then, 1  $\mu$ l of the (3-5 mM) ligand solution (prepared in 20 mM Tris-HCl pH 7.6) was added to the same drop containing the native crystals. The crystals of GANA and CANA complexes stayed in a good conditions after adding the ligand, while the crystals of FANA complex disappeared upon the addition of the ligand and reappeared in the next few days.

#### **4.2.2 Data collection**

All data were collected under cryogenic conditions using 20% glycerol in the crystallization buffer as the cryoprotector. Before starting data collections, reflection profiles were checked on the first recorded images and those crystals showed profiles with double maxima, substantially elongated or smeared out, which usually arises from crystal splitting, were disqualified and another crystal was used.

#### 4.2.2.1. Synchrotron data

Data of the complexes of: DANA, 3'-sialyllactose, and 6'-sialyllactose were collected on beamline ID14-2 at the ESRF, Grenoble, France. The beamline was equipped with a diffractometer comprising a single phi spindle goniometer with motorised z-translation and x/y sample alignment translations, ADSC Q4 CCD detector and an Oxford Cryostream for 100K data collection. A monochromatic x-ray beam with wavelength 0.933 Å and beamsize of 200 microns was used.

**DANA-complex:** The crystals were three months old, and diffracted to 1.8 Å resolution; 120 frames were collected with 1° oscillation and 15s exposure time.

**3'-Sialyllactose-complex:** Also, the three-month old crystals diffracted to 1.8 Å, and 120 frames were collected with 1° oscillation and 15s exposure time.

**6'-sialyllactose-complex:** crystals diffracted to 1.6 Å, and 240 frames were collected with 0.5° oscillation and 15s exposure time.

The exposure time 15s was used because of the small size of the mounted crystals.

#### 4.2.2.2. In-house data

Data for the FANA-, CANA-, and GANA-complexes were collected in-house using the rotating anode X-ray generator (Rigaku/MicroMax-007) with setting parameters of 40 KV, 20 mA and wavelength 1.5418 Å. The X-ray installation was equipped with the imaging plate area detector (R-AXIS IV++), and the cryogenic system (X-stream) that operates without liquid nitrogen. All data were collected under cryogenic conditions at 100 K, as for the first three complexes at Synchrotron.

**FANA-complex:** Crystals diffracted to 1.97 Å; 160 frames were collected with 0.5° oscillation and 15 min. exposure time.

**CANA-complex:** The crystal-to-detector distance was set to the resolution of 2.5 Å, 180 frames were collected using oscillation angle 0.5° and 15 min. exposure time, followed by another 220 frames with the crystal-to-detector distance set corresponding to 2.8 Å resolution.

**GAN-complex:** Crystals diffracted to 2.6 Å resolution, 300 frames were collected with exposure time 20 min. and using oscillation angle 0.5°.

### 4.2.3 Data processing

All data, except the data set of the 6'-sialyllactose complex, were processed using programs in HKL package (Otwinoski & Minor, 1997): DENZO is used for data reduction, which is the preliminary manipulation of the intensities and their conversion to a corrected more generally usable forms, and integration. SCALEPACK is used for merging and scaling (for symmetry-equivalent observed intensities). For the 6'-sialyllactose complex, MOSFLM (Leslie, 1992) was used for data reduction and integration, and SCALA (Evans, 1993) was used for merging and scaling. In typical data processing procedures, the following steps were conducted:

**DENZO** used a list of peaks from one image displayed in XDISPLAYF program in the HKL software package, which searched for the peaks in the frame and worked interactively with DENZO, for autoindexing the reflections in that image. The autoindexing routine resulted in the lattice and unit-cell distortion table with the possible 14-Bravais lattices listed in a decreasing order of symmetry. The goal was to find the lattice of the highest symmetry, which fits the data with minimal distortion; the primitive orthorhombic lattice was chosen. Agreement between the observed diffractions and the predictions indicated successful autoindexing; manual adjustment of mosaicity—

imperfect orientations of the unit cells relative to one another—and spot parameters was necessary for good prediction. Subsequent refinement was done after autoindexing starting at medium resolution and progressing to a higher resolution for the parameters: crystal rotations, beam position, crystal-to-detector distance, and unit cell dimensions. Once the first frame was indexed successfully, the rest of frames were processed, indexed and integrated in batch mode as a hands-off interactive session to spot any problems. The crystal and detector orientation parameters required refinement for each processed image. To calculate the diffraction intensity, the detector background had to be estimated and subtracted from the reflection profile. The accurate prediction of spot positions was necessary to achieve precise integration of Bragg peaks; DENZO applied profile fitting for the prediction and used  $\chi^2$ -test to check predictions. DENZO produced a list of *hkl*s and unscaled intensities; the log file was checked, more attention was paid to the mosaicity histogram for the reflections in each frame to assess the quality of the processing. Bad histogram indicated under- or overestimation of the mosaicity and required reprocessing with different mosaicity value.

**SCALEPACK** was used to merge and scale the symmetry-equivalent reflections in the set of frames, and to perform the global refinement of crystal parameters (postrefinement) in which separate refinement for the crystal orientation of each image was done, but with the same unit cell value for the whole data set. The program calculated single isotropic scale-factors and B-factors for each frame and applied them to  $I$  and  $\sigma(I)$ ; where  $\sigma(I)$  is the error of the intensity. The program allowed for the analysis of the quality of the diffraction data from the crystal employing many standards based on the resulting statistics from the merging and scaling. The criteria used to evaluate the data quality included:  $I/\sigma(I)$ , which is the signal to noise ratio,  $R_{merge}$  (a measure for the agreement

between symmetry related reflections), and percentage of the completeness in each resolution shells. The outer resolution limit was assigned to the resolution shell where  $I/\sigma(I)$  falls to 2.0, when  $R_{\text{merge}}$  reached 20-40% depending on the redundancy. It should be mentioned that for the CANA complex, the two collected data sets were processed separately to a cutoff resolution of 2.8 Å, then the two '.sca' files were merged and scaled together as will be discussed in the results section. The output file after scaling contained a list of *hkl*s, the scaled intensity *I*, and  $\sigma(I)$ . The intensity output was converted into structural factor amplitudes with TRUNCATE module in CCP4i package. Systematic absences indicated the space group to be  $P2_12_12_1$ .

**MOSFLM** was run interactively, a few images were read and autoindexed and resulted in the matrix of the crystal orientation and a list of possible Bravais lattices with given penalties, the solution with the  $P2_12_12_1$  symmetry was chosen; also the unit cell parameters were estimated. The selected images were integrated and the standard profile was inspected to check the integration parameters: box size and shape as well as the mosaicity; then postrefinement for unit cell parameters was carried out. Subsequently, the whole batch of collected images were integrated interactively with visual inspection of the profile, the agreement between the predicted and the observed peaks was considered as an indicator for successful indexing. The output file contained the list of *hkl* reflections and the integrated intensities. The output file (.mtz) was passed to the program SCALA for merging and scaling.

**SCALA** was used to scale and merge the output data of MOSFLM after being sorted using SORTMTZ program that arranged the data in certain order. The program calculated initial scale factors and B-factors for batches of the processed frames at intervals of 10°,

and refined the scale parameters; then generated the weighted mean of the observations after rejecting the outliers. Details of the statistical analysis were given in the output log file; parameters  $R_{\text{merge}}$ ,  $I/\sigma(I)$ , and percentage of completeness were checked to evaluate the data quality. Intensities in the output were passed to program TRUNCATE that produced an output file with a list of *hkl*s reflections and structure factor amplitudes.

#### 4.2.4 Molecular Replacement (MR)

The molecular replacement technique was applied to solve the 3D-structures of the DANA and CANA complexes. The program AMoRe (Navaza, 1994) in the CCP4i program suit (Collaborative Computational Project, 1994) was used.

In the case of DANA complex, the deposited structure of the free VCNA was extracted from the PDB database (pdb-code 1KIT) and employed as the search model. The final refined model of the DANA complex structure—the ligand removed—served as the model solution for the data of the complexes: 3'-sialyllactose, 6'-sialyllactose, and FANA, without *a priori* need for MR.

For CANA complex, the refined model, obtained from crystals of the same complex but turned to be empty of any bound ligand within the catalytic site (unsuccessful experiment), was used as the search model. The solved structure of the CANA complex (ligands not included) was in turn used as the phasing model for the data set of the GANA complex.

The following standard molecular replacement protocol was applied in the case of the two complexes, DANA and CANA:

**AMoRe** carried out the MR in steps applying several routines:

1) SORTFUN reformatted and packed the observed data ( $F_o$ ) in a larger triclinic  $P1$  unit cell for all  $hkl$  reflections within a resolution shell: 12-4 Å (for DANA complex) and 8-4 Å (in the case of CANA complex).

2) TABFUN tabulated the structure factors for the search model on a fine grid corresponding to a large unit cell (all angles  $90^\circ$ ); the model was fixed with no translation or rotation.

3) ROTFUN firstly calculated the structure factors for the model in a suitable cell and packed them in the same format as the output of SORTFUN. Secondly, the program calculated the spherical harmonics for the crystal and the model with  $hkl$  truncated to the assigned resolution range, and the integration was carried out over a sphere of radius 35 Å. Thirdly, the cross-rotation function was run searching for Patterson correlation using the spherical harmonic functions between radii 6 to 35 Å within which the interesting self vectors should lie. The output file of ROTFUN contained the rotational solutions given in terms of the Eulerian angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) and tabulated based on correlation coefficients.

4) TRAFUN calculated the translation function for a chosen solution from the output of the rotation function. A good solution was considered as the one with high correlation coefficient between the observed amplitude for the crystal and the calculated amplitude for the model, and with a low R-factor.

5) FITFUN carried out the rigid body refinement on the chosen solution from the rotation and translation search.

Finally, the successful solution from MR was applied to the search model to be used as the model structure for the refinement of the corresponding complex.

#### 4.2.5 Model refinement

For each complex, the model solution that fitted to the corresponding data set was improved by model refinement. A typical refinement protocol is described in the following paragraph.

Model refinement was applied using the maximum-likelihood method, with its function expressed in amplitudes (MLF), in the CNS program suit (Brunger et al., 1998) through its HTML interface. The input data files had to be prepared, in details: (i) The model-pdb file was reformatted to the suitable pdb-format for CNS and a molecular topology file (mtf) containing information about the atomic charges and masses was generated for the model structure. (ii) A cross-validation (cv) file was created in which a test array free R (5-10%) was setup using a random selection of data in the output file of data-processing. The rest of data were designated as a working set in the refinement process.

The refinement was carried out in several steps interleaved with model building and fitting of the ligands in the electron density map in the program O (Jones et al., 1991) as follows:

- 1) Two runs of rigid body refinement using data in the resolution ranges 8-4 Å and 30-3 Å, respectively, were carried out. The three subunits of the molecule corresponding to the central  $\beta$ -propeller and the two lectin-like domains were dealt with as three rigid bodies to allow for the relative displacement between the subunits.
- 2) The refined structures were subjected to a minimization run followed by simulated annealing applying torsional molecular dynamics and using slow-cooling temperature protocol; the default parameters of temperature and cooling rate were employed. Simulating annealing allowed for local conformation corrections that could be not attainable by minimization.

- 3) A run of restrained individual B-factor refinement was carried out using the whole resolution range.
- 4) A refinement run for all parameters was done using the whole resolution range, combined simulating annealing, energy minimization and B-factor refinement. The electron density maps,  $2F_o-F_c$  and  $F_o-F_c$ , were generated after each run and inspected searching for a defined density at the suspected binding sites using the molecular graphics in O.
- 5) The ligand(s) was/were fitted in the defined density in the electron density maps using graphics interface, and the pdb file of the refined model was edited to add the coordinates of the fitted ligand(s). Topology and parameter files for the ligand(s) were created for usage in the refinement. Several refine cycles for the model structure with ligand(s) interleaved by model adjustment using graphics in O were performed.
- 6) At the final stage of model building, water molecules were picked up using WATER-PICK module in CNS. Default criteria for water deleting in the program were used: minimum distance between water and any atom was set to 2.6 Å, maximum distance between water and any atom was defined as 4.0 Å, minimum H-bonding distance between water and O- or N-atom was set to 2.0 Å and maximum H-bonding distance was determined as 3.2 Å. Peaks greater than the minimum distance and less than the maximum distance from any atom were kept based on criteria of H-bonding.
- 7) The final built model with water added was examined visually using the graphical interface to correct any positional error that appeared in the electron density maps. Cycles of model building and refinement continued until a final model was obtained as judged by the values of R-factor and R-free. Through the whole procedures of refinement, R-factor and R-free were monitored, a successful refinement run should result in a decrease of both of them.

## 4.2.6 Model quality

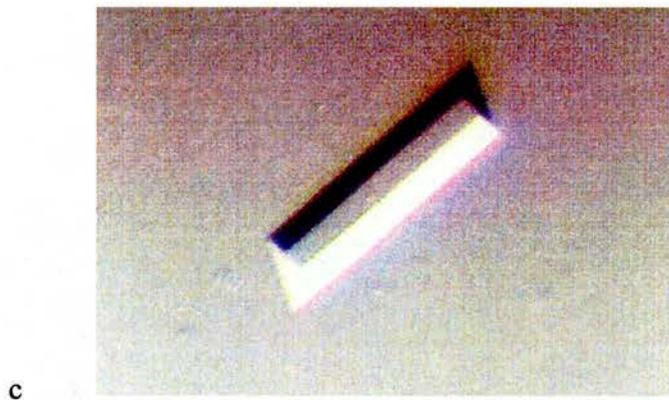
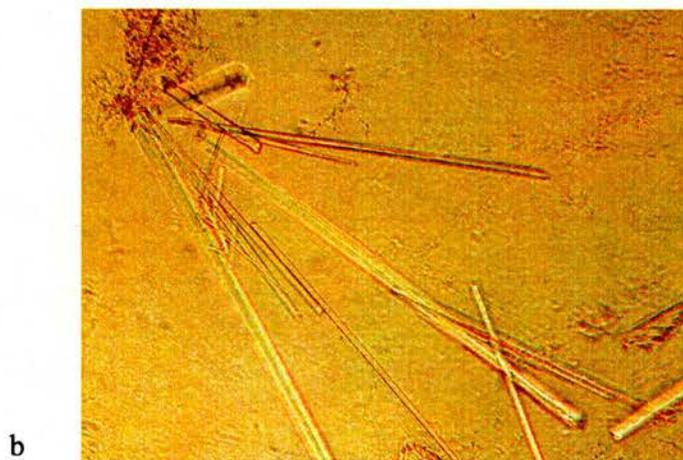
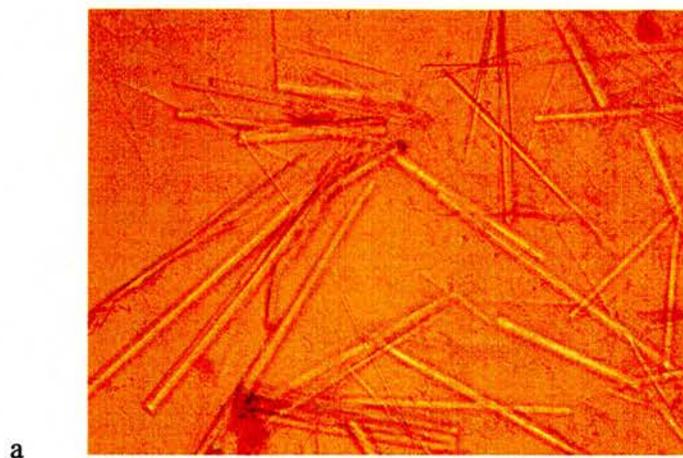
Various criteria were considered to assess the quality of the structures:

- 1) R-factor and R-free which are used as the global criteria to measure the agreement between the model and the experimental data.
- 2) The contacts and environment of amino acid residues were examined applying simple rules of thumb of the preference interactions between residues (hydrophobic, electrostatic, and hydrogen-bonding interactions).
- 3) The structures of the finally refined model were validated using the PROCHECK package (Laskowski et al., 1993) in which all bond distances and angles that deviated significantly from the database-derived reference values were flagged and checked further; this included deviations from planarity in aromatic rings and planar side-chain groups. The program also displayed main-chain dihedral angles ( $\phi, \psi$ ) as a two-dimensional Ramachandran plot, which classified each residue in one of three categories: allowed, partially allowed or disallowed conformations.

## 4.3 Results

### 4.3.1 Crystallization

Figure 4.1 displays sample pictures of the crystals of the six complexes that used for data collection. For the DANA and 3'-sialyllactose complexes needle-shape crystals were obtained. For the other four complexes, adding the sugar cocktail (sucrose, lactose, and maltose) to the crystallization conditions resulted in rod-shape crystals (Fig. 4.1c) of better quality (better looking and greater diffracting power).



**Figure 4.1.** Examples of the crystals used for data collection: (a) DANA-complex, (b) 3'-sialyllactose complex, (c) 6'-sialyllactose complex.

Trials to obtain cocrystals of the DANA-derivative inhibitors by just adding the ligand to the drop containing the protein and the precipitating agent, as in the case of the DANA crystals, did not work. Adding the ligand (FANA) to the drop containing the native crystals caused the crystals to dissolve; however, the crystals incorporating the ligands reformed again within few days. In the case of GANA and CANA ligands, adding the ligand solution to crystals of the native enzyme did not harm the crystals, which stayed in a condition suitable for shooting. The ligand concentration needed to be as low as 3 to 5 mM, since using higher concentration of the ligands dissolved the crystals, and no reformation could be observed. The crystals grown in presence of the ligand were found to have different packing from those grown in ligand-free conditions; this will be discussed in section 5.1.

### 4.3.2 Data collection and data processing

A summary of data collection for the six complex data sets is presented in table 4.1. Two data sets were collected for the same crystal of CANA-complex; the first data set obtained at resolution of 2.5 Å, scan range (20-110°). Processing this data set revealed that most of reflections were weak beyond 2.8 Å resolution shell, and completeness was only 68.8%. Therefore, we decided to collect a second data set to improve the completeness, this time the crystal-to-detector distance was set to 2.8 Å resolution, scan range (20-130°).

As shown in the table 4.1, the cutoff resolution of the data sets of the complexes of: DANA, 3'-sialyllactose, and FANA was set to a resolution lower than that at which data were collected. The criteria:  $I/\sigma(I) > 2$ , the completeness and  $R_{merge}$  of the top shell should be of acceptable values were used to set the resolution of the top shell.

**Table 4.1.** Summary of the collected data sets

Data set	Collected at	Diffraction Res.	Processing Res.
DANA-complex	Synchrotron	1.8 Å	1.9 Å
3-sialyllactose-complex	Synchrotron	1.8 Å	1.9 Å
6'-sialyllactose-complex	Synchrotron	1.6 Å	1.6 Å
FANA-complex	In-house	1.97 Å	2.1 Å
GANA-complex	In-house	2.6 Å	2.6 Å
CANA-complex (1)	In-house	2.5 Å	2.8 Å
CANA-complex (2)	In-house	2.8 Å	2.8 Å

#### 4.3.2.1 Space group and unit cell

The processed data showed that all crystals belong to the orthorhombic space group  $P2_12_12_1$  (the same space group reported for the native structure), which was confirmed by the systematic absences of reflections  $(h00)$  and  $(00l)$  for  $h=2n+1$  and  $l=2n+1$ , [( $0k0$ ) reflections were not measured]. The unit cell dimensions of the six complexes fall into two classes as can be shown from table 4.2, which summarises the statistics of data processing. The first class includes the complexes of: DANA, 3'-sialyllactose, 6'-sialyllactose, and FANA, with unit cell dimensions  $a$ ,  $b$ , and  $c$  not varying much between the four complexes, but quite different from those dimensions of the reported native structure, for which the data were collected at room temperature. The largest shift was observed in the  $c$ -dimension ( $\approx 13$  Å), and there was a shift of ( $\approx 4$  Å) in  $b$ , and ( $\approx 2$  Å) in  $a$ . The second class involves the other two complexes of GANA and CANA, whose cell dimensions are similar, and their unit cell dimensions are close to that of the room

**Table 4.2.** Statistics of data processing

Data set	VC-DANA	VC-3`Sialyllactose	VC-6`Sialyllactose
Cell (Å)			
a	70.4	70.3	70.5
b	74.6	74.9	75.1
c	151.0	151.6	151.6
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution (Å)	1.9	1.9	1.6
Number of observations	284,950	398,317	869,964
Number of unique reflections	63,722	63,699	106,040
Completeness (top shell) (%)	87.9(85.7)	91.4(94.8)	99.2(99.7)
R <sub>mege</sub> <sup>1</sup> (top shell) (%)	6.8(25.3)	10.3(32.8)	7.3(26.2)

Cell dimensions of the native crystal structure at room temperature:

$$a = 72.3 \text{ \AA}, b = 78.9 \text{ \AA}, c = 164.5 \text{ \AA}$$

**Table 4.2. (continued)** Statistics of data processing

Data set	VC-FANA	VC-CANA <sup>2</sup>	VC-GANA
Cell (Å)			
a	69.9	71.4	71.9
b	74.4	76.3	77.3
c	150.1	162.4	163.3
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution (Å)	2.1	2.8	2.6
Number of observations	377,212	32,454	601503
Number of unique reflections	46,599	22,682	28,826
Completeness (top shell) (%)	94.7(91.4)	76.7(48.1)	88.2(55.8)
R <sub>merge</sub> <sup>1</sup> (top shell) (%)	7.4(31.3)	7.5(13.8)	11.9(26.9)

Cell dimensions of the native crystal structure at room temperature:  
 $a = 72.3 \text{ \AA}$ ,  $b = 78.9 \text{ \AA}$ ,  $c = 164.5 \text{ \AA}$

<sup>1</sup>R<sub>merge</sub> =  $\sum_{hkl} \sum_i |\sum I_{hkl}^i - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i \langle I_{hkl}^i \rangle$  where the sum i is over all separate measurements of the unique reflections hkl.

<sup>2</sup>The statistics shown for CANA complex are for the merged two-scaled data sets with: 385635 & 461583 observations, and 23295 & 23189 unique observations for the 2.5 Å and the 2.8 Å data, in the same order.

temperature native structure rather than to the first four complexes. This change of cell dimensions between the two classes could be correlated with the way the crystals were grown. Crystals of the complexes in the first class were formed (or reformed) in presence of their ligands, whereas crystals of the second class were grown in absence of ligands and stayed in their integrated form upon adding the ligands to the drops containing the crystals (section 4.3.1). It is worthy to mention that crystals of the first class were packed differently from crystals of the second class, the latter showed similar crystal packing to the native structure, as will be shown in chapter 5. Interestingly, we had crystals of CANA-complex obtained in the same way as that of FANA-complex; a 2.4 Å data set collected for these crystals (at Daresbury Laboratory/UK) revealed the unit cell dimensions ( $a = 68.6$ ,  $b = 72.5$ ,  $c = 147.4$  Å) close to the crystals of the first class. Unfortunately, the active site of the structure that fitted this data set was found empty (perhaps because of the too much volume of the cryo solution added to the drop containing the crystals, which might wash out the bound ligand).

Analysis of the cell content in CCP4i and calculating Matthews' coefficient,  $V_M$  (Å<sup>3</sup>/Da), (Matthews, 1968), estimated one molecule per asymmetric unit for each complex, and solvent content ( $\approx 48-55$  %) in agreement with what was expected for protein crystals, see table 4.3.

#### **4.3.2.2 Scaled data**

Table 4.2 summarised the statistical analysis of the data processing. The reported data analysis for the CANA complex, shown in table 4.2, is extracted from the merged and scaled files of the two data sets collected for that complex and processed separately. The completeness of the data sets varied between the different complexes. Generally speaking,

**Table 4.3.** Cell content analysis

Content analysis	DANA complex	3`-sialyllactose complex	6`-sialyllactose complex
Cell vol. ( $\text{\AA}^3$ )	793,173.563	798,544.563	803,137.625
$V_M^a$ ( $\text{\AA}^3/\text{Da}$ )	2.4	2.4	2.4
Solvent-content <sup>b</sup> (%)	48.1	48.5	48.8
No.mol./ asy.unit <sup>c</sup>	1	1	1

**Table 4.3.(continued)** Cell content analysis

	FANA complex	CANA complex	GANNA complex
Cell vol. ( $\text{\AA}^3$ )	780604.125	884,726.0	907,600.188
$V_M^a$ ( $\text{\AA}^3/\text{Da}$ )	2.4	2.7	2.7
Solvent-content <sup>b</sup> (%)	47.3	53.5	54.7
No.mol./ asy.unit <sup>c</sup>	1	1	1

<sup>a</sup>  $V_M$  is the Matthews' coefficient

<sup>b</sup> Percentage of Solvent content in the unit cell

<sup>c</sup> Number of molecules per asymmetric unit

in each complex, the completeness reported in the table 4.2 was sufficient to show the extra density for the bound ligand(s) clearly at their pockets in the enzyme, as will be discussed in details in the next chapter. The most complete data set (99.2 % complete) was that of 6'-sialyllactose complex (1.6 Å). The least complete data set was that of CANA-complex with the highest resolution of 2.8 Å (76.7 % complete); the top shell had only (48.1 % complete). We decided to include the information from the last resolution shell (almost half of the required information) rather than leaving them out. We were satisfied by the final calculated electron density map, which showed the bound ligand at its binding site. The same argument can be said in the case of GANA- complex, which was 88.2% complete (top shell 55.8% complete).

To assess the quality of the processed data sets we show the statistics of the top resolution shells for each data set in the tables 4.4-4.11. The statistical parameters:  $I/\sigma(I)$ ,  $R_{merge}$  and the completeness in each resolution shell are given in the tables. Ideally, the intensities of 100% of the unique Bragg reflections should be measured and the data should be significant in terms of the  $I/\sigma(I)$  ratio throughout the resolution range (Dauter & Wilson, 2001). Also,  $R_{merge}$  that provides a measure of the distribution of symmetry-equivalent observed intensities is considered to judge the data quality. The condition of  $I/\sigma(I)$  being significant in the different resolution shells was met in the data sets of all complexes. The 1.6 Å resolution data set of 6'-sialyllactose was very close to the ideal situation in terms of the completeness. For the complexes of: DANA, 3'-sialyllactose, and FANA, the completeness throughout the resolution range was still high enough to show a good quality data. In the case of CANA- and GANA-complexes the outer resolution shells exhibited lower degree of completeness, compared to the other complexes. Yet, the incomplete shells contain useful structural information; and so it was better to include the hidden information in the outer shells. The lack of completeness in these two complexes

was mainly due to the large number of weak reflections in both cases, probably, because of the small size of the crystals used to collect the data. Also, during processing of the data sets of these two complexes a large number of partial reflections were observed; this was due to the large mosaic spread in the crystals. Closer investigation of tables 4.8-4.10, which present the statistics of the 2.5 Å and 2.8 Å resolution data sets of CANA-complex as well as the statistics of the merged two data, highlighted some differences between the two data sets:  $I/\sigma(I)$  is higher in the first data set compared to that of the second data because of the shorter crystal-to-detector distance; however, the completeness of the low-resolution shells in the first data set is less than that of the second data set. This can be explained by the occurrence of more overlaps between reflections when the detector was set closer to the crystal. By merging the two data sets together, the values of  $I/\sigma(I)$  as well as the completeness were improved. The low  $R_{\text{merge}}$  (13.8 %) of the top resolution shell in the merged data sets was due to the small number of observations. Ultimately, it should be emphasised that the quality of the final electron density map calculated for the different complex structures allowed the identification of the bound ligand(s) unambiguously.

### 4.3.3 Molecular replacement

MR was used to solve the structure of the DANA complex as a representative for the unit cell dimensions observed in the DANA, 3'-sialyllactose, 6'-sialyllactose and FANA complexes. Since the CANA and GANA complexes showed similar unit cell dimensions but different from those of the other complexes, MR was necessary to find the solution for one of the two complexes (CANA complex).

**Table 4.4.** Statistics of data processing in the ten resolution shells for DANA complex

Resolution shell (Å)	$I/\sigma(I)$	Completeness (%)	$R_{\text{merge}}$ (%)
40.00 – 4.09	56.1	85.4	3.0
4.09 – 3.25	46.1	89.5	3.4
3.25 – 2.84	27.4	90.0	4.9
2.84 – 2.58	17.0	90.0	7.8
2.58 – 2.39	12.5	88.9	11.8
2.39 – 2.25	10.7	88.2	11.8
2.25 – 2.14	9.4	88.4	13.8
2.14 – 2.05	7.6	86.8	16.1
2.05 – 1.97	6.1	86.7	19.7
1.97 – 1.90	4.6	85.7	25.3
Overall reflections	19.6	87.9	6.8

**Table 4.5.** Statistics of data processing in the ten resolution shells for 3'-sialyllactose complex

Resolution shell (Å)	$I/\sigma(I)$	Completeness (%)	$R_{\text{merge}}$ (%)
40.00 – 4.09	31.4	82.1	3.7
4.09 – 3.25	28.2	86.8	4.4
3.25 – 2.84	24.7	90.0	7.3
2.84 – 2.58	15.5	91.9	10.4
2.58 – 2.39	11.1	92.5	12.8
2.39 – 2.25	8.9	93.5	15.3
2.25 – 2.14	7.7	94.0	17.7
2.14 – 2.05	6.2	94.2	21.6
2.05 – 1.97	4.9	94.9	25.7
1.97 – 1.90	3.6	94.8	32.7
Overall reflections	14.6	91.4	10.3

**Table 4.6.** Statistics of data processing in the ten resolution shells for 6'-sialyllactose complex

Resolution shell (Å)	$I/\sigma(I)$	Completeness (%)	R <sub>merge</sub> (%)
∞ – 5.06	11.2	95.7	5.0
5.06 – 3.58	11.3	97.6	5.3
3.58 – 2.92	10.2	98.5	5.7
2.92 – 2.53	9.5	99.0	6.4
2.53 – 2.26	9.0	99.2	7.2
2.26 – 2.07	8.3	99.5	8.0
2.07 – 1.91	7.0	99.5	9.7
1.91 – 1.79	5.1	99.5	13.3
1.79 – 1.69	3.7	99.8	18.6
1.69 – 1.60	2.6	99.7	26.2
Overall reflections	6.6	99.2	7.3

**Table 4.7.** Statistics of data processing in the ten resolution shells for FANA complex

Resolution shell (Å)	$I/\sigma(I)$	Completeness (%)	R <sub>merge</sub> (%)
40.00 – 4.52	28.1	97.6	3.8
4.52 – 3.59	26.7	96.2	4.5
3.59 – 3.14	23.7	97.5	5.8
3.14 – 2.85	18.3	97.3	8.3
2.85 – 2.65	13.3	96.2	11.4
2.65 – 2.49	9.8	95.8	14.3
2.49 – 2.37	7.9	94.5	18.1
2.37 – 2.26	6.5	92.9	22.5
2.26 – 2.18	5.7	87.1	26.3
2.18 – 2.10	4.9	91.4	31.3
Overall reflections	17.7	94.7	7.4

**Table 4.8.** Statistics of data processing in the ten resolution shells for CANA complex (data set 1)

Resolution shell (Å)	$I/\sigma(I)$	Completeness (%)	$R_{\text{merge}}$ (%)
40.00 – 6.03	16.1	79.1	6.6
6.03 – 4.79	13.5	80.0	8.2
4.79 – 4.18	13.3	79.3	8.4
4.18 – 3.80	11.3	77.9	9.9
3.80 – 3.53	9.3	76.8	12.2
3.53 – 3.32	8.4	71.0	14.3
3.32 – 3.15	7.3	64.3	17.2
3.15 – 3.02	5.9	57.8	21.0
3.02 – 2.90	5.4	52.9	22.7
2.90 – 2.80	4.5	46.6	26.5
Overall reflections	10.5	68.8	11.2

**Table 4.9** Statistics of data processing in the ten resolution shells for CANA complex (data set 2)

Resolution shell (Å)	$I/\sigma(I)$	Completeness (%)	$R_{\text{merge}}$ (%)
40.00 – 6.03	13.4	94.6	8.9
6.03 – 4.79	10.7	93.1	11.1
4.79 – 4.18	11.2	88.8	12.2
4.18 – 3.80	8.9	86.2	14.0
3.80 – 3.53	7.4	82.2	17.3
3.53 – 3.32	6.9	71.1	19.3
3.32 – 3.15	5.9	63.5	21.7
3.15 – 3.02	4.9	55.1	25.2
3.02 – 2.90	4.6	47.6	26.3
2.90 – 2.80	3.7	40.6	30.9
Overall reflections	8.6	72.6	14.8

**Table 4.10** Statistics of data processing in the ten resolution shells for CANA complex (merged data)

Resolution shell (Å)	$I/\sigma(I)$	Completeness (%)	$R_{\text{merge}}$ (%)
40.00 – 6.03	18.6	94.4	4.6
6.03 – 4.79	14.4	94.1	5.9
4.79 – 4.18	15.7	89.7	6.9
4.18 – 3.80	12.7	87.9	7.6
3.80 – 3.53	10.5	84.8	9.2
3.53 – 3.32	9.8	77.7	8.8
3.32 – 3.15	9.0	68.7	9.9
3.15 – 3.02	6.9	62.3	11.2
3.02 – 2.90	5.7	56.5	12.3
2.90 – 2.80	6.1	48.1	13.8
Overall reflections	12.0	76.7	7.5

**Table 4.11** Statistics of data processing in the ten resolution shells for GANA complex

Resolution shell (Å)	$I/\sigma(I)$	Completeness (%)	$R_{\text{merge}}$ (%)
40.00 – 5.60	27.8	99.8	6.6
5.60 – 4.44	25.1	99.9	8.1
4.44 – 3.88	22.2	99.9	9.4
3.88 – 3.53	16.5	99.9	12.2
3.53 – 3.28	12.6	99.7	14.7
3.28 – 3.08	8.5	97.1	17.8
3.08 – 2.93	7.0	87.4	19.1
2.93 – 2.80	5.8	75.8	22.7
2.80 – 2.69	5.0	64.8	25.7
2.69 – 2.60	4.4	55.8	26.9
Overall reflections	14.1	88.2	11.9

In the case of DANA complex, applying the cross-rotational function of the MR in AMoRe to the search model (pdb-code 1KIT) resulted in a unique rotation solution as judged by the correlation coefficient between the observed and calculated amplitudes for the crystal and the model, respectively. The first solution corresponding to the highest Patterson peak had a correlation coefficient of 8.8%, and the next highest peak had a correlation coefficient of 1.4%. The top 10 solutions of the cross-rotation function are displayed in Table 4.12. The first two solutions from the rotation stage were passed to the next step in MR to apply the cross-translation function in order to locate the molecule in the unit cell. The cross-translation function clearly differentiated between the two rotation solutions. Translation of the first rotation solution gave a correlation coefficient of 32.2%, compared to 5.9% for the second rotation solution; see Table 4.13. The rigid body refinement of the best solution after rotation and translation, improved the correlation coefficient (from 32.2% to 43.2%) and slightly lowered the R-factor (from 48.6% to 45.4%), see table 4.14. The MR solution was then applied to the coordinates of the search model to work as the model solution in the refinement stage.

For CANA complex, a model structure of unsuccessful data set of 2.4 Å (Daresbury data), as mentioned in the methods section, was employed as the search model (this experiment is not discussed in the thesis). Applying the cross-rotation function resulted in a unique solution (only one solution) with correlation coefficient of 11.5%, see table 4.15. The cross-translation applied on the rotation solution resulted in two solutions as can be seen in the table 4.16. The correlation coefficient of the highest Patterson peak was 42.6% compared to 24.0% for the next peak. Rigid body refinement on the top solution (table 4.17) resulted from rotation and translation improved the correlation coefficient (42.6% to 44.4%) and lowered the R-factor (46.4% to 45.5%).

**Table 4.12.** Top 10 solutions of the cross-rotation function in MR, solutions passed to the cross-translation function are highlighted in magenta. (DANA complex)

Solution	$\alpha$	$\beta$	$\gamma$	<sup>a</sup> CC_F	<sup>b</sup> RF_F	<sup>c</sup> CC_I	<sup>d</sup> CC_P
1	99.13	40.07	79.53	8.8	57.3	15.3	14.3
2	112.4	49.13	38.33	1.4	59.3	3.2	4.7
3	128.50	58.36	25.73	1.0	59.4	1.7	4.4
4	164.00	33.08	148.50	1.0	59.7	2.5	4.9
5	76.94	59.37	278.41	0.8	59.6	1.7	4.4
6	51.00	72.19	110.11	0.8	59.8	2.5	5.6
7	114.73	81.91	145.38	0.7	60.0	1.2	4.5
8	58.50	23.50	204.00	0.7	59.9	1.9	4.4
9	168.95	57.52	291.00	0.6	59.6	0.9	4.4
10	67.93	76.50	232.39	0.5	59.7	1.3	4.4

<sup>a</sup>CC\_F, expressed in percentage (%), is the correlation coefficient between the observed amplitudes for the crystal and the calculated amplitudes for the model.

<sup>b</sup>RF\_F, expressed in percentage (%), is the classic R factor between the observed amplitudes for the crystal and the calculated amplitudes for the model.

<sup>c</sup>CC\_I, expressed in percentage (%), is the correlation coefficient between the observed intensities for the crystal and the sum of calculated intensities for all symmetry equivalents of the model.

<sup>d</sup>CC\_P, expressed in percentage (%), is the Patterson correlation coefficient between the crystal and the model pattersons evaluated within the defined sphere centred on the Patterson origin.

**Table 4.13.** Cross-translation function applied to the two solutions from the rotation stage (DANA complex)

Solution	$\alpha$	$\beta$	$\gamma$	TX	TY	TZ	CC_F	RF_F	CC_I
1	99.2	40.7	79.3	0.0818	0.2258	0.3552	32.1	48.5	32.7
2	112.54	49.18	38.39	0.0478	0.0930	0.0643	5.9	56.3	5.5

**Table 4.14.** Rigid-body refinement for the first solution after rotation and translation DANA complex

Solution	$\alpha$	$\beta$	$\gamma$	TX	TY	TZ	CC_F	RF_F	CC_I
1	99.9	39.3	80.7	0.0811	0.2310	0.3518	43.1	45.1	46.1

**Table 4.15** The solution of the cross-rotation function in MR (CANA complex)

Solution	$\alpha$	$\beta$	$\gamma$	CC_F	RF_F	CC_I	CC_P
1	-1.31	82.07	355.34	11.1	55.1	20.1	20.1

**Table 4.16.** Cross-translation function applied to the rotation solution (CANA complex)

Solution	$\alpha$	$\beta$	$\gamma$	TX	TY	TZ	CC_F	RF_F	CC_I
1	-1.39	82.7	355.34	0.0612	0.1252	0.4850	42.5	46.1	41.5
1	-1.39	82.07	355.84	0.0644	0.1233	0.0000	24.0	52.8	25.7

**Table 4.17.** Rigid-body refinement for the first solution after rotation and translation (CANA complex)

Solution	$\alpha$	$\beta$	$\gamma$	TX	TY	TZ	CC_F	RF_F	CC_I
1	-1.33	82.11	356.17	0.0652	0.1211	0.4857	44.1	45.5	44.1

#### 4.3.4 Model refinement

During the rigid body refinement the three domains of the VCNA enzyme structure: N-terminal domain (aa 25-216), central domain (aa 217-347, aa 543-777), and C-terminal domain (aa 348-542) were treated as three rigid bodies to allow the relative motions between the three subunits. Using the model as a rigid single unit did not drop the R factor. This was an indication for the relative movements between the domains, which could be resulted from the crystal packing, which was different from the native structure as discussed in section 5.2. Subsequent steps of refinement as described in material and methods section resulted in a decrease of R-factor and R-free. Visual inspection of the electron density maps,  $2F_o-F_c$  and  $F_o-F_c$ , clearly exhibited the extra densities for the bound ligands in the different complexes, as will be discussed in details in the next chapter. The refinement statistics for the six complexes are shown in table 4.18. The better the quality of the model, the lower the R-factor and R-free. Usually, during refinement R-free values are higher than R-factor, but in the final model they become more similar (Rhodes, 2000). The quality of the different complex structures are varied as can be judged by the values of R-factor and R-free. The model structures of the complexes of DANA (1.9 Å), 3'-sialyllactose (1.9 Å), and FANA (2.1 Å) are of good quality with (R-factor  $\leq$  0.20). Although the quality of the model structures of GANA (2.6 Å) and CANA (2.8 Å) were not as good as the quality of the higher resolution structures, the extra densities of the bound ligands were unambiguously determined in the two complexes as shown in chapter 5.

B-factor of the main-chain and side-chain protein atoms is another measure of the quality of the model. Higher values of B-factor reflect larger fluctuations of the atoms, which mean less accuracy in determining atom positions. The high quality 1.6 Å structure possessed the lowest B-factors, see table 4.18. The B-factors were higher for the medium

**Table 4.18** Refinement statistics

Complex structure	VC-DANA	VC-3'Sialyllactose	VC-6'Sialyllactose
Resolution range (Å)	100-1.9	100-1.9	100-1.6
R-factor <sup>1</sup>	0.18	0.18	0.20
R-free <sup>2</sup>	0.23	0.22	0.22
rms deviation in bond lengths(Å)	0.005	0.005	0.005
rms deviation in bond angles (°)	1.36	1.32	1.32
Number of non-hydrogen atom	5827	5827	5827
Number of water molecules	869	943	898
Average B-factor (Å <sup>2</sup> )			
Main-chain atoms	12.7	13.5	11.9
Side chain atoms	15.0	15.8	14.8
Average B-factor (Å <sup>2</sup> ) ligand:			
Dana	24.2	27.1	-
Sialic acid	22.6	29.5	33.3
Glycerol	-	-	33.0
Tris	-	-	26.9

**Table 4.18 ( continued)** Refinement statistics

Complex structure	VC-FANA	VC-CANA	VC-GANA
Resolution range	100-2.1	100-2.8	100-2.6
R-factor <sup>1</sup>	0.19	0.25	0.22
R-free <sup>2</sup>	0.24	0.32	0.28
rms deviation in bond lengths(Å)	0.005	0.008	0.007
rms deviation in bond angles (°)	1.33	1.44	1.39
Number of non-hydrogen atom	5827	5827	5827
Number of water molecules	683	67	199
Average B-factor (Å <sup>2</sup> )			
Main-chain atoms	20.9	55.4	39.7
Side chain atoms	22.4	55.6	40.4
Average B-factor (Å <sup>2</sup> ) ligand:			
FANA	21.7	-	-
FNANA <sup>3</sup>	39.8		
GANA	-		42.7
CANA	-	50.4	-

<sup>1</sup>R-factor =  $\sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ .

<sup>2</sup>R-free as R-factor but summed over a (5-10%) test set of reflections

<sup>3</sup>FNANA is the trifluoroacetyl-derivative of sialic acid (N-trifluoroacetylneuraminic acid)

resolution structures of GANA and CANA. Also, larger numbers of water molecules were picked up in the high resolution structures of DANA (869 waters), 3'-sialyllactose (943 waters), 6'-sialyllactose (898 waters), and FANA (686 waters). Only 199, and 67 water molecules were picked up in the case of GANA and CANA complexes, respectively.

The geometry of the model structures of the different complexes is in a good agreement with geometry of well-refined structures as judged by the values of rms (root mean square deviation) of both bond lengths and bond angles. Deviations of bond lengths  $< 0.02 \text{ \AA}$  and angle deviations  $< 4^\circ$  are expected for good models (Rhodes, 2000), which is the case in our model structures.

#### **4.3.5 Structure validation**

The refined structures of the six complexes were assessed by PROCHECK (Laskowski et al., 1993). Figures (4.2-4.7) display the 2D Ramachandran plots of the  $(\phi, \psi)$  angles as well as the parameters for the main- and side chains for the different refined structures. In the six complexes none of the protein residues but Asn51 in the 6'-sialyllactose complex structure existed in the disallowed region of the Ramachandran plot. Asn51 was located at the margin of the generally allowed region with  $(\psi = -119.1^\circ)$ ; the same residue was classified by the program in the generally allowed region in the 3'-sialyllactose complex with  $(\psi = -119.9^\circ)$ . The medium resolution structures of GANA and CANA complexes possessed less number of residues (80.6 and 73.1%, respectively) in the most favoured region compared to (86.3 to 86.9%) in the other four complexes.

### VC-DANA

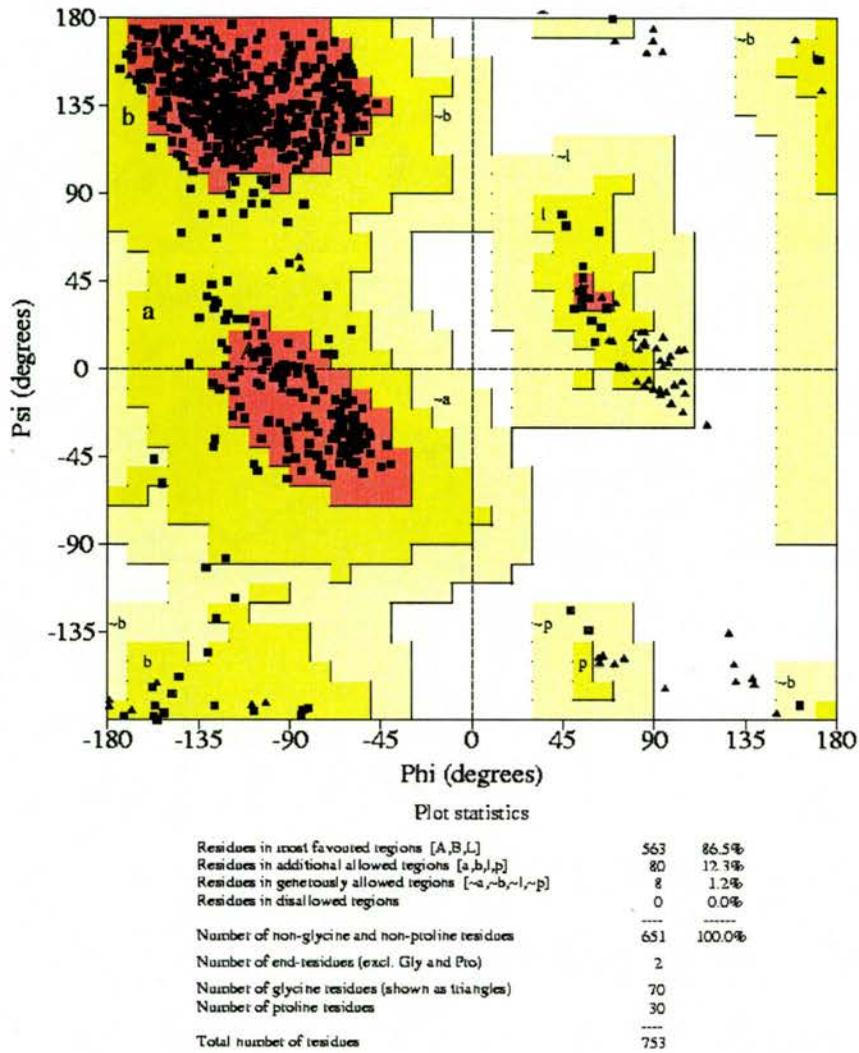
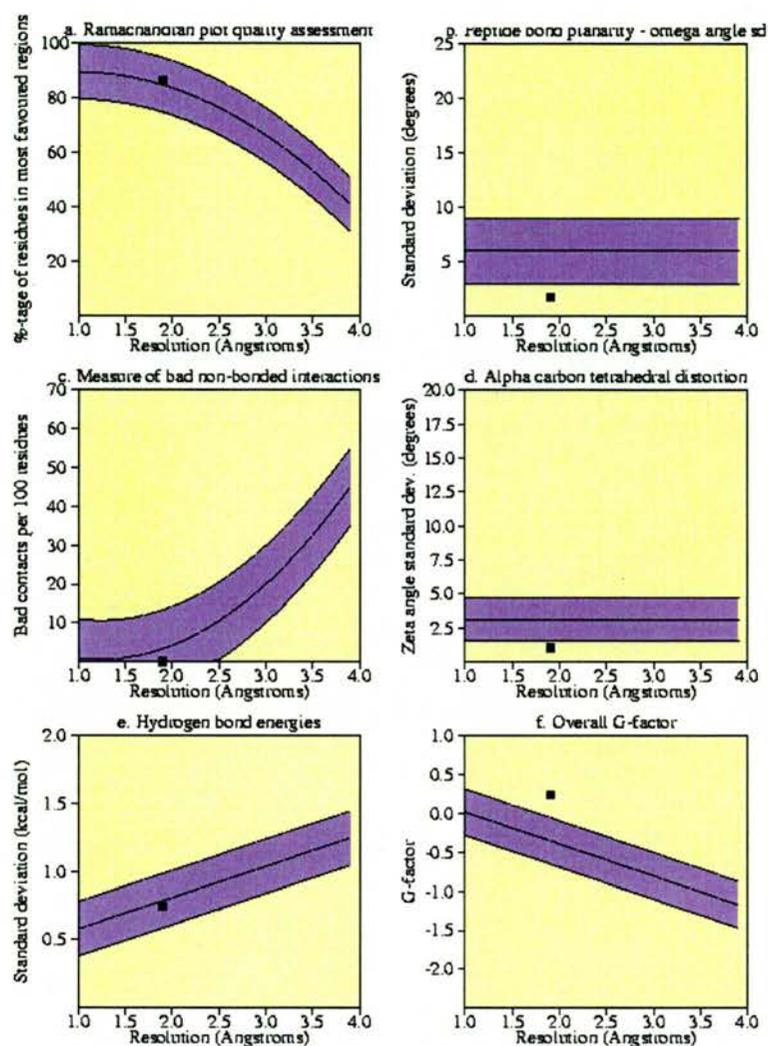


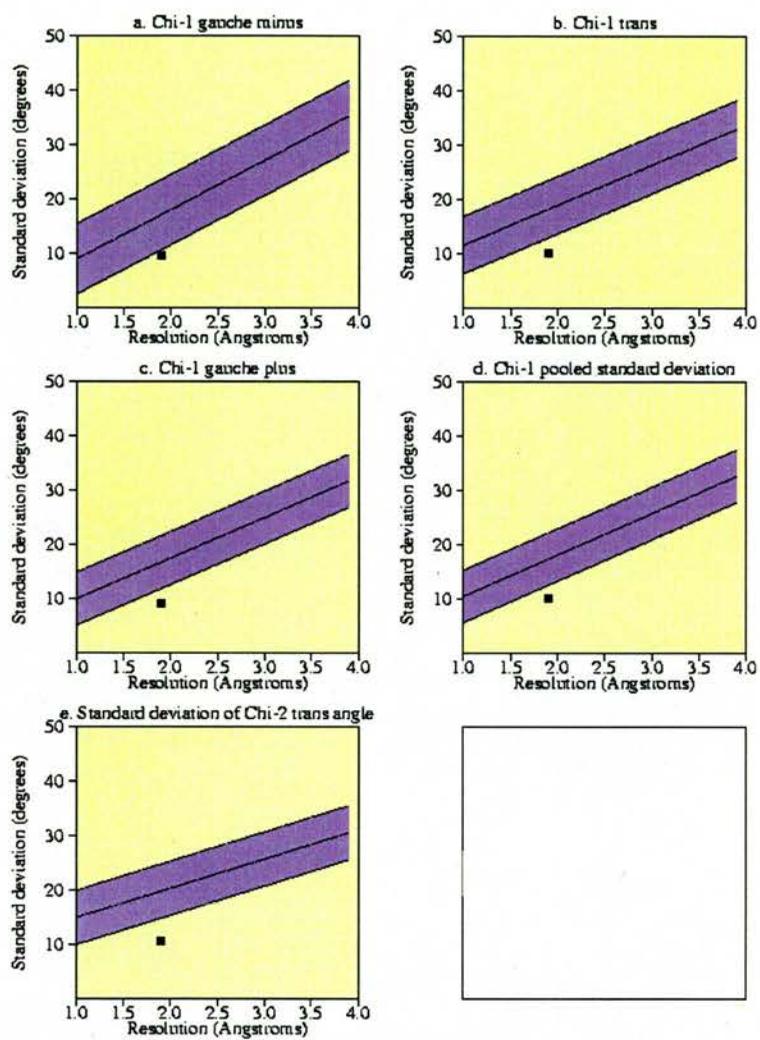
Figure 4.2a. Ramachandran plot for DANA complex



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. %age residues in A, B, L	651	86.5	84.9	10.0	0.2	Inside
b. Omega angle st dev	748	1.7	6.0	3.0	-1.4	BETTER
c. Bad contacts / 100 residues	0	0.0	3.3	10.0	-0.3	Inside
d. Zeta angle st dev	683	1.0	3.1	1.6	-1.3	BETTER
e. H-bond energy st dev	462	0.7	0.8	0.2	-0.2	Inside
f. Overall G-factor	753	0.2	-0.3	0.3	2.0	BETTER

Figure 4.2b. Main-chain parameters for DANA complex



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	125	9.6	17.2	6.5	-1.2	BETTER
b. Chi-1 trans st dev	197	10.1	18.3	5.3	-1.5	BETTER
c. Chi-1 gauche plus st dev	281	9.2	16.8	4.9	-1.5	BETTER
d. Chi-1 pooled st dev	603	10.1	17.4	4.8	-1.5	BETTER
e. Chi-2 trans st dev	181	10.8	19.9	5.0	-1.8	BETTER

Figure 4.2c. Side-chain parameters for DANA complex

VC-3'-sialyllactose

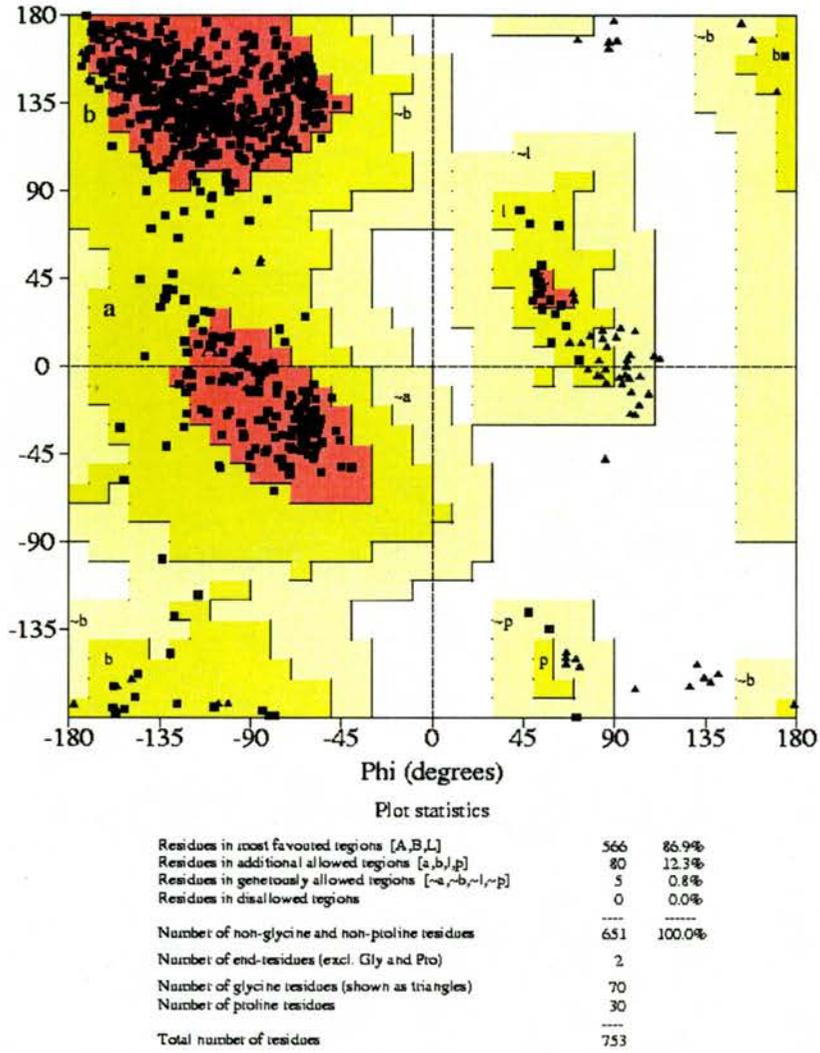
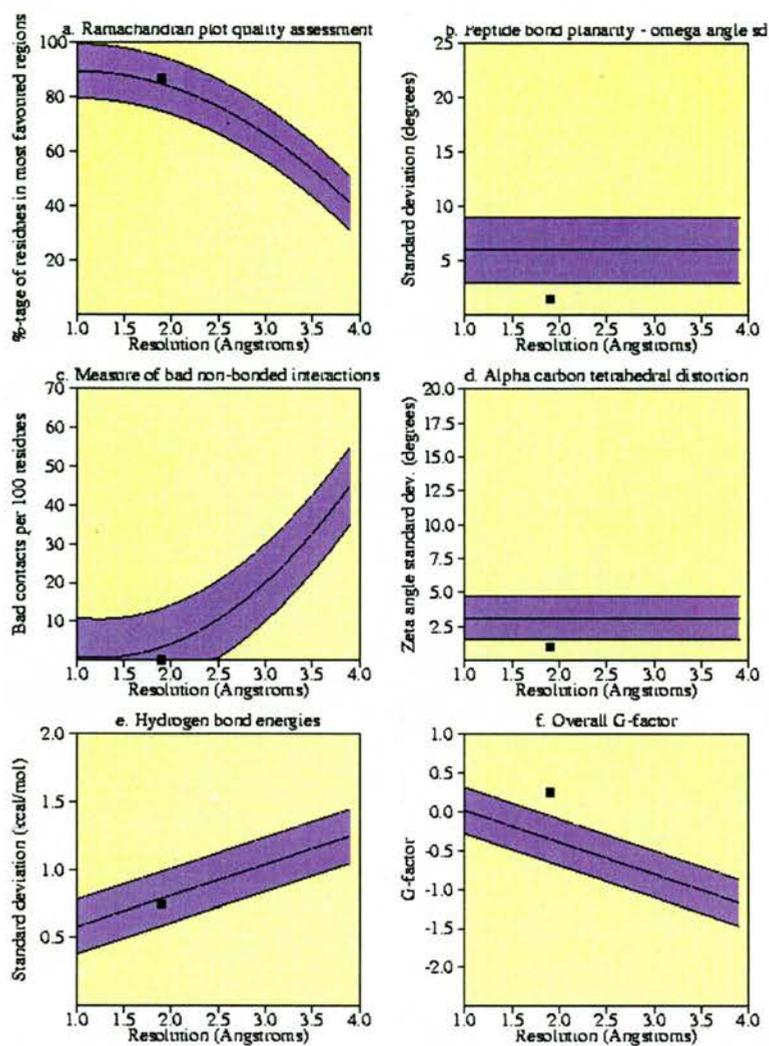


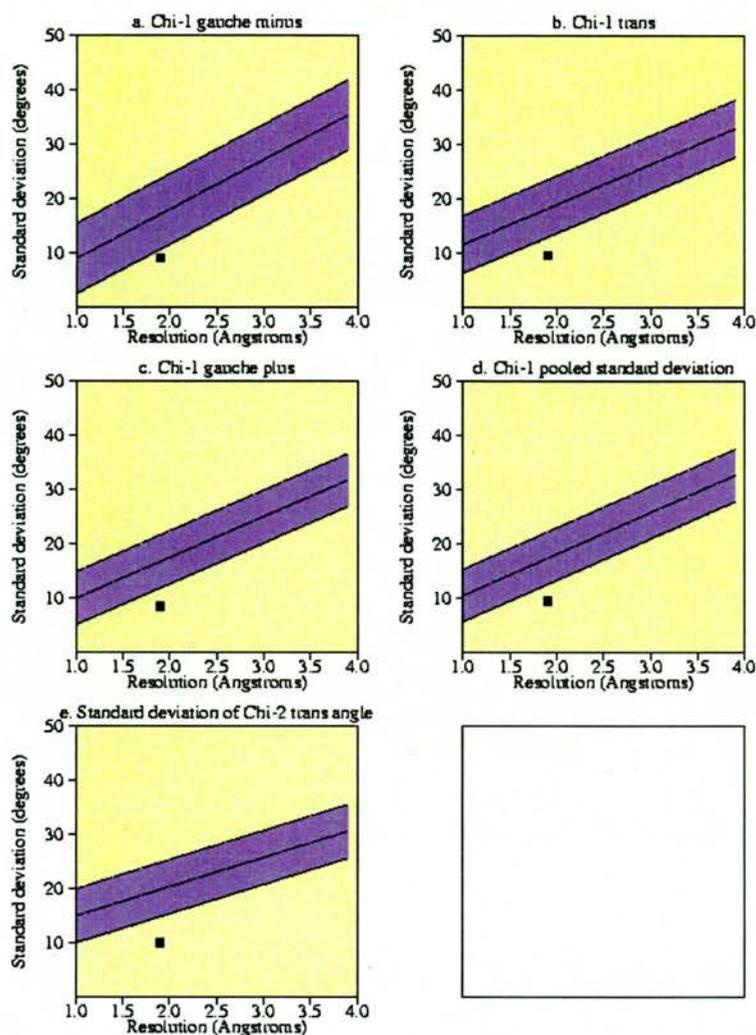
Figure 4.3a. Ramachandran plot of 3'-sialyllactose complex



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. %age residues in A, B, L	651	86.9	84.9	10.0	0.2	Inside
b. Omega angle st dev	748	1.5	6.0	3.0	-1.5	BETTER
c. Bad contacts / 100 residues	0	0.0	3.3	10.0	-0.3	Inside
d. Zeta angle st dev	683	1.0	3.1	1.6	-1.3	BETTER
e. H-bond energy st dev	465	0.7	0.8	0.2	-0.2	Inside
f. Overall G-factor	753	0.3	-0.3	0.3	2.0	BETTER

Figure 4.3b. Main-chain parameters for 3'-sialyllactose complex



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	126	9.1	17.2	6.5	-1.2	BETTER
b. Chi-1 trans st dev	193	9.7	18.3	5.3	-1.6	BETTER
c. Chi-1 gauche plus st dev	284	8.5	16.8	4.9	-1.7	BETTER
d. Chi-1 pooled st dev	603	9.6	17.4	4.8	-1.6	BETTER
e. Chi-2 trans st dev	179	10.1	19.9	5.0	-2.0	BETTER

Figure 4.3c. Side-chain parameters for 3'-sialyllactose complex

### VC-6'-sialyllactose

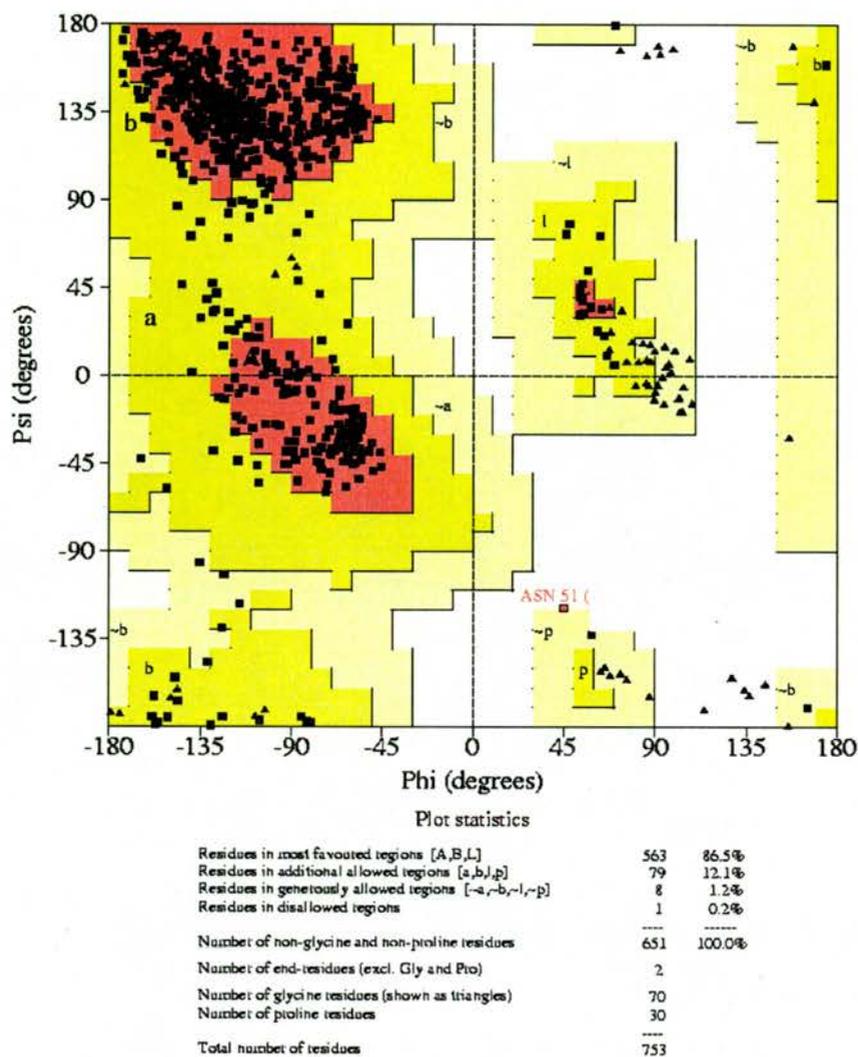
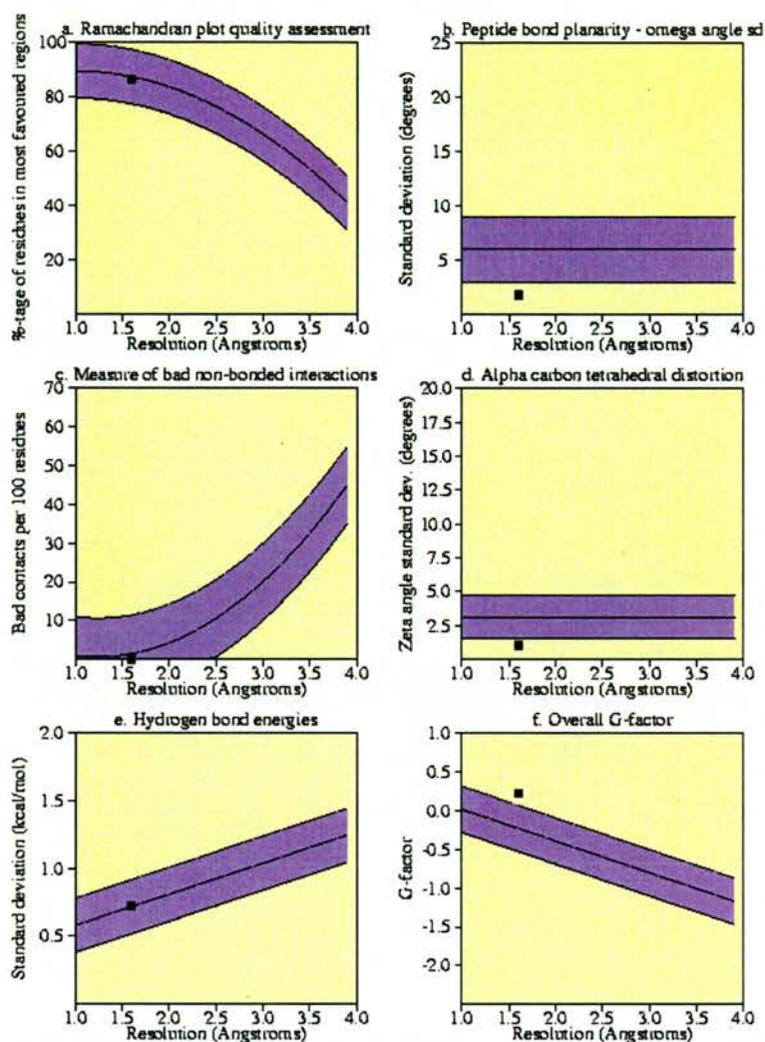


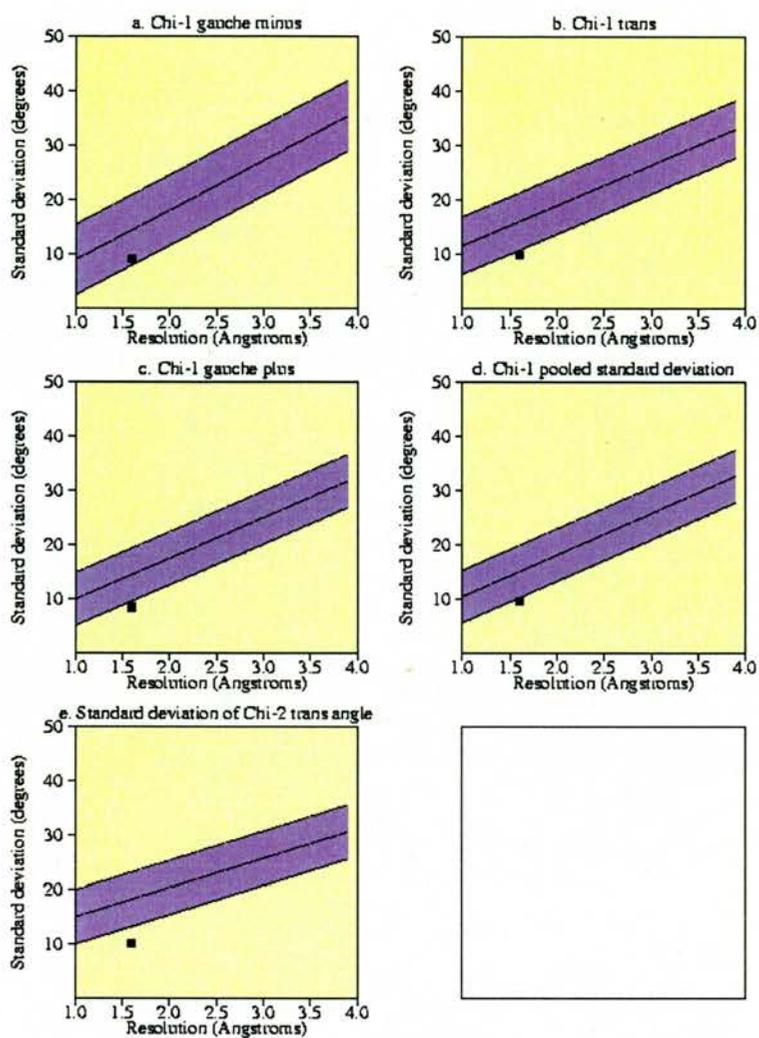
Figure 4.4a. Ramachandran plot of 6'-sialyllactose



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean
			Typical value	Band width	
a. %age residues in A, B, L	651	86.5	87.5	10.0	-0.1 Inside
b. Omega angle st dev	748	1.8	6.0	3.0	-1.4 BETTER
c. Bad contacts / 100 residues	0	0.0	1.4	10.0	-0.1 Inside
d. Zeta angle st dev	683	1.0	3.1	1.6	-1.3 BETTER
e. H-bond energy st dev	461	0.7	0.7	0.2	0.0 Inside
f. Overall G-factor	753	0.2	-0.2	0.3	1.5 BETTER

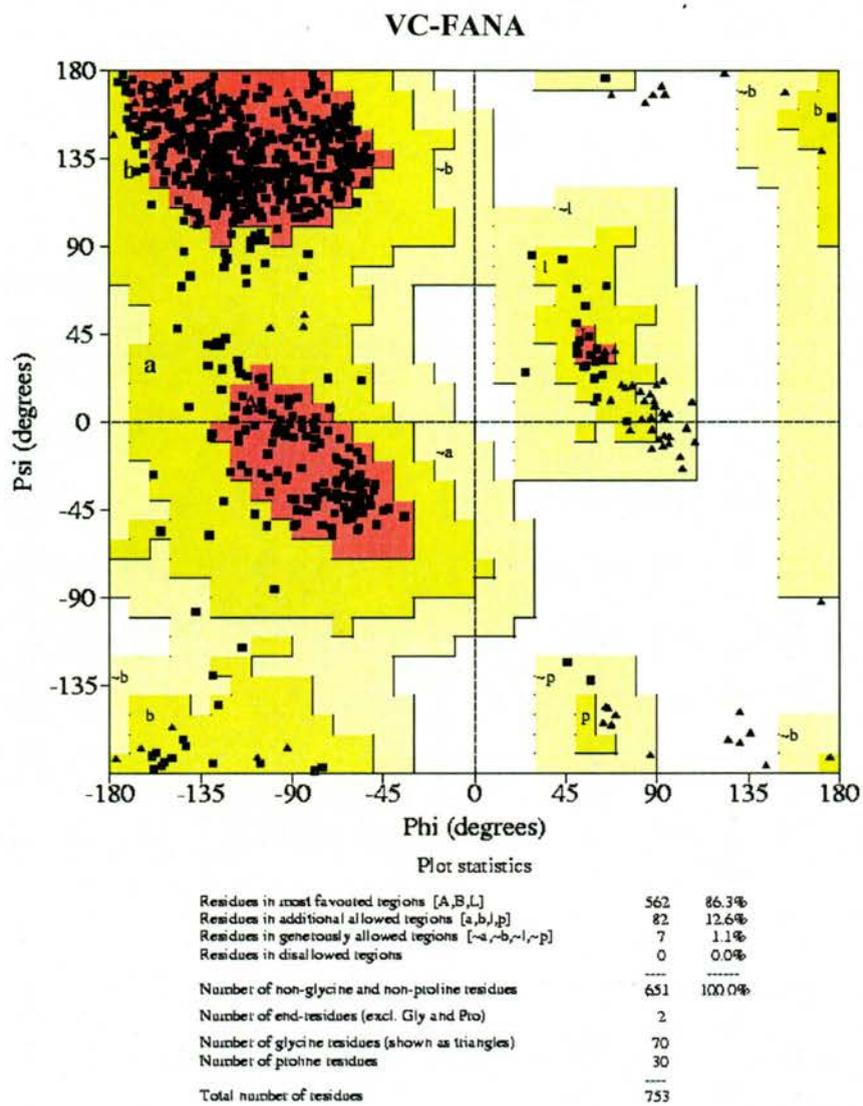
Figure 4.4b. Main-chain parameters for 6'-sialyllactose complex



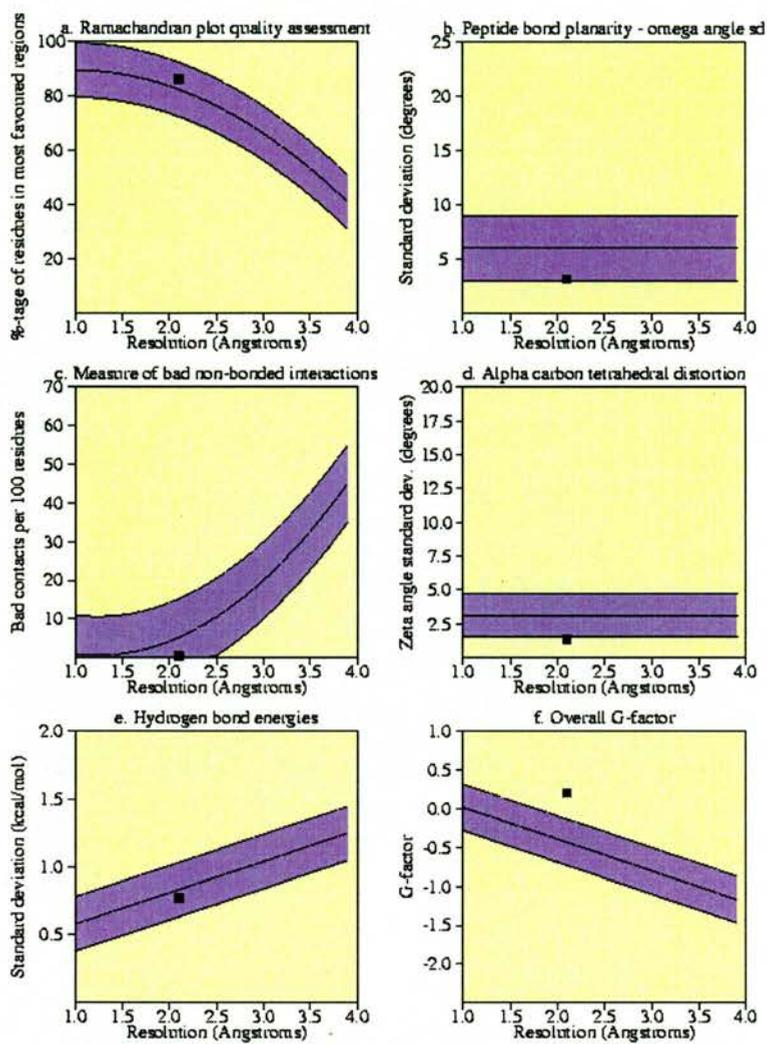
Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean
			Typical value	Band width	
a. Chi-1 gauche minus st dev	124	9.1	14.5	6.5	-0.8 Inside
b. Chi-1 trans st dev	199	10.0	16.1	5.3	-1.1 BETTER
c. Chi-1 gauche plus st dev	280	8.5	14.5	4.9	-1.2 BETTER
d. Chi-1 pooled st dev	603	9.6	15.1	4.8	-1.1 BETTER
e. Chi-2 trans st dev	182	10.1	18.2	5.0	-1.6 BETTER

Figure 4.4c. Side-chain parameters for 6'-sialyllactose complex



**Figure 4.5a.** Ramachandran plot of FANA-complex



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean
			Typical value	Band width	
a. %-tage residues in A, B, L	651	86.3	82.6	10.0	0.4 Inside
b. Omega angle st dev	748	3.2	6.0	3.0	-0.9 Inside
c. Bad contacts / 100 residues	3	0.4	5.2	10.0	-0.5 Inside
d. Zeta angle st dev	683	1.4	3.1	1.6	-1.1 BETTER
e. H-bond energy st dev	460	0.8	0.8	0.2	-0.3 Inside
f. Overall G-factor	753	0.2	-0.4	0.3	2.1 BETTER

Figure 4.5b. Main-chain parameters for FANA-complex

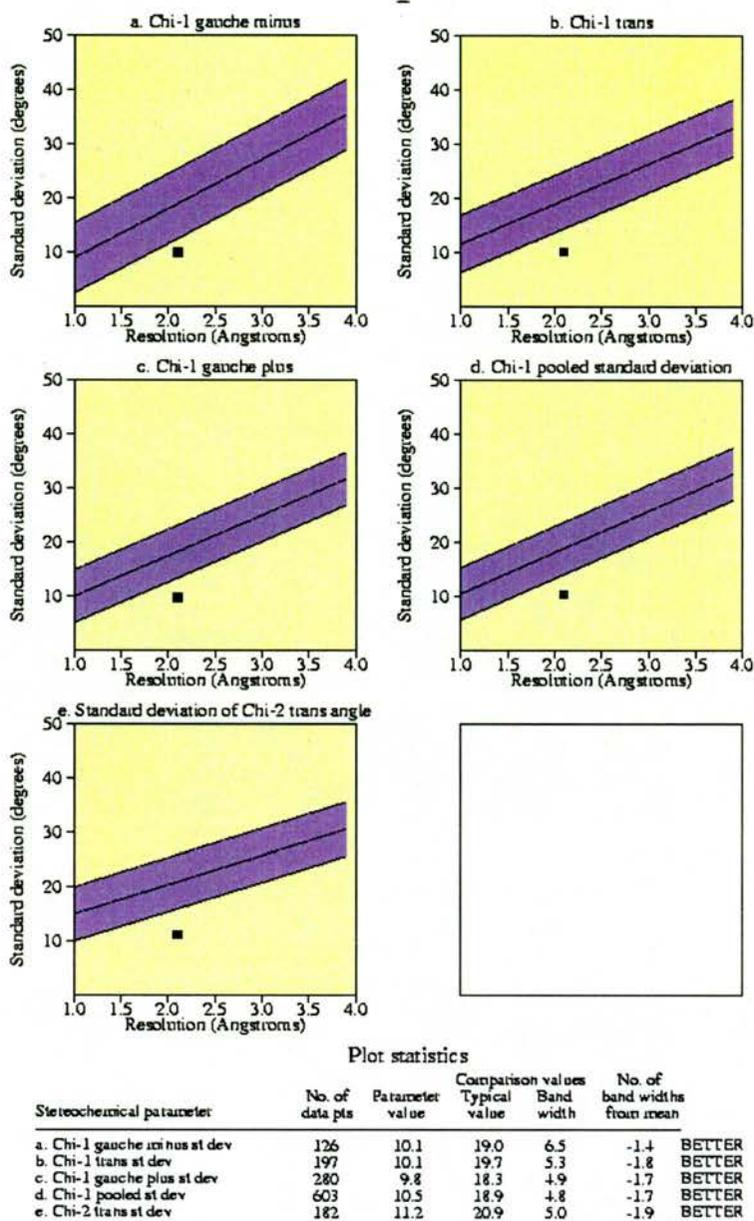
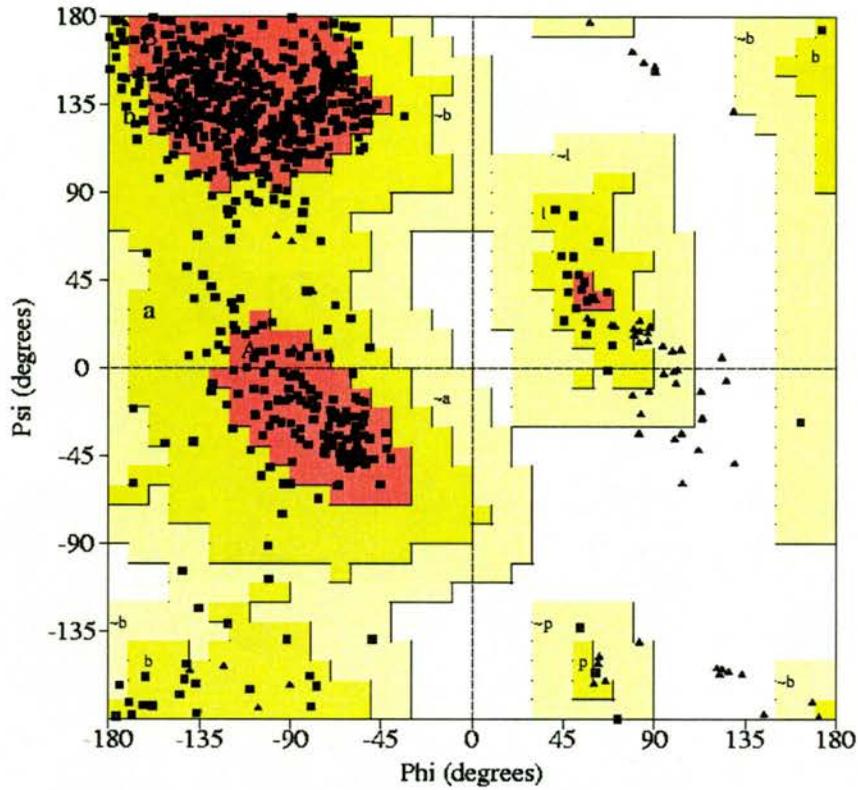


Figure 4.5c. Side-chain parameters for FANA-complex

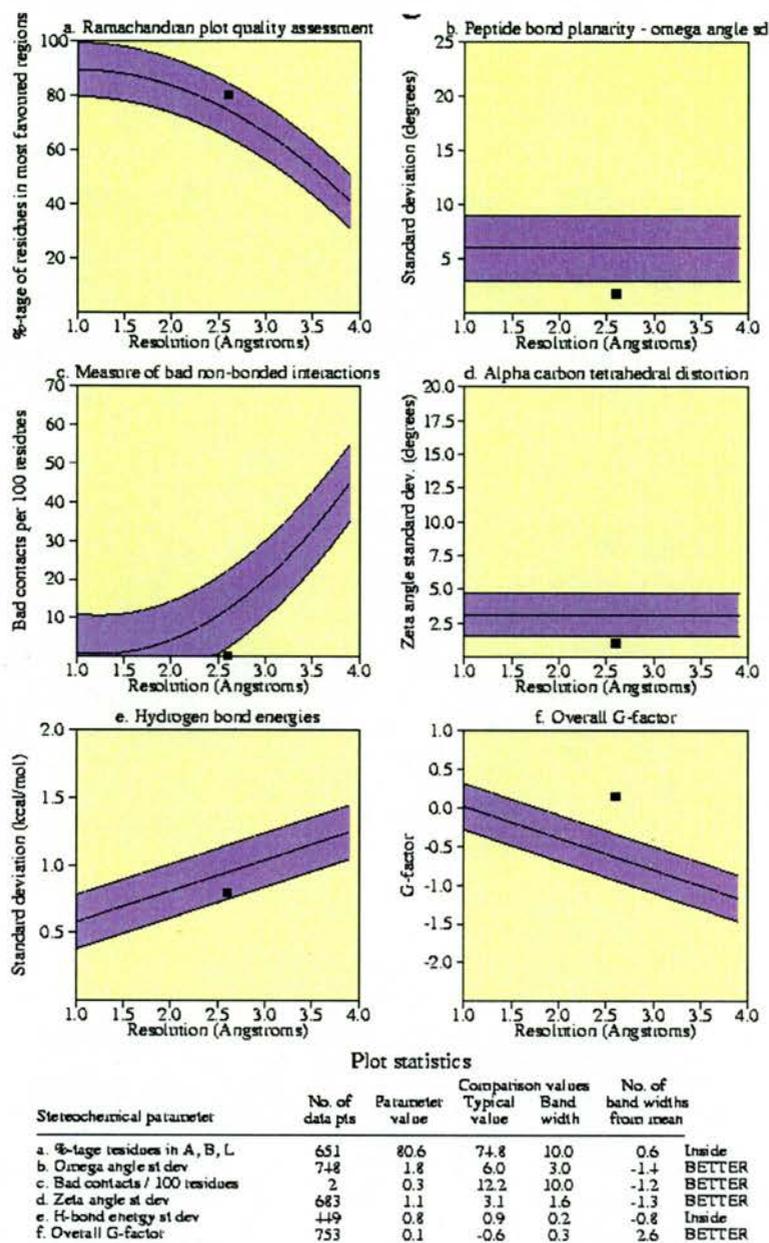
### VC-GANA



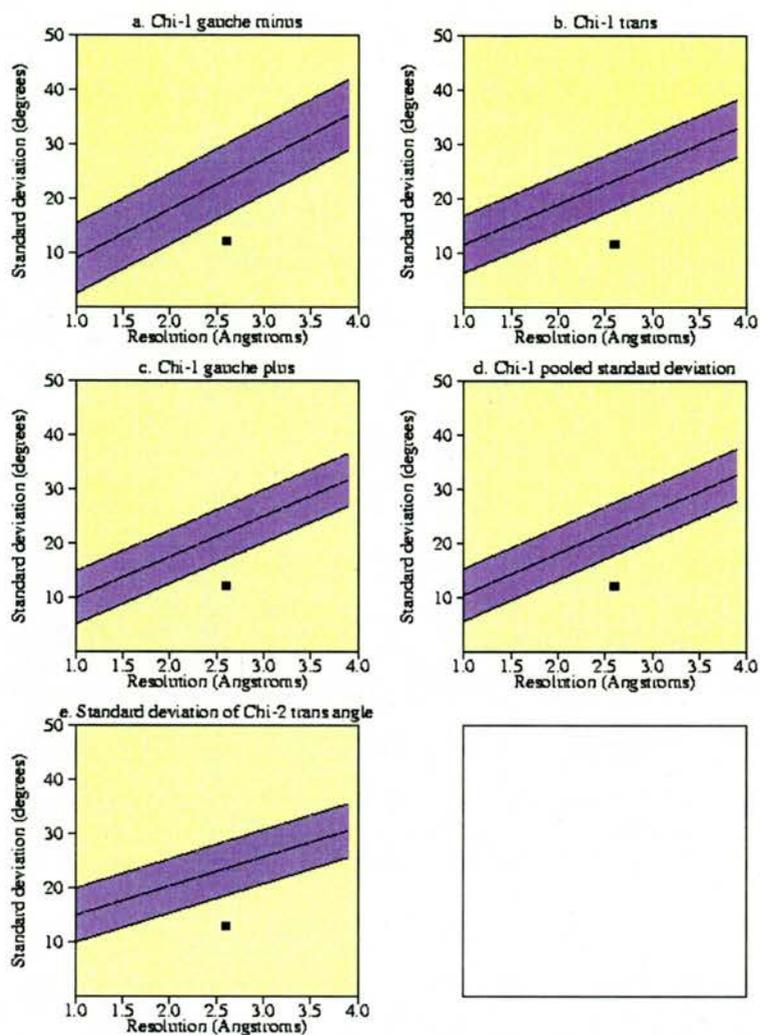
#### Plot statistics

Residues in most favoured regions [A,B,L]	525	80.6%
Residues in additional allowed regions [a,b,l,p]	113	17.4%
Residues in generously allowed regions [-a,-b,-l,-p]	13	2.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	651	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	70	
Number of proline residues	30	
Total number of residues	753	

Figure 4.6a. Ramachandran plot of GANA complex



**Figure 4.6b.** Main-chain parameters for GANA-complex



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison val Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	124	12.2	23.6	6.5	-1.8	BETTER
b. Chi-1 trans st dev	202	11.7	23.4	5.3	-2.2	BETTER
c. Chi-1 gauche plus st dev	277	12.2	22.0	4.9	-2.0	BETTER
d. Chi-1 pooled st dev	603	12.3	22.8	4.8	-2.2	BETTER
e. Chi-2 trans st dev	174	13.0	23.6	5.0	-2.1	BETTER

Figure 4.6c. Side-chain parameters for GANA-complex

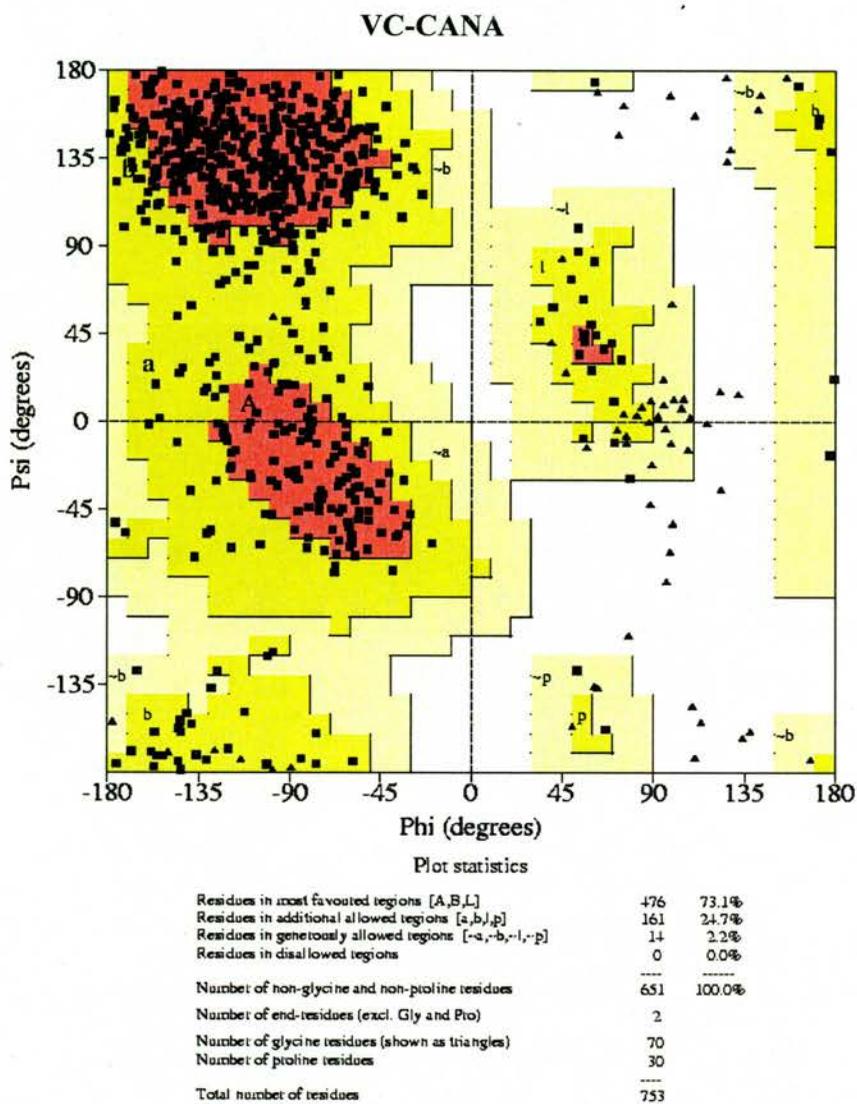
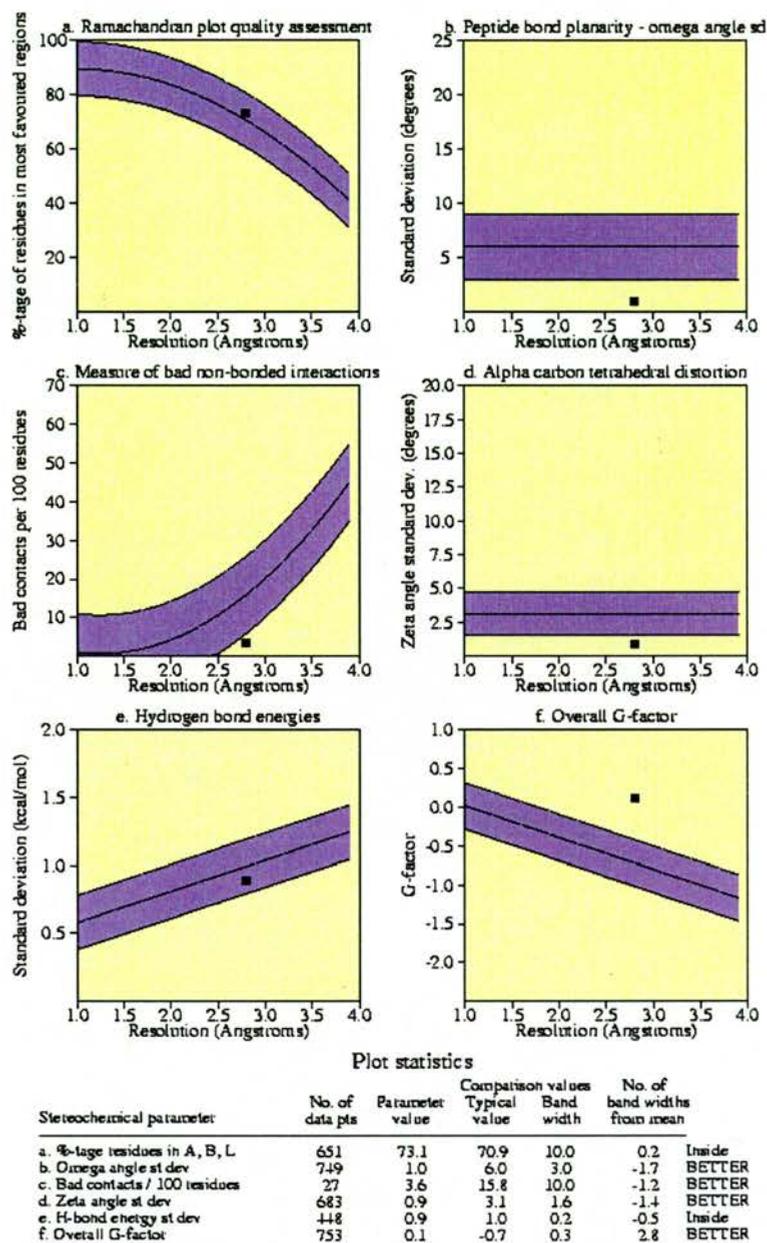
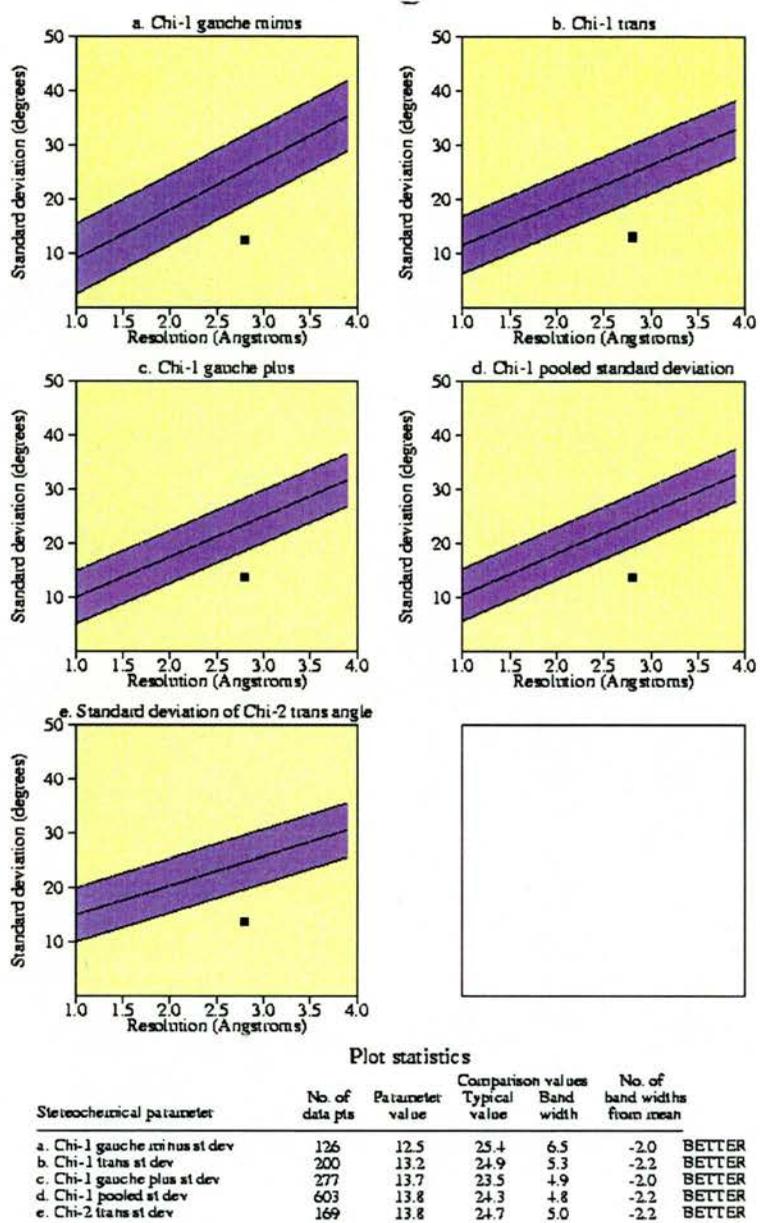


Figure 4.7a. Ramachandran plot of CANA complex



**Figure 4.7b.** Main-chain parameters for CANA-complex



**Figure 4.7c.** Side-chain parameters for CANA-complex

#### 4.4. Discussion

For a long time, only the small needle shape crystals were obtained when cocrystallizing VCNA with its ligands. Such small crystals were not suitable for the in-house data collection. Fortunately, the needle crystals of DANA- and 3'-sialyllactose-complexes diffracted well to a high resolution (1.8 Å) at ESRF. Meanwhile, trying to investigate the role of the lectin domains by cocrystallizing the enzyme with a series of available sugars, we observed that rod shape crystals were grown rather than needle crystals; not only that but the growth rate of the crystals was accelerated. Though, we did not obtain any positive result from this crystallographic sugar screening (not discussed in this thesis), we obtained benefits from including the sugar cocktail (sucrose, maltose, and lactose) of concentration (10 mM) of each disaccharide in the crystallization conditions. Not only were the crystals better looking, but with enhanced diffracting power as was seen in the data of the 6'-sialyllactose complex collected at the ESRF (1.6 Å resolution). At the same time, a new X-ray generator, which produces more intense focused beam, was installed in St-Andrews and enabled us to collect data in-house on the rod shape crystals of the other complexes.

Thereafter, we tried to cocrystallize the VCNA enzyme with the planned ligands in the same way as we did with the DANA-complex; however, we could not get any crystals. The nature of the ligands somehow affected the crystallization behaviour of the system. We have no explanation for why it was feasible to obtain cocrystals of DANA complex but not of the DANA-derivative complexes. When we tried to soak the crystals by adding the ligand to the same drop before transferring to the cryoprotector for mounting, the crystals in the case of FANA could not survive for long. In the case of GANA and CANA, crystals survived the soaking experiment. Fortunately, we observed that after the crystals of FANA-complex disappeared new crystals reformed (now in presence of the

ligand) within a few days. It should be emphasized that it was necessary to work with as low concentration of the ligands as 5 mM (or less). The new reformed crystals diffracted in-house, as described in the results section.

It is necessary to rotate the crystal so that all unique reciprocal lattice points pass through the Ewald sphere and the associated intensities are recorded on the detector. Ideally, the intensities of 100% of the unique Bragg reflections should be measured (Dauter & Wilson, 2001). In our case, this was not always possible but we considered that the crystallographic experiment would be counted as a success only if the electron density of the ligand(s) could be located unambiguously in the protein site(s) as well as the binding residues surrounding the ligand(s). In spite of the lack of completeness in the top resolution shell of the two complexes of CANA and GANA, the density of the ligands was identified clearly in the  $F_o - F_c$  and  $2F_o - F_c$  maps as will be shown in the next chapter.

Why did MR have to be used? Despite the crystals of the complexes and the native room temperature-structure belonging to the same space group ( $P2_12_12_1$ ), different unit cell dimensions were obtained. Because of this dimensional change, it was not possible to generate a difference Fourier electron density map of the complex structure using the native one as the phasing model; and MR was used to find the solution that fits the crystallographic data.

The unit cell dimensions of the complexes of: DANA, 3'-sialyllactose, 6'-sialyllactose, and FANA were close to each other. Therefore, finding the solution for one of these complexes was sufficient to find the model solution that can be used for the other three complexes. We had chosen DANA-complex structure not for any particular reason,

just because it was the first solved structure. CANA- and GANA-complexes possessed similar unit cell dimensions but different from the other four complexes. The structure of CANA-complex was solved first by MR and then used as the phasing model to solve the structure of GANA-complex directly.

For completeness, we have collected in-house data sets for crystals of VCNA grown in presence of lactose, heparin sulfate disaccharide sodium salt, and for crystals of the enzyme soaked in the thiosialoside, Neu5Ac- $\alpha$ -S-(2,6)-Gal $\beta$ Me. Unfortunately, we could not identify any binding of the putative ligands to the enzyme. These unsuccessful experiments are not discussed in the thesis.

## Chapter 5

### Structures of the VCNA-Complexes

#### 5.1 Introduction

In the previous chapter we discussed the different steps encountered in the X-ray experiments: crystallization, data collection and processing, molecular replacement, refinement and validation of the model structures for the complexes of VCNA with the six ligands: DANA, 3'-sialyllactose, 6'-sialyllactose, FANA, GANA, and GANA. In this chapter, the 3D-structures of the different complexes are discussed. The 3D structures of these complexes provide important information that can be used in structure-based drug design to find a blockade against the cholera enzyme.

In addition, as discussed in this chapter, the 3D structure of the DANA complex, surprisingly, unmasked the function of one of the two-lectin domains of VCNA where the N-terminal lectin domain revealed  $\alpha$ -sialic acid-binding site. This new finding was further confirmed in the structures of 3'-sialyllactose-, 6'-sialyllactose-, and FANA-complexes; full account of the new discovery is covered herein.

Also, discussed at the beginning of this chapter is the analysis of the two different crystal-packing patterns observed in the complex structures. Such analysis enabled us to pinpoint the origins of the observed variations.

## 5.2 Comparison between free and bound enzyme

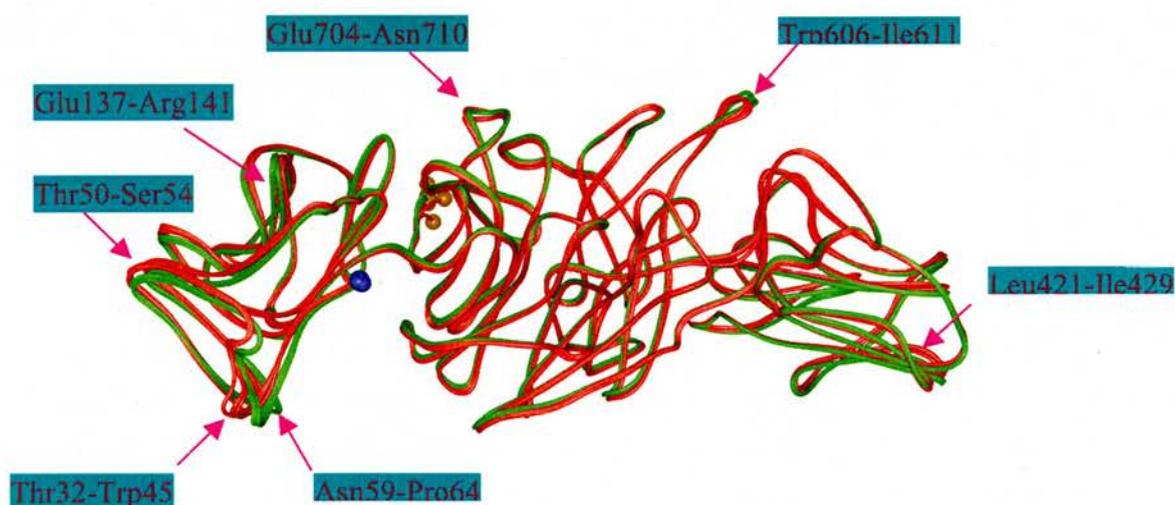
### 5.2.1 Overall structures

Crennell and co-workers (Crennell et al., 1994) previously described the native structure of VCNA solved at room temperature to a resolution of 2.4 Å (pdb code, 1KIT). The enzyme structure of 757 amino acids (aa 25-781) has three domains: central  $\beta$ -propeller domain, recognized in all neuraminidases, and two flanking lectin-like domains with identical topologies though they share only 23% sequence identity. The N-terminal lectin-like domain (wing1) runs from the amino terminus at residues 25 to residue 216. The other domain at the C-terminal (wing2) protrudes from the central domain and runs from residues 347-543. The two flanking domains were suggested to recognize sugar, though the exact function and the nature of carbohydrates they bound were to be established. Herein, the crystallographic structures of VCNA in complexes with six different ligands are being investigated.

Generally, the overall structures of the complexes were almost identical to the native structure and revealed the same features described earlier for the reported structure except of missing the last four residues (aa 778-781) from the complex structures; the electron densities were not clear enough in order to model these last four residues. The revelation of nearly identical structures for the free and bound enzyme is not unusual, as binding of small ligands at rather a rigid active site of VCNA was not expected to affect the enzyme structure dramatically. However, superimposing the native and the structures of the six complexes using the native structure as a template in the O program (Jones et al., 1991) gave different rmsd values for the superimposed  $C_{\alpha}$  (aa 25-777) in the different complexes as shown in the table 5.1. Figure 5.1 displays the six complex structures superimposed onto the native structure. As can be seen from the figure and the rmsd values in the table 5.1, there are deviations between the structure of the free enzyme and

**Table 5.1.** Superimposing the complex structures on the native VCNA structure as a template

Complex structure	rmsd (Å)
DANA	0.90
3'-sialyllactose	0.86
6'-sialyllactose	0.97
FANA	1.02
GANA	0.60
CANA	0.70



**Figure 5.1.** Displays the  $C_{\alpha}$  trace for the complex structures superimposed on the native 1KIT structure. The complex structures of DANA, 3'-sialyllactose, 6'-sialyllactose, and FANA are coloured in red (class I); whereas the complex structures of GANA and CANA as well as the native structure are coloured in green (class II). The regions of the structures, which showed large variations between the two groups are indicated by arrows and labelled with the amino acid sequences in the annotated figure. The  $C_{\alpha}$  at the N-termini and the C-termini are drawn as blue and orange spheres respectively.

the four complex structures: DANA-, 3'-sialyllactose-, 6'-sialyllactose-, and FANA-complexes with large rmsd close to 1 Å. Whereas the two structures: CANA- and GANA-complexes exhibit more similarity to the native structure with smaller rmsd value relative to the first four complex structures. The distinction between structures of the first four complexes (class I) on one hand and the other two complexes as well as the native structure (class II) on the second hand has been noted earlier during the discussion of the unit cell dimensions of the various structures (see section 4.3.2.1). As will be shown in the next section, the regions of the structures that apparently showed striking deviations between the superimposed structures (highlighted in figure 5.1) are involved in the crystal contacts, which differ dramatically between the two classes. Therefore, the large rmsd values observed for the complex structures of class I originated from the effect of the crystal packing. It is important to mention that a small angular displacement of the two lectin-like domains with respect to the central domain in the complex structures of class I was noticed during the rigid-body refinement step, while solving the structures (section 4.6). This displacement of the subunits, which could be ascribed to the crystal packing effect, was manifested in the large rmsd values between the structures of class I and the structure of the unbound VCNA.

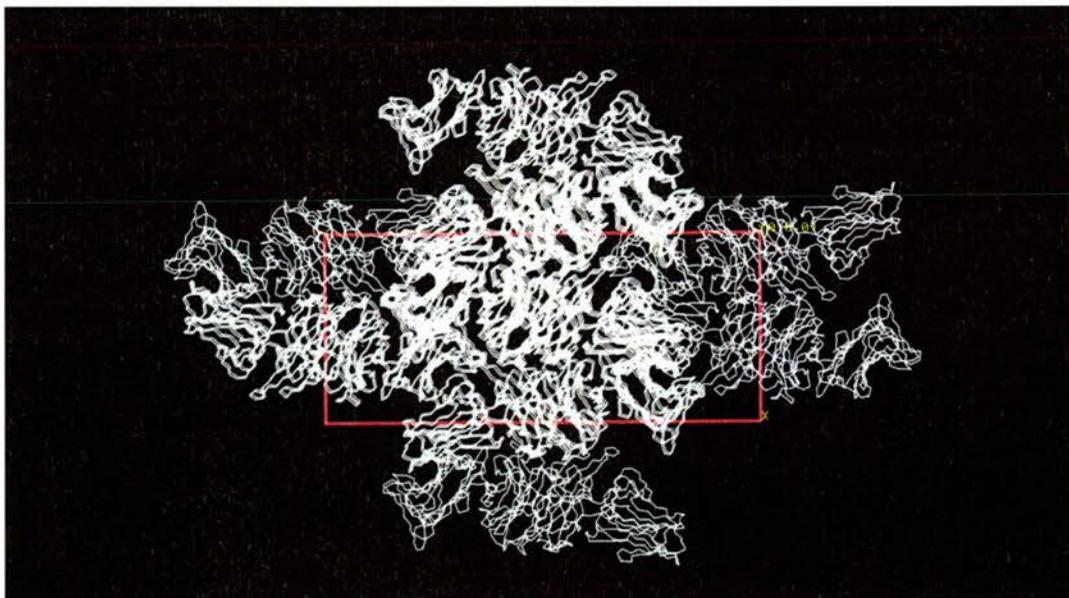
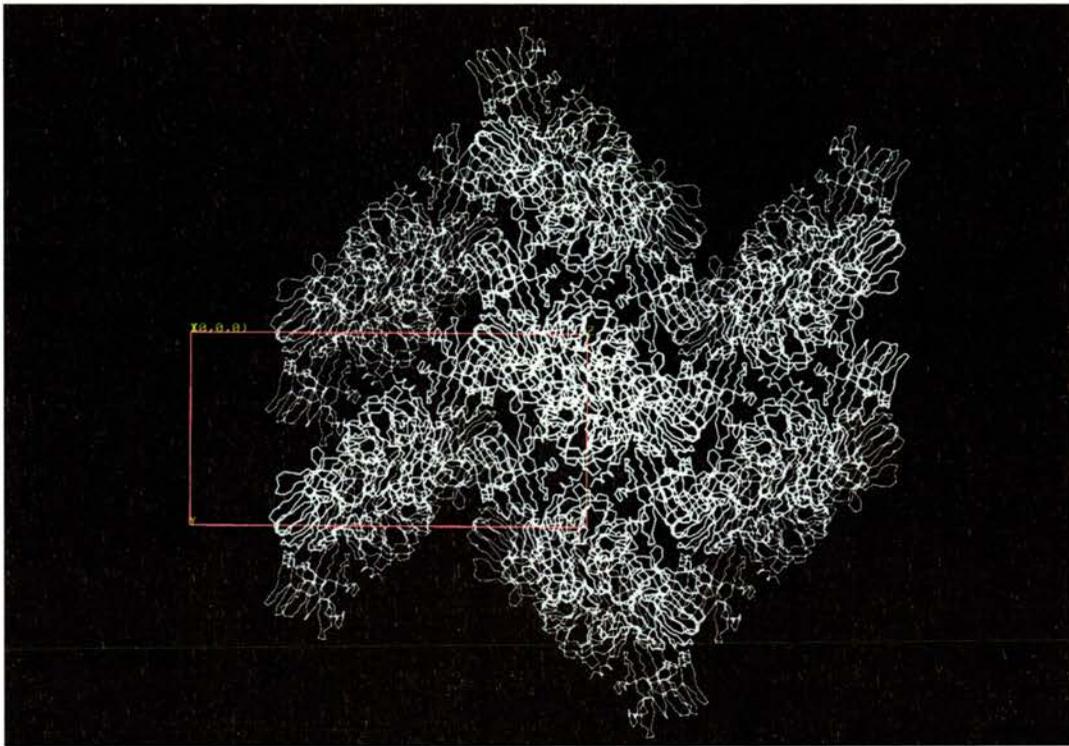
## **5.2.2 Crystal packing**

### **5.2.2.1 Packing pattern**

Proteins and other biomacromolecules have the natural potential to interact via hydrogen bonds, ionic, and van der Waals contacts. Such contacts are precisely those occurring in intermolecular packing within macromolecular crystals (Charron et al., 2001). Nevertheless, modifying the constituents in the crystallization drop can disturb the

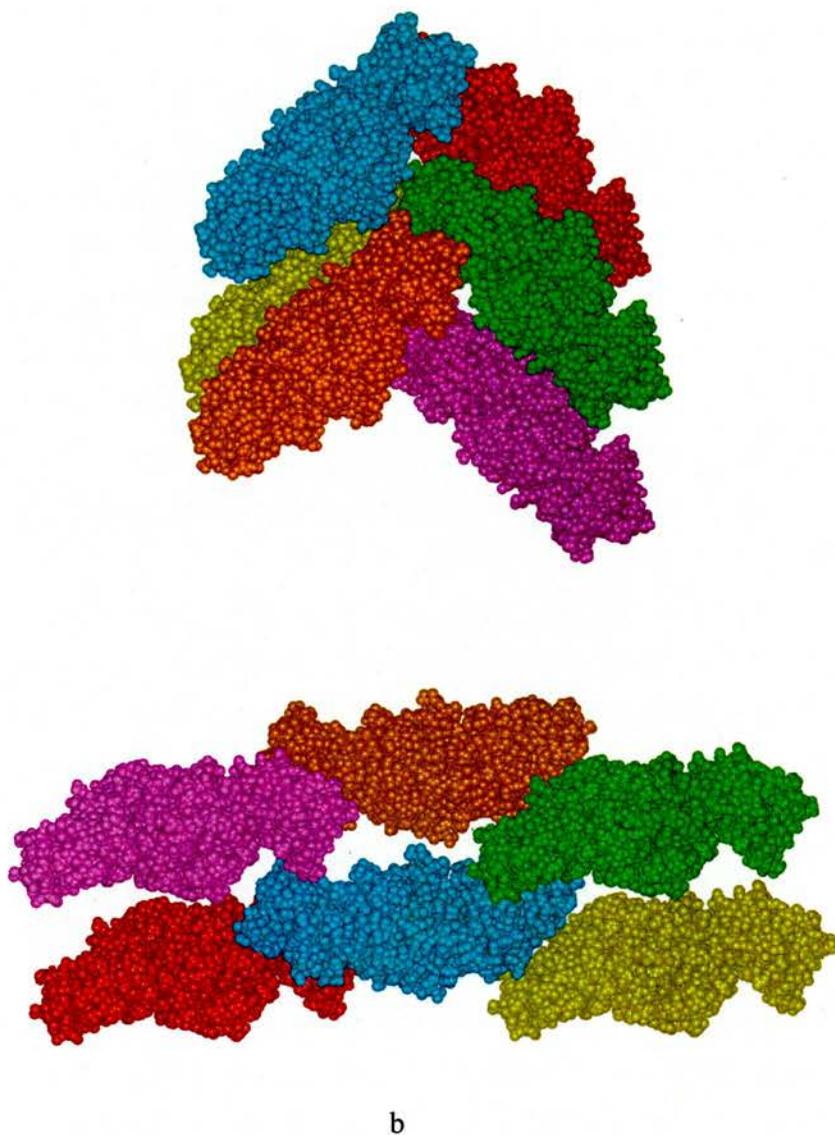
intricate interplay between these contact forces. Such disturbance might change the packing within the crystal dramatically, as was observed in our crystallographic study.

As mentioned previously (section 4.3.2.1), although the crystals of the six complex structures belonged to the same space group ( $P2_12_12_1$ ), they showed variations in the unit cell dimensions. On one hand, the unit cells of the DANA-, 3'-sialyllactose-, 6'-sialyllactose-, and FANA-complexes were more or less of similar dimensions, but smaller than the cell dimensions of the reported native structure, with the largest shrink in  $c$  ( $\approx 13$  Å); we called the four complex structures as class I. On the other hand, the unit cell dimensions of the GANA- and CANA-complexes were close to each other and to the cell dimensions of the native structure; the two complexes were put together and called class II. To figure out the origin of the observed variations in the unit cells of the two classes, the crystal packing of the six complex structures and the reported native structure was investigated using graphics and the symmetry related molecules were generated using the program O (Jones et al., 1991). Noticeably, two distinctive packing patterns were observed in the two classes despite the fact that the native and the ligand-bound proteins crystallized in the same space group ( $P2_12_12_1$ ). In the crystals of class I, the protein molecules were packed in a nest as shown in the top of figure 5.2a,b. Whereas in the crystals of class II, the biomolecules were arranged head-to head in parallel to each other, as displayed in the bottom of figure 5.2a,b; the same packing was exhibited in the crystal of the native structure. Solvent channels within crystals in the parallel arrangement were larger in diameter than that in the arrangement of class I. Clearly, the images in figure 5.2 revealed rather tight arrangements of the VCNA molecules in the crystals of class I compared to those of class II, in agreement with the solvent content and  $V_M$ -values (section 4.3.2.1, table 4.3). What was responsible for such a striking difference in the crystal packing between the two classes?



a

Figure 5.2.a (continued)



**Figure 5.2.** (continued) (a)  $C_{\alpha}$ -trace of VCNA illustrating the packing pattern within crystals of: (top) class I including crystals of the complexes of DANA, 3'-sialyllactose, 6'-sialyllactose, and FANA, (bottom) class II including crystals of the GANA and CANA complexes as well as the native structure crystal. (b) Space-filling model illustrating the packing in class I (top) and class II (bottom) shown in (a).

In fact, the difference in the packing pattern could be correlated to the way the crystals grown. Crystals in class I, except FANA-complex crystal, were grown as cocrystals with the ligands added to the crystallizing agents at the beginning of the crystallization process. In the case of FANA, crystals were obtained by adding the ligand to already grown crystals of the free-enzyme; however, the crystals dissolved upon adding the ligand and reappeared thereafter. In both cases, the crystals used for data collection were grown in presence of the ligands. Meanwhile, crystals in class II were grown in absence of the ligands and remained stable for X-ray shooting after adding the ligands. It should be noted that, all ligands used in the different complexes carried a negative charge on the sialic acid/or sialic acid-like moiety; at the same time, VCNA is characterized by a highly negatively charged surface. Under these circumstances of ionic environment, the negatively charged ligands disturbed the contact forces between the macromolecules within the crystals of FANA-complex causing the crystals to disaggregate, and somehow directed the molecules to arrange themselves as seen in the packing pattern of class I. Following the same reasoning, the negative ligands added at the beginning of crystallization of the other three complexes in class I, operated from the start to direct the packing in the observed manner. It should be mentioned that, crystals of CANA-complex, grown in the same way as that of FANA-complex, were shown to have the same unit cell dimensions and same packing observed in class I, [N.B. solving the structure for the data obtained from that crystal turned to be empty of any bound ligand, the reason for that could be the excess cryo-solution, added to the drop before mounting the crystal, washed the bound ligand out of the active site of VCNA, or the concentration of the added ligand was too low to bind significantly; data are not discussed in the thesis]. For crystals of class II, the concentration of the added ligands seemed to be not enough to disintegrate the crystals. Another possibility is that the slightly changed chemical nature of the

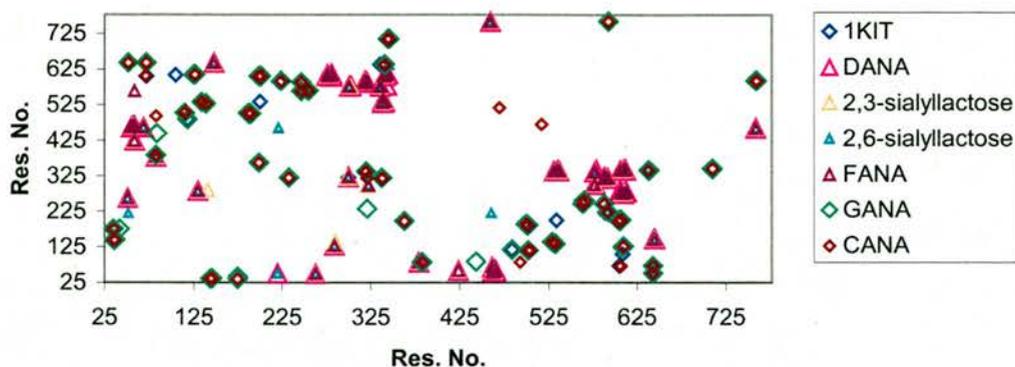
substituents at C5 of the sialic acid moiety [N-trifluoroacetyl (FANA), N-chloroacetyl (CANA), and N-glycolyl (GANA)] might affected the interaction of the GANA and CANA ligands with the contact surface of VCNA in the corresponding crystals slightly different.

### 5.2.2.2 Analysis of contacts within crystals

The contact residues involved in the protein-protein interactions within crystals of the different complexes and the native structure (1KIT) were analysed using the program CONTACT in the CCP4 suite package (Collaborative Computational Project, 1994). The program performed atom-atom automatic search with a cutoff distance 3.60 Å and extended the search to  $\pm 2$  lattice vectors in all directions. Figure 5.3 displays the chart mapping the contact residues in the different crystals of the native and complex structures. Clearly, from the chart, the pattern of the amino acids making intermolecular contacts in the crystals of class I can be distinguished from that of amino acids involved in the contacts of class II. The numbers of contacts as well as the symmetry operators used to generate the symmetry related molecules are summarized in table 5.2. As shown in the table 5.2, for the packing in class I, most of the contacts were observed between a molecule in position  $(x, y, z)$  and a symmetry related molecule in position  $(x+1/2, -y+1/2, -z)$ . Less number of contacts was observed between a molecule in position  $(x, y, z)$  and a molecule in position  $(-x+1/2, -y, z+1/2)$ . Very few contacts existed between a molecule in position  $(x, y, z)$  and molecules in positions  $(-x, y+1/2, -z+1/2)$  and  $(x, y, z)$ . Note that in the last contact the identity operator was combined with translation along the unit cell.

In the case of the packing observed in class II, most of the contacts were observed between a molecule in position  $(x, y, z)$  and molecules in positions  $(x+1/2, -y+1/2, -z)$  and

**Chart of contact residues**



**Figure 5.3.** Displays the chart of amino acids making the contacts between symmetry related molecules within crystals of the different structures. The legends used to label the different crystal structures are shown in the panel beside the chart. Structures of class I are designated by triangles with different colours and sizes; whereas structures of class II are designated with diamonds of different size and coloured differently.

$(-x+1/2, -y, z+1/2)$ , with more contacts with the molecule in the first position. Few contacts were observed between a molecule in position  $(x, y, z)$  and a molecule in position  $(-x, y+1/2, -z+1/2)$ , but slightly more than that observed in the packing of class I. None of the contacts between a molecule in position  $(x, y, z)$  and a molecule generated by the identity operator, observed in class I, was seen in class II. Altogether, as can be seen from the above analysis, the distributions of contacts between a molecule in position  $(x, y, z)$  and the symmetry related molecules varied between the two classes as expected for different patterns of packing. But, the total number of contacts was nearly the same in the structures of the two classes, the largest contact was observed in the structures of the DANA- and FANA-complexes. It should be mentioned that the number of contacts observed between a given pair of symmetry related molecules, varied across the different structures within the same class; this is not unusual as the residues at contact surfaces are flexible, and slight movement of the atoms can easily cause the contact involving

**Table 5.2.** Number of contacts between the symmetry related molecules in the different structures

Symmetry Operator	No. of contacts within crystals of						
	Class I				Class II		
	DANA	3'-sialylactose	6'-sialylactose	FANA	IKIT	GANA	CANA
$-x+1/2, -y, z+1/2$	40	28	20	36	56	44	54
$-x, y+1/2, -z+1/2$	4	2	4	2	10	12	10
$x+1/2, -y+1/2, -z$	102	84	98	110	60	70	72
$x, y, z$	2	4	2	2	-	-	-

the displaced atoms uncountable by the program. However, the trend of distributions of the contacts between the different pairs did not change within each class.

### 5.2.2.3 Nature of contact residues

Amino acids from VCNA making intermolecular contacts in the lattices of structures in the classes I and II are listed in tables 5.3 and 5.4, respectively. Due to symmetry reasons, the pairs of interacting residues were seen twice, for example, in class I the pairs: Asn51 $\leftrightarrow$ Arg263 and Arg263 $\leftrightarrow$ Asn51 existed between a molecule and its symmetry related ones, but only the first pair is listed in the table 5.3; this is applied for the whole list in tables 5.3 and 5.4. As can be seen from the two tables, the intermolecular contacts are completely different in the two classes. Only one contact (Ser300 $\leftrightarrow$ Arg321) was found in both packing patterns. As shown in figures 5.4 & 5.5, which display the space-filling model of VCNA showing the contact residues in the classes I & II, respectively, coloured differently. The contact residues in both classes are distributed mostly on the N-terminal domain (runs from aa 25-216) and the central domain (runs from aa 217-346 and aa 544-777); however, more residues at the N-terminal domain were involved in the contacts of class II compared to class I. Noticeably, three residues (Arg118, Gln188, and Thr199) participating in the sialic acid binding at the N-terminal lectin-like domain as will be discussed in due course, were making contacts in class II but not in class I. In fact, this could provide the reason for why in the presence of the ligand, during the growth of the crystals of class I, the resulting packing pattern was completely different from the pattern demonstrated by crystals in class II, which were fully grown in absence of the added ligands. In other words, the ligand interaction at the N-terminal domain disturbed the contacts involving the three residues mentioned above, and thereafter the crystallizing

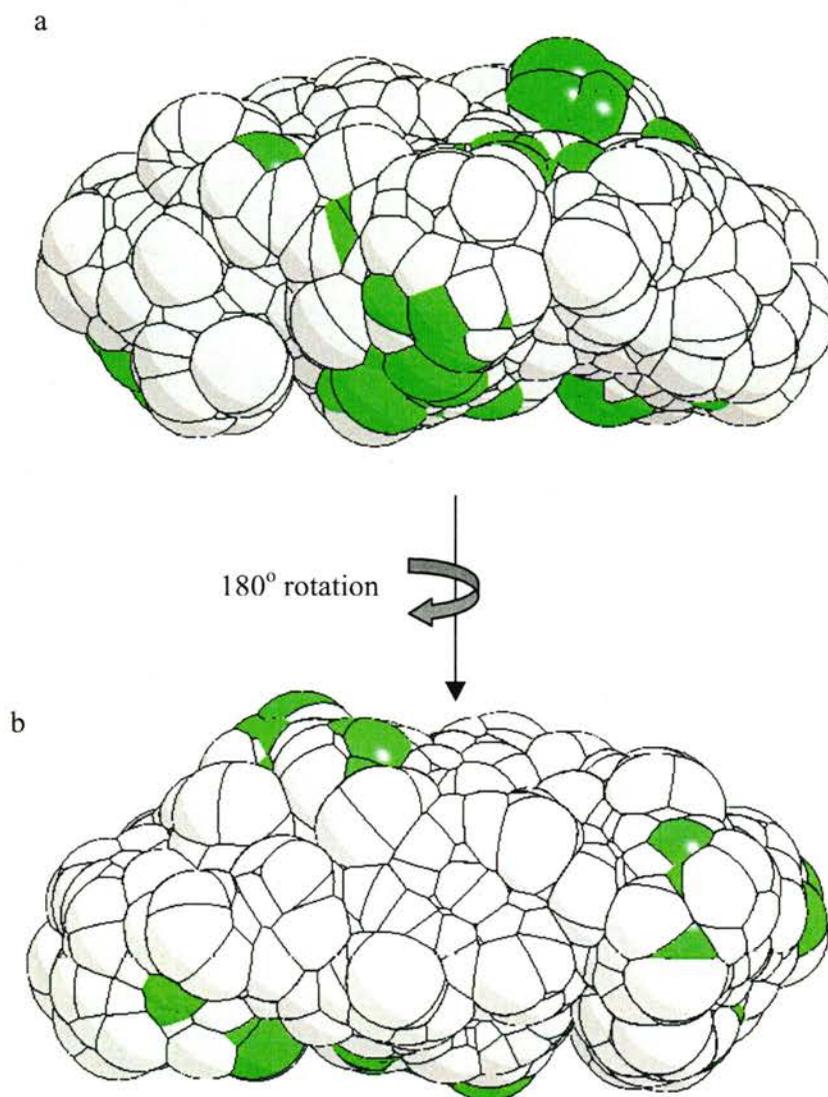
**Table 5.3.** Contact residues in class I

Residues in*	
molecule at origin	symmetry related molecule
Asn51 <sup>h</sup>	Arg263 <sup>h</sup>
Asn52 <sup>h</sup>	Arg220 <sup>h</sup>
Gly55 <sup>h</sup> , Ser54 <sup>w</sup> , Val56 <sup>w</sup>	Gln462 <sup>hw</sup> , Ile463 <sup>w</sup>
Leu57 <sup>h</sup>	Thr464 <sup>h</sup> , Thr465 <sup>h</sup>
Thr58 <sup>w</sup>	Thr465 <sup>w</sup>
Asn59 <sup>hw</sup>	Val424 <sup>w</sup> , Thr464 <sup>h</sup> , Thr465 <sup>hw</sup> , Trp466 <sup>w</sup>
Gln69 <sup>hw</sup>	Gln462 <sup>hw</sup>
Thr83 <sup>hw</sup>	Asp379 <sup>hw</sup>
Asn130 <sup>hw</sup>	Val284 <sup>hw</sup>
Arg141 <sup>h</sup>	Ser285 <sup>h</sup>
Gly147 <sup>w</sup> , Thr148 <sup>hw</sup>	Gly644 <sup>hw</sup> , Val645 <sup>w</sup>
Gly221 <sup>h</sup>	Gln461 <sup>h</sup>
Leu276 <sup>w</sup>	Ser608 <sup>w</sup> , Ser610 <sup>w</sup>
Asn277 <sup>h</sup> , Glu280 <sup>i</sup>	Lys607 <sup>hi</sup>
Gln281 <sup>hw</sup>	Arg605 <sup>h</sup> , Ile611 <sup>w</sup> , Glu613 <sup>hw</sup>
Ser300 <sup>hw</sup>	Arg321 <sup>h</sup> , Lys323 <sup>w</sup>
Ser301 <sup>hw</sup>	Arg321 <sup>h</sup> , Arg577 <sup>hw</sup>
Thr303 <sup>w</sup>	Phe578 <sup>w</sup>
Asn318 <sup>h</sup>	Asp589 <sup>h</sup>
Gly319 <sup>hw</sup>	Asp589 <sup>h</sup> , Ser592 <sup>w</sup>
Arg321 <sup>hwi</sup>	Asp298 <sup>i</sup> , Asp588 <sup>hw</sup> , Asp589 <sup>w</sup> , Gly590 <sup>w</sup>
Asp336 <sup>hw</sup>	Arg577 <sup>hw</sup> , Phe578 <sup>w</sup>
Ser339 <sup>hw</sup>	Lys531 <sup>h</sup> , Met536 <sup>w</sup> , Ser537 <sup>h</sup>
Asn341 <sup>hw</sup>	Thr535 <sup>h</sup> , Trp606 <sup>w</sup> , Ser610 <sup>h</sup>
Trp342 <sup>hw</sup>	Ser610 <sup>h</sup> , Ile611 <sup>w</sup> , Leu612 <sup>h</sup>
Gln343 <sup>w</sup>	Phe579 <sup>w</sup>
Ala344 <sup>h</sup>	Leu612 <sup>h</sup>
Asp459 <sup>hw</sup> , Gly460 <sup>h</sup>	Asn759 <sup>hw</sup>

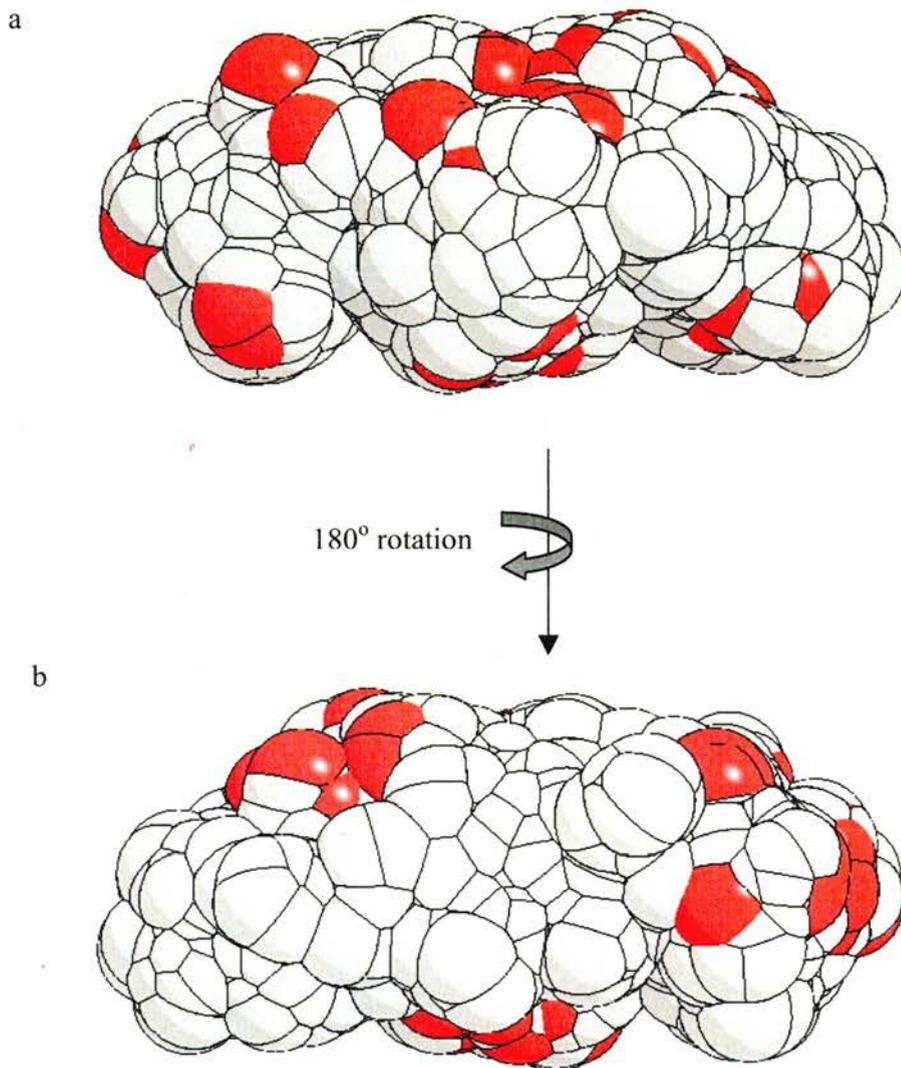
\* Interactions occur between residues of the same row. The type of interactions: hydrogen bonding (<sup>h</sup>), van der Waals (<sup>w</sup>), and ionic (<sup>i</sup>) are indicated as superscripts for the listed amino acids, same labels are used in table 5.4.

**Table 5.4.** Contact residues in class II

Residues in *	
molecule at origin	symmetry related molecule
Thr35 <sup>w</sup> , Glu36 <sup>hwi</sup> , Lys42 <sup>hw</sup>	Leu144 <sup>h</sup> , Lys175 <sup>hwi</sup>
Phe37 <sup>w</sup>	Ala145 <sup>w</sup> , Thr146 <sup>w</sup>
Asn52 <sup>hw</sup>	Asn643 <sup>hw</sup>
Ile71 <sup>w</sup> , Gly72 <sup>hw</sup>	Asn643 <sup>hw</sup> , Arg605 <sup>h</sup> , Lys607 <sup>w</sup>
Thr83 <sup>w</sup>	Asn383 <sup>w</sup>
Asn84 <sup>h</sup>	Glu444 <sup>h</sup>
Gly105 <sup>h</sup>	Ser609 <sup>h</sup>
Gly115 <sup>hw</sup>	Asn502 <sup>hw</sup>
Thr116 <sup>w</sup>	Glu505 <sup>w</sup>
Arg118 <sup>h</sup>	Gln484 <sup>h</sup>
Leu125 <sup>hw</sup>	Ser609 <sup>hw</sup>
Asp126 <sup>h</sup> , Ser127 <sup>w</sup>	Ser610 <sup>hw</sup>
Glu134 <sup>hw</sup>	Thr532 <sup>w</sup> , Gly533 <sup>h</sup>
Thr139 <sup>h</sup>	Lys529 <sup>h</sup>
Lys187 <sup>hw</sup>	His501 <sup>w</sup> , Asn502 <sup>h</sup>
Gln188 <sup>w</sup> , Met190 <sup>w</sup>	Gly500 <sup>w</sup>
Asn189 <sup>h</sup>	Thr497 <sup>h</sup> , Gly500 <sup>h</sup> , Asn502 <sup>h</sup>
Asn199 <sup>hw</sup>	Gly363 <sup>h</sup> , Arg605 <sup>w</sup> , Trp606 <sup>h</sup>
Thr200 <sup>h</sup>	Asn534 <sup>h</sup>
Asp201 <sup>hw</sup>	Arg605 <sup>hw</sup> , Trp606 <sup>hw</sup> , Lys607 <sup>w</sup>
Pro222 <sup>w</sup>	Asp589 <sup>w</sup>
Arg224 <sup>h</sup>	Ser592 <sup>h</sup>
Val233 <sup>w</sup>	Arg321 <sup>w</sup>
Gly247 <sup>h</sup>	Gln564 <sup>h</sup> , Asp588 <sup>h</sup>
Ser255 <sup>hw</sup>	Asn565 <sup>hw</sup> , Gly566 <sup>w</sup>
Ser300 <sup>h</sup>	Arg321 <sup>h</sup>
Gly319 <sup>w</sup> , Asp320 <sup>w</sup>	Ala338 <sup>w</sup>
Asp336 <sup>w</sup>	Phe638 <sup>w</sup>
Asn341 <sup>w</sup>	Phe638 <sup>w</sup> , Asn639 <sup>w</sup>
Ala344 <sup>w</sup> , Pro345 <sup>hw</sup> , Ile346 <sup>w</sup>	Asn710 <sup>hw</sup>
Asn593 <sup>hw</sup>	Asn759 <sup>w</sup> , Ser760 <sup>hw</sup>



**Figure 5.4.** (a) Space-filling model of VCNA with the contact residues observed in class I coloured in green. (b) Same as in (a) with the molecule rotated 180° around the vertical axis parallel to the plane of the paper to show the contact residues on the rear side of the molecule.



**Figure 5.5.** (a) Space-filling model of VCNA with the contact residues observed in class II coloured in red. (b) Same as in (a) with the molecule rotated 180° around the vertical axis parallel to the plane of the paper to show the contact residues on the rear side of the molecule.

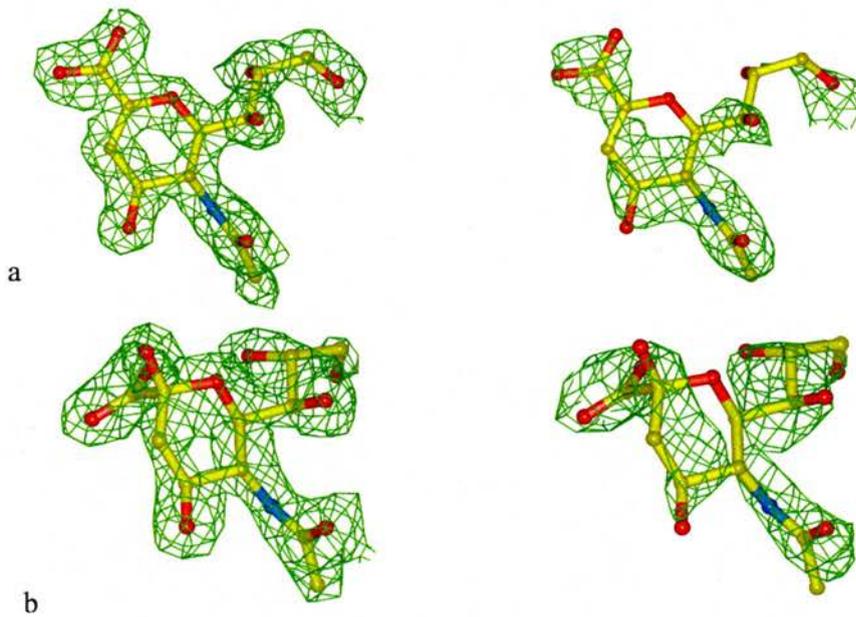
system preferred to arrange the macromolecules of the bound enzyme in a different way yielding the packing pattern seen in class I.

The types of interaction between the contact residues are indicated as superscripts for the listed residues in the tables 5.3 and 5.4. In both classes, the contact residues interact mostly via van der Waals forces and hydrogen bonding. The attractive ionic interactions does not play significant role in packing as only three pairs of ionic contact were observed: Glu280↔Lys607 & Arg321↔Asp298 in class I, and Glu36↔Lys175 in class II.

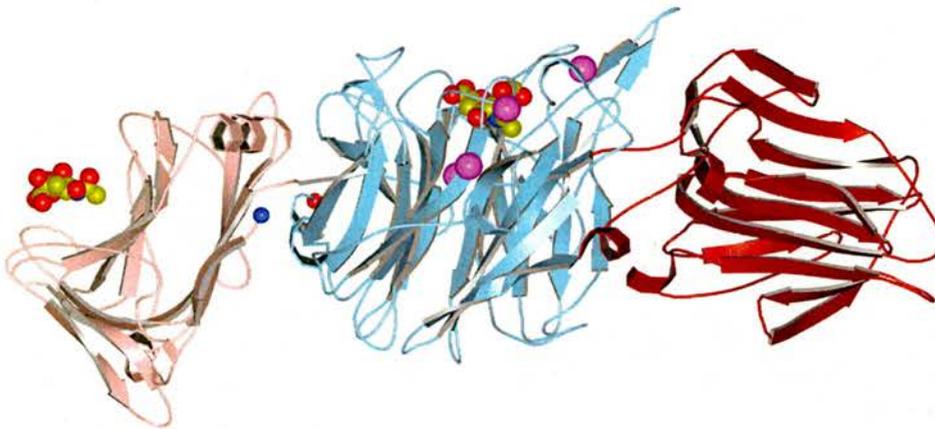
### 5.3 VCNA-DANA complex

The crystal structure of VCNA in complex with its inhibitor DANA [ $K_i = 26-30 \mu\text{M}$ , (Zbiral et al., 1989; Meindl et al., 1974)] resolved to 1.9 Å is reported herein. The electron density maps,  $F_o - F_c$  and  $2F_o - F_c$  shown in figure 5.6a, of the complex structure demonstrated the inhibitor bound at the catalytic site of the  $\beta$ -propeller neuraminidase domain of VCNA. Moreover, and surprisingly, the electron density maps (figure 5.6b) of the crystallographic structure, unambiguously, revealed  $\alpha$ -sialic acid lying in the main cleft of the N-terminal lectin-like domain of VCNA. Figure 5.7 displays schematically the structure of the enzyme with the two ligands bound at their corresponding sites. Additionally, as shown in the figure 5.7, two more calcium ions that were not reported in the native structure (Crennell et al., 1994), appeared in the 1.9 Å resolution structure.

It should be emphasized that neither the crystallization conditions nor the purification protocol of the enzyme included any source of sialic acid. Nevertheless, sialic acid could be originated from two sources:



**Figure 5.6** (a) The electron density maps:  $2F_o-F_c$  contoured at  $1.2\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the ligand DANA at the catalytic site. (b) the  $2F_o-F_c$  at  $1.2\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the  $\alpha$ -sialic acid at the N-terminal lectin domain. All maps were calculated before adding the ligands to the refined model of the VCNA-DANA complex.

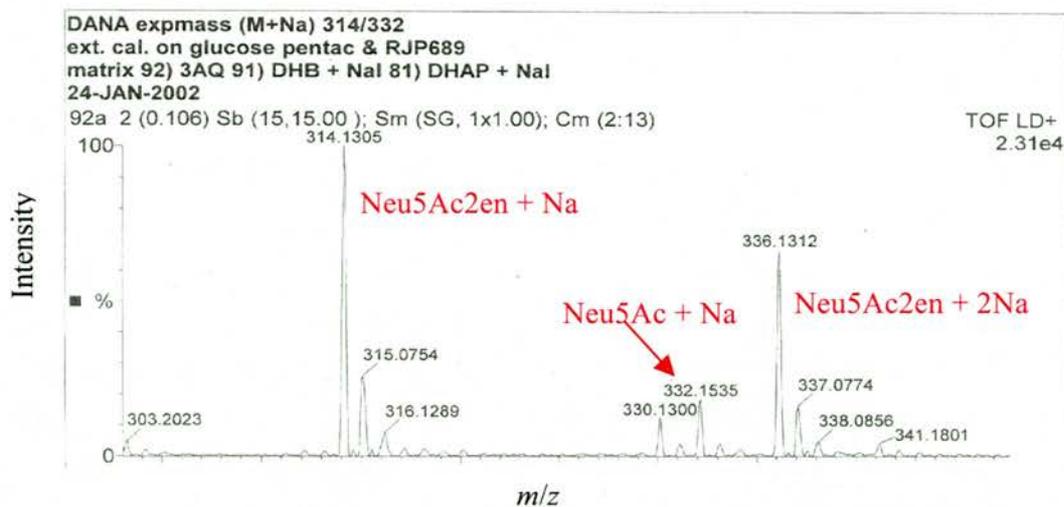


**Fig. 5.7** Schematic representation of VCNA with the ligands: DANA in the active site and NANA in the sialic acid-binding site shown in space-filling model. The  $Ca^{+2}$  ions are represented as magenta spheres;  $C_\alpha$  at the N-terminus and the C-terminus are drawn as blue and red spheres respectively. Figure prepared with BOBSCRIPT (Esnouf, 1997).

(1) It could be formed via the hydrolytic action of the enzyme upon DANA as deduced from our results of the NMR experiments (section 8.4.1). Even if the amount of sialic acid produced from the hydrolysis reaction of DANA by VCNA was too small, as inferred from the NMR, the relatively long incubation period of the enzyme with DANA before the crystals were being formed could result in accumulation of a reasonable quantity of sialic acid sufficient to be picked by the enzyme; it took one month to grow crystals of VCNA-DANA complex at 18 °C (section 4.2.1).

(2) Additionally, sialic acid might be formed via degradation of DANA in the drop having the crystallizing agent over their incubation period at 18 °C. Especially, the mass spectrum of DANA solution used in the crystallization experiment (stored in the fridge at 4 °C), showed a small peak corresponding to the mass of sialic acid; the spectrogram is shown in figure 5.8. It is more likely that sialic acid was formed as a degradation product of DANA rather than pre-existed as a contaminant, since the mass spectrum of a fresh sample of DANA did not show any presence of sialic acid in the mass spectrum.

Clearly, the fact that VCNA picked up sialic acid in its low-populated  $\alpha$ -form, at equilibrium sialic acid solution contains 5%  $\alpha$ -sialic acid and 95%  $\beta$ -sialic acid (Dabrowski et al., 1979), from its surrounding in absence of any direct source for the sialic ligand, anticipated high binding affinity of  $\alpha$ -sialic substrates towards the N-lectin-like domain of the enzyme. It is turned to be true as revealed by the analysis of the interactions at the binding-site, and confirmed further by measuring the affinity of the newly discovered binding site towards an  $\alpha$ -sialic acid-containing substrate using ITC ( $K_d = 30 \mu\text{M}$ ), to be discussed in chapter 7.

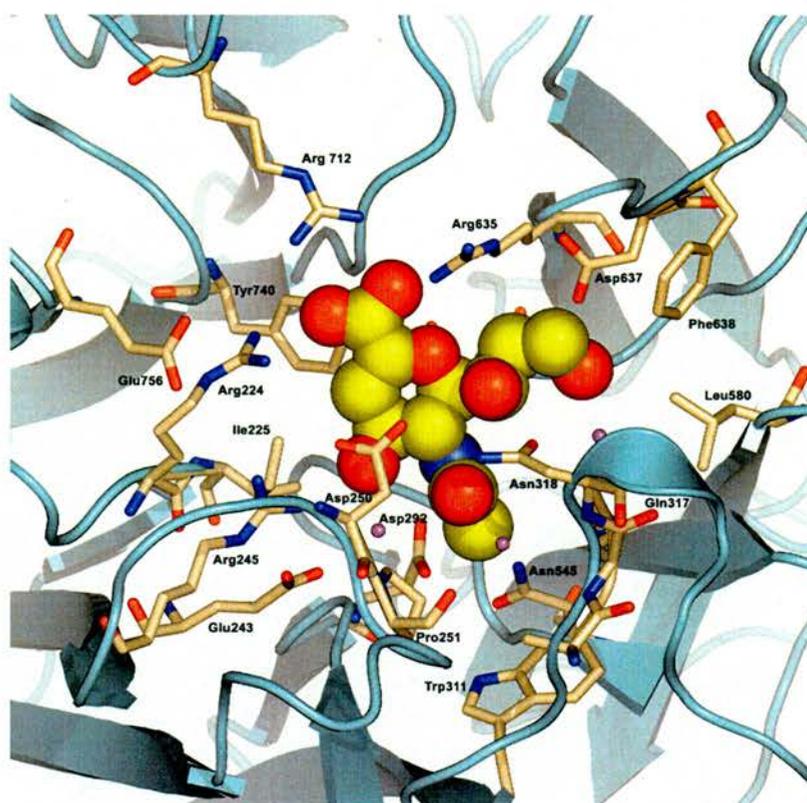


**Figure 5.8.** The mass spectrum of DANA solution that was employed in the crystallization of VCNA-DANA complex. The main peaks identified for DANA as well as the suspected peak of sialic acid are labelled in the spectrogram.

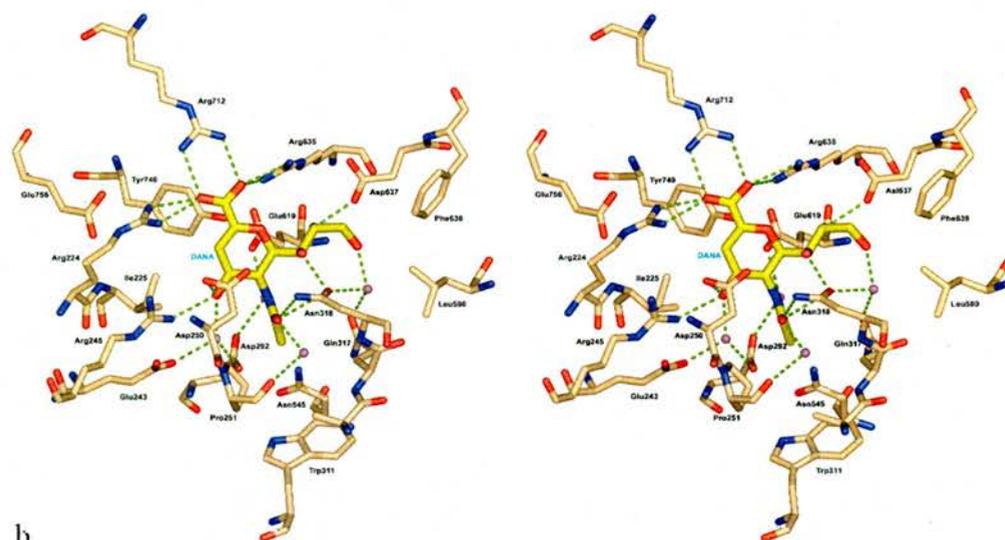
### 5.3.1 Catalytic site

The catalytic site of VCNA was described previously (Crennell et al., 1994); moreover, its interaction with DANA has been analysed based on the 4.5 Å resolution structure of the VCNA-inhibitor complex. This analysis of the interaction was guided by the close similarity between the active site in VCNA and that of *Salmonella typhimurium* neuraminidase (STNA) whose structure with the inhibitor bound was available by then (Crennell et al., 1993). The 1.9 Å resolution structure provided more information and exhibited the details of the direct and water-mediated protein-inhibitor interactions. Apparently, as discussed below, these interactions are important for the binding process.

As shown in figure 5.9 displaying DANA bound at the catalytic cleft, the carboxylate group interacts with the arginine triad (224, 635, 712) via strong charge-charge interaction, as can be deduced from the distances between the interacting atoms of the inhibitor/enzyme listed in table 5.5. Note the significant arrangement of the guanidine

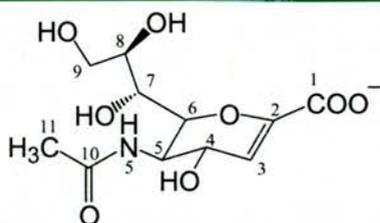


a



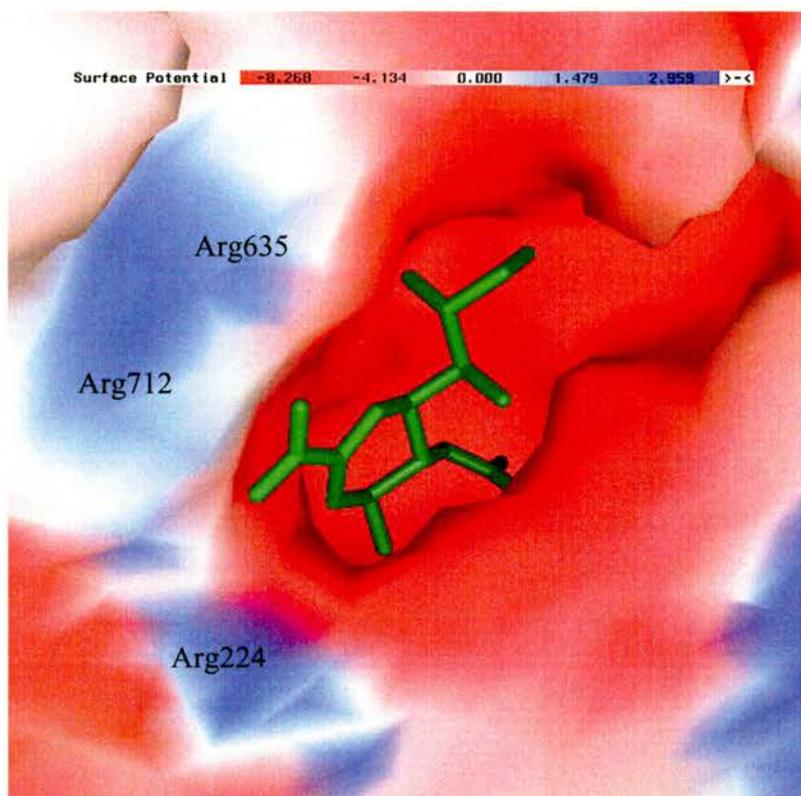
b

**Figure 5.9.** (a) Ribbon representation of the active site of VCNA with the inhibitor, DANA, bound (drawn in CPK model), the active site residues are labelled and drawn as sticks. (b) Stereo view for the active site, the hydrogen bonding interactions drawn as dotted lines; water molecules are represented as magenta spheres. Figure prepared with PyMOL (Delano, 2002).

**Table 5.5.** List of interactions between VCNA and the bound DANA

Hydrogen bonding			Non-bonding interaction *		
DANA-atom	VCNA-atom	Distance (Å)	DANA-atom	VCNA-atom	Distance (Å)
O1A	Arg224-N <sup>n1</sup>	2.86	C1	Tyr740-O <sup>n</sup>	2.90
O1A	Arg224-N <sup>n2</sup>	2.92	C2	Tyr740-O <sup>n</sup>	2.76
O1A	Arg712-N <sup>n1</sup>	2.86	C2	Ile225-C <sup>δ</sup>	5.29
O1B	Arg712-N <sup>n2</sup>	2.89	C3	Ile225-C <sup>δ</sup>	4.09
O1B	Arg635-N <sup>n1</sup>	3.27	C9	Phe638-C <sup>ζ</sup>	4.29
O1B	Arg635-N <sup>n2</sup>	3.16	C9	Leu580-C <sup>δ1</sup>	4.87
O4	Arg245-N <sup>n2</sup>	2.87	C9	Asn637-C <sup>β</sup>	4.45
O4	Asp250-O <sup>δ1</sup>	3.18	C11	Trp311-C <sup>ζ2</sup>	4.69
O7	Asn318-O <sup>δ1</sup>	2.69	C11	Pro251-C <sup>γ</sup>	3.88
O8	Asp637-O <sup>δ2</sup>	2.71	C11	Gln317-C <sup>β</sup>	4.15
O10	Asn318-N <sup>δ2</sup>	3.36			
O10	Asn318-O <sup>δ1</sup>	3.30			

\* For the non-bonding interaction, except for C1, C2 ↔ Tyr740-O<sup>n</sup> interactions, the distance was measured between the ligand atom and the nearest atom of the interacting protein residue.



**Figure 5.10.** Electrostatic potential at the active site of VCNA with the bound ligand (DANA), shown in a stick model and coloured in green. Note the positive potential accommodating the carboxylate group of the ligand that is created by the arginine triad. The electrostatic potential surface is colour coded ranging from  $-8.26$  KT (red) to  $2.95$  KT (blue). Figure prepared with GRASP (Nicholls et al., 1991).

group of Arg712 and the inhibitor-carboxylate group lying in the same plane and forming a “six-membered ring structure” that enhanced the interaction between the two groups. The basic arginine residues provide the required positive potential to accommodate the negative carboxylate group of the inhibitor with a remarkable complementarity as illustrated in figure 5.10.

Beneath the carboxylate group lies residue Try740 with close contacts between its hydroxyl group and atoms C1 & C2 of DANA (distances  $2.90$  and  $2.76$  Å, respectively). Proximity of this tyrosine to C1–C2 bond is thought to play a significant role in

stabilizing the developing positive charge on the oxocarbenium ion of the transition state during the enzymatic hydrolysis (Chong et al., 1992); the nearby residue Glu619 below the sugar ring enhances the role of Tyr740 as have been implicated in the catalytic mechanism of influenza NA (Burmeister et al., 1993). Additionally, C3 and C4 of the sugar ring have a hydrophobic interaction with Ile225.

The O4 hydroxyl interacts via hydrogen bonding with a conserved Arg245. Also, O4 forms direct and water-mediated interactions with Glu243 and Asp292. The interaction of this glutamate is peculiar for VCNA as this residue is replaced by aspartate, with a shorter side chain, which is not suitable to mediate the interaction with the ligand in other neuraminidases. The structural water mediating this interaction was persistently observed in all VCNA-complex structures discussed in this chapter. It is worthy to note that DANA analogues with bulky substituents at C4 have reduced binding affinity for VCNA (Holzer et al., 1993; Schreiner et al., 1991a).

For the N-acetyl substituent at C5 of the sugar ring, the carbonyl oxygen interacts with the amide side chain of Asn318 through hydrogen bonding and embraced in a water-mediated interaction with the side chain amide N-atom of Asn318 and the backbone O-atom of Pro251. A bridging-water located below the sugar ring intervenes between the N5 of the acetyl group and the carboxylate of Glu619 and Asp292. Again, these two structural waters mediating the interaction of the N-acetyl group were observed in the other VCNA complexes. The methyl group has hydrophobic interactions with the side chains of Pro251, Trp311, and Gln317. Interestingly, substituting the methyl group with trifluoromethyl (FANA), chloromethyl (CANA), and glycolyl (GANNA) groups resulted in improved inhibition activity for the synthesized analogues (Wilson et al., 2000), as discussed in due course.

The three hydroxyls of the glycerol side chain of DANA are engaged in the interaction with active site residues, in details: the hydroxyls at O7 and O8 form H-bonds with Asn318 and Asp637 respectively, also water-mediated interactions involving O8/Gln317 and O9/Asn318 contribute to the protein-ligand interaction. In addition the hydrophobic residues Phe638 and Leu580 as well as the aliphatic part of Asp637 are forming a hydrophobic lip close enough to C9 and O9 to provoke the hydrophobic interaction. This potential of C9 to form hydrophobic interaction might explain the earlier observation of the slightly enhanced inhibition of the 9-deoxy-DANA analogue relative to DANA; on the other hand disrupting the interactions involving O7 and O8 with the active site residues by synthesizing the 7-, and 8-deoxy analogues of DANA resulted in decreased inhibition activity for VCNA (Zbiral et al., 1989).

It is conceivable that dissecting the interactions between DANA and VCNA as described above provided invaluable information, which can be used in designing more specific inhibitors for the cholera enzyme. Furthermore, comparing the VCNA-DANA interactions with those observed in other viral and bacterial neuraminidases/transneuraminidases provides better understanding for the structural determinants of neuraminidase inhibition as discussed in the next chapter.

### **5.3.2 Sialic acid-binding site**

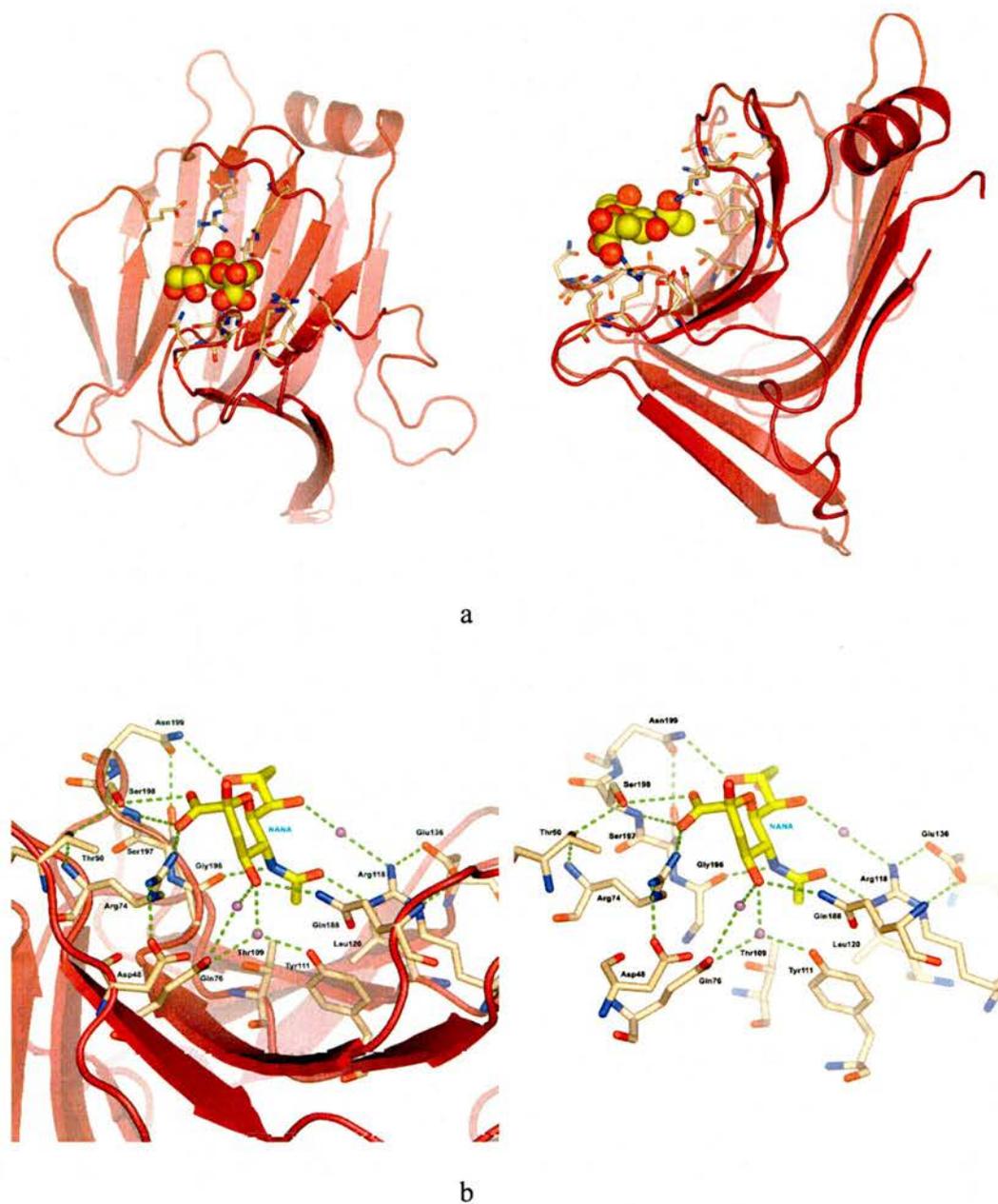
The crystal structure of VCNA-DANA complex, indisputably, revealed  $\alpha$ -sialic acid bound to the main cleft in one of the two lectin-like domains at the N-terminal, see figure 5.6b. The binding pocket is located at a distance of  $\sim 52$  Å apart from the catalytic site at the central domain, and approximately on the same side of the enzyme with respect to the

catalytic site. Such arrangement of the two sites might be of significance to facilitate the enzyme function as discussed in section 5.8. Exquisitely, the finding of a sialic acid-binding site at the N-terminal lectin-like domain enabled us to assign the long-unknown function of this carbohydrate-recognition domain, and elucidated the multiple functionality of the cholera enzyme. The newly discovered function seems to be important for the enzyme to elicit its *in vivo* functions as will be shown when discussing the model action of VCNA. Interestingly, identifying the function of this lectin domain provides a plausible explanation for an early observed phenomenon that VCNA remains attached to the membrane of neuraminidase-treated cells while retaining its activity, even after repeated four times washing (Sedlacek & Seiler, 1974). In the light of the new finding, this can be accounted for by the usage of the enzyme for its lectin domain to attach itself to the cell surface, at the same time the access to the catalytic site is retained open and capable to process a new substrate molecule. Additionally, some other results from early as well as recent works done on VCNA become better understood based on our results. In previous work it has been shown that the release of sialic acid by the action of VCNA on gangliosides in the micelles form was up to 30-fold more efficient compared to the sialic acid release from the free substrate (Venerando et al., 1982). Most recently, Thobhani and co-workers demonstrated that VCNA is able to hydrolyse polyvalent substrates with much greater efficiency than their monovalent counterparts (Thobhani et al., 2003). The enhanced catalytic efficiency in both cases can be ascribed to the role of the sialic acid-binding domain, which facilitated the access of the catalytic domain to the micelle surface or to the polyvalent substrate. It should be mentioned that for the enhanced activity of the polyvalent substrate, the authors ascribed this effect to the recognition of D-galactose residue (a component of the applied substrate); this interpretation is in disagreement with our results. Therein, the authors based their argument on the interference of D-galactose

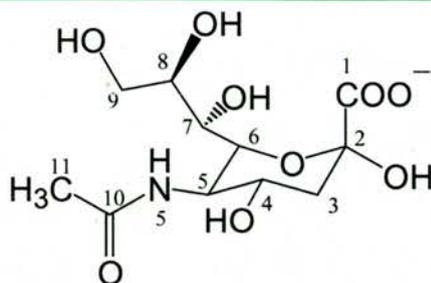
with the measured catalytic activity of the enzyme. However, our crystallographic and NMR (chapter 8) studies did not show any evidence for galactose recognition by VCNA. Perhaps, the interference of the galactose with the neuraminidase activity could be due to indirect effect on the system under study or non-specific binding of the sugar to the surface of the enzyme. Clearly, based on our results, the interpretation of this polyvalency enhancement is more straightforward.

### 5.3.2.1 Interactions at the sialic acid-binding site

As shown in figure 5.11,  $\alpha$ -sialic acid is sited in the binding site at the main groove of the N-terminal domain with its carboxylate pointing to the protein and the anomeric hydroxyl sticks out to the solvent. All substituents at the sugar ring, except the anomeric hydroxyl and O9 of the glycerol side chain, both are exposed to the solvent, contribute to the interaction with the protein. Table 5.6 lists the interactions between sialic acid and the residues at the binding site that are demonstrated in figure 5.11. The carboxylate group forms H-bonds with the side chain of Arg74 and with the hydroxyl group as well as the backbone N-atom of Ser198. The binding residues Arg74 and Ser198 are maintained in their binding positions by receiving help from the secondary shell residues: Asp48 is H-bonded to Arg74-N<sup>H2</sup> atom and Thr50 mediate a H-bonding interaction with Arg74-N atom, additionally the side chain of Thr50 packs against Ser198. Obviously, the interactions of the carboxylate group at the N-terminal binding site are not as strong as that described previously for the carboxylate interactions at the catalytic site. The arginine triad at the catalytic site provides efficient interaction with the ligand carboxylate, which forces the ring flattening in a step required for the hydrolysis mechanism (Burmeister et al., 1993). However, at the binding site the carboxylate interactions are softer, but suitable



**Figure 5.11.** Sialic acid binding site: (a) Ribbon diagram for the N-terminal lectin domain;  $\alpha$ -sialic acid (NANA), drawn as CPK model, is laid in the binding pocket, top view (left) and side view (right), the important residues at the binding cavity are represented as sticks. (b) (left) Zoom-in view of the binding site showing the ribbon diagram combined with the stick-model of the binding site residues and the ligand (NANA), the H-bonding interactions are depicted as dotted green lines. (right) the same view as on the left without displaying the ribbon diagram for clarity. Residues are annotated in the figures, ordered water molecules are represented as spheres coloured in magenta.

**Table 5.6.** List of interactions between VCNA and the bound NANA

Hydrogen bonding			Non-bonding interaction*		
NANA-atom	VCNA-atom	Distance (Å)	NANA-atom	VCNA-atom	Distance (Å)
O1A	Arg74-N <sup>n1</sup>	2.83	C11	Tyr111-C <sup>ζ2</sup>	3.73
O1A	Ser198-N	2.99	C11	Leu120-C <sup>δ2</sup>	4.18
O1B	Ser198-O <sup>γ</sup>	2.91	C11	Thr109-C <sup>γ2</sup>	4.39
O4	Gln188-N <sup>ε2</sup>	3.18			
O4	W↔Gln76-O <sup>ε1</sup>	2.71			
O4	W↔Gln76-N <sup>ε2</sup>	2.80			
	and Tyr111-O <sup>n</sup>				
N5	Gly196-O	2.77			
O7	W↔Arg118-N <sup>n2</sup>	2.66			
O8	Asn199-N <sup>δ2</sup>	3.17			
O10	Arg118-N <sup>n1</sup>	2.82			

\* For the non-bonding interaction interactions, the distance was measured between the ligand atom and the nearest atom of the interacting protein residue.

for the binding function. The axial orientation of the carboxylic group seems to be more complementary to the lectin's binding site; this explains why the lectin domain picked up  $\alpha$ - rather than  $\beta$ -sialic acid, or even DANA the predominant species in solution, from the surrounding environment. Actually, the apparent importance of the free carboxylic group in determining the specificity of the recognized sialic acid has been demonstrated for other sialic acid-binding lectins as discussed in section 6.3.

The O4 hydroxyl has weak interactions with the protein; it forms a H-bond with Gln188, and participates in water-mediated interactions with the residues Gln76 and Tyr111. Such weak interactions might suggest a minor role of O4 in the substrate binding. Indeed, large variations for the role of O4 in sialic acid recognition has been reported for other lectins, see section 6.3.

The N-acetyl group acts as a bridge between two opposite sides of the binding cleft: from one side it forms a H-bond involving its N5 and the backbone O-atom of Gly196; at the other side a H-bond is formed between its carbonyl oxygen and Arg118. This arginine is kept in its binding orientation by strong H-bonding interaction with an important glutamate (Glu136). Moreover the methyl group takes part in the hydrophobic interaction with the nearby residues Tyr111, Thr109, Leu120. The strong interactions of the N-acetyl group suggested that it is a key determinant for the sialic recognition at the lectin-like domain. This was supported further by the results from the STD NMR experiment on an  $\alpha$ -sialic acid containing substrate (thiosialoside) that showed the major contribution to the protein-ligand interactions was due to the N-acetyl group at C5 (section 8.4.2.2). The important role of the N-acetyl group in the sialic acid recognition has been observed in other sialic acid-specific binding-lectins as shown in section 6.3.

The glycerol side chain forms a H-bond between its O8 and the side chain amide N-atom of the solvent exposed Asn199, which is held in the binding position by interacting

with the hydroxyl group of the neighbouring Ser197. In addition, O7 interacts with Arg118 through a bridging-water that is at a H-bond distance from the carbonyl O-atom of the N-acetyl group. The fact that O9 is pointing away from the active site implied that it is non-relevant to the sialic recognition by the lectin domain.

It is important to emphasize that although the overall topology of the binding site remained almost the same like the unbound enzyme, there are some minor changes observed at the sialic acid-binding site, relative to the free enzyme, upon binding of the ligand: (1) the side chain of Arg74 moved such that its C<sup>γ</sup> and C<sup>δ</sup> back against the side chain of Gln76, (2) the torsional angle,  $\chi^3$ , of Gln188 rotated to bring its N<sup>e2</sup>-atom to the proper position required for the interaction with O4, (3) the  $\psi$  angle of Gly196 and Asn199 rotated by ( $\approx 40 - 50$  degrees) to be in the right position for the interaction with N5 of the N-acetyl group and O8 of the glycerol side chain, respectively. Actually, such a little conformational change upon binding is in agreement with what is known about lectins as they undergo few, if any, changes in conformation upon binding to sugar (Weis & Drickmer, 1996).

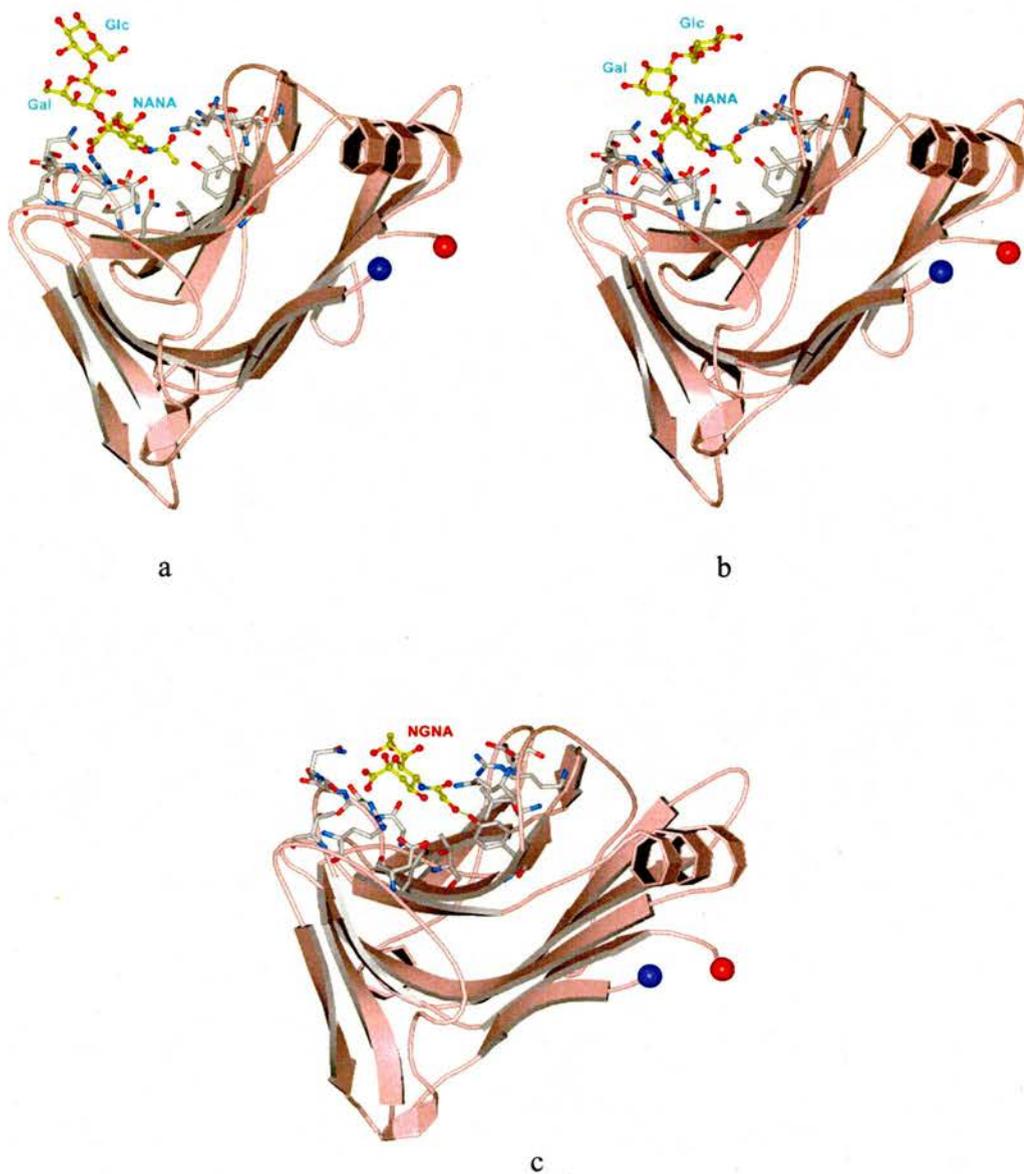
### **5.3.2.2 Specificity of the sialic acid binding**

As demonstrated in the crystallographic model (figure 5.11), it is apparent from the complementarity between the carboxylate and the binding residues that an axially oriented carboxyl group is a prerequisite for the sialic acid binding at the N-lectin domain. In fact, in most lectins studied, the negatively charged carboxylate group has been proved to be critical for recognition (Varki, 1997) [e.g., influenza haemagglutinin (HA) (Weis et al., 1988), ovine placenta (OPL I) (Troncoso et al., 2000)].

Because of the lack of any interaction of the anomeric hydroxyl at C2 of the bound  $\alpha$ -sialic acid, it might be suggested that the binding site at the N-terminal lectin-like domain of VCNA would recognize  $\alpha$ -sialic acid with different glycosidic linkages,  $\alpha(2\rightarrow3)$ , or  $\alpha(2\rightarrow6)$ . Indeed, the NMR study on thiosialosides with  $\alpha(2\rightarrow6)$  *S*-glycosidic linkage demonstrated the binding of the ligand at the sialic binding-site (section 8.4.2.2). This is illustrated by modelling  $\alpha$ -2,3- and  $\alpha$ -2,6-sialyllactose at the sialic acid-binding site as exhibited in figures 5.12a,b. The modelled ligands are expected to sit comfortably in the lectin cavity without interference with the protein residues. Indeed, this tolerance in the specificity for the sialic acid glycosidic linkages is not uncommon for sialic acid-binding lectins. Several examples of lectins recognizing different linked sialic acid have been reported [e.g., lectins from: the snail *Cepaea hortensis* (CHA) (Brossmer et al., 1992), and *Tritrichomonas foetus* (TFA) (Babal et al., 1999)].

As appeared in the crystal structure, the O4 undergoes moderate interaction with the binding site. The role of O4 varies among other lectins from being essential for binding [e.g., wheat germ agglutinin, WGA, (Wright, 1990)], to playing no role in most of the cases [e.g., CHA (Brossmer et al., 1992), influenza HA (Weis et al., 1988)].

The significant interaction of the N-acetyl group at C5 suggested that it is a key determinant for the ligand recognition. It should be mentioned that the N-acetyl group is indispensable for sialic binding in large number of lectins, the specificity of which have been studied in detail (Troncoso et al., 2000) [e.g., WGA (Wright, 1990), CHA (Brossmer et al., 1992), and *Limax flavus* (LFA)(Miller et al., 1982)]. Beside that and based on the crystal structure, it is not unlikely that N-glycolyl neuraminic acid (NGNA), a variant carries  $-\text{CH}_2\text{OH}$  group instead of  $-\text{CH}_3$  of the N-acetyl group, would be in an ideal position to H-bond with the side chain of Tyr111 as expected from the modelling of the



**Figure 5.12.** Modelling of (a)  $\alpha$ -2,3 sialyllactose, (b)  $\alpha$ -2,6-sialyllactose and (c) N-glycolyl neuraminic acid (NGNA) in the sialic acid binding site of the N-terminal lectin domain in VCNA. The binding residues at the binding site are shown in stick model and the modelled ligands are represented as ball-and-stick model. Note that in the case of (NGNA) an extra H-bond would be formed between the hydroxyl of the glycolyl side chain and the hydroxyl group of Tyr111.

putative ligand at the binding groove of the lectin domain (figure 5.12c).

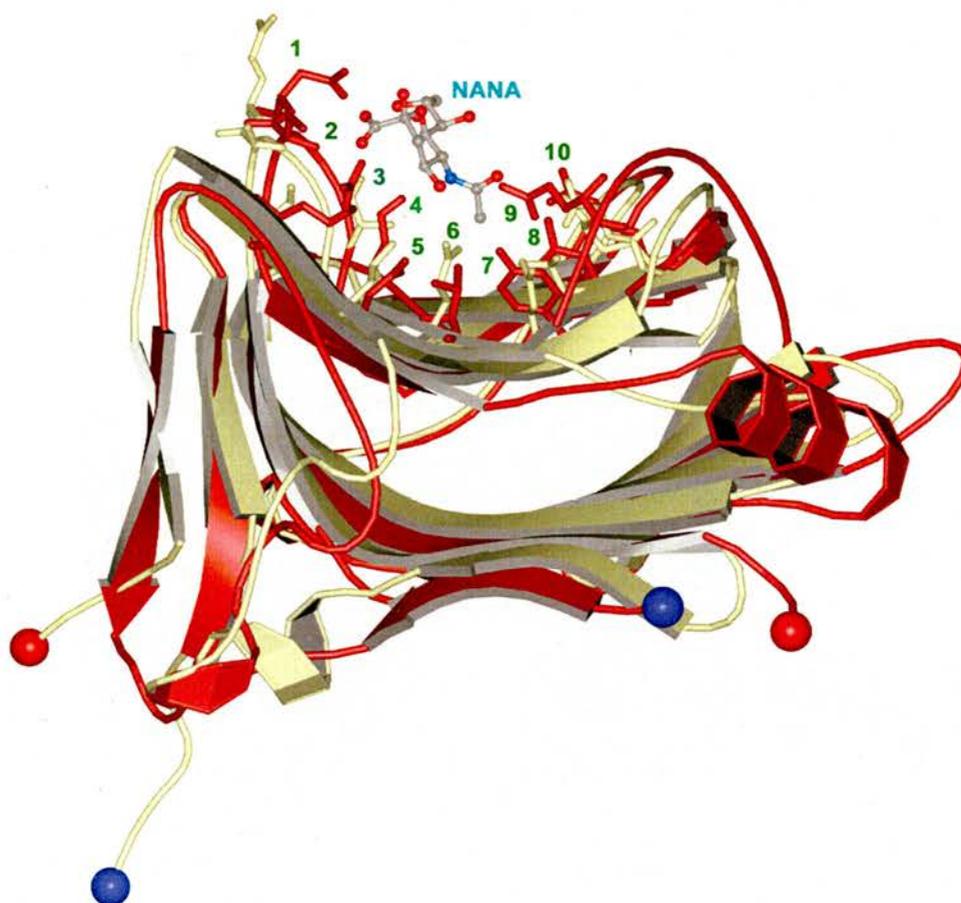
The glycerol group of the bound ligand is pointing to the exit of the binding groove with its O7 and O8 engaged in water-mediated and H-bonding interactions with the protein. Clearly, as apparent from the complex structure, the role of the trihydroxyl side chain in binding seems not to be of the same importance as the other functional groups at C2 and C5. Moreover, the lack of interaction of O9 implied the tolerance of the lectin site to bind sialic acid acetylated at O9. This is further supported by the results from the NMR experiment with Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me (section 8.4.2.2); it was deduced that the 9-O-acetyl group has no significant contribution to the ligand binding at the sialic acid-binding site. This tolerance to bind sialic acid with substituted hydroxyl group at C9 was shown before in some other lectins, e.g., lectin from the snail of *Cepeae hortensis* (Brossmer et al., 1992), and (OPL I) (Troncoso et al., 2000). On the other hands, in other studied lectins, e.g. influenza haemagglutinin (HA) (Weis et al., 1988), and sialoadhesins (May et al., 1998), the O9 was shown to be required for the sialic binding.

### 5.3.2.3 What about the C-terminal lectin domain?

The topology of the C-terminal domain is very similar to that of the N-terminal, although they only share 23% sequence identity (Crennell et al., 1994). Moreover, the relative orientations of the two domains with respect to the central domain are quite different (see figure 5.7); also the buried surface area (2015, 1998 Å<sup>2</sup> for the N- and C-lectin domains respectively) and the nature of the interactions at the interfaces are quite different: mainly polar at the N-lectin domain while it is hydrophobic at the C-terminal domain (Crennell et al., 1994). The electron density map of the complex structure with

DANA showed no indications for any binding at the C-terminal domain. Superimposition of the two wings based on their 3D structures demonstrated that the side chains of the amino acids corresponding to the binding residues at the N-lectin domain are not in the proper positions to mediate interactions with the ligand — assuming the ligand would bind to the C-lectin domain in the same way. As can be seen in figure 5.13, the residues interacting with the ligand in the N-lectin domain are, either replaced by non-equivalent residues or, if the equivalent residues existed, shifted in space from the correct positions for ligand binding. The residues interacting with the carboxylate group of the sialic acid in the N-lectin domain (Arg74, Ser198) are replaced by residues Ser365 and Ala483, with Ser365 pointing away from direction of Arg74. The C-terminal residue Glu473 corresponding to Gln188, which interacts with O4 of sialic acid, is shifted in space from the right distance required to interact with O4. Also, Leu367 at the C-terminal replaces Gln76 that is involved in water-mediated interaction with O4. The residue Arg118, H-bonded to the carbonyl oxygen of the N-acetyl group, has an equivalent residue Lys404; yet the lysine residue is not at the optimum H-bond distance. Similarly, Gly196 that interacts with N5 has equivalent Ala481 at the C-terminal, nonetheless the main-chain O-atom is away from the equivalent position in the glycine residue. Residues Asn199, Tyr111, Thr109, and Leu120 in the N-terminal domain corresponds to residues: Gln484, Gln397, Gln395, and Val406, in the same order. Note the distinct displacement of Gln484 from the position of Asn199 resulting in the loss of the H-bond with O8.

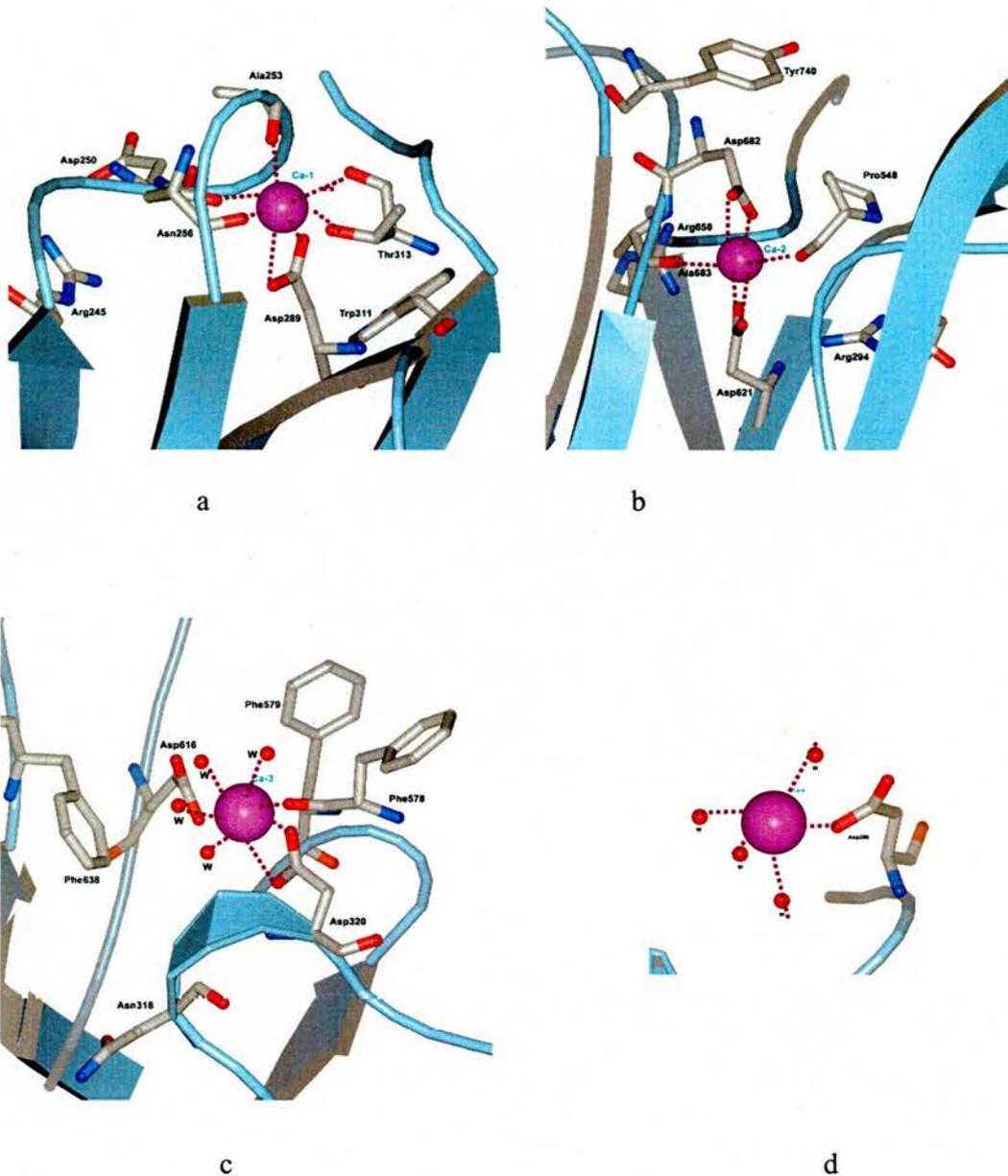
Clearly, the key interactions involving the carboxylate, N-acetyl, O4, and the glycerol groups would be lacking in the case of the C-terminal domain; this might explain why we did not see sialic acid binding in this domain. Still remains to be determined are the function of the C-terminal lectin domain and its true ligand.



**Figure 5.13.** Schematic representation of the two domains, N-terminal lectin domain coloured in brown-red and C-terminal domain coloured in gold, superimposed with the sialic acid (NANA) bound to the N-lectin domain. The residue pairs (the binding residues in the N-lectin domain and the equivalent residues in the C-terminal domain), are labelled with numbers from 1 to 10; the first residue of the pair corresponds to the N-terminal lectin domain: 1.Asn199/Gln484, 2.Ser198/Ala483, 3.Arg74/Ser365, 4.Gly196/Val406, 5.Gln76/Leu367, 6.Thr109/Gln395, 7.Tyr111/Gln397, 8.Leu120/Val406, 9.Gln188/Glu473, 10.Arg118/Lys404.  $C_{\alpha}$  at the N-termini and the C-termini of the two domains are represented as blue and red spheres respectively. Figure prepared with BOBSCRIPT (Esnouf, 1997).

### 5.3.3 Calcium ion and its role

The enzyme activity of VCNA is stimulated by calcium ions, with maximum stimulation at 1 mM CaCl<sub>2</sub> and is inhibited by EDTA at 10 mM or greater concentration (Ada et al., 1961). Two calcium ions were identified in the reported structure (Crennell et al., 1994). The electron density map (8-12 $\sigma$  level) of the DANA-complex structure revealed extra two calcium ions in addition to the previously identified ones. One of the two reported calcium ions (Ca<sup>+2</sup>-1, B = 6.93 Å<sup>2</sup>) is on the surface of the catalytic domain stabilizing the positions of important residues in the active site: Asp250, Arg245, and Trp311 (figure 5.14a). This calcium ion is coordinated to seven O-atoms (backbone and side-chain atoms) with coordinate bond distances in the range 2.24 to 2.62 Å; table 5.7 lists the coordinating atoms and distances. The second calcium ion (Ca<sup>+2</sup>-2, B = 5.95 Å<sup>2</sup>) is located beneath the sugar ring of the ligand, ~11 Å apart from the carboxylate group, and coordinated to six protein O-atoms (table 5.7). This calcium plays a significant role to stabilize the architecture of the catalytic site by neutralizing Asp621 and Asp682 in the vicinity of Tyr740 and Glu619 involved in the mechanistic hydrolysis (figure 5.14b). The third calcium ion (Ca<sup>+2</sup>-3, B = 9.16 Å<sup>2</sup>), not reported previously, is located on the surface of the central domain not far from the catalytic site and about 19 Å apart from Ca<sup>+2</sup>-1. As shown in figure 5.14c, this calcium ion has a coordination number of seven, four of the coordinate O-atoms are from ordered water molecules (table 5.7). The importance of this calcium arises from supporting the loop carrying Asn318, which interacts with the bound ligand (section 5.2.1.1). Additionally, this calcium ion is not far from the aromatic ring of Phe638 (distance 4.40, 4.73 Å from C<sup>δ</sup> and C<sup>ε2</sup>, respectively), which is involved in the hydrophobic interaction with C9 of DANA. The close proximity to the phenyl ring suggests a possible role in stabilization of the solvent exposed side chain through cation- $\pi$  interaction (Gallivan & Dougherty, 1999). Finally, an electron density (8 $\sigma$  level) at a



**Figure 5.14.a to d** Display the coordination of the four calcium ions. The calcium ion is shown as large sphere coloured in magenta and the coordinating residues are drawn as sticks; coordinating water molecules are represented as red spheres.

**Table 5.7.** Lists of the coordinate atoms for calcium ions and the corresponding bond distances

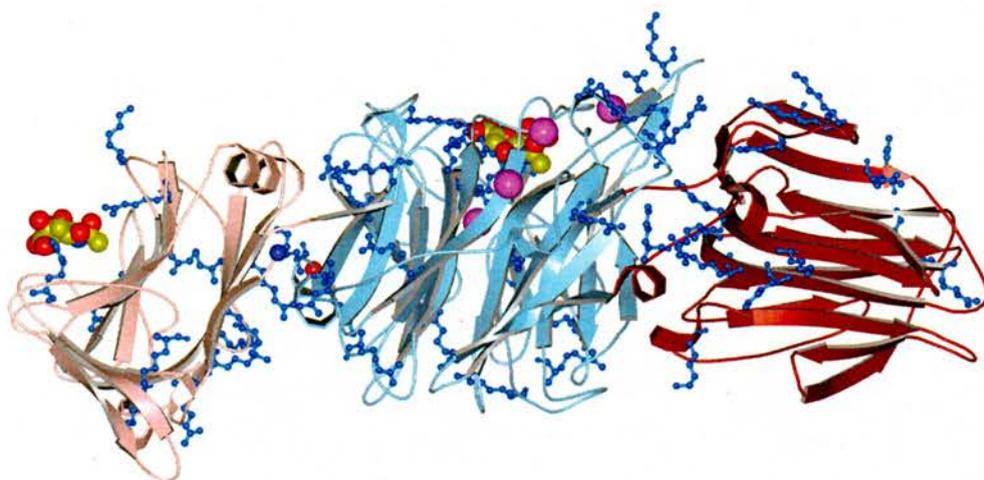
Ca <sup>+2</sup> -1		Ca <sup>+2</sup> -2		Ca <sup>+2</sup> -3		Ca <sup>+2</sup> -4	
Coordinate atom	Distance (Å)						
Asn256-O	2.24	Pro548-O	2.35	Asp320-O <sup>δ1</sup>	2.66	Asp286-O <sup>δ1</sup>	2.40
Asn256-O <sup>δ1</sup>	2.26	Ala683-O	2.34	Asp320-O <sup>δ2</sup>	2.47	OH <sub>2</sub> 971	2.43
Asp289-O <sup>δ1</sup>	2.62	Asp682-O <sup>δ1</sup>	2.39	Phe578-O	2.30	OH <sub>2</sub> 1009	2.48
Asp289-O <sup>δ2</sup>	2.23	Asp682-O <sup>δ2</sup>	2.37	OH <sub>2</sub> 908	2.39	OH <sub>2</sub> 1030	2.64
Ala253-O	2.30	Asp621-O <sup>δ1</sup>	2.46	OH <sub>2</sub> 1767	2.32	OH <sub>2</sub> 1049	2.36
Thr313-O <sup>γ1</sup>	2.37	Asp621-O <sup>δ2</sup>	2.42	OH <sub>2</sub> 1768	2.51		
Thr313-O	2.52			OH <sub>2</sub> 1769	2.48		

\*The numbering of coordinating water molecules is that used in the pdb file.

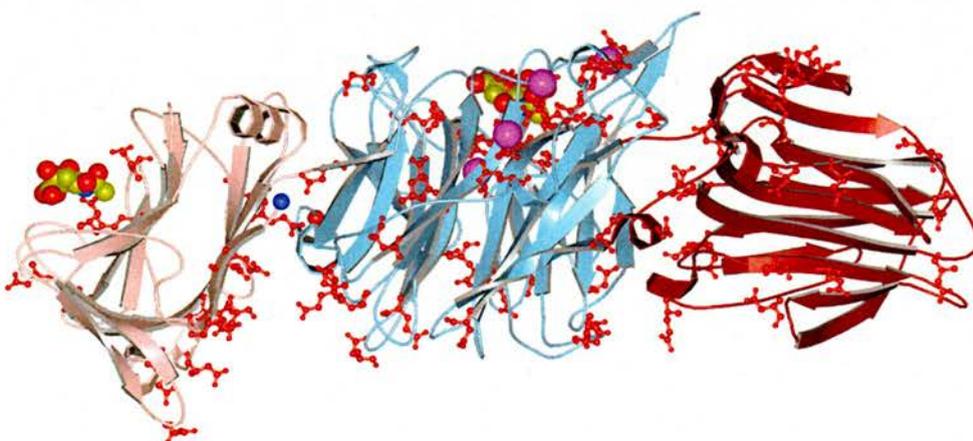
distance of 2.40 Å from Asp286-O<sup>δ1</sup> and surrounded by four water molecules within distances of 2.36 to 2.64 Å was interpreted tentatively as Ca<sup>+2</sup> ion (Ca<sup>+2</sup>-4, B=13.03 Å<sup>2</sup>), see figure 5.14d. It is worthy to mention that the most common coordination number of Ca<sup>+2</sup> is 6 or 7; nevertheless, calcium ion with higher (8, 9) and lower (5) coordination number has been documented (Nayal & Di Cera, 1994). This calcium ion is positioned on the surface of the catalytic domain ~11 Å from Ca<sup>+2</sup>-1. Although, there is no clear structural role for this last calcium ion, it could be used, beside the other three ions, to create a positive potential counteracting the highly repulsive negative potential on the protein surface opposite to the catalytic site as discussed in the next section.

#### **5.3.4 Surface charge: implication for sialic acid recognition**

The distribution of charged groups in VCNA was analysed and revealed to be asymmetrical with a high negative charge on the face of the protein opposite to the catalytic site. Figure 5.15a,b exhibit the distributions of the positively (Arg, Lys) and negatively (Asp, Glu) residues in VCNA; figure 5.16 maps the electrostatic potential on the surface of the enzyme. The formal charge on VCNA is -19 (taking into account the charge of the 4×Ca<sup>+2</sup> ions). Similar charge distributions were observed for the neuraminidase from *M. viridifaciens* (Gaskell et al., 1995) and trans-neuraminidase from *M. decora* (Luo et al., 1998). As suggested for the *M. decora* trans-neuraminidase, these enzymes could utilize this electrostatic feature (Luo et al., 1998), to increase their efficiency by avoiding being misoriented and trapped on the anionic sialoglycoconjugate-rich cell surface. Moreover, the high negative repulsive potential at the cleft between the N-terminal and central domains, figure 5.16, might help to keep the orientation of the N-terminal domain in optimum position to bind the cell surface facilitating the enzyme



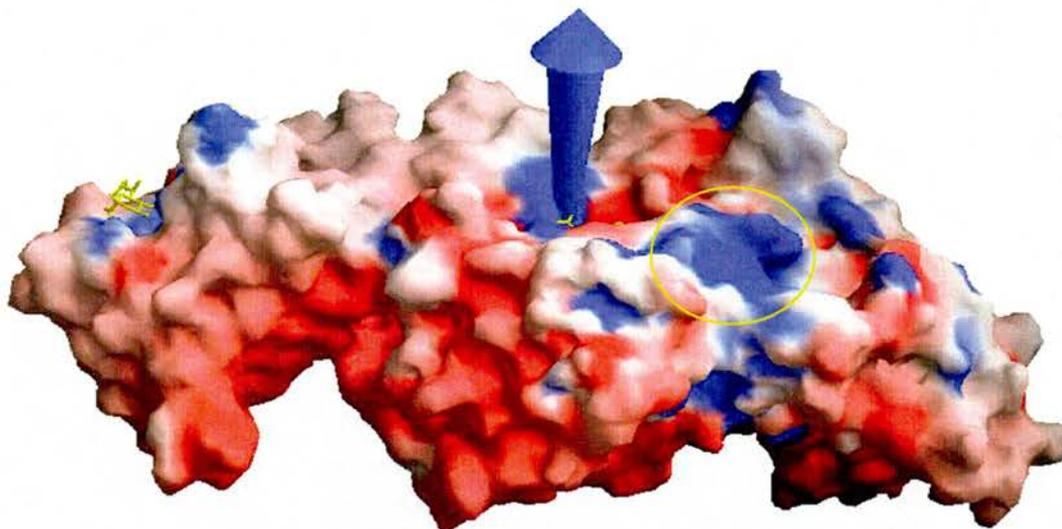
a



b

**Figure 5.15.** Distribution of the charged residues in VCNA: (a) positively charged residue of types Arg and Lys, coloured in blue, (b) negatively charged residue of types Asp and Glu, coloured in red; residues are drawn as ball-and-stick. The ligands, DANA and NANA, are shown as CPK model. The spheres coloured in magenta represents calcium ions. Figure prepared with BOBSCRIPT (Esnouf, 1997).

Surface Potential    -9.850    -4.925    0.000    1.835    3.670    >-<



**Figure 5.16.** The electrostatic potential mapped onto the surface of VCNA, colour coded from  $-9.65$  KT in red to  $3.67$  KT in blue. The dipole moment of the protein is represented as an arrow, which pass along the location of the carboxylate group of the ligand. The ligands DANA and NANA are represented as capped-stick model and coloured in yellow. The positive pocket enclosed within the yellow circle might be a potential site that recognizes a negatively charged carbohydrate. Figure prepared with GRASP (Nicholls et al., 1991).

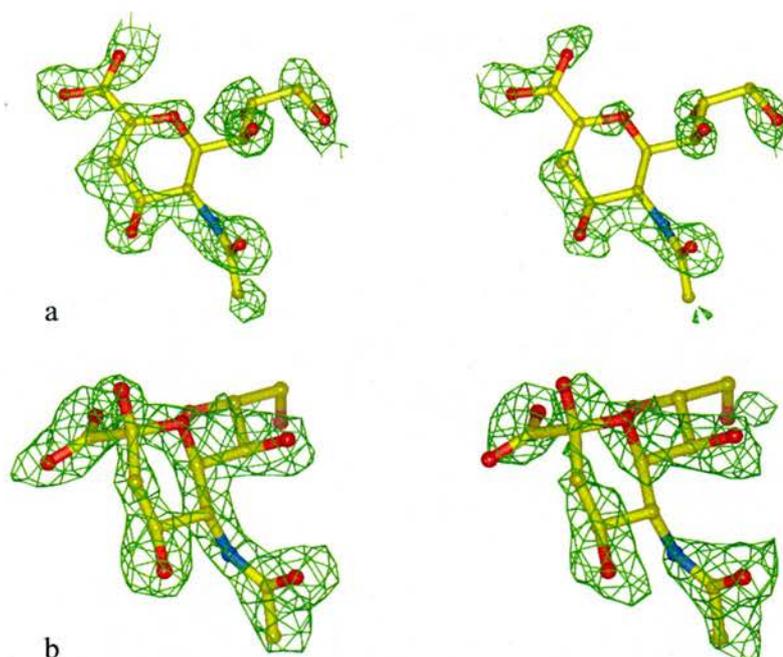
function. Interestingly, in addition to the positive attractive potential created by the arginine triad that guides the approaching negatively charged sialic-substrate towards the enzyme active site, the calculated dipole moment of the protein is pointing directly along the carboxylate group of the ligand with the right orientation to attract the ligand to its receptor. Indeed, the positions of the four calcium ions close to the face of the protein where the catalytic site located, and more specifically the  $\text{Ca}^{+2}$ -2 beneath the ligand position, might be of significance in creating a counteracting positive potential to balance the high negative potential on the bottom surface of the enzyme.

Another point to mention, is that the positive pocket at the interface between the C-terminal and central domains could be a potential binding pocket for a negatively charged carbohydrate has yet to be determined. In fact, we thought that this suspected pocket might recognize heparin sulfate as a substrate; however, the idea did not work, see section 8.4.3.

#### **5.4 VCNA-3'-sialyllactose complex**

The structure of VCNA in complex with the substrate,  $\alpha$ -2,3-sialyllactose solved at 1.9 Å, reproduced the same results obtained from the crystal structure of the DANA-complex. The electron density maps,  $2F_o-F_c$  and  $F_o-F_c$  (figure 5.17b), as expected, revealed a clear density for  $\alpha$ -sialic acid ( $B_{av} = 29.54 \text{ \AA}^2$ ) bound to the N-terminal lectin domain exactly in the same way as that observed in the DANA-complex structure, confirming the new finding of the sialic acid-binding site. However, the maps exhibited no density for the glucose or galactose residues of the substrate. This could be due to the free mobility of the two carbohydrate moieties, as both residues would be hanging free in the solvent with no contact with the protein, see section 5.3.2. Another possibility still exists is that the sialic acid seen bound at the lectin domain could be the one produced by the enzymatic hydrolysis of the substrate rather than the intact-sialic acid in the unprocessed  $\alpha$ -2,3-sialyllactose. The intact substrate might be prevented from binding because of the unfavourable contact with symmetry related macromolecules in the crystal lattice. We will not discuss the interactions of the sialic acid at the N-terminal lectin domain, as it would be a repeat of what was presented previously in section 5.3.2.1.

Additionally, it was expected to find the substrate or the products of the hydrolysis

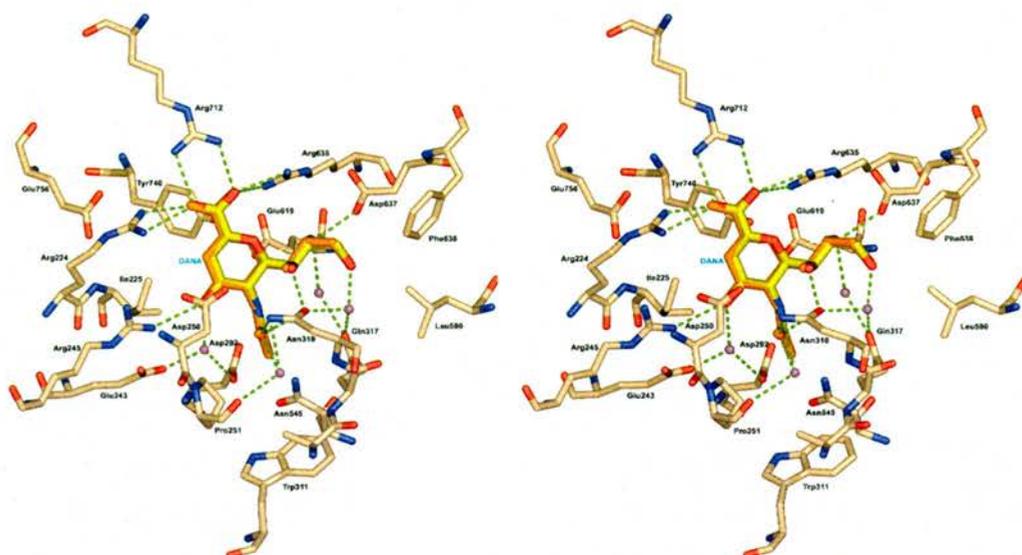


**Figure 5.17.** (a) The electron density maps for the 3'-sialyllactose-complex:  $2F_o-F_c$  contoured at  $1.2\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the ligand DANA at the catalytic site. (b) the  $2F_o-F_c$  at  $1.2\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the  $\alpha$ -sialic acid at the N-terminal lectin domain. All maps were calculated before adding the ligands to the refined model of the VCNA-3'-sialyllactose complex.

reaction at the catalytic site; however, surprisingly, the 3'-sialyllactose complex structure exposed the inhibitor DANA ( $B_{av} = 27.11 \text{ \AA}^2$ ) laid at the catalytic domain as judged from the electron density difference maps ( $2F_o-F_c$  and  $F_o-F_c$ ) shown in figure 5.17a. The features of the electron density of the ligand at the catalytic site were almost identical to that observed in the DANA complex. The absence of density corresponding to the anomeric hydroxyl and the fitting of the half-chair DANA conformation, with the carboxylate group in the same plane formed by the atoms O6-C1-C2-C3, to the observed electron density computed before adding any ligand indicated the presence of DANA in the  $\beta$ -propeller fold. However, the electron density of C6 is missing from the calculated difference maps, also the electron density around the atoms O6 and O1B (close to

Arg635) are slightly distorted and extends in space pointing to the direction above the sugar ring. These little imperfections of the electron density could be due the low occupancy of the identified ligand as a result of the low concentration of the formed inhibitor. Actually, the average B-factor of DANA is higher than that of the protein atoms ( $B_{av} = 27.1, 14.6 \text{ \AA}^2$ , for DANA and protein atoms, respectively; section 4.3.2.2, table 4.3). Indeed, adding the solvent to the calculated electron density maps ( $F_o - F_c$  and  $2F_o - F_c$ ), before incorporating the ligand, revealed water molecules at the exact positions corresponding to the oxygen atoms of DANA. Further evidence for the formation of VCNA of its own inhibitor (DANA) during the hydrolysis of the substrate was disclosed by the  $H^1$  NMR results as discussed in section 8.4.1. The NMR study demonstrated the reversible formation of DANA during the enzymatic hydrolysis of  $\alpha$ -2,3-sialyllactose by the action of VCNA. The higher affinity of DANA towards the catalytic site [ $K_i = 26\text{-}30 \mu\text{M}$  (Zbiral et al., 1989; Meindl et al., 1974)], might explain the observation of DANA at the catalytic domain despite its small concentration. Figure 5.18 displays the ligand (DANA) bound at the catalytic site as revealed from the structure of the 3'-sialyllactose complex. As shown in figure 5.18, DANA binds at the catalytic site in the exact way observed previously in the VCNA-DANA complex, this is evident from superimposing the two structures.

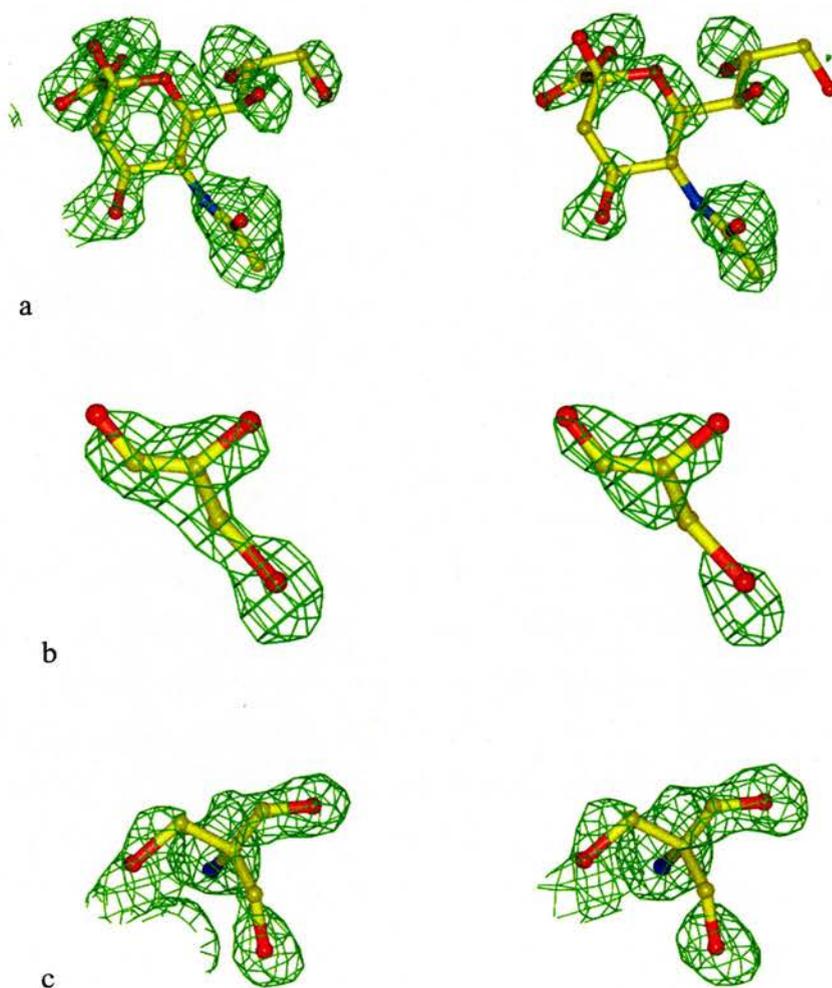
All together, the crystal structure of the VCNA-3'-sialyllactose complex demonstrated the ability of the bacterial neuraminidase to form its own inhibitor as shown before in the case of the viral neuraminidase from influenza (Burmeister et al., 1993), and more recently in the *T. rangeli* neuraminidase (Buschiazzo, 2002).



**Figure 5.18.** Stereo view of the catalytic site with DANA bound, coloured in: yellow (C), red (O), and blue (N), as revealed by the structure of the 3'-sialyllactose complex. The protein-ligand interactions are depicted as dotted green lines. The ligand DANA (coloured in orange) from the VCNA-DANA complex structure is superimposed on the bound DANA from the 3'-sialyllactose complex structure to show the exact mode of binding of the ligand in both structures.

## 5.5 VCNA-6'-sialyllactose complex

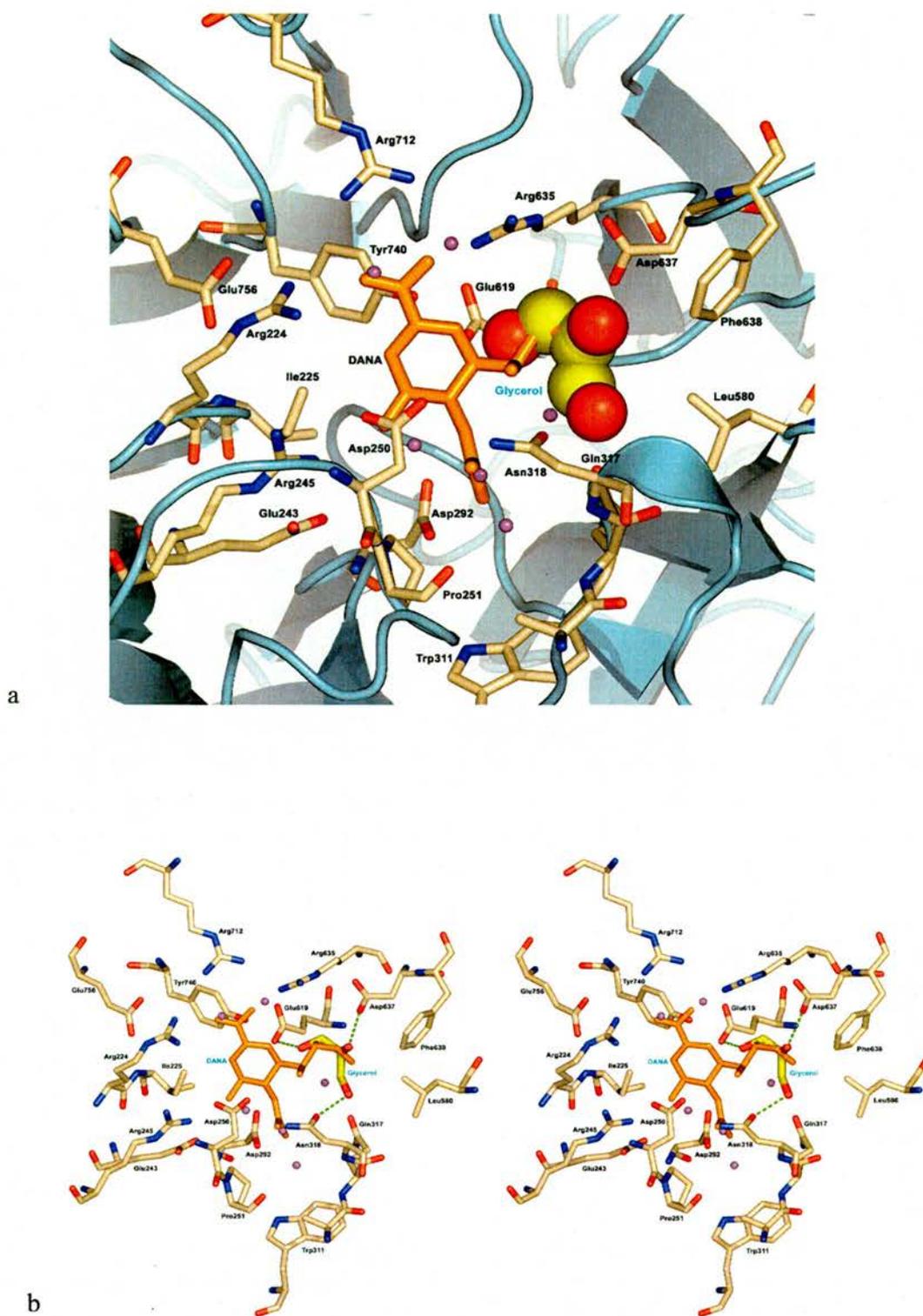
As reported in earlier work, VCNA releases sialic acid from  $\alpha$ -2,3-sialyllactose more rapidly than its release from the  $\alpha$ -2,6-sialyllactose substrate (Corfield et al., 1983; Friebolin et al., 1984). We were interested in the crystal structure of VCNA in complex with the substrate  $\alpha$ -2,6-sialyllactose for comparison with the results obtained from the 3'-sialyllactose-complex. The 6'-sialyllactose complex structure was solved to 1.6 Å resolution. The electron density maps ( $2F_o - F_c$  and  $F_o - F_c$ ), figure 5.19a, displayed  $\alpha$ -sialic acid bound at the N-terminal lectin domain as shown before in the DANA- and 3'-sialyllactose complexes. However, the electron density is not as clear as that observed for  $\alpha$ -sialic acid in the previous two complexes; this might suggest lower occupancy of



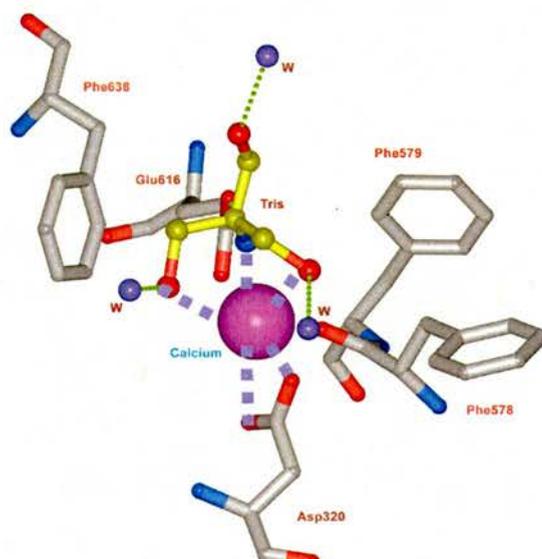
**Figure 5.19.** (a) The electron density maps for the 6'-sialyllactose-complex:  $2F_o-F_c$  contoured at  $1\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for  $\alpha$ -sialic acid at the N-terminal lectin domain. (b)  $2F_o-F_c$  at  $1.2\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the glycerol molecule at the catalytic site. (c)  $2F_o-F_c$  at  $1.2\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the tris molecule identified near  $Ca^{+2}$ -3 ion. All maps were calculated with the ligands omitted from the VCNA-6'-sialyllactose complex model structure.

the bound ligand, especially the B-factor of the ligand was higher than that of the protein atoms ( $B_{av} = 33.3, 13.3 \text{ \AA}^2$ , for sialic acid and protein atoms, respectively; section 4.3.2.2, table 4.3). Similarly, as observed in the 3'-sialyllactose complex, no density could be observed for the glucose and galactose carbohydrate residues of the substrate. As mentioned before in the case of the 3'-sialyllactose complex, this could be due to the mobility of the unbound sugar residues, or the bound sialic acid was the product of the hydrolysis reaction rather than the intact substrate.

In contrast to the results obtained for the 3'-sialyllactose complex, no electron density was observed for the substrate (or the inhibitor) at the catalytic domain. Instead, a glycerol molecule (sourced from the cryoprotector solution) was identified at the active site as judged by the observed electron density, figure 5.19b. The trihydroxyl group forms H-bond with Gln317, Asn318, Asp619, and Asp637, as shown in figure 5.20. Moreover, the polar group participates in the hydrophobic interaction with Phe638, Ile580, and the aliphatic part of the side chain of Asp637. The position of the glycerol group is slightly different from that of the glycerol substituent at C5 of the bound DANA in the DANA- and 3'-sialyllactose complexes. Interestingly, when superimposing the two structures from the 6'-sialyllactose- and DANA-complexes as shown in figure 5.20, the position of the glycerol molecule suggests the possibility of replacing the glycerol side chain of DANA with a butyl group. The extra hydroxyl group is expected to mediate the interaction observed for one of the hydroxyls in the glycerol molecule. It is important to mention that the electron density map revealed the presence of structural waters at positions corresponding to the carboxylate oxygens, O4, and the oxygen atom of the N-acetyl group of DANA as noted in the previous two complexes, see figure 5.20. This is in agreement with the fact that the neuraminidase catalytic site is rigid, and consequently the residues at the active site are pre-formed for recognition of the recognizable carbohydrate.



**Figure 5.20.** (a) Ribbon diagram for the active site of VCNA with the bound glycerol moiety depicted as CPK model, with yellow C-atoms, in the 6'-sialyllactose structure, and the ligand DANA, coloured in Orange, from the structure of DANA-complex. (b) Stereo view for the active site showing the interactions of the glycerol side chain. Structure waters are represented as spheres in magenta.



**Figure 5.21.** Coordination of  $\text{Ca}^{+2}$ -3 ion by the Tris molecule as revealed in the 6'-sialyllactose complex structure.

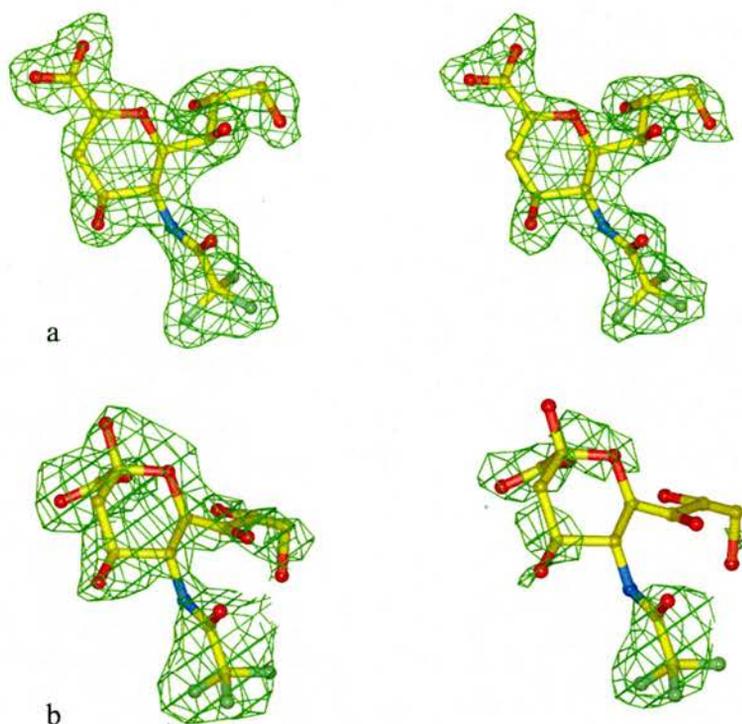
The absence of the true ligand from the catalytic site could be attributed to the lower affinity of the  $\alpha$ -2,6-sialyllactose substrate towards VCNA compared to the  $\alpha$ -2,3-sialyllactose analogue. Another feature revealed by the high resolution structure is the observation of an electron density that fits a Tris molecule, or Tris[hydroxymethyl]aminomethane (a constituent of the buffer system), chelating the  $\text{Ca}^{+2}$ -3 ion. In this case, the  $\text{Ca}^{+2}$ -3 is coordinated to two oxygen atoms and the nitrogen atom of the Tris molecule (figure 5.21), and the coordinating water molecules and the Phe578-O atom are replaced, see the discussion in section 5.3.3. Whether the Tris molecule was also the real chelating agent in the DANA- and 3'-sialyllactose complex structures, yet could not be observed due to the limitation of the resolution, is uncertain.

## 5.6 VCNA complexes with DANA-derivatives varied at C5

Large numbers of sialic acid-based inhibitors have been synthesised and tested against VCNA; most of these inhibitors are derivatives of the transition-state-analogue DANA (Chan et al., 1997; Zbiral et al., 1989; Holzer et al., 1993; Khorlin et al., 1970; Meindl et al., 1974; Meindl & Tuppy, 1969; Nohle et al., 1985; Schreiner et al., 1991a; Schreiner et al., 1991b; Wilson et al., 2000). Of these inhibitors, FANA (Meindl et al., 1974), which has N-trifluoroacetyl instead of the N-acetyl group at C5 of DANA, is the most potent inhibitor for VCNA reported so far, a 10-fold increase in inhibition over DANA was reported [ $K_i = 1.8\text{-}2.5 \mu\text{M}$  (Meindl et al., 1974; Schreiner et al., 1991b)]. Also, the derivatives (GANA and CANA), both have the core of DANA with modifications at C5, showed improved inhibition over DANA against VCNA, yet not as much as FANA (Nohle et al., 1985; Wilson et al., 2000). For these N-acyl derivatives, molecular modelling study anticipated favourable interactions of the substituents at C5 relative to the N-acetyl group of DANA within the catalytic site of VCNA (Wilson et al., 2000). Also, from the crystal structure of DANA complex (section 5.3.1) the improved inhibition of FANA, GANA, and CANA might be predicted. For a deeper understanding of the observed preferences of VCNA to bind these three inhibitors we investigated the 3D-structures of the inhibitors bound at the catalytic domain of the enzyme as discussed in the following sections.

### 5.6.1 VCNA-FANA complex

The FANA-complex structure was resolved at resolution 2.1 Å; clear electron density was observed in the electron density maps ( $2F_o - F_c$  and  $F_o - F_c$ ) for the inhibitor bound at the catalytic domain of VCNA, as shown in figure 5.22a. Unexpectedly, the maps ( $2F_o - F_c$



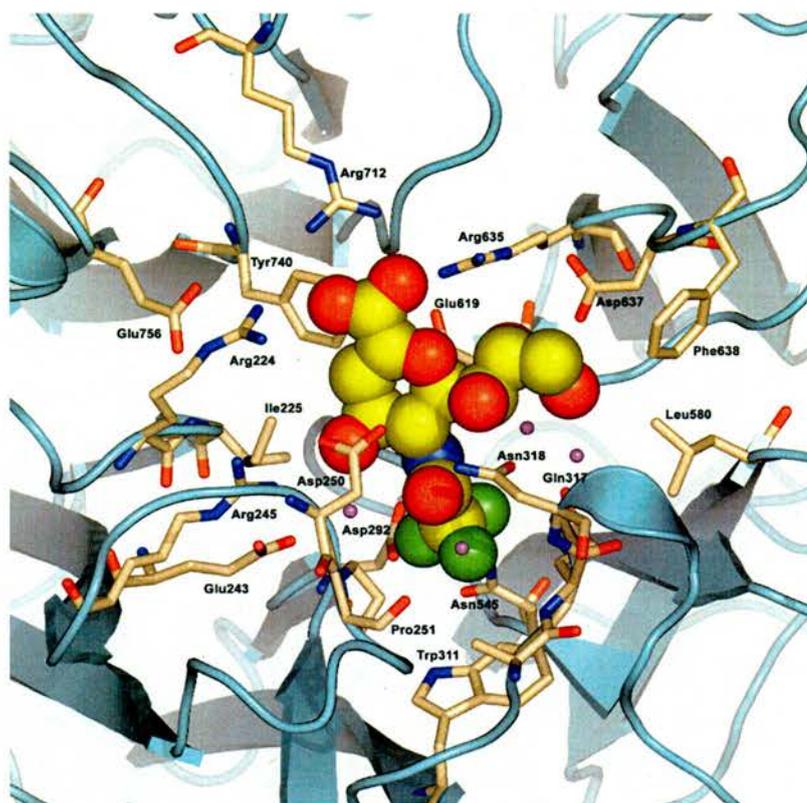
**Figure 5.22.** (a) The electron density maps for the FANA-complex:  $2F_o-F_c$  contoured at  $1.2\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the ligand FANA at the catalytic site. (b)  $2F_o-F_c$  at  $0.9\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the FNANA at the N-terminal lectin domain. All maps were calculated before adding the ligands to the refined model of the VCNA-FANA model structure.

and  $F_o-F_c$ ) revealed an electron density at the sialic acid-binding site. This electron density was interpreted as the  $\alpha$ -sialic acid analogue of FANA (5-N-trifluoroacetyl-neuraminic acid, or FNANA), figure 5.22b. Actually, the quality of the electron density is relatively poor, however, fitting the assumed ligand to the observed density showed a high electron density (contour level  $4.6\sigma$ ) at the location of the fluorine atoms; and the density of the oxygen atoms of the carboxylate group, O4, and O6 can be identified at  $3.0\sigma$  level. This suggested that FNANA (the hydrolytic product of the C2–C3 double bond of the inhibitor FANA) was laid in the sialic acid-binding site at the N-terminal domain similar

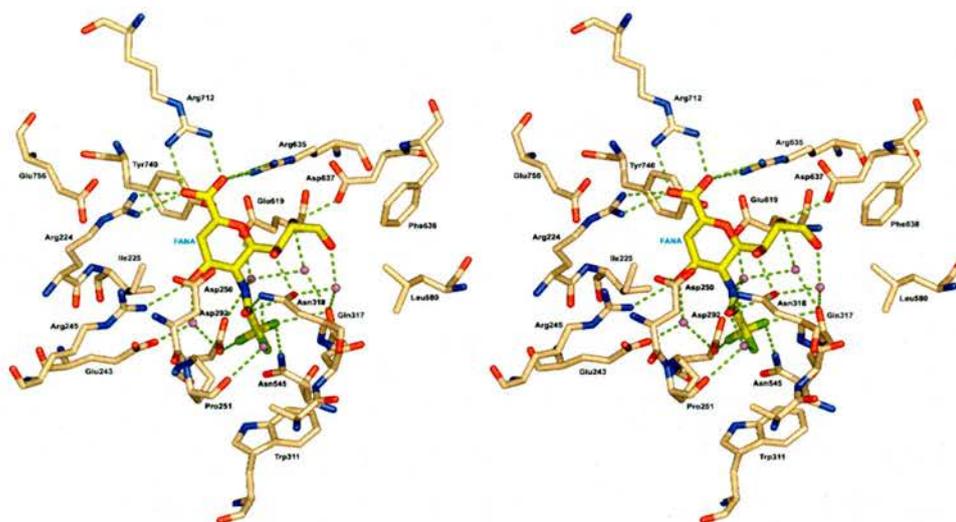
to the earlier observation. In particular, the ligand fitted the observed density without steric clashes with the binding site residues. In fact, our previous argument (section 5.3) that VCNA can act on DANA to produce sialic acid, which was further supported by the evidence from the NMR study (section 8.4.1), is applicable to the case of FANA. This means that the observed ligand (FNANA) at the sialic acid-binding site could be produced by the action of VCNA on FANA. Nevertheless, the possibility that this ligand could have pre-existed as a contaminant during the synthesis of the compound, or as a degradation product of FANA could not be escaped. Finally, the analysis of the crystal packing in the different complexes suggested that the ligand binding at the N-terminal domain might account for the different packing observed among these complexes (section 5.2.2.3). We showed that the FANA-complex was packed in the same way as the complexes of DANA, 3'-sialyllactose, and 6'-sialyllactose, in which the sialic acid ligand was observed at the N-terminal lectin domain. Consequently, it is not unlikely that the presence of a bound ligand at the sialic acid-binding site of FANA-complex was the cause for the same observed packing pattern.

#### **5.6.1.1 Binding of FANA at the catalytic site**

The bound FANA exhibited the same binding features as that observed earlier for DANA in its complex structure (section 5.3.1.1), except for the interaction of the  $-\text{CF}_3$  group. The interactions between FANA and the protein residues at the catalytic cleft are depicted in figure 5.23; table 5.8 lists the protein-inhibitor interaction and the corresponding interatomic distances. As displayed in the figure 5.23, the fluorine atoms undergo direct and water mediated H-bonding interactions with nearby residues. The fluorine atom (F11A) is involved in a direct H-bond with Asp292, and in a strong water-

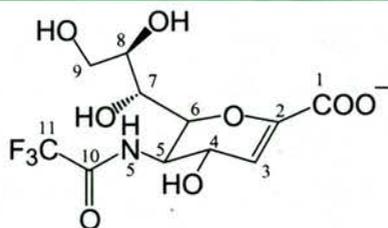


a



b

**Figure 5.23.** (a) Ribbon representation of the active site of VCNA with the inhibitor, FANA, bound (drawn in CPK model), the active site residues are labelled and drawn as sticks. (b) Stereo view for the active site, the hydrogen bonding interactions drawn as dotted lines; water molecules are represented as magenta spheres.

**Table 5.8.** List of interactions between VCNA and the bound FANA

Hydrogen bonding			Non-bonding interaction		
FANA-atom	VCNA-atom	Distance (Å)	FANA-atom	VCNA-atom	Distance (Å)
O1A	Arg224-N <sup>η1</sup>	2.99	C1	Tyr740-O <sup>η</sup>	2.94
O1A	Arg224-N <sup>η2</sup>	3.03	C2	Tyr740-O <sup>η</sup>	2.88
O1A	Arg712-N <sup>η1</sup>	2.88	C4	Ile225-C <sup>δ</sup>	4.09
O1B	Arg712-N <sup>η2</sup>	2.80	C9	Phe638-C <sup>ζ</sup>	4.18
O1B	Arg635-N <sup>η1</sup>	3.37	C9	Leu580-C <sup>δ1</sup>	4.73
O1B	Arg635-N <sup>η2</sup>	3.20	C9	Asn637-C <sup>β</sup>	4.41
O4	Arg245-N <sup>η2</sup>	3.03	F11A	Trp311-C <sup>ζ2</sup>	3.94
O4	Asp250-O <sup>δ1</sup>	2.88	F11B	Pro251-C <sup>γ</sup>	3.69
O7	Asn318-O <sup>δ1</sup>	2.67	F11C	Gln317-C <sup>β</sup>	3.17
O8	Asp637-O <sup>δ2</sup>	2.84			
O10	Asn318-N <sup>δ2</sup>	3.30			
O10	Asn318-O <sup>δ1</sup>	3.29			
F11A	Asp292-O <sup>δ1</sup>	2.88			
F11B	Gln317-N <sup>δ2</sup>	2.98			
F11B	Asn545-N <sup>δ2</sup>	3.05			

**Table 5.9.** Inhibition data against VCNA for DANA and its N-fluoro-derivatives\*

R	CH <sub>3</sub> <sup>a</sup>	CH <sub>2</sub> F <sup>b</sup>	CHF <sub>2</sub> <sup>c</sup>	CF <sub>3</sub> <sup>d</sup>
IC <sub>50</sub> (10 <sup>-5</sup> M)	3.0	0.95	0.32	0.25

\* Reference: (Meindl et al., 1974)

<sup>a</sup>R=CH<sub>3</sub>: 2-deoxy-2,3-didehydro-5-N-acetylneuraminic acid (DANA)

<sup>b</sup>R=CH<sub>2</sub>F: 2-deoxy-2,3-didehydro-5-N-monofluoroacetylneuraminic acid

<sup>c</sup>R=CHF<sub>2</sub>: 2-deoxy-2,3-didehydro-5-N-difluoroacetylneuraminic acid

<sup>d</sup>R=CF<sub>3</sub>: 2-deoxy-2,3-didehydro-5-N-trifluoroacetylneuraminic acid

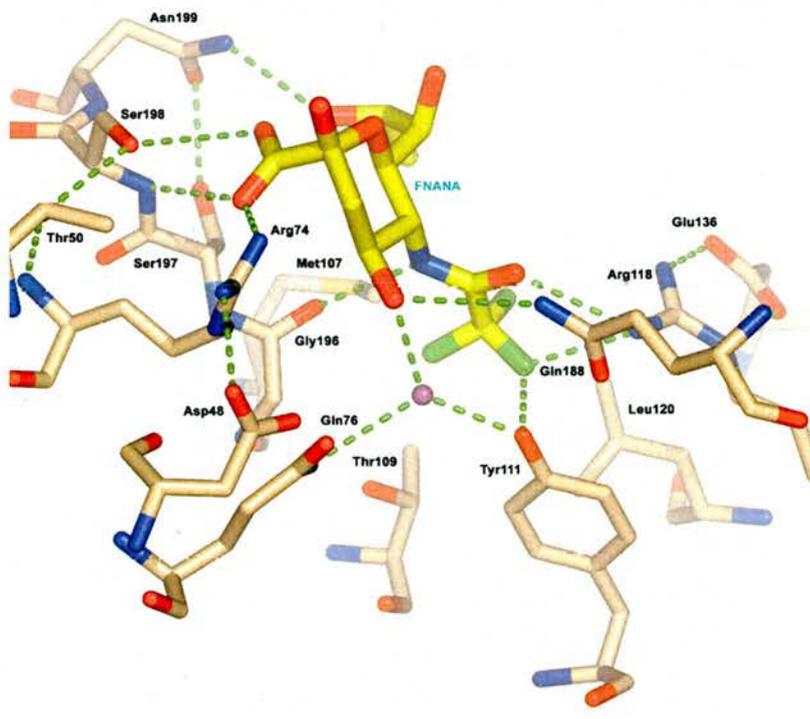
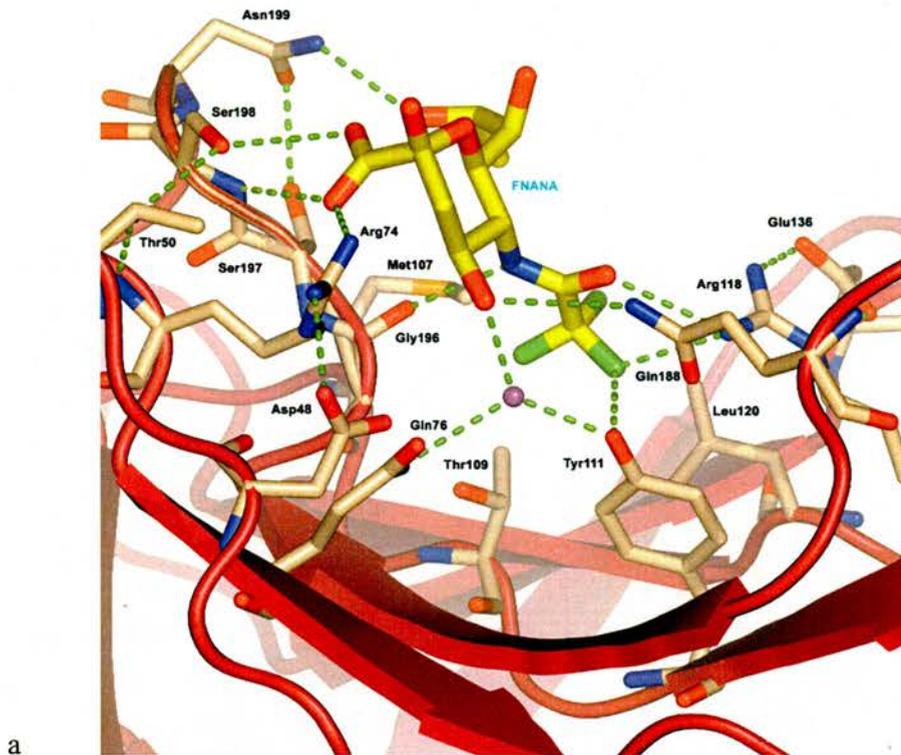
mediated interaction with Glu243. Two H-bonds are formed between the fluorine atom (F11B) and the two residues, Gln317 and Asn545. The third fluorine atom (F11C) mediates a H-bond with Asn318 through a bridging-water; additionally, this fluorine atom is in close proximity to Trp311, Pro251, and side chain of Gln317 indicating a strong hydrophobic contact.

Clearly, the extra interaction introduced to the DANA inhibitor by substituting the methyl group of the acyl substituent at C5 with the trifluoromethyl group, that is capable of making H-bonding interactions with residues at the binding pocket, can account for the enhanced potency of FANA over DANA against VCNA. It should be mentioned that the FANA analogues with one and two fluorine atoms at the methyl group also showed improved inhibition compared to DANA (Meindl et al., 1974). Indeed, the enhanced inhibition due to replacing two hydrogen atoms of the methyl group by two fluorine atoms had almost the same effect as that produced by replacing the third hydrogen by a fluorine atom. This can be seen in the table 5.9, which presents the inhibition data (IC<sub>50</sub>) determined by Meindl et al. This is in agreement with the structural data, which showed interactions, inferred from the number of direct H-bonds, for the two fluorine atoms (F11A, and F11B) compared to the third atom (F11C).

### 5.6.1.2 Binding at the sialic acid-binding site

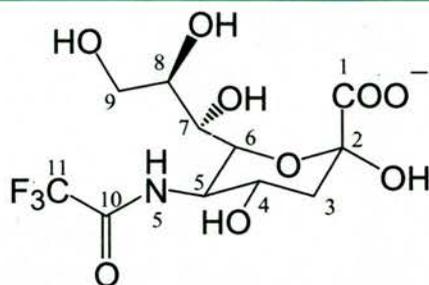
Figure 5.24 shows the interaction of FNANA at the sialic acid-binding site of the N-terminal domain. The ligand demonstrated the  $\alpha$ -conformation of sialic acid, and interacts at the binding groove with its carboxylate group down towards the protein surface and the anomeric hydroxyl pointing to the solvent. The observed protein-ligand interaction is similar to that described previously for the binding of  $\alpha$ -sialic acid described in section (5.3.2.1), except for the interaction of the extra trifluoromethyl group of the N-acyl at C5. Table 5.10 lists the contacts of the ligand within the binding site. The fluorine atoms displayed favourable H-bonding and hydrophobic interactions with the binding site residues. One of the fluorine atoms (F11A) forms two H-bonds with the hydroxyl group of Tyr111 and the side chain of Arg118. Additionally, this fluorine atom is in close contact with Tyr-C<sup>52</sup> (distance 3.20 Å), indicating a hydrophobic interaction with the tyrosine ring. Note that, Tyr111-OH was not involved in any direct interaction in the case of sialic acid binding. The second fluorine atom (F11B) forms a hydrophobic interaction with residues Met107 and Thr109. The third fluorine atom (F11C) contributes to a hydrophobic interaction with the nearby residue Leu120.

It is conceivable that replacing the methyl hydrogen atoms of the N-acetyl group at C5 by fluorine atoms enhanced the interaction with the protein via extra contacts. This suggests higher binding affinity of the unnatural fluorinated sialic acid towards the N-lectin domain of VCNA.



**Figure 5.24.** (a) Ribbon diagram combined with the stick-model of the binding site residues and the ligand (FNANA), (b) the same view as in (a) without displaying the ribbon diagram for clarity. The dotted lines represent H-bonds.

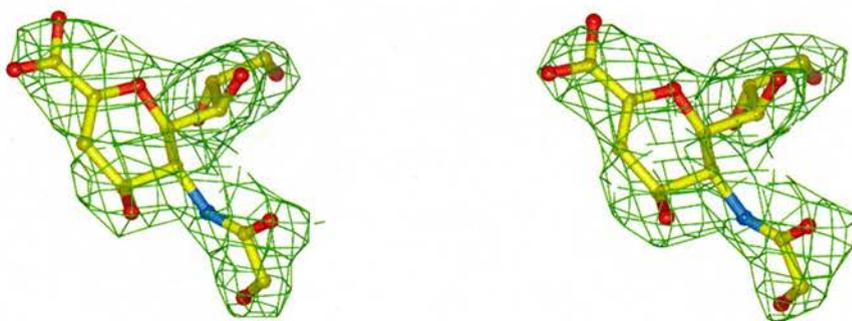
**Table 5.10.** List of interactions between VCNA and the bound FNANA



Hydrogen bonding			Non-bonding interaction*		
NANA-atom	VCNA-atom	Distance (Å)	NANA-atom	VCNA-atom	Distance (Å)
O1A	Arg74-N <sup>n1</sup>	2.92	F11A	Tyr111-C <sup>ζ2</sup>	3.20
O1A	Ser198-N	3.07	F11B	Met107-C <sup>ε</sup>	3.27
O1B	Ser198-O <sup>γ</sup>	3.13	F11B	Thr109-C <sup>γ2</sup>	3.49
O4	Gln188-N <sup>ε2</sup>	3.51	F11C	Leu120-C <sup>δ2</sup>	3.31
O4	W↔Gln76-O <sup>ε1</sup>	2.71			
O4	W↔Gln76-N <sup>ε2</sup>	2.80			
	and Tyr111-O <sup>n1</sup>				
N5	Gly196-O	2.82			
O7	W↔Arg118-N <sup>n2</sup>	2.66			
O8	Asn199-N <sup>δ2</sup>	2.89			
O10	Arg118-N <sup>n1</sup>	3.18			
O10	Asn318-O <sup>δ1</sup>	3.03			
F11A	Arg118-N <sup>n1</sup>	2.97			
F11A	Tyr111-O <sup>n1</sup>	2.96			

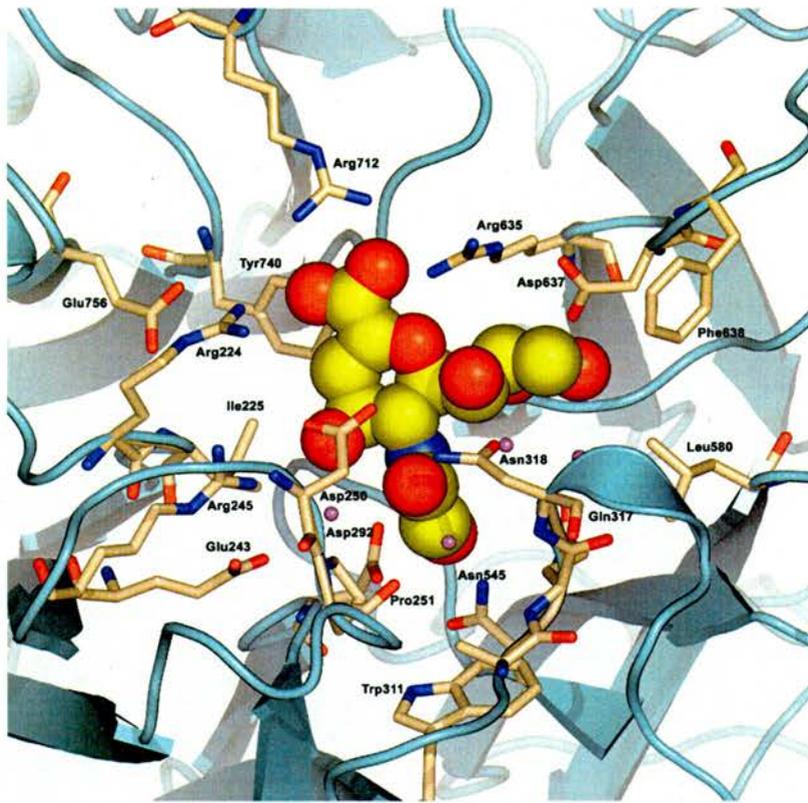
### 5.6.2 VCNA-GANA complex

The structure of VCNA in complex with GANA was solved to 2.6 Å resolution. The electron density maps ( $F_o - F_c$  and  $F_o - F_c$ ), figure 5.25, clearly revealed the presence of the ligand bound at the catalytic groove. Expectedly, no evidence for ligand binding at the N-terminal lectin domain could be observed; this is distinct from the DANA- and FANA-complexes where sialic acids ligand were observed at the sialic acid-binding site. As presented previously (section 5.2.2.3), this distinction could be ascribed to the different behaviour of the crystallizing agent during crystallization.

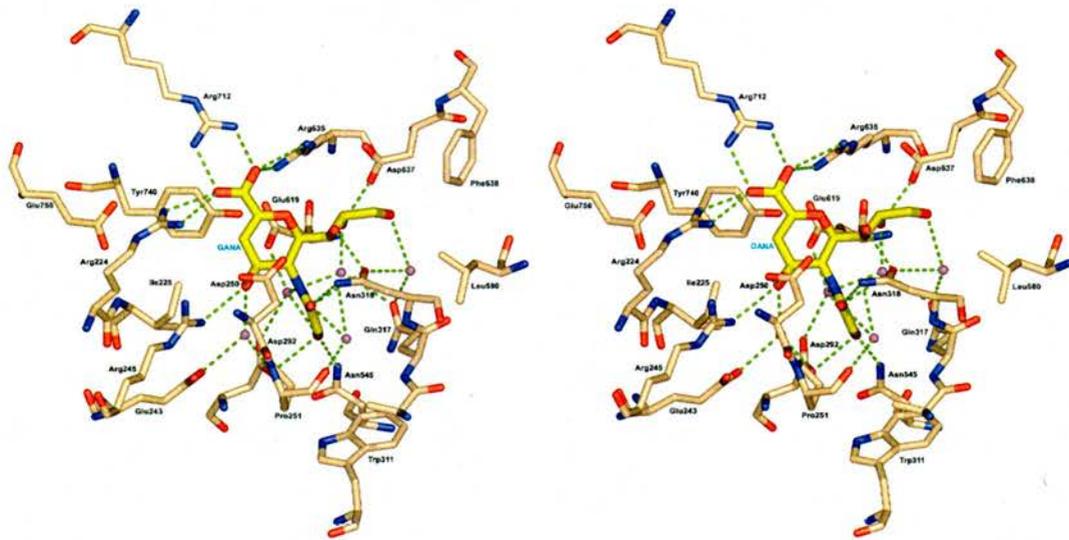


**Figure 5.25.** The electron density maps for the GANA-complex:  $2F_o - F_c$  contoured at  $1.2\sigma$  level (left) and the  $F_o - F_c$  at  $2.5\sigma$  level (right) for the ligand GANA at the catalytic site. The maps were calculated before adding the ligand to the refined model structure of the complex.

As displayed in figure 5.26, the ligand binds at the catalytic site in the same way as the parent ligand DANA; the only difference between the two ligands is around the hydroxyl group of the N-glycolyl at C5 in GANA that replaced the N-acetyl group of DANA. Table 5.11 lists the protein-ligand contacts at the catalytic site. The hydroxyl of the N-glycolyl is oriented anti to the carbonyl O-atom of the same group pointing down to the floor of the binding pocket. This hydroxyl group forms two H-bonds with Asp292 and Asn545, and a water-mediated interaction with Glu619. The moderate interaction of the

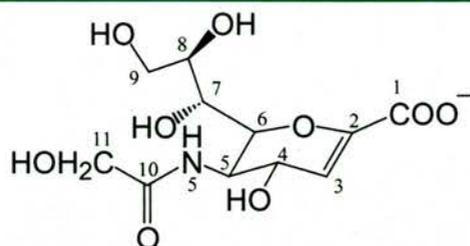


a



b

**Figure 5.26.** (a) Ribbon representation of the active site of VCNA with the inhibitor, GANA, bound (drawn in CPK model), the active site residues are labelled and drawn as sticks. (b) Stereo view for the active site, the hydrogen bonding interactions drawn as dotted lines; water molecules are represented as magenta spheres.

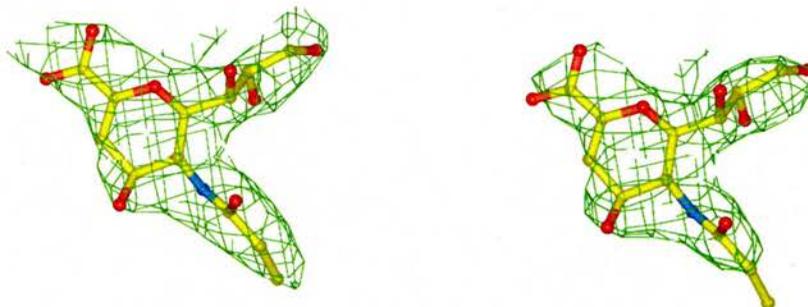
**Table 5.11.** List of interactions between VCNA and the bound GANA

Hydrogen bonding			Non-bonding interaction		
GANA-atom	VCNA-atom	Distance (Å)	GANA-atom	VCNA-atom	Distance (Å)
O1A	Arg224-N <sup>η1</sup>	2.95	C1	Tyr740-O <sup>η</sup>	3.20
O1A	Arg224-N <sup>η2</sup>	2.84	C2	Tyr740-O <sup>η</sup>	3.02
O1A	Arg712-N <sup>η1</sup>	2.71	C4	Ile225-C <sup>δ</sup>	4.17
O1B	Arg712-N <sup>η2</sup>	2.73	C9	Phe638-C <sup>ζ</sup>	4.24
O1B	Arg635-N <sup>η1</sup>	3.45	C9	Leu580-C <sup>δ1</sup>	4.64
O1B	Arg635-N <sup>η2</sup>	3.48	C9	Asn637-C <sup>β</sup>	4.22
O4	Arg245-N <sup>η2</sup>	3.10	C11	Trp311-C <sup>ζ2</sup>	4.84
O4	Asp250-O <sup>δ1</sup>	2.71	C11	Pro251-C <sup>γ</sup>	3.90
O7	Asn318-O <sup>δ1</sup>	2.90	C11	Gln317-C <sup>β</sup>	4.36
O8	Asp637-O <sup>δ2</sup>	2.90			
O10	Asn318-N <sup>δ2</sup>	3.20			
O10	Asn318-O <sup>δ1</sup>	3.30			
O11	Asp292-O <sup>δ1</sup>	3.33			
O11	Asn545-N <sup>δ2</sup>	3.29			

N-glycolyl hydroxyl, provides an explanation for the modest enhanced potency of inhibition of GANA relative to DANA, Nohle et al. reported  $IC_{50}$  values against VCNA of 10  $\mu$ M for GANA, and 15  $\mu$ M for DANA; (Nohle et al., 1985).

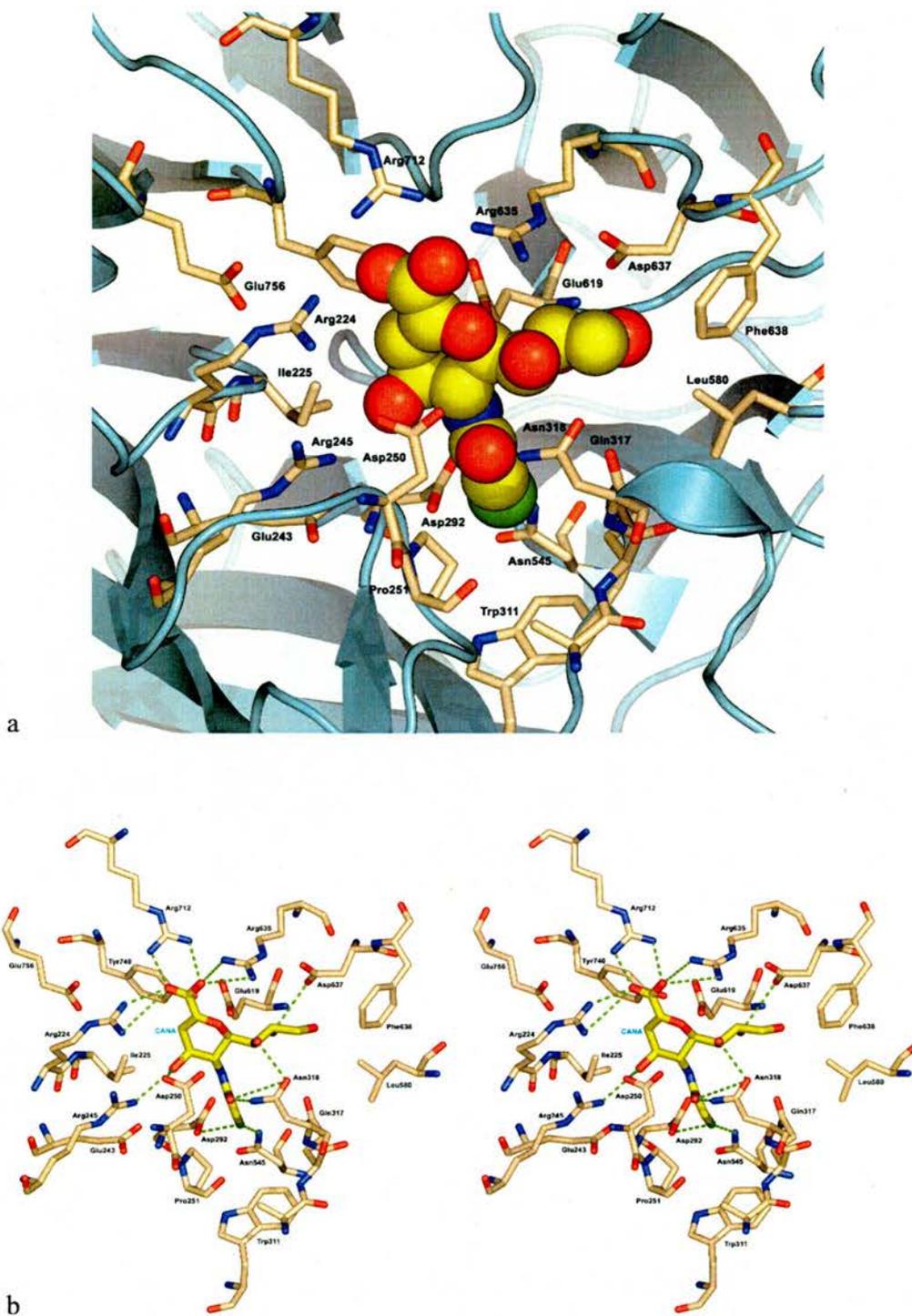
### 5.6.3 VCNA-CANA complex

The complex structure of VCNA with the inhibitor CANA was solved to 2.8 Å. Figure 5.27 shows the electron density maps ( $2F_o-F_c$  and  $F_o-F_c$ ), which clearly demonstrates the bound CANA at the catalytic site, despite the modest quality of the collected data on this complex, see section 4.3.2.2.

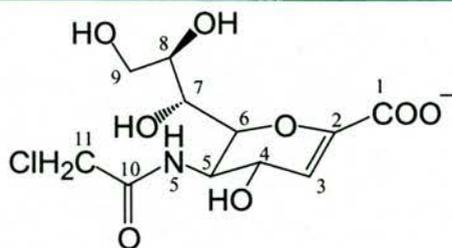


**Figure 5.27.** The electron density maps for the CANA-complex:  $2F_o-F_c$  contoured at  $1.2\sigma$  level (left) and the  $F_o-F_c$  at  $2.5\sigma$  level (right) for the ligand CANA at the catalytic site. The Maps were calculated before adding the ligand to the refined model structure of the complex.

Similar to the previous two complexes of FANA and GANA, the inhibitor (CANA) binds at the catalytic site nearly in the same mode as the parent inhibitor DANA and undergoes the same interactions except for the interactions of the chlorine atom (Cl11) of the N-chloroacetyl group at C5, see figure 5.28. Table 5.12 lists the interaction of CANA with the active site residues. The sugar ring is in a slightly different orientation from that observed in the previous complexes, as the carboxylate group does not fall in the same plane of the guanidine group of Arg712. This little difference in the orientation is indicated in the interatomic distances in table 5.12 compared to the other complexes. This could be ascribed to the large size of the chlorine atoms or the lower quality of the diffraction data. The chlorine atom mediates two H-bonds with Asp292 and Asp545, and participates in a hydrophobic interaction with Trp311. The structural data is in agreement with the reported inhibition data, which demonstrated a little enhancement for the potency of CANA relative to DANA against VCNA; however the efficiency of inhibition for CANA is less than that reported for GANA and FANA (Wilson et al., 2000). This can be accounted for by the less number of interactions of CANA compared to the other two inhibitors.



**Figure 5.28.** (a) Ribbon representation of the active site of VCNA with the inhibitor, CANA, bound (drawn in CPK model), the active site residues are labelled and drawn as sticks. (b) Stereo view for the active site, the hydrogen bonding interactions drawn as dotted lines; water molecules are represented as magenta spheres.

**Table 5.12.** List of interactions between VCNA and the bound CANA

Hydrogen bonding			Non-bonding interaction		
CANA-atom	VCNA-atom	Distance (Å)	CANA-atom	VCNA-atom	Distance (Å)
O1A	Arg224-N <sup>η1</sup>	3.82	C1	Tyr740-O <sup>η</sup>	3.50
O1A	Arg224-N <sup>η2</sup>	2.74	C2	Tyr740-O <sup>η</sup>	3.34
O1A	Arg712-N <sup>η1</sup>	2.41	C4	Ile225-C <sup>δ</sup>	4.11
O1B	Arg712-N <sup>η2</sup>	2.94	C9	Phe638-C <sup>ζ</sup>	4.48
O1B	Arg635-N <sup>η1</sup>	4.23	C9	Leu580-C <sup>δ1</sup>	4.68
O1B	Arg635-N <sup>η2</sup>	3.88	C9	Asn637-C <sup>β</sup>	4.25
O4	Arg245-N <sup>η2</sup>	2.94	C11	Pro251-C <sup>γ</sup>	3.67
O4	Asp250-O <sup>δ1</sup>	2.93	C11	Gln317-C <sup>β</sup>	4.51
O7	Asn318-O <sup>δ1</sup>	2.85	C111	Trp311-C <sup>ζ2</sup>	4.05
O8	Asp637-O <sup>δ2</sup>	3.07			
O10	Asn318-N <sup>δ2</sup>	2.63			
C111	Asp292-O <sup>δ1</sup>	3.07			
C111	Asn545-N <sup>δ2</sup>	2.77			

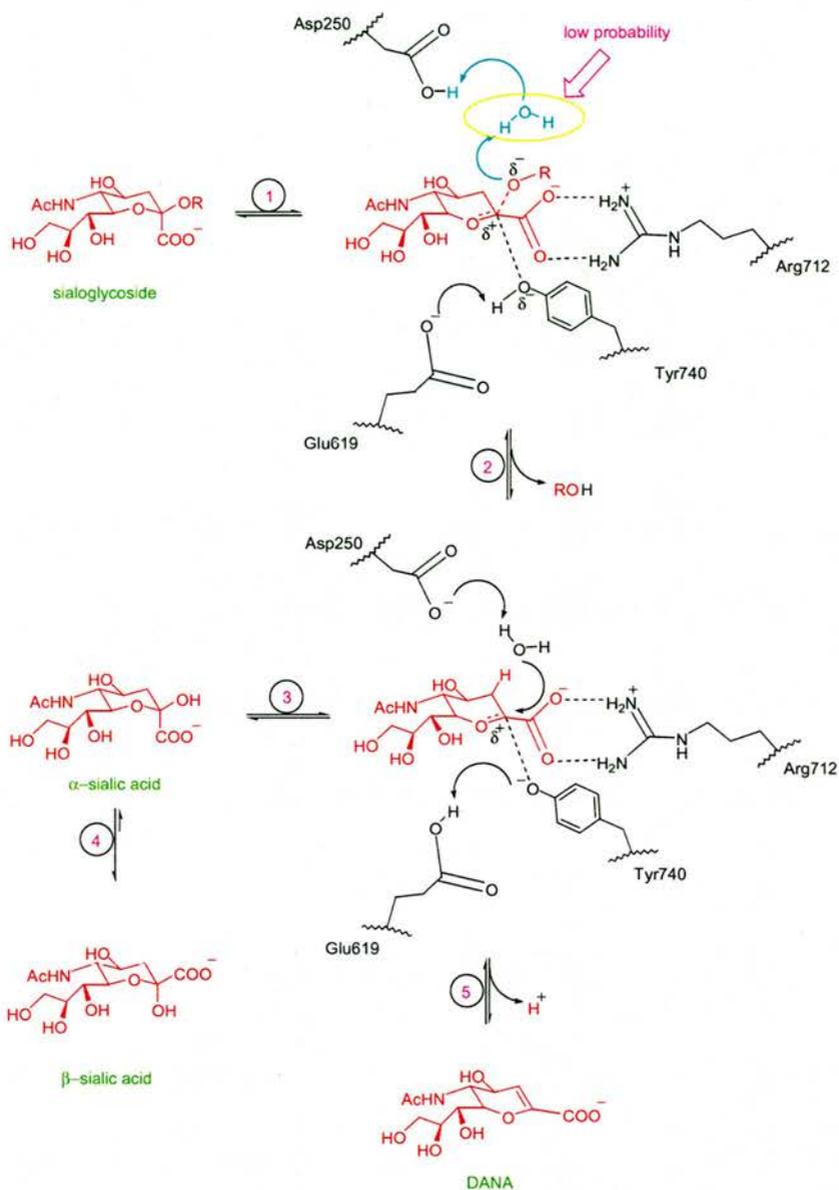
## 5.7 Implications of structures for the mechanism

The cleavage of  $\alpha$ -ketosidically linked sialic acid by VCNA has been shown to proceed with full retention of configuration (Guo & Sinnott, 1993). Also, NMR studies in our lab (section 8.4.1) and by others (Friebolin et al., 1984; Friebolin et al., 1981; Wilson et al., 1999) demonstrated the release of  $\alpha$ -sialic acid as the first product of the enzymatic hydrolysis, which subsequently mutarotates to the more stable  $\beta$ -sialic acid. The mechanism of this catalysed cleavage by neuraminidases has been studied in most detail for influenza virus enzyme (Chong et al., 1992; Janakiraman et al., 1994). There are two characteristic features for the proposed mechanism: (1) the sugar ring of  $\alpha$ -sialic acid undergoes a considerable distortion from its chair conformation upon binding and adopts a planar arrangement for the atoms O6-C2-C3; the distortion brings the carboxylate group to be equatorial to the sugar ring, stabilized by the nearby arginine triad around the carboxylate (Varghese et al., 1992). (2) The ring distortion induces the formation of the sialosyl cation with a developed positive charge on the reaction centre at C2 in the transition state; the conserved tyrosine and glutamate below the sugar ring at the catalytic site of neuraminidases are implicated in stabilizing the formed positive charge (Chong et al., 1992). These two features are widely accepted for all neuraminidases.

However, Sinnott and co-workers (Guo & Sinnott, 1993) in their kinetic-isotope-effect study of the catalysis by VCNA demonstrated the unusual low charge development at the reaction centre compared to the influenza enzyme. Additionally, they showed that there is little or no proton donation to the leaving group in the transition state, in which the sugar ring is largely flattened to perhaps the half-chair conformation with a largely cleaved glycosidic C–O bond. These outlined features in conjunction with the data from the complex structures of VCNA presented in this chapter can be used to depict the

following picture for the mechanistic pathway of the cholera enzyme as shown in figure 5.29.

The sugar ring of the sialic acid is flattened to adopt a distorted half-chair conformation (the reactive form of the bound sialoglycoconjugate), very close to the conformation of DANA. The energy cost of the distortion is compensated by the favourable interaction of the substituents on the ring in the adopted half-chair conformation, especially the strong electrostatic interaction of the carboxylate group with the arginine triad. Moreover, the environment below the anomeric centre, where the bulky Tyr740 and the negatively charged Glu619, forces the carboxylate group to be equatorial to the sugar ring enhancing the ring flattening step. This ring distortion, very likely, is the driving force for the cleavage of the glycosidic bond accompanied by the formation of the sialosyl cation; in other words there is no need to protonate the glycosidic oxygen before the cleavage step. Indeed, the aspartate (Asp250) sitting above the sugar ring is further away from the anomeric centre to be the proton donor to the glycosidic oxygen. Nonetheless, there is a little chance that under acidic conditions of low pH, Asp250 might mediate proton-transfer via a bridging water molecule. It should be mentioned that this enzyme works at conditions of alkaline pH (~8-8.5) in the intestine. Evidently, as shown in the VCNA-complexes structures, the two residues (Tyr740 and Glu619) play a significant role in neutralizing the developed positive charge in the formed oxocarbenium ion in the transition state. In a subsequent step, a water molecule receiving help from the deprotonated Asp250 above the sugar ring perhaps acts as the nucleophile to attack the anomeric centre and produce  $\alpha$ -sialic acid as the first product. The gained energy from bonding with the nucleophilic water might be used to bring the carboxylate group to be axial to the sugar ring and consequently exiled from the active site due to the steric clash with Tyr740 and the repulsive electrostatic interaction with Glu619.



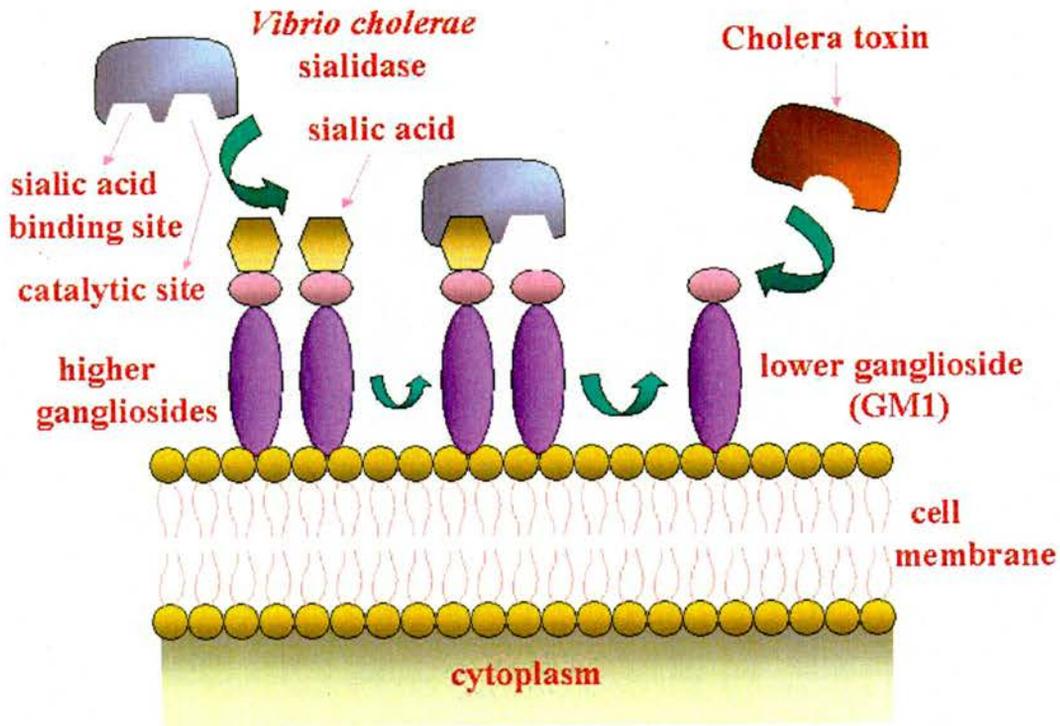
**Figure 5.29.** Outlines of the mechanism of hydrolysis by VCNA. In reaction 1, the substrate changes its conformation from chair to half-chair driving the bond cleavage of the aglycon; the developed oxocarbenium ion is stabilized by residues Tyr740 and Glu619. The role of water to protonate the glycosidic oxygen is unnecessary. In reaction 2, the sialosyl cation in the transition state is attacked by a water molecule, the general base, enhanced by residue Asp250. Reaction 3, represents the product formation of  $\alpha$ -sialic acid which mutarotates to the more stable  $\beta$ -form in reaction 4. The sialosyl cation in the transition state can proceed through reaction 5, with a little chance, and eliminate a  $H^+$  to form DANA.

Another point to mention is that during the reaction course the sialosyl cation might proceed through a different pathway to eliminate a proton from C3 to form DANA as demonstrated by the crystal structure of VCNA-3'-sialyllactose and further confirmed by NMR. In fact, the actual step used to produce DANA is unclear.

### **5.8 Model of the VCNA action: the multifunctional enzyme**

Since disclosing the structure of VCNA, the two lectin-domains flanking the central  $\beta$ -propeller domain have been suggested to play additional role(s) beside the neuraminidase function of the enzyme (Crennell et al., 1994). These lectin domains have been predicted to be involved in sugar recognition promoting adhesion to epithelial cells while the enzyme carries out its catalytic function, as it is known that *V. cholerae* secretes several adhesins allowing it to colonize the severe environment of the small intestine (Finkelstein et al., 1992). The secreted VCNA, as part of the mucinase complex, plays a role in the degradation of the mucin layer to allow the organism to reach the intestinal microvilli (Stewart-Tull et al., 1986). At the host cell surface, VCNA prunes the higher gangliosides by removing the terminal sialic acid producing GM1, the receptor of the B-subunit of cholera toxin (Galen et al., 1992). Additionally, VCNA has been demonstrated to remain attached to the membranes of cells treated with neuraminidase, even after several washings (Sedlacek & Seiler, 1974).

The finding of the sialic acid-binding site at the N-terminal lectin domain described herein, unequivocally, established the biological function associated with this lectin domain. It is conceivable that the lectin domain acts as a hook to attach the cholera enzyme to the cell surface of the host, thereby facilitating the access of the catalytic domain to its receptors presented on the cell surface. Consequently, this results in increa-



**Figure 5.30.** Diagram illustrates the dual functions of VCNA: upon approaching the cell surface, the enzyme attaches itself to the ganglioside receptor using the sialic acid-binding site and catalytically cleaves the terminal sialic moiety from a neighbour ganglioside to produce the lower ganglioside, GM1, the receptor of the CT. The enzyme remained attached to the cell surface to do its action on other gangliosides diffusing through the cell membrane.

-sing the effective concentration of the available substrate nearby the catalytic domain of the enzyme to carry out its function. The outlined interplay between the catalytic and the binding domains allowing the cholera enzyme to elicit its biological function is depicted diagrammatically above in figure 5.30.

It should be mentioned that the multi-functionality of neuraminidases has been demonstrated before for the viral enzyme HN-NDV (Crennell et al., 2000). However, the HN enzyme possesses a sialic acid-binding site, separate from the catalytic site, on the surface of the dimeric protein that performs the binding function rather than having a

whole domain to carry out the binding (Zaitsev et al., 2004). On the other hand, the influenza virus has a separate membrane protein (haemagglutinin, HA) (Weis et al., 1988), for sialic acid binding in addition to the influenza neuraminidase carrying out the catalytic function; both proteins are presented on the viral surface facilitating its function.

## 5.9 Discussion

In this chapter we presented the main results of this thesis, and described the solved structures of the cholera enzyme in complex with various inhibitors and substrates. In these various complexes, we have observed two distinct patterns of crystal packing; as a result of this crystal packing effect the rmsd among the diverse complexes showed large deviations. Analysing the intermolecular contacts within the crystals enabled us to reason the observed variations in the crystal packing. Interestingly, this observed phenomenon demonstrated the minor but subtle effect that could be introduced by adding the ligand to the crystallizing system. Slightly different ways for growing crystals lead to drastic changes in the packing within crystals.

The crystallographic structures provided the detailed protein-ligand interactions at the submolecular level between VCNA and its counterparts. Indeed, the yielded information forms a good base for designing more potent blockade for the cholera enzyme hoping to find a cure for the cholera disease. Moreover, the crystallographic structures unmasked the presence of a sialic-acid binding site at the N-terminal domain and solved its puzzling function. Disclosing the function of this lectin domain provided explanations for the results obtained from earlier as well as recent works by other laboratories. In fact, the established new function depicted a deeper view for how the cholera enzyme carries out its *in vivo* function.

Meanwhile, the structure of VCNA in complex with 3'-sialyllactose demonstrated the ability of the cholera enzyme to synthesize its own inhibitor as shown previously for the neuraminidases from influenza virus and *T. rangeli*. Not only that, but the cholera enzyme could act on the double bond of the transition-state-analogues, as demonstrated for DANA and FANA, to produce the corresponding  $\alpha$ -sialic acids.

In addition, the 3D-structures of VCNA with DANA and the three DANA-based inhibitors: FANA, GANA, and CANA enabled us to understand, at the atomic level, the observed trends of the inhibiting power of these compounds. The crystallographic structures suggested that FANA, the best cholera inhibitor, could be the core to synthesize other derivatives with variations at the glycerol side chain since there is enough room facing this group. We anticipated that replacing the glycerol group by a longer side chain (e.g., butyl group) might induce favourable interactions with the active site residues.

Not only do the crystallographic structures have implications for the structure-based drug design but also for understanding the mechanistic action of VCNA. The 3D structures enabled us to depict a picture for the hydrolysis mechanism of this enzyme based on the kinetic and crystallographic data; consequently, a better view and deeper insight into the enzymatic action of VCNA was obtained.

## Chapter 6

### Comparison of VCNA with other neuraminidases and lectins recognizing sialic acids

#### 6.1 Introduction

As shown in chapter 5, solving the complex structure of VCNA with the inhibitor (DANA) delineated the detailed protein-ligand interactions between the bound ligand and the catalytic site within the central  $\beta$ -propeller domain. In addition, this complex structure unmasked a second sialic acid-binding site at the N-terminal lectin domain of VCNA, which promotes the attachment of the cholera enzyme to the surface of the hosting cells. The finding of the second site has been confirmed subsequently in the complex structures of VCNA with the substrates  $\alpha$ -2,3/6-sialyllactose.

In this chapter, we compare the recognition pattern of the bound DANA in VCNA with other neuraminidases/trans-neuraminidases from bacterial and non-bacterial sources, whose structures are determined to date. Interestingly, the active sites of the neuraminidases used in the comparison study have been identified crystallographically through complexes with DANA. It is important to mention that DANA is not a good inhibitor for all neuraminidases. Comparing the interactions of DANA with the different neuraminidases helps to clarify the differential inhibition ability of DANA for the various enzymes and deepens our understanding for the structural determinants required for designing potent inhibitors for the investigated enzymes.

Beside that, herein we compare the recognition pattern of sialic acid by the lectin domain in VCNA with other specific sialic acid-binding lectins. The comparison with

other lectins provides a clear view of how important is the sialic acid binding for the cholera neuraminidase.

## 6.2 Comparison of VCNA-DANA binding with other neuraminidases

The following 3D structures of the different neuraminidase complexes used in the comparison were extracted from the PDB database (Berman et al., 2000): *Salmonella typhimurium* [pdb code 2SIM, (Crennell et al., 1996)], *Micromonospora viridifaciens* [1EUS, (Gaskell et al., 1995)], *Trypanosoma rangeli* [1N1T, (Amaya et al., 2003)], influenza virus [1NSD, (Burmeister et al., 1993)], Newcastle disease virus haemagglutinin-neuraminidase, or HN-NDV, [1E8V,(Crennell et al., 2000)], *Macrobodella decora* [1SLI, (Luo et al., 1998)], *Trypanosoma cruzi* [1MS1, (Buschiazzo et al., 2002)].

It is worthy to note that the active sites of the above neuraminidases were identified crystallographically through complexes with the inhibitor DANA, which inhibits most neuraminidases with a  $K_i$  in the micromolar range (Taylor, 1996). Nevertheless, it is wrong to assume that DANA is a potent inhibitor for all neuraminidases, as it inhibits *S. typhimurium* neuraminidase, or STNA, with a  $K_i$  of 380  $\mu\text{M}$  (Hoyer et al., 1991), and 140  $\mu\text{M}$  for *T. rangeli* neuraminidase, or TRNA, (Amaya et al., 2003). Indeed no significant inhibition was observed for the trans-neuraminidases (Taylor, 1996).

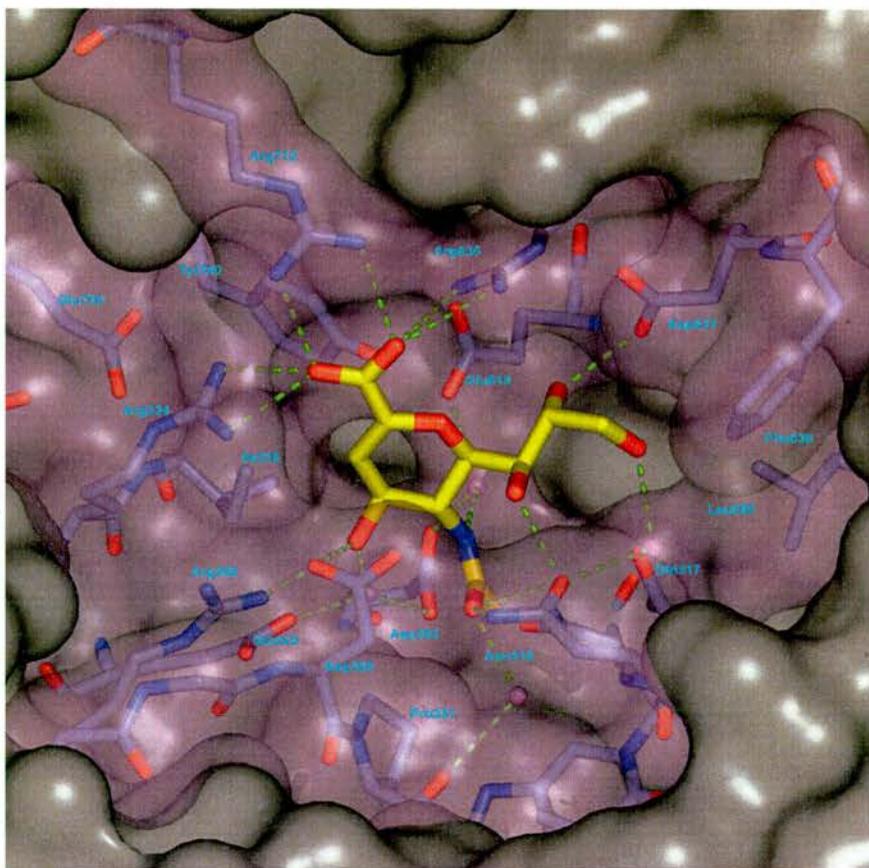
Despite the low sequence identity between VCNA and other neuraminidases (16-24% identity) (Crennell et al., 1994), the topology of the catalytic domain is conserved and the active sites share many features (Taylor, 1996). Such a similarity is manifested in the binding mode of DANA within the catalytic site of the various neuraminidases. The catalytic cavities within the canonical  $\beta$ -propeller domain, containing the bound DANA, of the investigated neuraminidases/trans-neuraminidases are displayed in the figures 6.1

to 6.8; the figures elucidate the detailed interactions of the ligand with the important residues within the catalytic clefts. Dissecting the different substituents on the sugar ring of DANA (2-carboxylate, O4, 5-N-acetyl, and 6-glycerol side chain), and analysing the surrounding environment of the various functional groups within the active sites highlighted distinct features of the ligand recognition by these enzymes. Table 6.1 lists the protein residues involved in the interactions with the various functional groups of DANA in the enzymes being investigated. The detailed analyses of the interactions of the four substituents on the sugar ring of DANA are described in the following paragraphs.

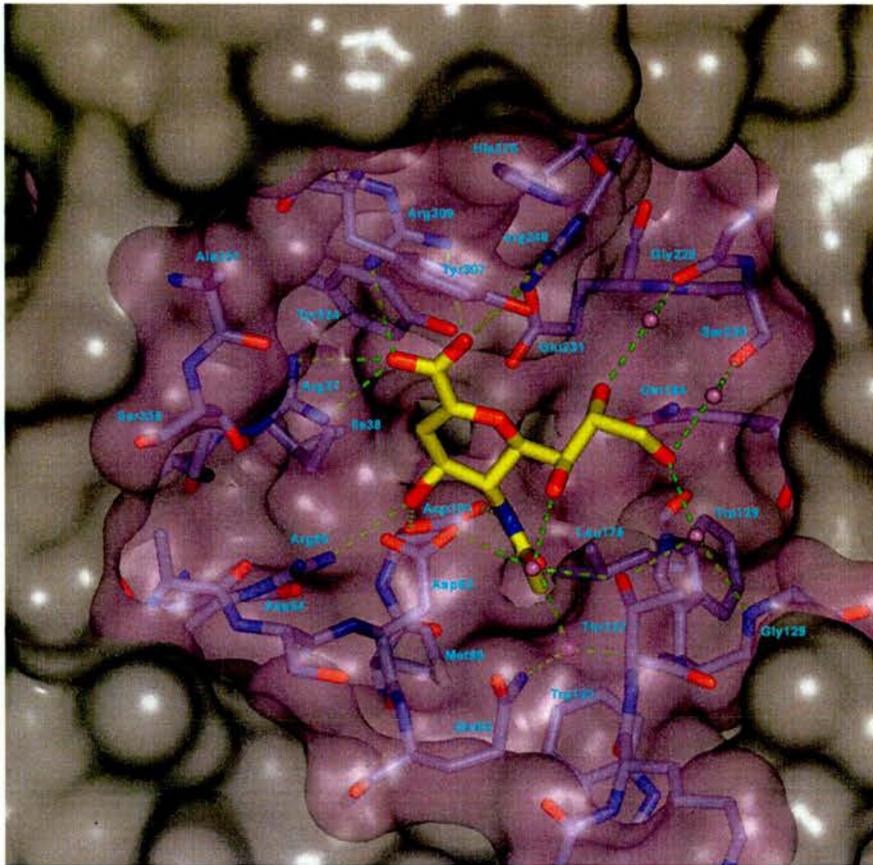
### **6.2.1 Interactions of carboxylate group**

As shown in the figures 6.1-8, the invariant arginine triad (224, 635, 712) in the vicinity of the carboxylate group, the numbering refers to VCNA unless stated otherwise, provides the positive electrostatic potential to accommodate the negatively charged group. This feature with the architecture of the catalytic cleft force a similar mode of binding of the inhibitor in all these neuraminidases; a conserved Glu756 stabilizes the first arginine. In fact this electrostatic interaction between the negatively charged group on C2 of DANA and the arginine triad has been proved to be highly significant. As shown for *S. typhimurium* (Crennell et al., 1996), substituting the carboxylate group of DANA by a phosphonate group revealed a similar binding mode to DANA for two oxygens of the phosphonate group with the arginine triad.

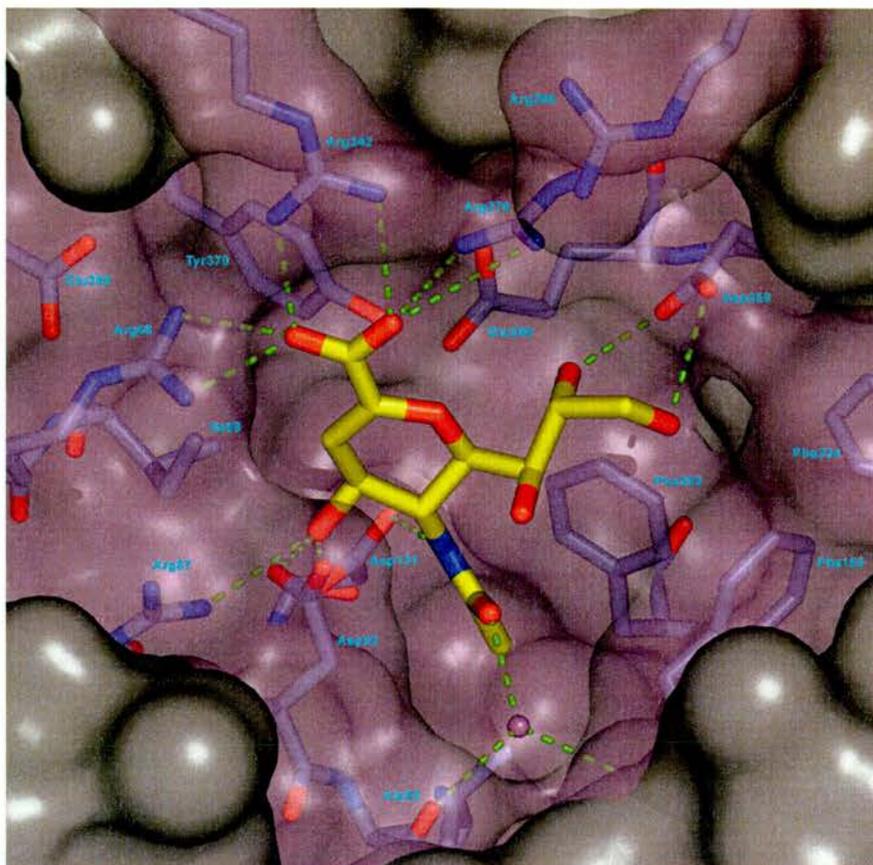
Also conserved is a short contact between the hydroxyl group of Tyr740 and C1 & C2 of the sugar ring; this tyrosine and the nearby residue Glu619, common in all neuraminidases, are implicated in the catalytic mechanism (Burmeister et al., 1993).



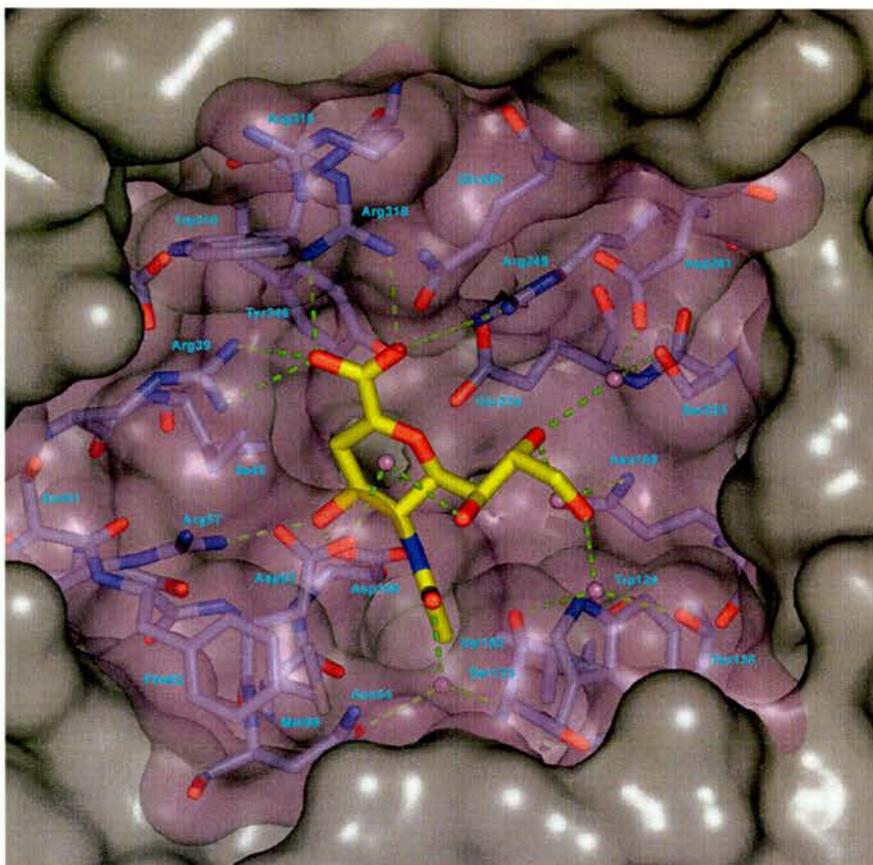
**Figure 6.1.** Catalytic cleft of VCNA with DANA bound: The interacting residues at the active site are drawn as capped-sticks model and seen through a transparent surface coloured in magenta for the displayed residues; the opaque grey surface represents the boundaries from the non-displayed protein residues. The ligand model is represented with thicker sticks and coloured atoms [yellow (C), red (O) and blue (N)]. The protein/ligand interactions are depicted as dotted green lines. The spheres coloured in magenta represent the ordered water molecules within the active site.



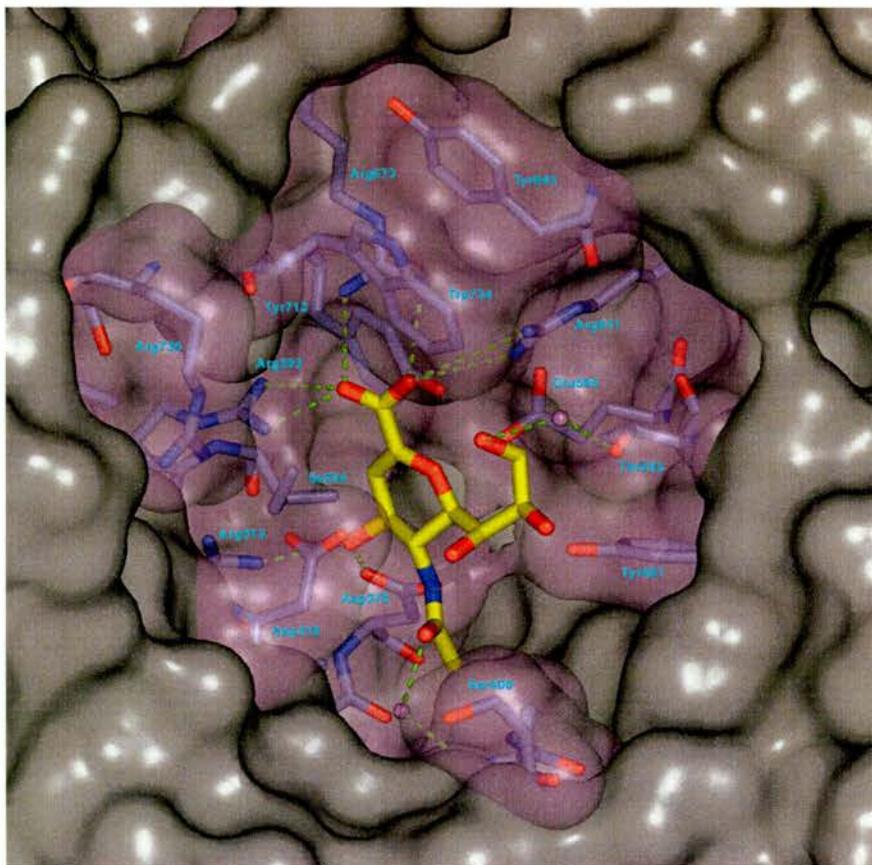
**Figure 6.2.** Catalytic cavity of the bacterial neuraminidase from *S. typhimurium* (pdb code, 2SIM) represented in the same way shown in figure 6.1. As can be observed in the figure, there is a large cavity facing the glycerol side chain of the ligand, consequently, direct interactions of this group with the enzyme are absent; the same has been observed for the *T. rangeli* neuraminidase (figure 6.4). Note the bulky Tyr307 at the opening of the cleft, which might hinder the ligand binding.



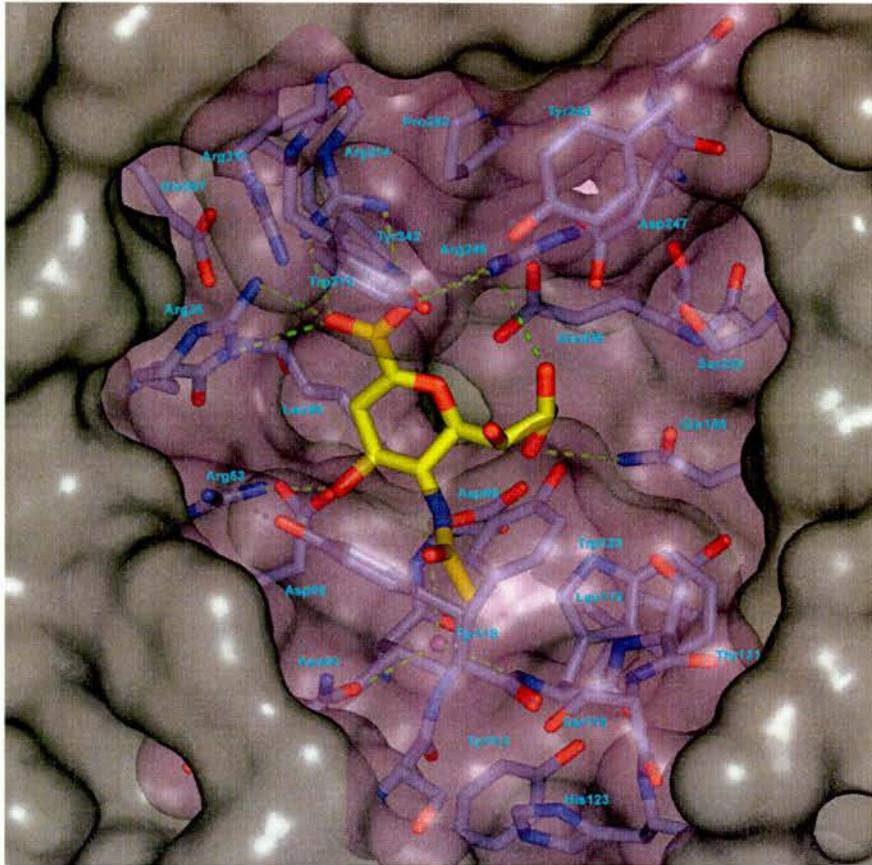
**Figure 6.3.** Catalytic cavity of the bacterial neuraminidase from *M. viridifaciens* (pdb code, 1EUS) represented in the same way described in figure 6.1. Note the direct interaction of O4 with the Asp92 sitting above the sugar ring, this interaction has been observed in VCNA but not in other neuraminidases/trans-neuraminidases.



**Figure 6.4.** Catalytic cavity of the *T. rangeli* neuraminidase (pdb code, 1N1T) shown in the same way as described in figure 6.1. Note the large cavity facing the glycerol group, no direct interactions involving the trihydroxyl group with the protein residues similar to *S. typhimurium* neuraminidase (figure 6.2). Also, note the bulky Trp316 at the entrance of the cleft, which as in STNA might be the cause for the low affinity towards DANA.

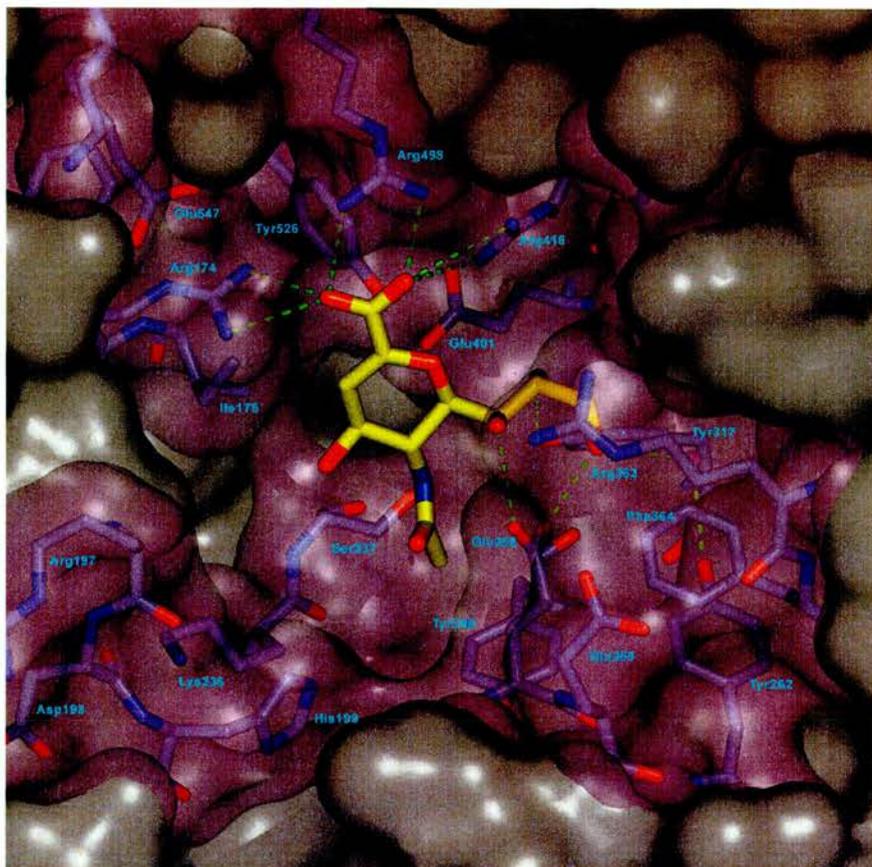


**Figure 6.5.** Catalytic cavity of the trans-neuraminidase from *M. decora* (pdb code, 1SLI) displayed in the same way as in figure 6.1. Note the conformations of the glycerol side chain, which is quite different from the other conformations of the same group observed for the bound DANA in the other enzymes, suggesting a different mechanism for the hydrolysis reaction as mentioned in the text. Also, the bulky Trp734 above the carboxylate group acts as an obstacle for the ligand binding.



**Figure 6.6.** Catalytic cavity of the trans-neuraminidase from *T. rangeli* (pdb code, 1MS1). The catalytic groove is very deep compared to the other neuraminidases. Similar to the trans-neuraminidase from *M. decora*, the bulky Trp312 sitting at the opening of the cleft might act as an obstacle for the ligand binding.





**Figure 6.8.** Catalytic cavity of the viral neuraminidase HN from NDV (pdb code, 1E8V). The large space around O4 is characteristic of this neuraminidase; also the strong interaction of the glycerol group at C6 is a key determinant for ligand binding to this neuraminidase. Note the position of Asp198 pointing away from above the sugar ring, the dynamic loop carrying this acidic residue is assumed to move bringing the acidic residue to sit above the sugar ring similar to the equivalent aspartate in all neuraminidases.

**Table 6.1.** List of residues involved in the interaction of DANA within the catalytic sites of the different neuraminidases.

Protein (pdb <sup>a</sup> )	Carboxylate	O4	NHAc	glycerol
VCNA	Arg-triad (224, 635, 712), Tyr740, Glu619, Ile225, Glu756	Arg245, Glu243, Asp292, Asp250	Asp292, Asn317, Asn318, Trp311	Gln317, Asn318, Leu580, Asp637, Phe638
2SIM	Arg-triad (37, 246, 309), Tyr342, Glu231, Ile38, Glu361	Arg56, Asp100	Asp100, Met99, Trp121, Trp128, Leu175	Gln63, Ser230, Thr127, Trp128
1EUS	Arg-triad (68, 276, 342), Tyr370, Glu260, Ile69, Glu386	Arg87, Asp131, Asp92	Asp131, Val148, Phe155, Phe203	Phe155, Phe234, Asp259
1N1T	Arg-triad (39, 249, 318), Tyr346, Glu234, Glu361, Ile40	Arg57, Asp100, Asp63	Asp100, Met99, Phe117, Trp124, Val180	Trp124, Thr125, Ser233
1SLI	Arg-triad (293, 611, 673), Tyr713, Glu595, Ile294, Glu729	Arg312, Asp375	Asp375, Ile374, Ser400, Val538	Tyr561, Thr593
1MSI	Arg-triad (35, 245, 314),	Arg53	Asp96, Val95, Trp120,	Gln195, Trp120, Arg245

	Tyr342, Glu230, Ile36, Glu375	Leu176	
INSD	Arg-triad (115, 291, 373), Tyr408, Glu275, Glu116, Glu427	Arg222, Glu225, Trp176, Ser177	Arg222, Ala244, Glu274, Asn293
IE8V	Arg-triad (174, 416, 498), Tyr526, Glu401, Ile175, Glu547	Ser237, Tyr299, Tyr317	Glu258, Tyr262, Tyr317, Phe364

<sup>a</sup> Full names and references of the enzymes with the given abbreviations and pdb codes: VCNA, *Vibrio cholerae* neuraminidase, *S. typhimurium* neuraminidase [pdb code, 2SIM, (Crennell et al., 1996)], *M. viridifaciens* neuraminidase [1EUS, (Gaskell et al., 1995)], *T. rangeli* neuraminidase [1N1T, (Amaya et al., 2003)], *M. decora* intramolecular trans-neuraminidase [1SLI, (Luo et al., 1998)], *T. cruzi* trans-neuraminidase [1MS1, (Buschiazzo et al., 2002)], influenza virus neuraminidase [1NSD, (Burmeister et al., 1993)], New castle disease virus haemagglutinin-neuraminidase, or HN-NDV, [1E8V, (Crennell et al., 2000)].

Another important residue is Ile225 that is conserved in all neuraminidases but not in influenza NA, which has Glu116 instead. Both the isoleucine and the aliphatic part of the glutamate participate in hydrophobic interactions with the sugar ring as well as with the aromatic ring of the catalytic tyrosine.

Beside these constant environment, in neuraminidases from: *S. Typhimurium* (Crennell et al., 1996) and *T. rangeli* (Amaya et al., 2003) as well as in trans-neuraminidases from *M. decora* (Luo et al., 1998) and *T. cruzi* (Buschiazzo et al., 2002) there is a bulky residue, either tyrosine or tryptophan, sitting at the entrance of the catalytic cleft on the same side facing the carboxylate group. Perhaps such a bulky residue acts as an obstacle for the binding of the inhibitor (or substrate). This could in part be responsible for the low binding affinity of DANA towards *S. Typhimurium* and *T. rangeli* neuraminidases and the non-significant affinity for the two trans-neuraminidases.

### 6.2.2 Interactions of O4

The environment at the vicinity of the hydroxyl group at C4 is less conserved compared to that of the carboxylate group. Such variations modulate the binding strength of the inhibitor with the active site residues near to 4-position. In the non-viral neuraminidases and trans-neuraminidases, the O4 interacts with a conserved arginine (Arg245) and forms either direct or water-mediated interactions with a conserved aspartic acid residue (Asp292). Noticeably, and characteristic for VCNA, the O4 interacts with Glu243; this interaction is absent in the other non-viral enzymes as a result of replacing the glutamic by aspartic residue with a shorter side chain. In the case of viral neuraminidases, the O4 is facing a cavity and has no direct interactions with the protein residues. Indeed, in the HN-NDV the O4 is devoid of any interaction as displayed in

figure 6.8 and a large cavity accommodates O4. Interestingly, this large cavity is assumed to be necessary to allow Arg174 to swing in a switching mechanism, the alteration between catalytic and binding functions of the same single site (Crennell et al., 2000). Note that, His133 in influenza and Ile192 in HN-NDV are in equivalent positions to Arg245. Worthy to note that in the influenza enzyme the side chain of Arg153 sits not far from the position of Arg245 yet further away to mediate an interaction with O4. It should be mentioned that in the case of influenza, the cavity around O4 has been successfully exploited for the design of novel inhibitors with bulkier substituents at the C4 (von Itzstein et al., 1993); note that in VCNA such substitutions reduced the inhibiting power of the designed ligands (Holzer et al., 1993; Schreiner et al., 1991a,b). Whilst describing the environment around the 4-position, it is important to mention the solvent-exposed aspartic acid (Asp250) sitting above the sugar ring from the side of O4 yet apart from interacting with O4 in all of the cases but VCNA and *M. viridifaciens* neuraminidase (figure 6.1 and 6.3), (Gaskell et al., 1995). This acidic residue is conserved in these neuraminidases; however as shown in figure 6.8, in HN-NDV, the equivalent residue (Asp198) is pointing away from the ligand, nevertheless this aspartate is hypothesized to undergo a back-and-forth movement between this position and the position above the sugar ring similar to other neuraminidases (Crennell et al., 2000). This acidic aspartic residue presumably acts as a proton donor to the glycosidic oxygen, but only at low pH, or could be involved in stabilizing a proton-donating water molecule participating in the hydrolysis reaction (Burmeister et al., 1993; Chong et al., 1992), as discussed in section 5.9 for the hydrolysis mechanism in VCNA.

### 6.2.3 Interactions of N-acetyl group

The N-acetyl moiety of DANA participates in a H-bonding interaction through its amide N- and O-atoms, and in a hydrophobic interaction using its methyl group. As mentioned before (section 5.2.1), in VCNA the carbonyl oxygen atom is engaged in direct H-bonds with Asn318, beside that it participates in water-mediated interaction with the same residue and the backbone O-atom of Pro251. In the other enzymes, but not viral ones, this direct interaction with Asn318 is absent. Instead, water mediated-interactions involving the carbonyl O-atom and asparagine (1N1T, 1MS1) or glutamine (2SIM) residues sitting on the other side of the carbonyl group relative to the location of Asn318 are observed. In the two enzymes from *M. viridifaciens* and *M. decora*, the carbonyl group interacts via a bridging water with the backbone O-atom of Ser319 (1SLI) and Ala93 (1EUS). The situation is different in the case of the viral neuraminidases; similar to VCNA the influenza enzyme possesses a direct H-bond between the protein (Arg149 in influenza) and the carbonyl O-atom. Whereas, in HN-NDV there is no interaction involving the carbonyl oxygen mediated with the protein. Regarding the N-atom joining the acetyl moiety to the sugar ring, a water-mediated H-bonding interaction is observed between the N-atom and Asp292 in VCNA; the corresponding interaction in the other non-viral enzymes is direct rather than water-mediated. This water-mediated interaction is also present in the influenza, which has Glu225 sitting below the sugar ring in equivalent position to Asp292 of VCNA, and in HN-NDV with residues: Ser237, Tyr317, and Glu401 (Crennell et al., 2000) [this is according to the authors as the extracted pdb file, 1E8V, does not include any water molecules below the sugar ring near the listed residues that could mediate such interaction].

In addition to the H-bonding, the methyl group of the N-acetyl moiety contributes to the non-bonding interactions with hydrophobic residues facing the alkyl group. These

hydrophobic residues are mainly tryptophan, phenylalanine and/or tyrosine with varying distributions around the interacting group of the ligand across the different neuraminidases. More or less, the pattern of interaction looks similar among all these enzymes with the lining hydrophobic residues occupying positions equivalent to Trp311 and Asn317 of VCNA. However, the hydrophobic pocket is less extensive in HN-NDV, perhaps to allow the hypothesized movement of residues in the vicinity of the N-acetyl group during the proposed functional-switching-mechanism for HN-NDV.

It is evident from dissecting the contacts of the N-acetyl group of DANA that the sum of interactions of its constituents imparts similar contribution to the ligand affinity towards the different enzymes apart from HN-NDV.

#### **6.2.4 Interactions of glycerol group**

Variation of the environment close to the glycerol side chain of DANA among the different neuraminidases and trans-neuraminidases is more pronounced than that surrounding the other substituents at C2, C4, and C5. Several features could be recognized from the contacts of the glycerol group in the different enzymes. In the viral enzyme from NDV, the interaction is extensive compared to the other viral and non-viral enzymes, where O7, O8, and O9 are engaged in direct interactions with the protein; this suggested that the glycerol side chain is a key determinant in the substrate/inhibitor recognition (Crennell et al., 2000). In VCNA only O7 and O8 have direct H-bonds with the protein, and additionally O8 & O9 mediate water-mediated interactions. The interaction in other enzymes, except the trans-neuraminidase from *M. decora*, is similar to VCNA. Nevertheless, in *S. typhimurium* and *T. rangeli* neuraminidases, (Crennell et al., 1996; Amaya et al., 2003), the interaction is a rather weak as the glycerol group faces a

large cavity and being involved only in water-mediated H-bonds. Noteworthy, in *M. decora* the glycerol side chain adopts an alternative conformation compared to the other enzymes. In this conformation the glycerol group has no interaction with the protein, and O7 is assumed to be the nucleophile, which attacks the carbocation centre of the transition state in a unique mechanism to release 2,7-anhydro-Neu5Ac instead of Neu5Ac (Luo et al., 1998). A rather common observation in all of these enzymes is the non-bonding interaction of the glycerol group with lining hydrophobic residues in the catalytic cavity.

All together, the above analysis of the DANA-enzyme interactions can be used to rationalize the differences of the inhibitor affinities towards the various enzymes. In the enzymes from: *V. cholera*, *M. viridifaciens*, influenza virus, and HN-NDV, where the DANA binds in a shallow catalytic cleft, the different substituents of DANA contributes to the protein-ligand interaction to various degrees. Nonetheless, the overall interactions of the different groups are more or less alike in these enzymes. This is reflected in the small variations of the inhibition constants of DANA to the four enzymes ( $K_i \approx 10^{-6}$  M). In contrast, in the trans-neuraminidases (*T. cruzi*, and *M. decora*), in spite of the observed interactions of the functional groups of DANA with the protein, no significant inhibition could be observed. This could be attributed to the binding of DANA in a deep pocket with bulky tryptophan at the opening of the catalytic cleft, which might hinder the ligand binding. Similarly, in neuraminidases from *T. rangeli* and *S. typhimurium* the inhibitor binds in a deep pocket with bulky residues at the inlet of the binding pocket; however, in the former enzyme the bulky tryptophan is shifted in space from the corresponding positions seen in trans-neuraminidases, and in the latter a tyrosine rather than tryptophan is sitting at the entrance of the catalytic groove, see figures 6.2, 6.4. This, in addition to the weak interaction of the glycerol group of DANA in the case of 2SIM, can account for

the observed weak inhibition of DANA for these two neuraminidases [ $K_i = 340$ , and  $140 \mu\text{M}$  for *S. typhimurium* (Hoyer et al., 1991), and *T. rangeli*, respectively, (Buschiazzo et al., 2002)].

### **6.3. Comparison with other sialic acid-binding lectins**

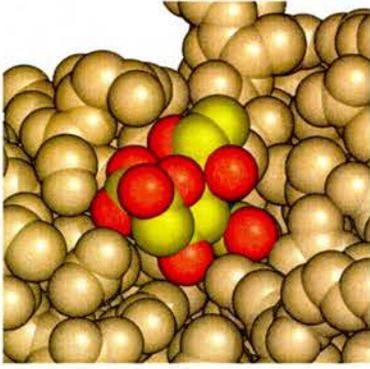
Lectins represent a large family of proteins or glycoproteins, which bind carbohydrates with great specificity and selectivity (Brook et al., 2002); they are involved in a variety of biological processes, such as cell-cell and host-pathogen interactions, serum glycoproteins turnover, and innate immune response (Lis & Sharon, 1998; Vijayan & Chandra, 1999). The terminal position of sialic acids at the end of oligosaccharide chains means that they are appropriately presented as convenient binding partners for lectins. Most lectins that bind sialic acid are selective as to the linkage of the sialic acid they recognize, and many also recognize sub-terminal monosaccharides (Brook et al., 2002). Many sialic acid-binding lectins from diverse sources, ranging from plants and animals to viruses and bacteria, have been described and for some of them the 3D structures are known. Herein, the pattern of the sialic acid binding in the N-terminal lectin domain of VCNA is compared to ten different structures of lectins in complex with sialic acid-containing substrates from various origins: bacterial, viral, animal, and plant lectins. The lectins with the following pdb entries were extracted from the PDB database (Berman et al., 2000): cholera toxin, or CT, [2CHB, (Merritt et al., 1997)]; pertussis toxin, or PT, [1PTO, (Stein et al., 1994)]; tetanus neurotoxin, or TeNT, [1DFQ, (Emsley et al., 2000)]; influenza HA [5HMG, (Weis et al., 1990)]; virus coat protein of murine polyomavirus [1SIE, (Stehle & Harrison, 1996)]; capsid spike protein of rotavirus [1KQR, (Dormitzer et al., 2002b)]; sialoadhesins [1QFO, (May et al., 1998)]; selectins [1G1R, (Somers et al.,

2000)]; wheat germ agglutinin [2CWG, (Wright, 1990)]; legume lectin, or MAL, [1DBN, (Imberty et al., 2000)]. Generally, the sialic acid ligand binds in its  $\alpha$ -conformation in the investigated structures; but in 1DFQ, the binding site has  $\beta$ -sialic acid; however, in that case sialic acid binding site is adjacent to the N-acetylgalactosamine (NGA) site, suggesting that sialic acid is not the primary carbohydrate recognized by the carbohydrate recognition domain (CRD) of the neurotoxin.

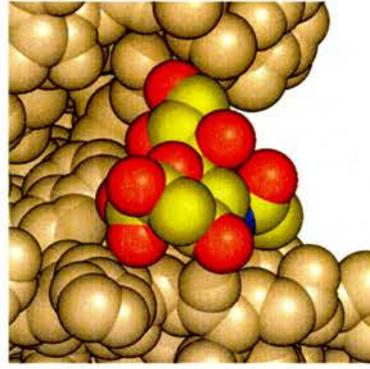
General investigation of these complex structures revealed that the way sialic acid fits into the site seem to fall into three classes: (1) in most structures including (VCNA, 2CHB, 1PTO, 1DFQ, 1QFO, 5HMG, and 1KQR), the sialic acid contacts the protein surface with one side of the ring and the other side of the ring is exposed to the solvent; (2) in the structures of (1SIE, 2CWG, and 1DBN), sialic acid inserts into a groove with residues on both sides of the groove facing the carbohydrate ring; (3) the third case, in the structure of (1G1R), sialic acid has few contacts with the protein through its peripheral side chains while the two faces of the carbohydrate ring are exposed to the solvent. The three different situations are illustrated in figure 6.9a-k, which displays the space-filling model for the sialic acid ligand bound to its receptor. Dissecting the interactions of the various parts of the sialic acid ligand recognized by the lectins is discussed in the following section.

### **6.3.1 Interactions of carboxylate group**

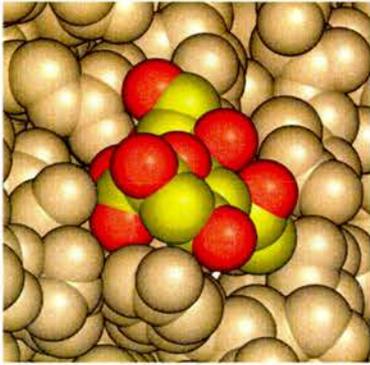
For comparison, figure 6.10a-k schematically represents the lectin $\leftrightarrow$ sialic acid-moiety interactions in the different lectin complexes as well as in VCNA. The residues interacting with the various functional groups of the sialic acid ligand in the different structures are



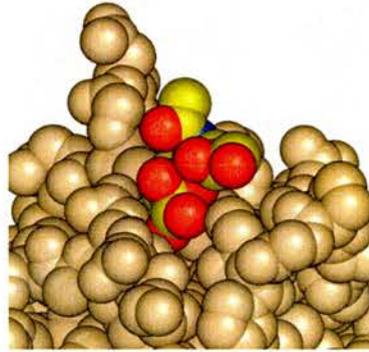
a. VCNA



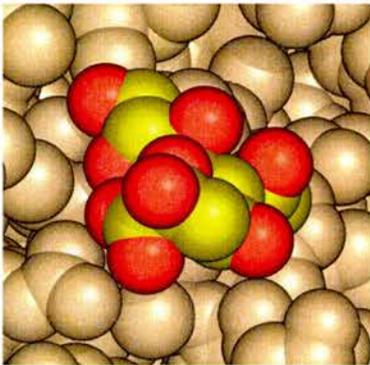
b. (2CHB, cholera toxin)



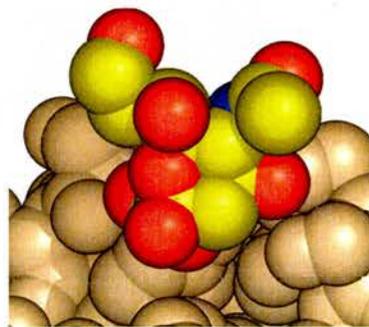
c. (1PTO, pertussis toxin)



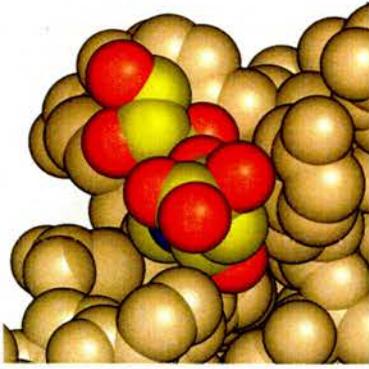
d. (1DFQ, tetanus neurotoxin)



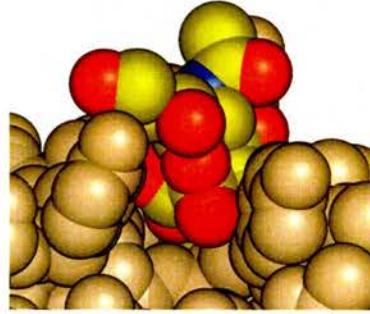
e. (1QFO, sialoadhesin)



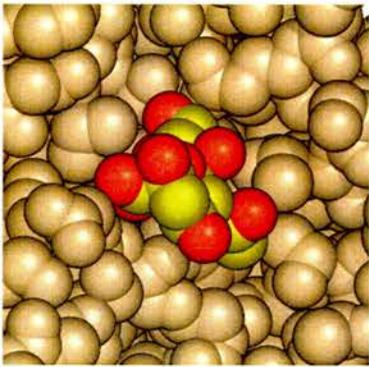
f. (1G1R, selectin)



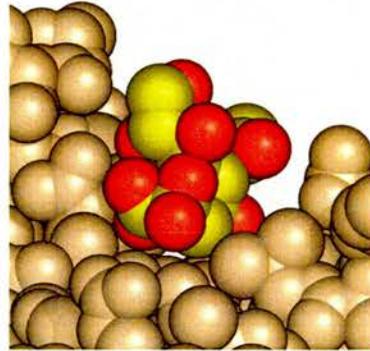
g. (2CWG, wheat germ agglutinin)



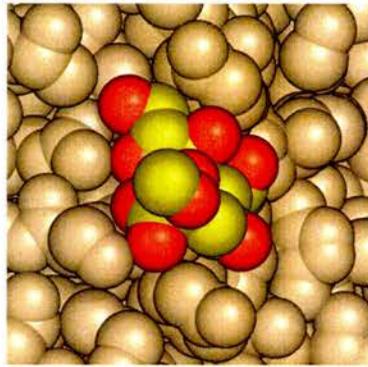
h. (1DBN, legume lectin)



i. (5HMG, influenza virus haemagglutinin)



j. (1SIE, VP1 polyomavirus)



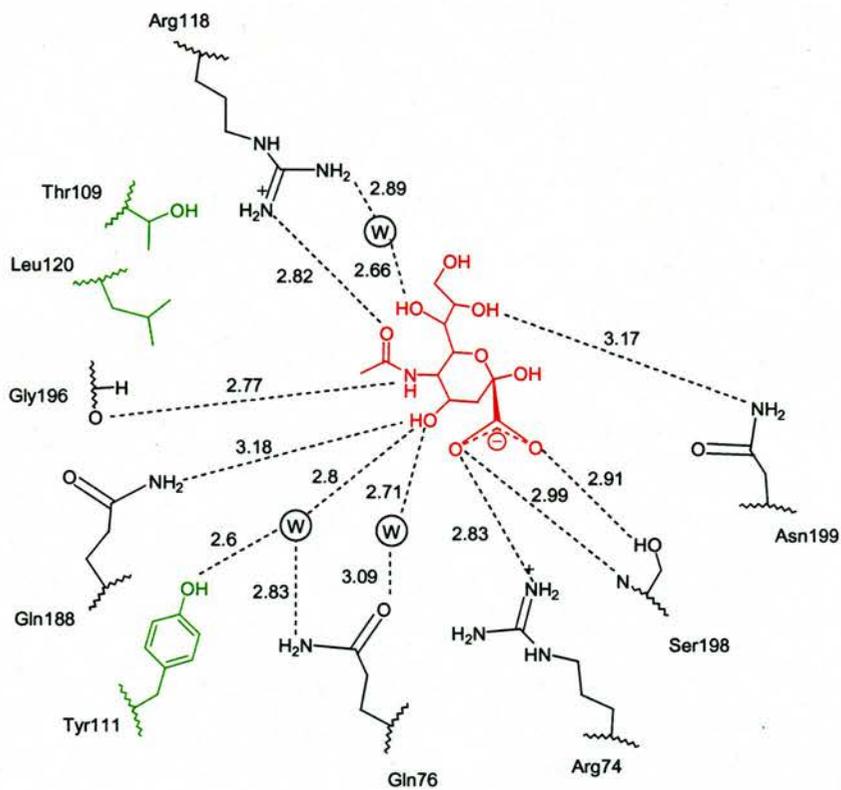
k. (1KQR, capsid spike rotavirus)

**Figure 6.9.** Space-filling-model of the sialic acid ligand bound at its receptors of the different lectins, pdb codes annotated the pictures from a to k. As mentioned in the text, in the investigated lectins, sialic acid bind in three modes: contacting the protein with one side of the ring as in (a, b, c, d, e, i and k), inserting into a groove as in (g, h, and j), or hanging into the solvent as in f.

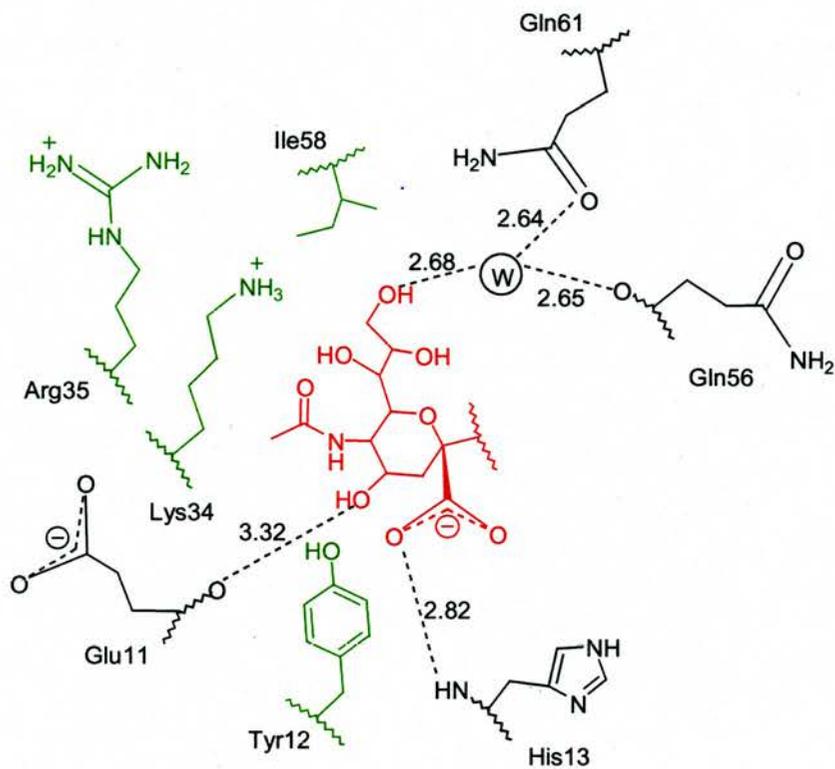
summarized in table 6.2. Although charge-charge interaction between the carboxylate and positively charged amino acids might be expected to be important to recognition, in most cases the carboxylate group interacts with the main-chain amide groups, polar side chains (especially serine), and structural waters rather than fully-charged side chains. In (VCNA, IPTO, and IDBN) only one O-atom of the carboxylate interacts with basic residues: arginine (one N-atom of the guanidine side chain), or lysine. Sialoadhesin (1QFO) is an exception where the charged group has strong interaction with Arg97.

Clearly, the extent of the interactions of the carboxylate group varies to different degrees among these lectins. Nonetheless, this recognition of the charged carboxylate seems to be the only persistent feature among these lectins as can be seen from table 6.2. Indeed, the constant participation of the carboxylate group in sialic acid recognition does make sense, since it is the carboxylate group that apparently discriminate between sialic acid and other carbohydrate sugars. In other words, the C2-carboxylate represents a recognition fingerprint of sialic acid by lectins.

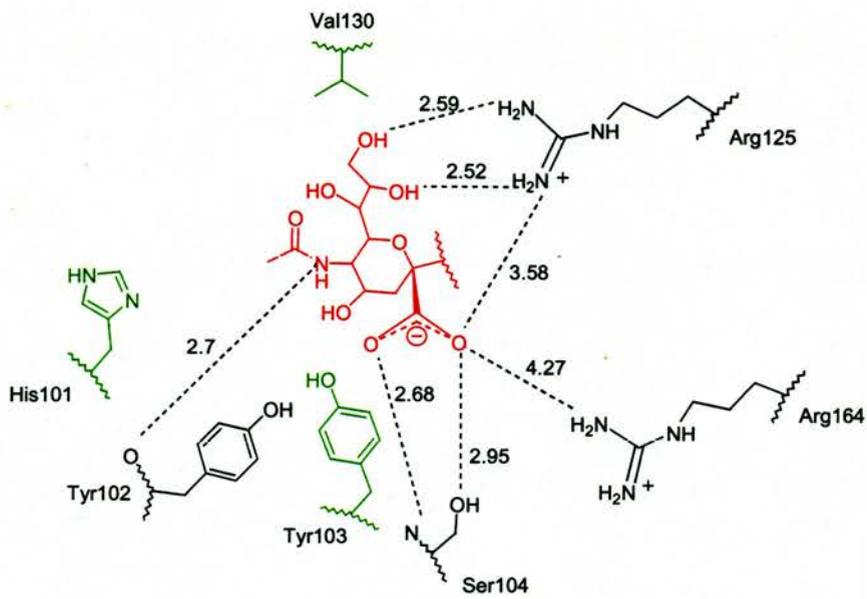
Apparently, the interaction of the negatively charged group in sialic acid-binding lectins is less strong than in neuraminidases (section 6.2.1). This is can be understood in the light of the different roles played by this interaction. In the case of lectins, the interaction with the carboxylate takes part in the recognition phenomenon, while in neuraminidases the strong interactions with the charged group is required to force the conformational change of the sugar ring in a step towards the hydrolytic cleavage of the terminal sialic acid.



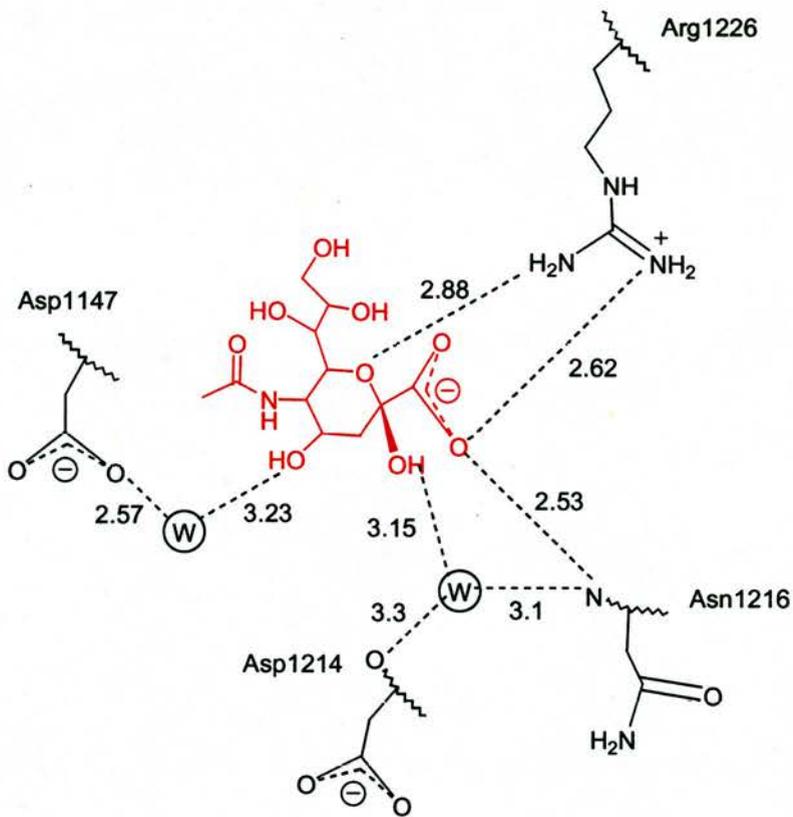
a (VCNA)



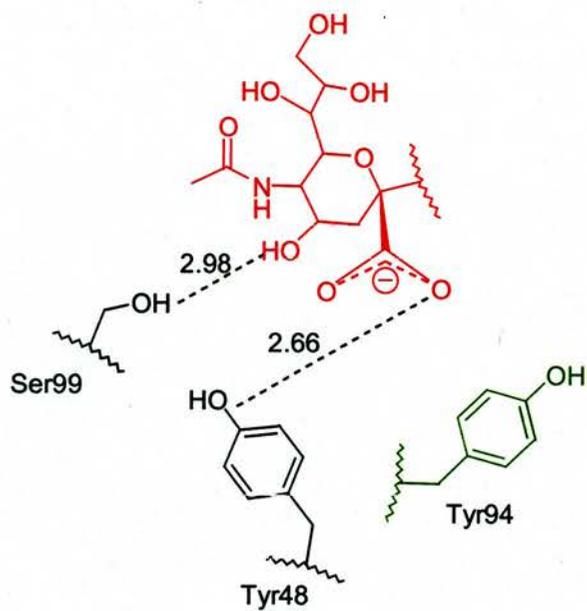
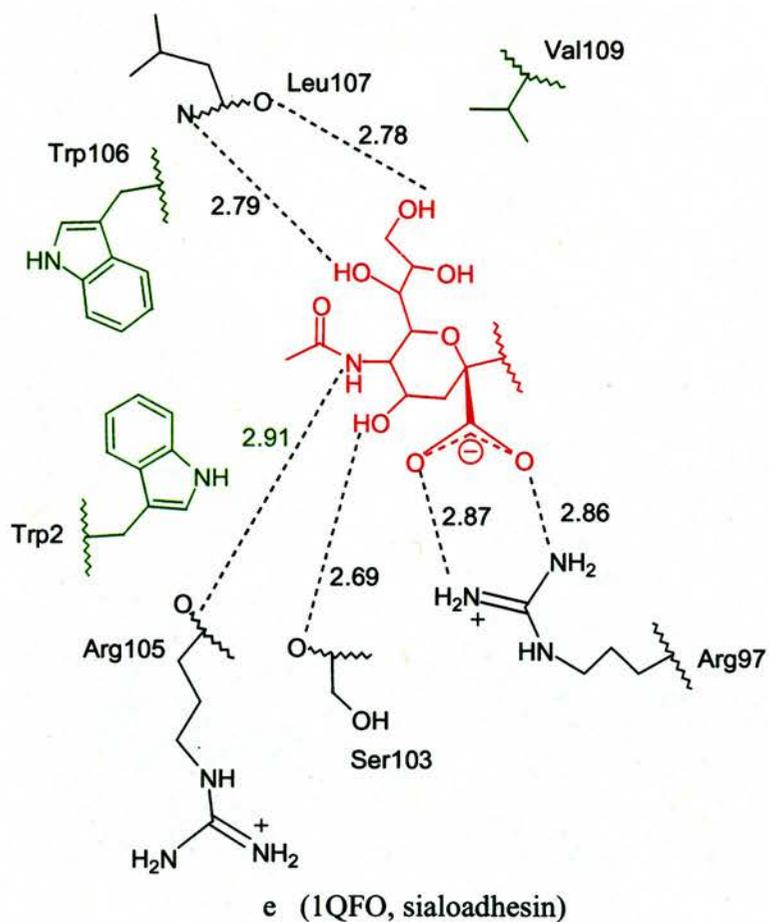
b (2CHB, cholera toxin)



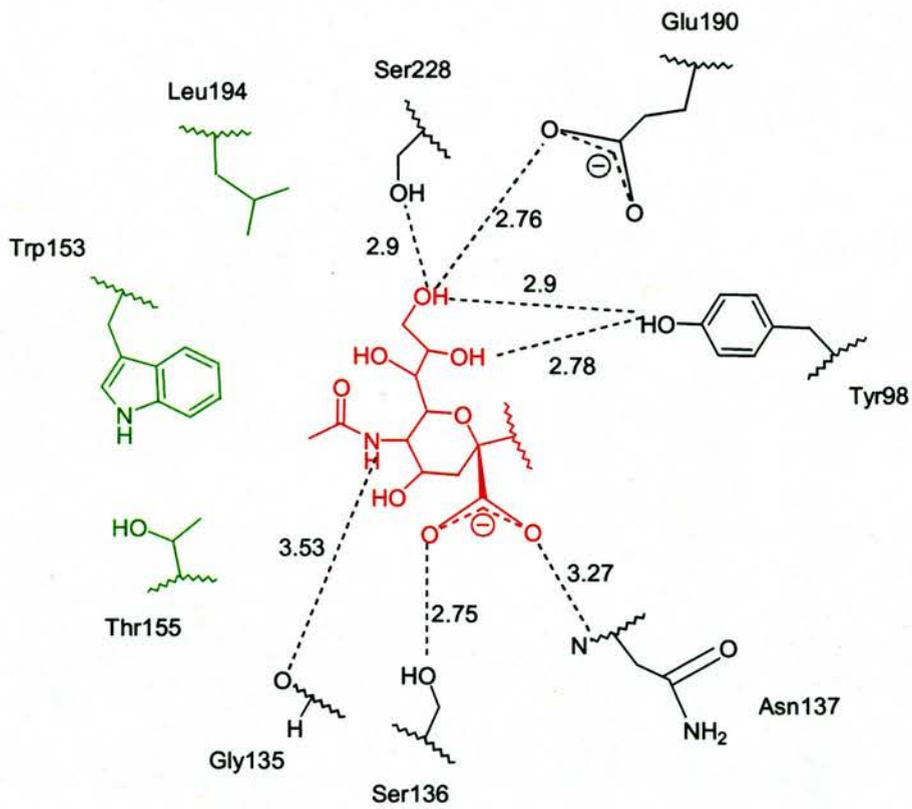
c (IPTO, pertussis toxin)



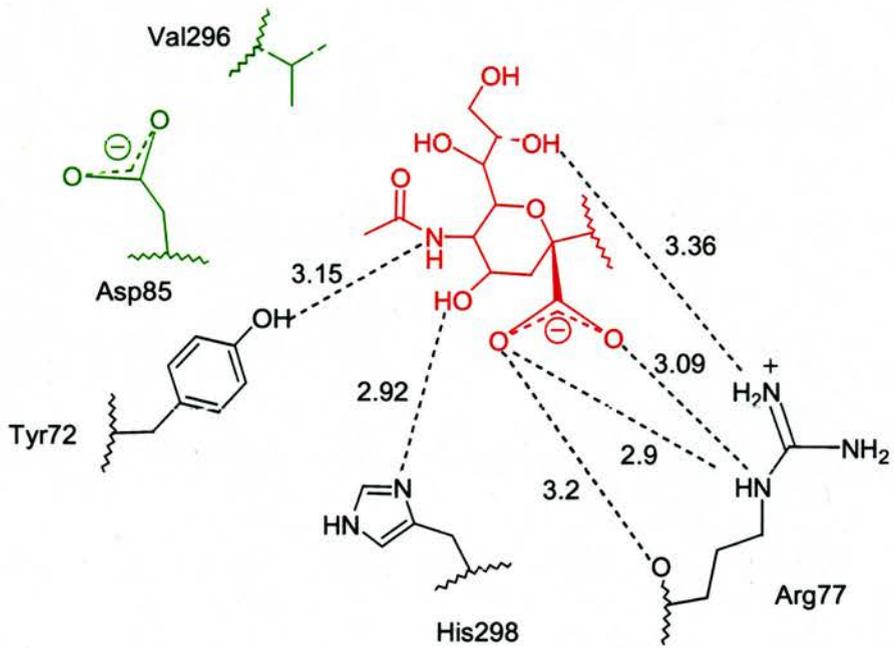
d (1DFQ, tetanus neurotoxin)



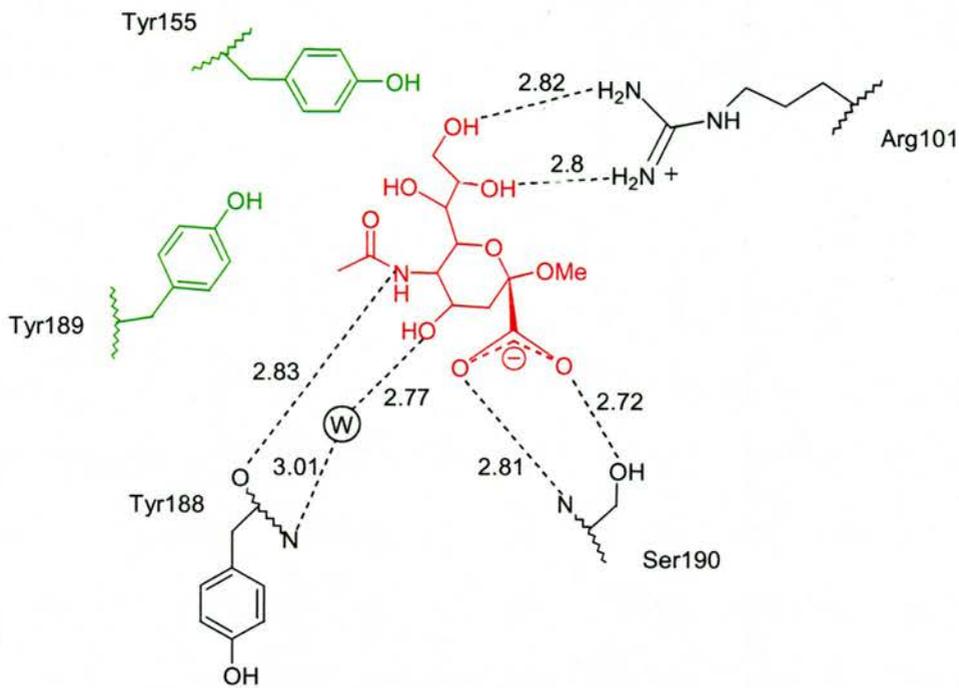




i (5HMG, influenza haemagglutinin)



j (1SIE, virus coat protein VP1)

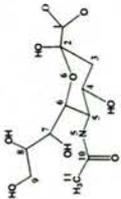


k (1KQR, capsid spike protein of rotavirus)

**Figure 6.2.** Schematic representations of the interactions of sialic acid in the sialic acid-binding lectins: (a)VCNA, *vibrio cholerae* neuraminidase; (b)2CHB, cholera toxin; (c)1PTO, pertussis toxin; (d)1DFQ, tetanus neurotoxin; (e)1QFO, sialoadhesin; (f)1G1R, selectins; (g)2CWG, wheat germ agglutinin; (h)1DBN, legume lectin (MAL); (i)5HMG, influenza HA; (j)1SIE, virus coat protein VP1 of polyomavirus; (k)1KQR, capsid spike protein of rotavirus. Sialic acid residue is drawn in red; residues forming hydrophobic interaction are coloured in green; residues interacting through H-bonds, depicted as dotted lines, are shown in black with the distances between atoms indicated; W represents bridging waters.

**Table 6.2.** Summary of the sialic acid interactions in VCNA and the sialic acid-binding lectins used in the comparison study

Protein <sup>a</sup> (PDB)	Ligand	Carboxylate	O2	O4	O6	NHAc	Glycerol	Ring
VCNA	Sialyllactose	Arg74, Ser198	-	Gln188, Tyr111, Gln76	-	Arg11, Gly196, Tyr111, Leu120, Thr109	Arg118, Asn199	-
2CHB	GM <sub>1</sub>	His13	-	Glu11	-	Glu11, Lys34, Arg35	Gln56, Gln61, Ile58	Tyr12
1PTO	Biantenary undecasaccharide	Ser104, Tyr102, Arg125	-	-	-	Tyr102, Tyr103, His101	Arg125, Val130	Tyr102
1DFQ	Sialic acid	Arg1226, Asn1216	Asn1216, Asp1214	Asp1147	Arg1226	-	-	-
1QFO	Sialyllactose	Arg97	-	Ser103	-	Arg105, Trp106, Trp2	Leu107	-
1GIR	Sialyl Lewis <sup>x</sup>	Tyr48	-	Ser99	-	-	-	-
2CWG	Sialoglyco- peptide	Thr6-D, Tyr7-D (from ligand)	-	Tyr73; Ser43, Ser114	-	Glu115, Ser62, Tyr73	Tyr64	Tyr66
1DBN	Sialyllactose	Lys107	-	-	-	Ser104	-	Tyr45
5HMG	Sialyllactose	Ser136, Asn137	-	-	-	Gly135	Tyr98, Glu190,	-



1SIE	Sialyllactose	Arg77	-	His298	-	Tyr72, Asp85	Ser228, Trp153, Leu194 Arg77	Val296
1KQR	2-OMe sialic	Ser190	-	-	-	Tyr188, Tyr189	Arg101, Tyr155	-

<sup>a</sup> Full names and references of the lectins for the given pdb codes: VCNA, *Vibrio cholerae* neuraminidase; cholera toxin, or CT, [2CHB, (Merritt et al., 1997)]; pertussis toxin, or PT, [1PTO, (Stein et al., 1994)]; tetanus neurotoxin, or TeNT, [1DFQ, (Emsley et al., 2000)]; sialoadhesins [1QFO, (May et al., 1998)]; selectins [1G1R, (Somers et al., 2000)]; wheat germ agglutinin [2CWG, (Wright, 1990)]; legume lectin, or MAL, [1DBN, (Imberty et al., 2000)]; influenza HA [5HMG, (Weis et al., 1990)]; virus coat protein of murine polyomavirus [1SIE, (Stehle & Harrison, 1996)]; capsid spike protein of rotavirus [1KQR, (Dormitzer et al., 2002b)].

### **6.3.2 Interactions of the anomeric hydroxyl and the ring oxygen**

The anomeric hydroxyl O2 has no interaction with the lectin proteins, except in 1DFQ. Note that in 1DFQ sialic acid binds in its  $\beta$ -conformation in contrast to the other lectins, probably sialic acid is not the true ligand for this particular lectin, especially it is next to the NGA binding site (Emsley et al., 2000). Similarly, the O6 of the ring does not take any part in the interactions of the investigated lectins except in 1DFQ; this seems to be an individual case rather than a common interaction pattern involved in sialic acid recognition.

### **6.3.3 Interactions of O4**

The role of O4 varies greatly in these lectins. Interactions of O4 with main-chain amide groups, polar side chains or ordered waters are observed in these lectins but not in (1PTO, 1DBN, and 5HMG). Apparently, the interaction of O4 is more pronounced in (2CWG), where it is essential for binding (Wright, 1990). Also, the interactions of O4 with Gln188 and via bridging waters with the protein residues have been described previously for VCNA (section 5.3.2).

### **6.3.4 Interactions of N-acetyl group**

The N-acetyl group is involved in the interaction in all the investigated lectins except in (1DFQ and 1G1R), in both of these cases sialic acid is not the chief recognizable carbohydrate (Emsley et al., 2000) (Somers et al., 2000). Evidently, the interaction is extensive in the case of VCNA and 2CWG (Wright, 1990), where in these two proteins the N-acetyl group is indispensable for recognition. The extent of the interaction of the N-

acetyl group varies among the different lectins. It should be noted that the N5 is engaged in H-bonding more frequently than the carbonyl oxygen of the group. The formed H-bond could be mediated with either main-chain amide groups or polar side chains; no ordered water molecules are caught up in these interactions. In addition, the methyl group participates in hydrophobic interactions, mainly with tyrosine, tryptophan, apolar residues, and the aliphatic part of residues with long side chain such as arginine and lysine.

### **6.3.5 Interactions of glycerol group and sugar ring**

The glycerol group plays a major role in sialic recognition in the lectins (1PTO, 1QFO, 5HMG, and 1KQR), where the hydroxyl groups are engaged in H-bonding interaction mainly with polar side chains; additionally, the group participates in hydrophobic interactions. On the contrary, in (1DFQ, 1G1R, and 1DBN) the trihydroxyl group has no interaction with the protein. For the rest of these lectins including VCNA the interaction with the glycerol group is less marked than in the first case. It is noted that for those lectins interacting with the glycerol group, none of them form simultaneous H-bonds with the three hydroxyls: O7, O8 and O9. In many cases O8 or/and O9 participate in H-bond formation, O7 is less frequently involved in the interaction.

Beside the interactions of the different substituents at the carbohydrate ring, in some lectins (2CHB, 1PTO, 2CWG, 1DBN, and 1SIE) the sugar ring has a non-polar interaction with residues, tyrosine or valine. The packing of the aromatic ring of tyrosine is frequently described as stacking; this interaction is most likely driven by the proximity of the aliphatic protons of the sugar ring, which carry a net positive charge, and the  $\pi$ -system of the aromatic ring (Weis & Drickmer, 1996).

It is evident from the above analysis that the protein-ligand interactions vary greatly among the studied lectins ranging from extensive interactions as seen in VCNA to few contacts with the protein as in 1G1R. Complementary to the qualitative analysis outlined above, the number of H-bonds (counted from figure 6.2), and the buried surface areas [calculated in GRASP (Nicholls et al., 1991)] were used as descriptors to characterize the lectin-sialic acid interactions, see table 6.3. The dissociation constants of the interactions of these lectins with sialic acid-containing substrates are also included. As can be seen from the table, the dissociation constants of the various lectins fall into one of three ranges: (1) high affinity with  $K_d$  below micromolar range including the cholera toxin (2CHB) and the pertussis toxin (1PTO), (2) moderate affinity with  $K_d$  in the micromolar range including VCNA, cholera toxin (2CHB) and the legume lectin 1DBN, and (3) low affinity with  $K_d$  in the millimolar range including the rest of the studied lectins. It is important to emphasize that the high affinity of the neuro- and pertussis toxins is not because of the strong interaction involving the sialic acid moiety of the substrates used to measure the affinity, but due to the fact that these toxins recognize additional sialic acid(s) in the gangliosides (GD1b, Gt1b, and GQ1), or other carbohydrates in the Feutin glycoproteins used in measuring the reported affinity for the neurotoxin and the pertussis toxin, in the same order (Krell et al., 2003). Indeed, both of the feutin and asialofeutin bind to the pertussis toxin with similar affinity. Of those lectins binding sialic acid substrates with micromolar affinity, the cholera toxin, a pentameric lectin, binds to the oligo-gangliosides, and it is known that the cooperative effect enhances the binding of the CT to its substrate (Schon & Freire, 1989). Moreover, CT recognizes the penultimate sugar of the ganglioside GM1. The binding constant of the legume lectin 1DBN was measured using the Biacore (Surface Plasmon Resonance) method using oligosaccharide substrates (Haseley et al., 1999). However, the arrangement of substrates on the surface

**Table 6.3.** Number of H-bonds and buried surface area as well as the dissociation constants of the sialic acid binding in the different lectins

Protein/Pdb code	No. of H-bonds <sup>a</sup>	Buried SA <sup>c</sup> (Å <sup>2</sup> )	K <sub>d</sub> (M)	Ligand
VCNA	11	503	~30×10 <sup>-6</sup>	Thisisialoside <sup>(ref1)</sup>
2CHB	4	525	~1.0×10 <sup>-6</sup>	Oligo-GM1 <sup>(ref2)</sup>
1PTO	7	476	≤1.0×10 <sup>-6</sup>	Fetuin <sup>(ref3)</sup>
1DFQ	6	376	≤1.0×10 <sup>-6</sup>	Gangliosides <sup>(ref3)</sup>
1QFO	6	405	~1.0×10 <sup>-3</sup>	Oligosaccharides <sup>(ref4)</sup>
1G1R	2	187	~1.0×10 <sup>-4</sup>	SLe <sup>x</sup> <sup>(ref5)</sup>
2CWG <sup>b</sup>	6	486	~1.0×10 <sup>-4</sup>	Sialyllactose <sup>(ref6)</sup>
1DBN	3	404	~1.0×10 <sup>-6</sup>	Oligosaccharides <sup>(ref7)</sup>
5HMG	7	473	~1.0×10 <sup>-3</sup>	Sialyllactose <sup>(ref8)</sup>
1SIE	6	360	~1.0×10 <sup>-4</sup>	Sialyloligosaccharide <sup>(ref9)</sup>
1KQR	6	608	~1.0×10 <sup>-3</sup>	Sialyllactose <sup>(ref10)</sup>

<sup>a</sup> Numbers of H-bonds are obtained from figure 6.2, H-bonds are counted as those connecting the binding residues with either the sialic acid ligand or water molecules.

<sup>b</sup> For wheat germ agglutinin, the H-bonds formed between the sialic acid moiety and residues of the sialoglycopeptide ligand are not counted.

<sup>c</sup> Buried surface areas were calculated in GRASP (Nicholls et al., 1991).

<sup>ref1</sup> Chapter 7

<sup>ref2</sup> (Schon & Freire, 1989)

<sup>ref3</sup> (Krell et al., 2003)

<sup>ref4</sup> (May et al., 1998)

<sup>ref5</sup> (Poppe et al., 1997)

<sup>ref6</sup> (Kronis & Carver, 1982)

<sup>ref7</sup> (Haseley et al., 1999)

<sup>ref8</sup> (Sauter et al., 1989)

<sup>ref9</sup> (Stehle & Harrison, 1996)

<sup>ref10</sup> (Dormitzer et al., 2002a)

during the measurement could indirectly enhance the binding compared to the binding in solution. On the contrary, the micromolar affinity in the case of VCNA was determined using the mono substrate in solution (by ITC, chapter 7), additionally, STD NMR (chapter 8) showed that the sialic acid of the thiosialoside is the chief binding residue for the substrate recognition. Therefore, the high affinities of the lectins: 2CHB, 1PTO, 1DFQ, and 1DBN relative to VCNA do not necessarily indicate stronger interaction between the lectins and the sialic acid moiety of the substrate. Note that, the binding constants of the wheat germ agglutinin and the capsid spike protein of rotavirus determined using sialyllactose as a substrate were two fold weaker than that of VCNA. From the above argument, perhaps the interactions of sialic acid with the lectin domain of VCNA is the strongest among the other lectins included in our study, this is supported by the qualitative analysis, discussed above, which showed that VCNA possesses extensive interactions with sialic acid compared to the other lectins.

In addition, the number of H-bonds formed between sialic acid and the binding site of VCNA (11 H-bonds) is almost as twice as that formed in the other lectins listed in Table 6.3. Nonetheless, it is difficult to correlate the affinity of the lectins with the number of H-bonds, as it is not only the number of H-bonds that determine the strength of binding, but also the strength of the formed H-bonds is an important factor to consider, which is not available. Also, the calculated buried surface areas do not show any trend that could be correlated with the measured affinity. Nevertheless, in some cases, they are in agreement with the qualitative analysis, as in the case of 1G1R that turned to have the smallest buried surface area in accord with the least number of contacts with the sialic acid. In contrast, 1KQR demonstrated the largest buried surface area despite its weak binding affinity. Indeed, it is important to consider the contribution of entropy to the  $\Delta G$  of

binding. This will include both the entropic gain of water molecules released upon ligand binding and the entropic loss of immobilizing a flexible ligand in the lectin-binding site.

In conclusion, from the above analysis it is evident that the protein-ligand interactions vary greatly among the studied lectins ranging from extensive interactions as seen in VCNA to few contacts with the protein as in 1G1R. Noticeably, there are no specific residues that characterize the sialic acid-binding site in the various lectins. That is why, many of the observed interactions involving main-chain atoms. This suggests that what is important is the availability of the right environment suitable to form H-bonds with the different substituents at the sugar ring, which can be provided by the H-donor and acceptor atoms from the side- and/or main-chain, and the presence of the complementary topographic features on the surface of the proteins. In these complexes the different lectins bind to the same ligand by recognizing different regions of the molecule. Nonetheless, the carboxylate group is involved in the sialic acid $\leftrightarrow$ lectin interactions to various extents in all the investigated lectins.

## **6.4 Discussion**

Without doubt the 3D X-ray structure of the protein-ligand complex provides a wealth of information about the interaction of the protein and its counterpart at the molecular level, despite the fact that these information are of qualitative nature. Herein, such a qualitative description of the protein-ligand interactions deepens our understanding for the recognition of DANA by the various neuraminidases/trans-neuraminidases whose structures are available to date. This deep understanding is of great importance for designing more potent inhibitors for the enzyme of interest in this study (VCNA) as well as for other neuraminidases. As it is known that DANA is not an absolute inhibitor for all

these enzymes, analysing the 3D structures of the neuraminidase/trans-neuraminidase-DANA complexes including the complex of the cholera enzyme enabled us to rationalize the differential binding affinity of DANA towards the various enzymes. Indeed, this comparison is useful in underlying the roles of the different parts of the ligand molecule in its interaction with the protein.

Evidently, from the analysis of the different DANA-complexes the constant protein environment around the carboxylate group, formed by the arginine triad and well conserved residues, is necessary to ensure the same binding mode of the transition-state-analogue molecule in the catalytic groove. This information can be extrapolated to the binding of the sialic acid substrate, where this unchanging environment forces the sugar ring to distort its conformation from the pyranoside chair conformation to adopt a geometry that maximizes the interaction of the ligand carboxylate with the arginine triad. In fact, trials to trap the substrate in its  $\alpha$ -conformation within the active site of neuraminidases are unsuccessful so far, and either DANA or a distorted geometry derived from the chair conformation was found instead as shown in our case of VCNA and as reported previously for the enzymes from influenza (Janakiraman et al., 1994) and *T. rangeli* (Amaya et al., 2003).

On the other hand, the environment around the other substituents of the sugar ring, varying to different extents, manifests itself in modulating the binding affinity of DANA to the various neuraminidases. In terms of the substrate binding, such varying environment could be implemented in tuning the efficiency of the catalytic activity of these enzymes; actually the turnover number differs greatly for the different neuraminidases (e.g., STNA has a turnover number of  $2700\text{ s}^{-1}$  (Hoyer et al., 1991), whereas it is only  $9\text{ s}^{-1}$  for influenza enzyme (Janakiraman et al., 1994)). Also, the variations seen at the active sites of neuraminidases might dictate the specific pathway of

the hydrolysis reaction as shown in the case of *M. decora* trans-neuraminidase, which follows a different mechanism compared to the other enzymes (Luo et al., 1998).

Another important factor, which can explain the absence of inhibition of DANA for the trans-neuraminidases from *M. decora* and from *T. rangeli*, and the low affinity towards neuraminidases from *S. typhimurium* and *T. rangeli*, is the presence of residues with bulky side-chains (tyrosine, tryptophan) at the opening of the catalytic cavity in these four enzymes. The bulky residues prevent the easy access of DANA to the active sites of these enzymes to different extents, in agreement with the observed measured affinities as discussed in section 6.2.

It is obvious from the comparative study that designing a potent inhibitor for a particular neuraminidase based on the chemistry of the transition-state-analogue requires exploitation of different parts of the sugar molecule. In the influenza enzyme, exploitation of DANA at C4 resulted in the synthesis of powerful inhibitors with bulky substituents at the 4-position (von Itzstein et al., 1993). The same idea has been exploited in the design of specific inhibitor of the viral enzyme HN-NDV, where there is a large cavity around 4-position. However, the case is different in the other neuraminidases in which the hydroxyl group at C4 is accommodated in a limited space. For VCNA, there is enough space around the glycerol group that could be exploited to improve the inhibition of DANA-based molecules; this is applicable for the *S. typhimurium* enzyme as well. As discussed in chapter 5, exploitation of the substituent at C5 resulted in a more potent inhibitor (FANA) for VCNA, where the introduced trifluoromethyl at the N-acyl group mediated favourable interactions with Asn545, Asn318, and Asp292 (see section 5.3.7.1). There is enough room around the N-acetyl group to be exploited in most of these enzymes. Finally, the concluding remark from the comparison study is that the different enzymes

require different strategies when designing inhibitors that interfere with its function; the availability of the 3D structure of the enzyme of interest in complex with the sought-after ligand renders the rational tackling of the problem easier.

In this chapter we compared the recognition pattern of sialic acid by the lectin domain in the multifunctions VCNA with that observed in other sialic acid-binding lectins. The polyhydroxy sialic acid mediates its interaction with the diverse lectins through direct or water-mediated H-bonding and hydrophobic interactions. The various parts of the molecule participate in these types of interaction to various extents among the lectins investigated. The negatively charged carboxylate group of sialic acid is involved in the interactions with all of these lectins; nonetheless, the strength of the interaction differs among them. It should be mentioned that the carboxylate group does not interact with charged residues in all cases; more frequently is the H-bonding interaction with polar residues like serine.

There are no specific residues or motifs that could be considered as identifiers for the sialic acid-binding sites in these lectins. The lectin-ligand interaction involves side- and main-chain atoms capable of forming H-bonds with the ligand, and residues that can participate in hydrophobic interactions.

The sialic acid interaction is extensive in the case of VCNA compared to the other lectins. Indeed, this is reflected in the high affinity of the N-terminal lectin domain in VCNA towards the sialic acid substrate ( $\sim 30 \mu\text{M}$  range, whereas the affinity of the monosaccharide towards the other lectins is usually in the mM range (Lis & Sharon, 1998)). In fact, the extensive interaction in the case of the cholera lectin indicates the importance of the sialic acid recognition by VCNA. As discussed in section 5.9, the lectin domain attaches the cholera enzyme to the host cell surface, and thereby, facilitates the

access of the catalytic domain to its receptor presented on the cell surface in order to carry out its neuraminidase function. In fact, the high affinity of the cholera lectin to the sialic acid substrates could be of use for biological applications as was the case for other lectins that are used as biological probes to identify and isolate the carbohydrates marking the different cell surfaces (Vijayan & Chandra, 1999).

The last point to mention is that lectins including most of those investigated herein use their multimeric association to enhance the binding affinity towards the recognizable carbohydrate (Lis & Sharon, 1998); whereas, the cholera enzyme does not show such multimeric feature. This might explain why the cholera enzyme had to develop such extensive interaction with its substrate at the lectin domain to compensate for the absence of the enhancement that would be received from other monomers.

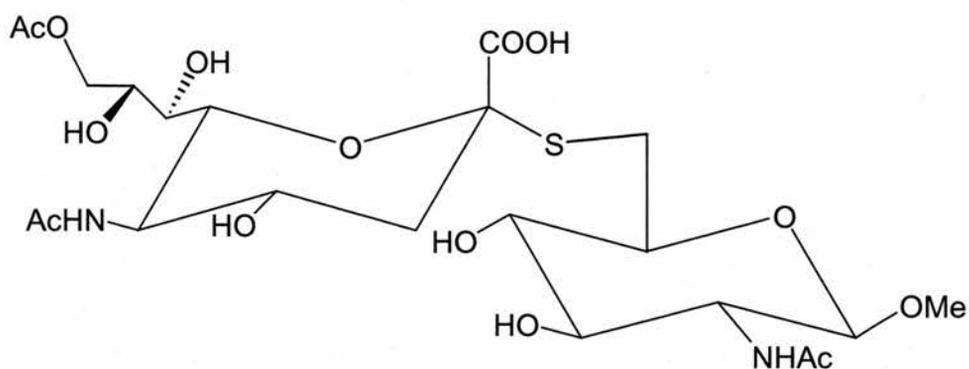
## Chapter 7

### Isothermal Titration Calorimetric study on VCNA

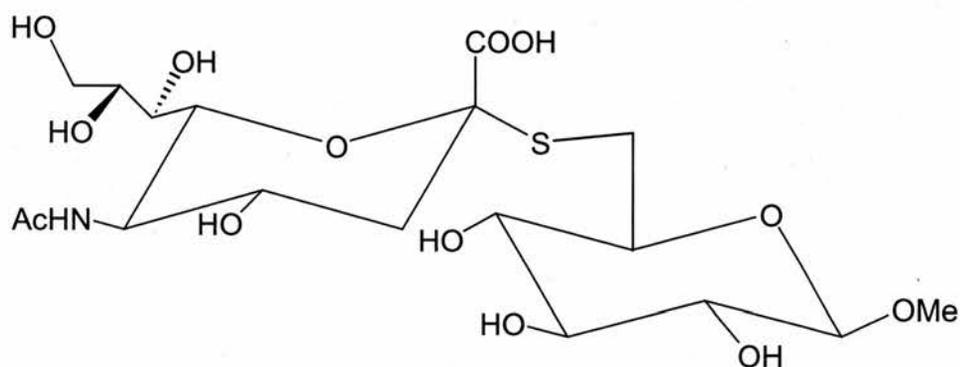
#### 7.1 Introduction

The crystallographic study of the complexes of VCNA revealed a new sialic acid-binding site at the N-terminal lectin-like domain as described in chapter 5. However, there is a limitation for the structural data, in that it provides only qualitative information about the protein-ligand interaction. We were interested in quantifying the interaction between VCNA and its binding ligand at the new sialic acid binding-site. There are many available techniques used to measure the ligand affinity towards its receptor (Connors, 1987) including: spectroscopic, chromatographic, kinetic, dialysis, electrical, and thermal techniques. One of those techniques is Isothermal Titration Calorimetry (ITC). The ITC is peculiar among the other techniques in that it can measure the thermodynamic parameters, the enthalpy ( $\Delta H$ ) and the entropy ( $\Delta S$ ) changes in one experiment, in addition to the determination of the binding constant and stoichiometry (Leavitt & Freire, 2001). Herein we applied the ITC technique to determine the affinity of the two thiosialoside ligands: Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me **1** and Neu5Ac-2-S- $\alpha$ -(2,6)-Glc $\beta$ Me **2** (scheme 7.1) towards VCNA.

Thiosialosides have been shown to be stable towards the hydrolytic action by VCNA, and to have inhibition activity against the enzyme (Khorlin et al., 1970; Sabesan et al., 1994). Wilson co-workers (Wilson et al., 1999) in their NMR study demonstrated that thiosialosides with sialic acid coupled to methyl  $\beta$ -galactose via  $\alpha$ (2 $\rightarrow$ 3) and  $\alpha$ (2 $\rightarrow$ 6) S-ketosidic linkages were not hydrolysed by VCNA; however, the non-hydrolysis of the thiosialosides shown by the NMR, in absence of information from other experiments, did



1 Neu5,9Ac<sub>2</sub>-2-S-α-(2,6)-GlcNAcβMe



2 Neu5Ac-2-S-α-(2,6)-GlcβMe

**Scheme 7.1** Thiosialosides used in the ITC experiments.

not prove or disprove that the ligands bind at the active site of the enzyme; no inhibition data were reported by the authors in their study. In our case, herein a competitive inhibition study was performed to check the binding of thiosialosides **1** and **2** at the active site of VCNA. The information from the inhibition study supported the interpretation of the results from the ITC experiments.

A brief account of the principles of the ITC technique and its applications in studying the protein-ligand interaction is given at the beginning of this chapter. Descriptions of the ITC and the competitive inhibition experiments are given in the materials and methods section. The results from the two experiments and the discussions are presented in two following sections.

## 7.2 Isothermal Titration Calorimetry: Overview

Isothermal Titration Calorimetry (ITC) technique, sometimes referred to as isothermal titration microcalorimetry, measures directly the heat changes and associated thermodynamics for biomolecular processes (Blandamer, 1998). ITC provides the direct measurement for the enthalpy change ( $\Delta H$ ) for protein-ligand interaction and the determination of its equilibrium binding constant ( $K_a$ ) that is directly related to the free energy change ( $\Delta G$ ) of the reaction (Equation 7.1); from ( $\Delta H$ ) and ( $\Delta G$ ), the entropy change ( $\Delta S$ ) of the system is calculated (Equation 7.2); additionally, the reaction stoichiometry ( $n$ ) can be estimated in the ITC experiment.

$$\Delta G = -RT \ln K_a \quad (7.1)$$

$$\Delta G = \Delta H - T\Delta S \quad (7.2)$$

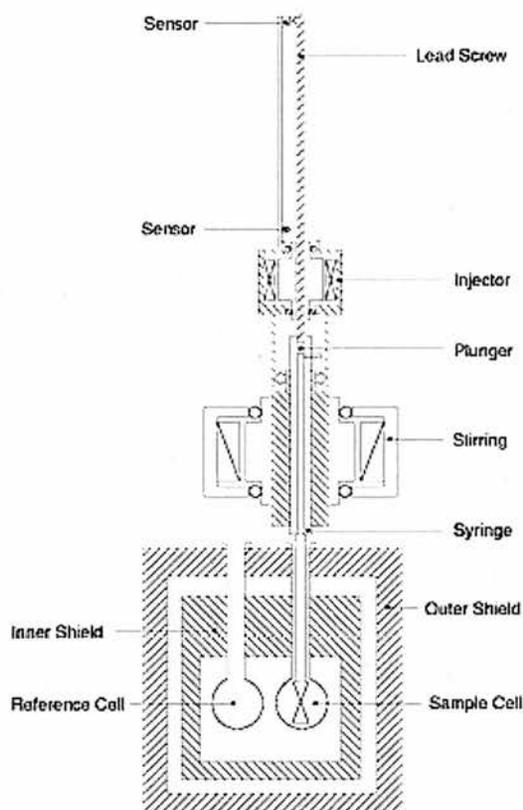
$R$  = universal gas constant of magnitude 8.31 J/mol K

$T$  = absolute temperature in Kelvin

Thereby, a single ITC experiment provides a complete thermodynamic profile of the molecular interaction. This characterization of the energetics of the biomolecular system extends our understanding for the biomolecular process (Ladbury & Chowdhry, 1996). That is why ITC became the method of choice to study the biomolecular interaction,

covering a whole range of applications: protein-ligand interaction (Dam & Brewer, 2002; Leavitt & Freire, 2001), protein-protein interaction (Weber & Salemme, 2003), DNA interactions (Barcelo et al., 2001), structure-based drug design (Ladbury, 2001; Muraki et al., 2002; Pickens et al., 2002), and pharmaceutical applications (Gaisford & Buckton, 2001).

The essential features of the ITC are shown in figure 7.1. In a typical ITC experiment, a syringe coupled to a motor stirring under computer control injects a small volume of solution containing one of the reactants (usually the ligand) into a solution in the sample cell (usually the protein) (Freire et al., 1990; Wiseman et al., 1989). The calorimeter records the accompanying heat of the  $i^{\text{th}}$  injection ( $q_i$ ) as an output peak, which can be



**Figure 7.1.** Diagrammatic representation of ITC

either exo- or endothermic. As the name implies, the ITC experiment is conducted under isothermal conditions. ITC is a differential technique having both a reference and a sample cell; to maintain the temperature difference between the two cells at constant value, a compensating power is used, continuously, regulating the amount of heat applied to the titration cell (O'Brien et al., 2001). The amount of compensated power ( $\mu\text{cal/s}$ ) is the output signal of the instrument and a power-time curve is produced (Leavitt & Freire, 2001). The heat after each injection is therefore obtained by integrating the area under each peak. The amount of heat after each injection ( $q_i$ ) is proportional to amount of ligand that binds to the protein in the cell and the characteristic binding enthalpy ( $\Delta H$ ), as given in Equation (7.3).

$$q_i = v \times \Delta H \times \Delta L_i \quad (7.3)$$

where  $v$  is the volume of the titration cell, and  $\Delta L_i$  is the increase of the concentration of bound ligand after the  $i^{\text{th}}$  injection.

Because the amount of uncomplexed protein available progressively decreases after each successive injection, the magnitude of the peaks becomes progressively smaller until complete saturation is achieved. Once this situation is reached, subsequent injections produce similar peaks corresponding to dilution or mechanical effects that need to be subtracted from all injection peaks before data analysis. The corrected area under each peak given by (Equation 1) is used to analyse the data (Tame et al., 1998).

Since heat is a ubiquitous physical property, calorimetry is a non-specific technique and, in principle, can be applied to all types of systems and processes as long as the monitored processes are not enthalpically neutral and the produced heat within the limit of detection. However, in complex systems the non-specificity makes it difficult to interpret the results from ITC experiments and more information is required from techniques applying more specific methods for adequate interpretation (Wadso, 2001).

The final ITC output, the isothermal binding curve, is usually obtained by plotting the integrated heat corrected for the heat of dilution vs. the molar ratio of the ligand. Data are analysed by fitting to a pre-assumed model employing some minimization procedures to evaluate the observed parameters of the binding constant ( $K_{\text{obs}}$ ), the enthalpy change ( $\Delta H_{\text{obs}}$ ), and the reaction stoichiometry ( $n$ ) (Freire et al., 1990). The free energy change ( $\Delta G_{\text{obs}}$ ) is obtained from relation (7.1), and the entropy change ( $\Delta S_{\text{obs}}$ ) is calculated from Equation (7.2). Note that, the derived parameters are often better defined as observed, since the ITC experiment is based on measuring the total enthalpy change (Tame et al., 1998).

## 7.3 Materials and methods

### 7.3.1 ITC experiment

**Sample preparation:** The enzyme VCNA was produced using the new expression system as described previously in chapter 3. The produced enzyme sample was prepared in the buffer system composed of: (20 mM Tris-HCl pH 7.6 + 150 mM NaCl + 10 mM CaCl<sub>2</sub> + NaN<sub>3</sub>). Before running the ITC experiment, the enzyme sample was dialyzed exhaustively for 24 hours against the ITC buffer containing (50 mM Tris-HCl pH 7.6 + 150 mM NaCl + 10 mM CaCl<sub>2</sub> + NaN<sub>3</sub>). The concentration of the dialyzed enzyme sample was determined by Bradford assay (Bradford, 1976)(see section 3.2.4) and UV-absorbance at  $\lambda_{280}$  using the extinction coefficient  $\epsilon_{280}$  of magnitude  $8.3 \times 10^4 \text{ cm}^{-1} \text{ M}^{-1}$ ; see Appendix 1 for the determination of the extinction coefficient. The measured concentrations from the two methods were close to each other. Bradford assay gave a protein concentration of (10 mg/ml), and the spectroscopic method gave a concentration of (11 mg/ml) for the same sample. We used the concentration obtained from the

spectroscopic method (Scopes, 1974), which is commonly used in ITC for its accuracy (Pace et al., 1995; O'Brien et al., 2001). The molar concentration of the protein sample was calculated using the molecular weight of the protein as 82,975 g/mol.

Samples of ligands **1** and **2** (gifts from Dr. K. Melton, Institute for Glycomics, Griffith University, Queensland, Australia) were prepared by dissolving the required amount of the ligand(s) in the appropriate volume of the ITC buffer (same buffer of the protein sample) to give the concentration needed for each experiment. Using the same buffer for the protein and the ligand(s) samples helps to minimize errors in the titration experiment (O'Brien et al., 2001).

**Titration run:** Titration experiments were performed using the VP-ITC calorimeter (Microcal Inc., Northampton, MA, U.S.A). All titrations were conducted under isothermal conditions at 37 °C. The titration cell and the syringe were washed thoroughly with water and sample buffer before loading the protein and ligand samples. The samples of the protein and the ligand(s) were degassed; the protein sample was loaded into the titration cell and the ligand sample was loaded into the 250 µl syringe. For the ligand **1**, the isothermal binding curve was obtained by titrating a 0.6 mM ligand solution (loaded in the syringe) into a 27 µM VCNA solution (loaded in the 1.4 ml titration cell). Aliquots of the ligand in the syringe (stirred at 300 rpm) were injected into the titration cell, 1 µl for the first injection followed by 25 injections of 10 µl each. The heat released or absorbed was obtained by integrating the heat burst peaks of the power-time curve. The duration of each injection was 10s (2s for the first injection) and the delay between injections was 150s and the initial delay prior injection was 60s. Heats of dilution were determined by titrating 0.6 mM ligand into the ITC buffer, and the observed heats of binding were corrected for heats of dilution prior to data analysis. The normalized heats of binding for the injections were plotted versus the molar ratio (ligand : protein) to produce the

isothermal binding curve. Raw data were integrated and processed with the ORIGIN 5.0 analysis software (Microcal). The non-linear least squares method was used for fitting the observed data to the two built-in models accompanying the software for binding: single set of binding sites and two sets of binding sites; the first data point corresponding to 1  $\mu$ l injection was removed from the fitting. The ITC experiment for ligand **1** was repeated for the required reliability of the estimated parameters.

For ligand **2**, the isothermal binding curve was obtained in the same way as in the titration of ligand **1**, using a VCNA sample (15  $\mu$ M) and a ligand sample (0.5 mM), the titration consisted of 21 injections, 1  $\mu$ l for the first injection followed by 20 injections, 10  $\mu$ l per injection. For the blank titration, only 10 aliquots of ligand **2** were injected into the buffer, 1  $\mu$ l for the first injection and 9 injections of 10  $\mu$ l each. Unfortunately, there was no enough material of **2** to repeat the experiment.

### **7.3.2 Competitive inhibition of VCNA**

The competitive inhibition of the ligands **1** and **2** were studied by measuring the neuraminidase activity (Myers et al., 1980; Potier et al., 1979) of VCNA (see section 3.2.3), in presence and absence of the ligand(s), using the substrate 2'-(4-methylumbelliferyl)- $\alpha$ -D-N-acetylneuraminic acid sodium salt (Sigma Chemicals Co.). The inhibitor FANA was used as a positive inhibition test.

Three samples of VCNA were prepared in 100 mM sodium acetate buffer, pH 5.5, containing 10 mM CaCl<sub>2</sub>. Each of the protein samples was incubated with one of the ligands **1**, **2**, or FANA for 10 minutes before adding the substrate. Each of the samples contained a final concentration of 10 mM of the incubated ligand, and 10 nM of VCNA. A control sample was prepared, without adding any ligands, with the same protein

concentration (10 nM). The substrate was added to the protein samples to a final concentration of 50  $\mu$ M; a total volume in each sample was 100  $\mu$ l. The samples were incubated at 37 °C for 5 min.; adding 3 ml of 250 mM solution of glycine-NaOH, pH 10.5, quenched the reaction. The fluorescence of the solutions was then measured using the spectrophotometer TD-700 Laboratory Fluorometer (Turner Designs). The same procedures were repeated using the same ligand concentration (10 mM), but a 10-fold less protein concentration (1 nM). The inhibition activity of the ligand was expressed as the percentage of the decrease in the readout (amount in ng/ml of the released 4-methylumbelliferone) from the Fluorometer measured in presence of the ligand, compared to the readout of the control sample and referenced to the latter one. All measurements were done in duplicate for reliability.

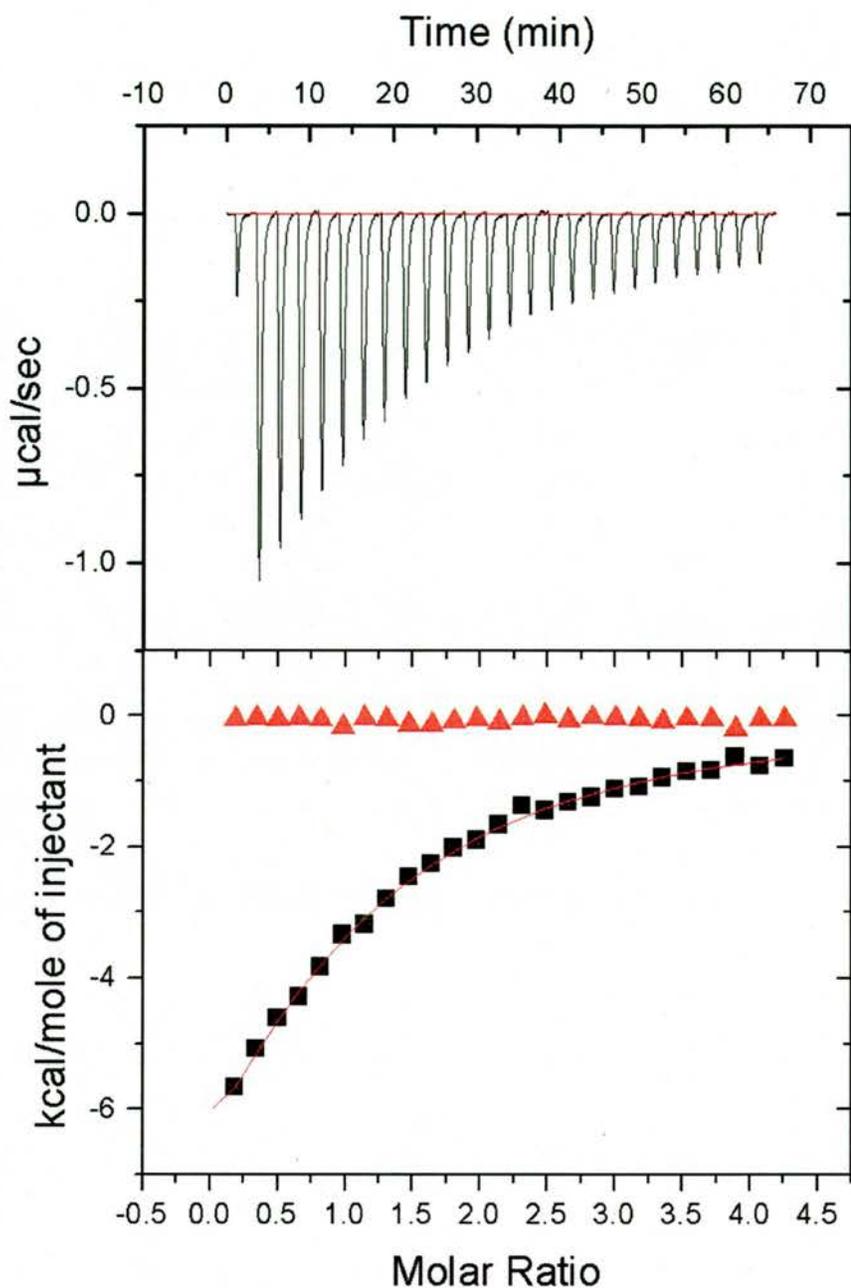
## 7.4 Results

### 7.4.1 Binding affinity of the thiosialosides towards VCNA

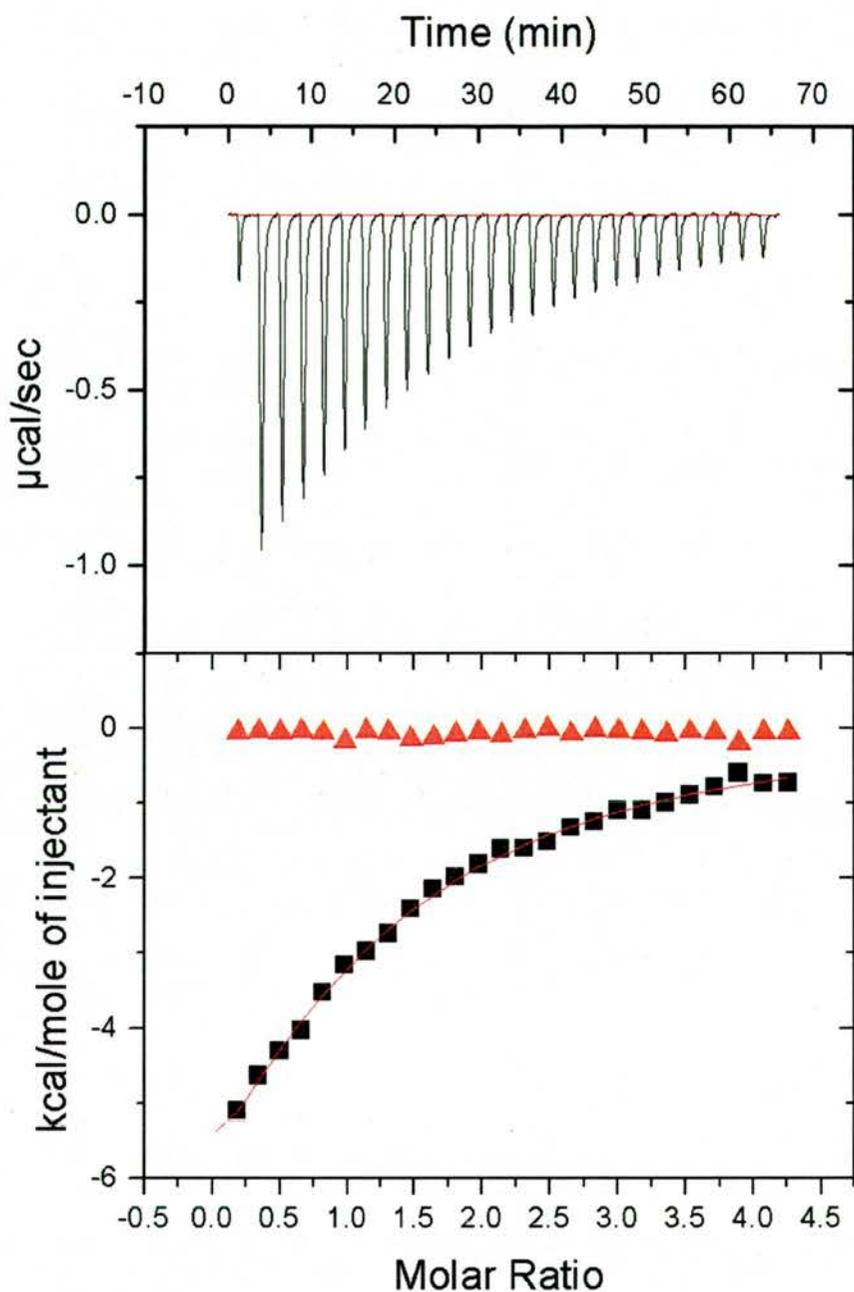
Our crystallographic study unmasked the sialic acid binding-site at the N-terminal lectin domain of VCNA. Analysis of the 3D structure of this site revealed an extensive interaction between sialic acid and the binding site residues, comparable to other lectins recognising sialic acids (see chapter 6). The qualitative characterisation of the interaction at the new site anticipated its high affinity for ligands containing  $\alpha$ -sialic acid. The ITC technique was applied to quantify the affinity using thiosialosides **1** and **2** as ligands (scheme 7.1). The two ligands were slightly different: in ligand **1** sialic acid was acetylated at C9 and linked through the  $\alpha(2\rightarrow6)$  *S*-ketosidic linkage to methyl  $\beta$ -N-acetylglucosamine residue, whereas in ligand **2** sialic acid was coupled to methyl  $\beta$ -glucose residue via the *S*-ketosidic  $\alpha(2\rightarrow6)$  linkage.

Figures 7.2 and 7.3 display the isothermal profiles for thiosialoside **1**, from the two ITC experiments of ligand **1**, obtained by injecting the ligand (0.6 mM) into the titration cell containing the enzyme solution (27  $\mu$ M) buffered to pH 7.6, at 37  $^{\circ}$ C. The isothermal profile for the titration of thiosialoside **2** (0.5 mM) into VCNA (15  $\mu$ M) at pH 7.6 and temperature 37  $^{\circ}$ C is shown in figure 7.4. Unfortunately, due to a lack of material of ligand **2** we were unable to repeat its titration experiment for the required statistical reliability; however, the interpretation of the data was in agreement with information obtained from other experiments, as will be presented in due course. The upper panels of the three figures represent the raw data of the power-time curve showing the exothermic peaks of the heat released from reactions of the thiosialosides **1** & **2** with VCNA. The height of the exothermic peaks decreased progressively upon the addition of more ligand into the titration cell, since the amount of the free binding site(s) available became less. The lower panels show the normalized integrated heats, corrected for the dilution effect, and plotted against the molar ratio, the total concentration of ligand(s) to protein in the titration cell. Injections of the ligand(s) into the buffer produced a small amount of heat compared to the ligand(s)-into-protein titration as shown in the top parts of the lower panels. In the blank injection of ligand **2**, only 10 injections were used since there was not enough ligand; these injections were enough to get an average value for the heat of dilution.

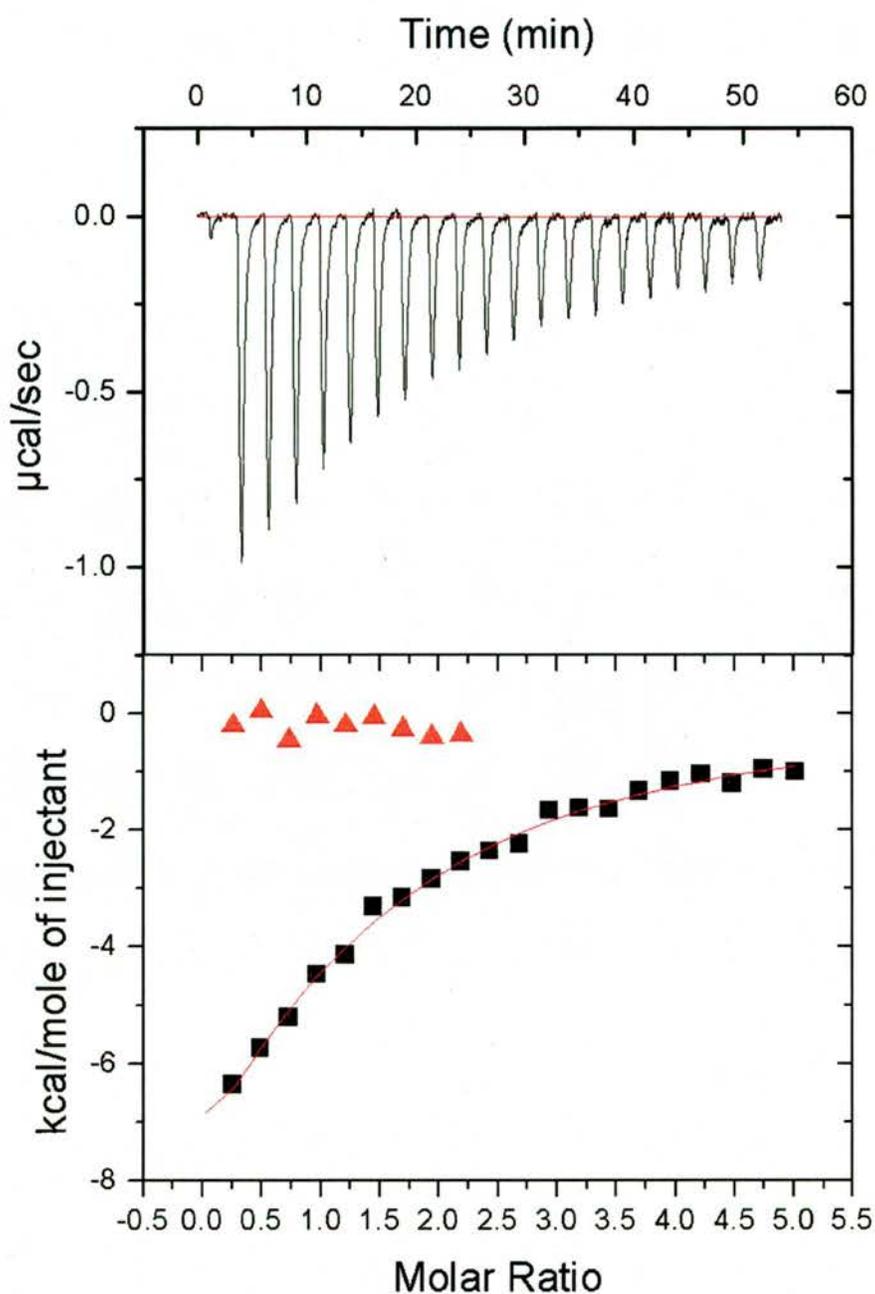
To obtain the thermodynamic characterization of the protein-ligand interaction, the ITC data had to be processed. The non-linear least squares method was used to fit the data with three independent variables,  $K_a$ ,  $\Delta H$ , and stoichiometry ( $n$ ). The single-set of sites model best fitted the data obtained for the two ligands **1** & **2**, as shown in the lower panels of figures 7.2, 7.3, and 7.4. The variable parameters were best fitted with values given in table 7.1. The thermodynamic parameters, from the first fitting to the single-set model



**Figure 7.2.** ITC of thiosialoside **1** (0.6 mM) titrated against VCNA (27  $\mu$ M) in 50 mM Tris-HCl buffer, pH 7.6, containing 150 mM NaCl and 10 mM CaCl<sub>2</sub>, at 37 °C. (Upper panel) Raw data of **1** injected into the titration cell with the VCNA sample, 1  $\mu$ l for the first injection followed by 25 injections of 10  $\mu$ l each. (Lower panel) Integrated and normalized injection heats of **1** into VCNA titration (black square), and the blank injection of **1** into buffer (red triangle) are plotted as a function of the molar ratio (ligand : protein) in the titration cell. The fit obtained from the single-set of sites model is drawn as solid red line; the first data point was removed from the fitting.



**Figure 7.3.** ITC profile for the repeat experiment of titration of thiosialoside **1** (0.6 mM) into VCNA (27  $\mu\text{M}$ ) performed under the same conditions of the first run. (Upper panel) Raw data of **1** injected into the titration cell with the VCNA sample, 1  $\mu\text{l}$  for the first injection followed by 25 injections of 10  $\mu\text{l}$  each. (Lower panel) Integrated and normalized injection heats of **1** into VCNA titration (black square), and the blank injection of **1** into buffer (red triangle) are plotted as a function of the molar ratio (ligand : protein) in the titration cell. The fit obtained from the single-set of sites model is drawn as solid red line.



**Figure 7.4.** ITC of thiosialoside **2** (0.5 mM) titrated against VCNA (15  $\mu$ M) in 50 mM Tris-HCl buffer, pH 7.6, containing 150 mM NaCl and 10 mM CaCl<sub>2</sub>, at 37 °C. (Upper panel) Raw data of **2** injected into the titration cell having the VCNA sample, 1  $\mu$ l for the first injection followed by 20 injections of 10  $\mu$ l each. (Lower panel) Integrated and normalized injection heats of **2** into VCNA titration (black square), and the blank injection of **2** into buffer, only 10 injections, (red triangle) are plotted as a function of the molar ratio (ligand : protein) in the titration cell. The fit obtained from the single-set of sites model is drawn as solid red line; the first data point was removed from the fitting.

(table 7.1), indicated that both of the thiosialosides **1** and **2** were bound only to one site with stoichiometry ( $n$ ) very close to the value of 1, and with an average apparent dissociation constants: [ $K_d = 1/K_a = 3.7 \times 10^{-5} \text{ M}$ , or 37  $\mu\text{M}$  for ligand **1**; and  $K_d = 2.6 \times 10^{-5} \text{ M}$  or 26  $\mu\text{M}$ , for ligand **2**]. Clearly, the binding affinity of the two thiosialosides did not differ greatly, which might suggest that the major role in the binding was played by the sialic acid moiety with a small contribution from the second sugar residues. Indeed, the results from the STD NMR experiment on ligand **1** showed that sialic acid was the chief residue responsible for the binding of **1** towards its target; see section 8.4.2. It is important to emphasize that it is hard to say that ligand **2** possessed slightly higher binding affinity over ligand **1**; this was due to the lower confidence in the estimated parameters from the ITC experiment of ligand **2**, this point is discussed further in the next paragraph. Nonetheless, this should not be understood as an underestimation of the derived parameters for ligand **2**. It is just safer to say that both of the two ligands bound to the enzyme VCNA with dissociation constants in the same micromolar range. The ligand-protein interaction, for both of the two thiosialosides, was driven by the favoured enthalpy changes ( $\Delta H = -14.3, -17.7 \text{ kcal/mol}$  for ligands **1** and **2**, respectively) that compensated the unfavourable negative entropy changes ( $T\Delta S = -7.9, -11.1 \text{ kcal/mol}$  for ligands **1** and **2**, respectively), which resulted in binding free energies ( $\Delta G = -6.3, -6.6 \text{ kcal/mol}$  for ligands **1** and **2**, respectively). The differences in values of the enthalpy and entropy changes between the two ligands might have some basis in the details of the interaction of the two ligands with the enzyme, which are not yet available, such as the different number of water molecules involved in the protein-ligand interactions. A sheath of solvent molecules surrounding the ligand as well as solvent molecules occupying the binding site need to be rearranged or displaced upon the binding event. This

**Table 7.1.** Fitting and thermodynamic parameters for the titration of thiosialosides **1** and **2** obtained using the single-set of sites model

Exp. No.	Ligand	$n$	$K_a/10^4$ ( $M^{-1}$ )	$\Delta G$ (kcal/mol)	$\Delta H$ (kcal/mol)	$T\Delta S$ (kcal/mol)
1	Thiosialoside <b>1</b>	$0.89 \pm 0.07$	$2.79 \pm 0.21$	-6.3	$-15.0 \pm 1.5$	-8.7
2		$0.99 \pm 0.09$	$2.57 \pm 0.22$	-6.3	$-13.3 \pm 1.5$	-7.0
	average	$0.94 \pm 0.08$	$2.68 \pm 0.22$	-6.3	$-14.3 \pm 1.5$	-7.9
1	Thiosialoside <b>2</b>	$1.08 \pm 0.17$	$3.83 \pm 0.52$	-6.6	$-17.7 \pm 3.4$	-11.1

rearrangement of solvents might contribute differently to the enthalpy and entropy changes of the interaction of a protein with different ligands (Cooper et al., 2001; Ladbury & Chowdhry, 1996).

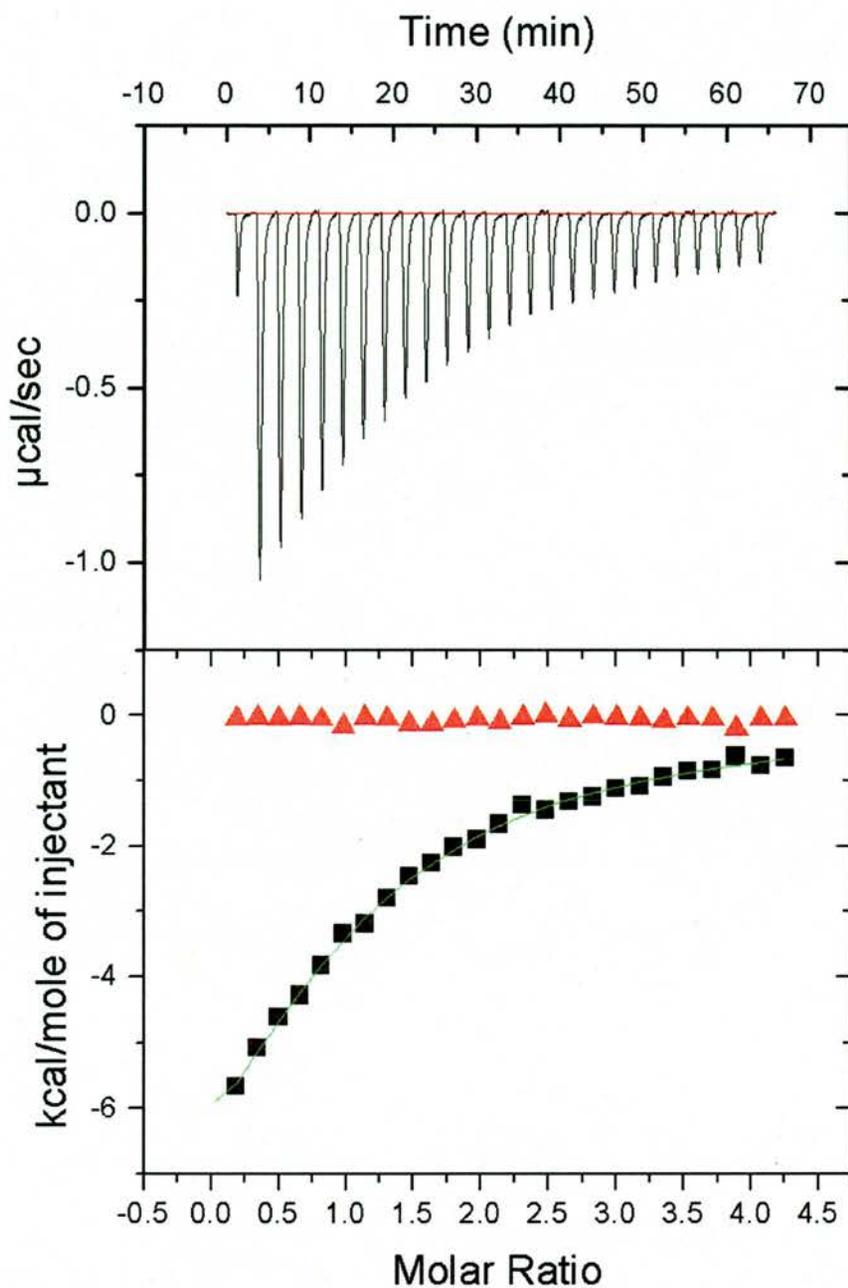
It should be mentioned that the quality of the isothermal binding curve obtained for ligand **2** titrated into VCNA (lower panel of figure 7.4) was not as good as that obtained for the titration of ligand **1** (lower panels of figures 7.2 & 7.3). Inspecting the curves by eyes revealed that there were more deviations of the data points from the fitting line for ligand **2** than the observed deviations from the fitting line in the case of ligand **1**. This appeared in the relatively larger statistical errors for the derived parameters of ligand **2** compared to the derived parameters of ligand **1** (table 7.1). This could be explained by the usage of lower concentration of the enzyme in the ITC experiment of ligand **2**. As the employed concentration of the enzyme (15  $\mu M$ ) did not satisfy the rule of thumb stating that better ITC signals are obtained when the applied concentration of the macromolecule gives values of the quantity  $C$ , which affects the shape of the binding curve, falling within the range (1 to 500) in the following Equation:

$$C = K_a [M]_{\text{total}} \quad (7.4)$$

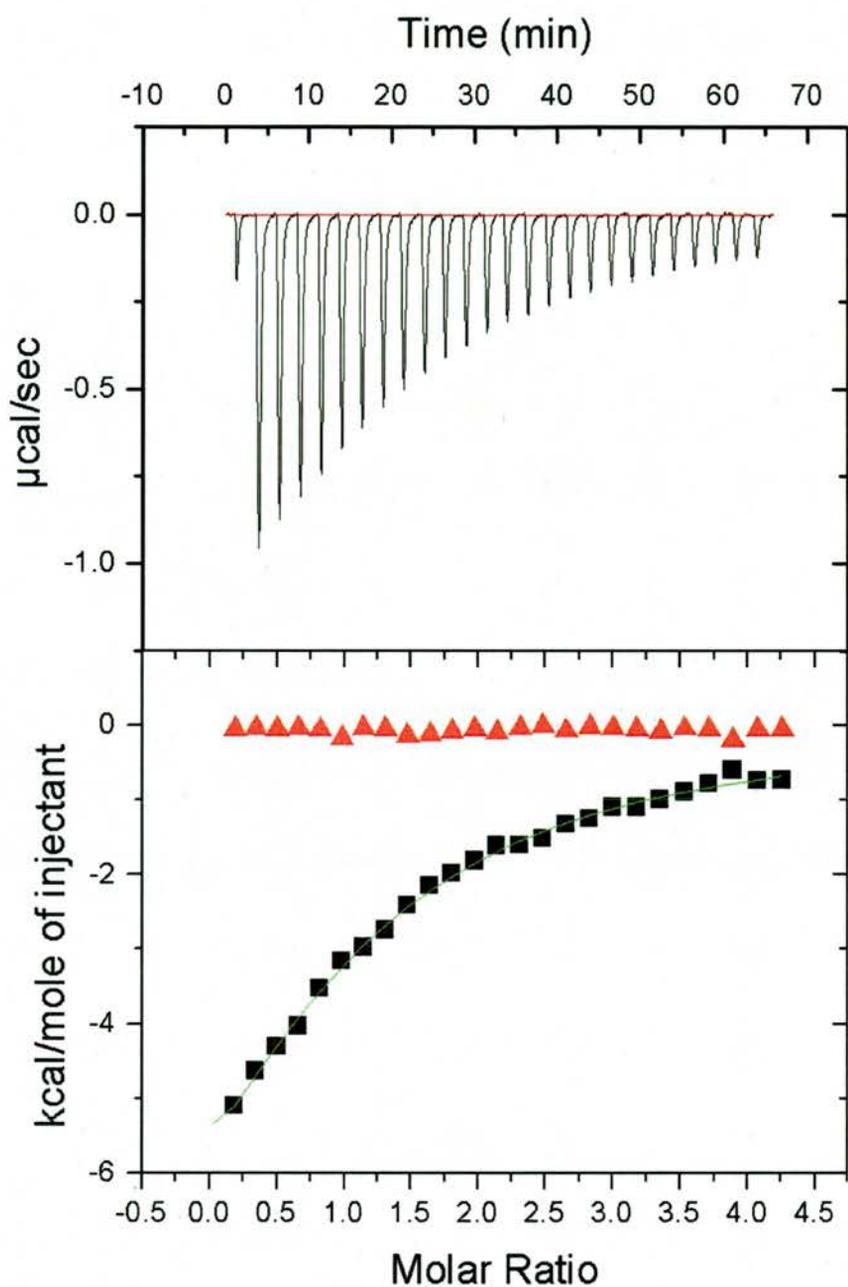
where  $C$  is a dimensionless quantity in the range 1:500,  $K_a$  is the association constant and  $[M]_{\text{total}}$  is the total concentration of the macromolecule. In the case of ligand **1** titration,  $C$  was close to 1 ( $C \approx 0.9$ ), whereas for ligand **2** titration,  $C$  was less than 1 ( $C \approx 0.5$ ). In fact, with  $C > 0.1$ , the titration plot should be still informative, but with less confidence in the estimates of  $K_a$  and  $\Delta H$  (Blandamer, 1998).

Fitting the data to the two-sets of sites model, shown in figures 7.5 and 7.6 for ligand **1**, and figure 7.7 for ligand **2**, with fixed values of the stoichiometric parameters ( $n_1 = n_2 = 1$ ) showed an apparent fitting similar to the single-set model but with very poor statistics, values of the fitting parameters are given in table 7.2; only the values quoted from the fitting of the data from one experiment of ligand **1** are shown in the table. Similarly, the fitting from the repeat experiment of ligand **1** gave the same sort of parameter values as those shown in the table 7.2. Allowing the stoichiometric parameters ( $n_1$  and  $n_2$ ) to vary during the fitting resulted in highly floating values for  $n_1$  and  $n_2$ , which did not make any sense, and the fitting did not converge to a minimum; an indication that the two-sets of sites model was not the appropriate model.

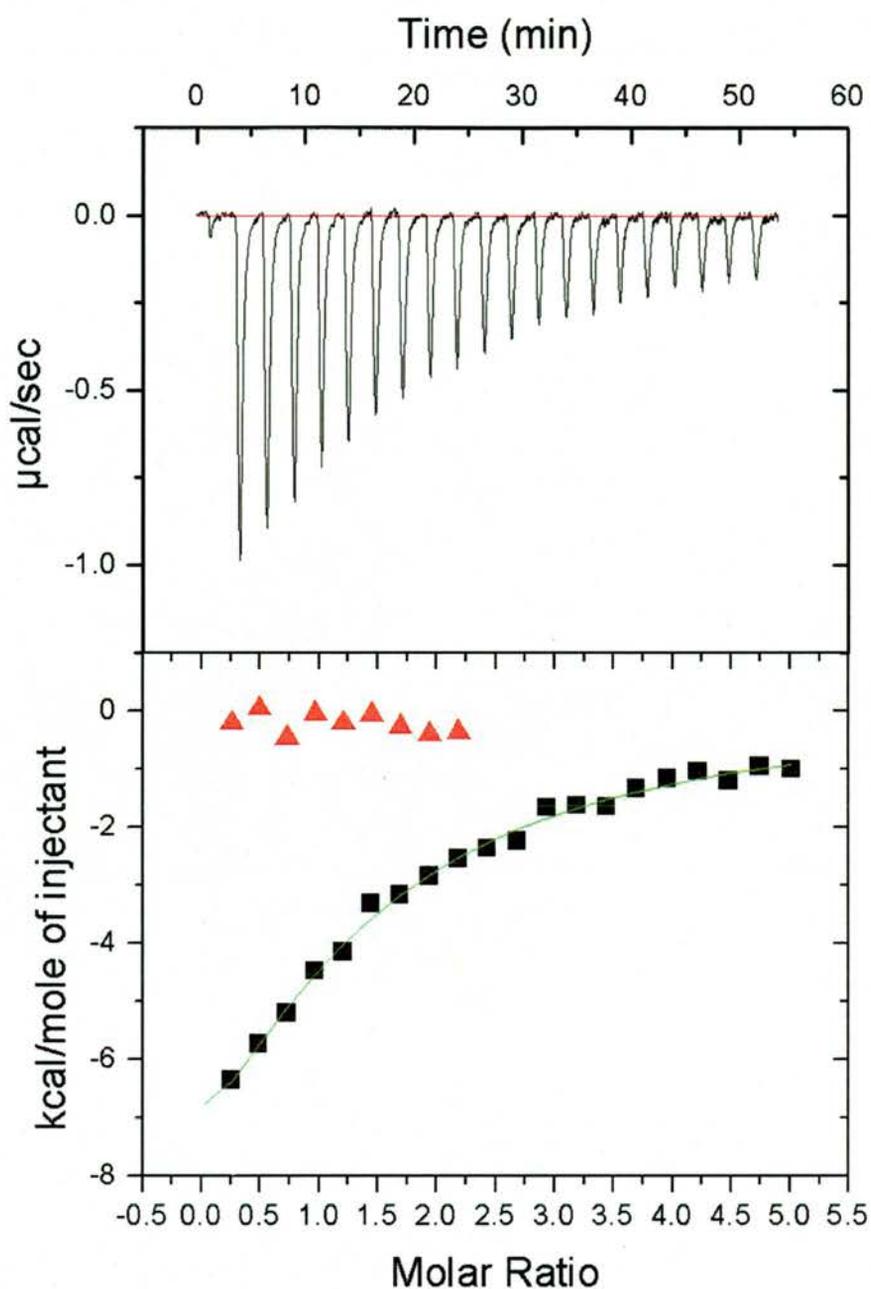
So far, only the thermodynamic characterization of the VCNA-thiosialosides interactions has been highlighted; yet, further investigation is needed to fully understand the ITC results. Despite the presence of two sites in VCNA, the catalytic and the binding sites, the data fitted best to the single-set of sites model with stoichiometry ( $n = 1$ ). This implied that the thiosialosides **1** and **2** interacted with one site only. Indeed, we were inclined to believe that the isothermal binding curves of the two ligands had originated from their recognition by the N-terminal lectin domain because of two reasons:



**Figure 7.5.** The same ITC profile displayed in figure 7.2 showing the fitting of the observed data (green solid line in the lower panel) using the two-sets of binding site model. The apparent fitting is almost the same as that obtained from the single-set of sites model, but with poor statistics, see table 7.2.



**Figure 7.6.** The same ITC profile for the repeat experiment displayed in figure 7.3 showing the fitting of the observed data (green solid line in the lower panel) using the two-sets of binding site model. The apparent fitting is almost the same as that obtained from the single-set of sites model, but with poor statistics, see table 7.2.



**Figure 7.7.** The same ITC profile displayed in figure 7.4, for the ITC experiment of thiosialoside 2, showing the fitting of the observed data (green solid line in the lower panel) using the two-sets of binding site model. The apparent fitting is almost the same as that obtained from the single-set of sites model, but with poor statistics as shown in table 7.2.

**Table 7.2.** Fitting parameters for the two-sets of sites model used for fitting the observed data from the ITC experiments of the two thiosialosides **1** and **2**

Parameter	Thiosialoside <b>1</b>	Thiosialoside <b>2</b>
$n_1, n_2$	$n_1 = n_2 = 1$	$n_1 = n_2 = 1$
$K_{a1}$ ( $M^{-1}$ )	$(0.02 \pm 7.90) \times 10^3$	$(7.20 \pm 5.82) \times 10^4$
$K_{a2}$ ( $M^{-1}$ )	$(3.87 \pm 2.09) \times 10^4$	$(1.44 \pm 0.75) \times 10^4$
$\Delta H_1$ (cal/mol)	$(-0.03 \pm 95.3) \times 10^8$	$(-1.28 \pm 0.67) \times 10^4$
$\Delta H_2$ (cal/mol)	$(-1.14 \pm 0.24) \times 10^4$	$(-7.15 \pm 7.76) \times 10^3$
$\Delta S_1$ (cal/mol K)	$-1.08 \times 10^4$	-19.1
$\Delta S_2$ (cal/mol K)	-15.7	-4.04

(1) The crystallographic structure of  $\alpha$ -sialic acid bound at the N-terminal domain anticipated less restriction on the conformations and the type of the glycosidic linkage adopted by the sugar linked to sialic acid compared to the catalytic site (chapter 5). As shown in section 5.3.2.2, both  $\alpha(2 \rightarrow 3)$  and  $\alpha(2 \rightarrow 6)$  *O*-ketosidically linked sialic acids can have access to the N-terminal lectin domain. By analogy, the glycomimetics **1** and **2** having *S*-ketosidic  $\alpha(2 \rightarrow 6)$ -linked sialic acid were more likely to behave in a similar way. On the other hand, for the catalytic domain, it has been shown that VCNA hydrolyse substrates with  $\alpha(2 \rightarrow 6)$ -linked sialic acid less efficiently compared to substrates with  $\alpha(2 \rightarrow 3)$ -linked sialic acid (Friebolin et al., 1984). Hence, ligands **1** and **2** were expected to have limited access towards the catalytic site of the enzyme. However, the fact that the nature of the penultimate sugar of the substrate might affect the binding could not be escaped (Corfield, 1992).

(2) It was reported that a thiosialoside analogue to ligand **2**, with  $\beta$ -galactose instead of the methyl  $\beta$ -glucose residue, has no inhibition activity for VCNA (Sabesan et al., 1994), which means there was no binding of that ligand at the catalytic domain of the enzyme.

Along the above reasoning we wanted to get an experimental evidence to prove or disprove that the thiosialosides **1** and **2** were unable to bind at the enzyme active site, under the given experimental conditions, by checking their inhibition activities for VCNA as discussed in the following section.

#### **7.4.2 Competitive inhibition of thiosialosides 1 and 2**

The binding of thiosialosides, known to be resistant to enzymatic hydrolysis by VCNA, at the catalytic site would manifest itself in the inhibition of the neuraminidase activity. Synthesis of thiosialosides, glycomimetics to the natural substrate of neuraminidase enzymes, possessing moderate inhibition activity against VCNA has been reported (Sabesan et al., 1994). In addition, simple aryl thiosialosides were shown to inhibit VCNA with inhibition constant in the low millimolar range (Khorlin et al., 1970). Herein, we checked the inhibition activity of the thiosialosides **1** and **2** against VCNA. This was carried out by monitoring the catalysed liberation of the fluorescent 4-methylumbelliferone from the substrate 2'-(4-methylumbelliferyl)- $\alpha$ -D-N-acetylneuraminic acid in absence and presence of the ligands as described in the methods (section 7.3.2). In this experimental setup, the thiosialosides and the methylumbelliferyl-substrate are mutually exclusive, *i.e.*, the enzyme active site would not bind to the ligand(s) and the substrate at the same time and a competitive inhibition would be established (Engel, 1977). For comparison, a known inhibitor FANA ( $K_i=1.8 \mu\text{M}$ ) (Meindl et al., 1974) was used as a positive inhibition test.

Table 7.3 summarises the results for the activity measurements. It is evident from these results that the presence of 10 mM thiosialosides (200 times excess over the 50  $\mu\text{M}$

**Table 7.3.** Inhibition of VCNA with FANA and thiosialosides 1 & 2

Enzyme conc. (nM)	Release of 4-methylumbelliferone (ng/ml) in presence of			
	Control	FANA	Thiosialoside 1	Thiosialoside 2
10	433	3.20	431	434
10	432	2.90	432	433
1	266	1.20	199	218
1	268	0.80	201	220
<b>Inhibition (%)<sup>1</sup></b>				
10 nM VCNA	-	99	0	0
1 nM VCNA	-	99	25	18

<sup>1</sup> Inhibition is expressed in (%)

substrate) did not affect the enzyme activity measured in absence of the ligands. On the contrary, the use of FANA (10 mM) almost blocked the enzyme activity completely (99% inhibition). Clearly, in spite of the high ligand excess, the thiosialosides did not compete with the substrate, for which VCNA has a  $K_m$  value of 0.24 mM (Myers et al., 1980), to sit in the enzyme active site as evident from the lack of any inhibition activity of the thiosialoside ligands. Even so, these results did not exclude the possibility of binding of the thiosialosides at higher concentration.

It was necessary to test the inhibition activity of the two thiosialosides at higher concentration. Unfortunately, there were not enough of the ligands, so the inhibition test was carried out using the enzyme concentration (1 nM), one-tenth of that used in the

previous experiment. As shown in table 7.3, this resulted in about 25% inhibition for thiosialoside **1** and 18% inhibition for thiosialoside **2**; and as expected, FANA nearly inhibited the enzyme activity totally (99% inhibition). This weak inhibition of **1** and **2** could be accounted for if the ligands had dissociation constants in the range of high millimolar or so. Presumably, 10 mM ligand concentration in the first experiment was not enough to exhibit substantial inhibition activity in the range of detection. In addition, this view was in agreement with results reported in a previous study on thiosialosides mimicking the natural substrate of neuraminidases (Sabesan et al., 1994). Therein, it was shown that glycomimic thiosialosides need to be in certain conformation to access the catalytic site of VCNA and other neuraminidases included in the study. A thiosialoside analogue to ligand **2**, which possessed  $\beta$ -galactose instead of methyl  $\beta$ -glucose, exhibited no activity for VCNA at 8 mM concentration. However, forcing that particular thiosialoside to adopt certain conformation (tg-rotamer conformation) by methylating C6 of the galactose sugar resulted in an inhibitor for VCNA with  $K_i$  of value 0.35 mM (Sabesan et al., 1994). As mentioned in the same study, in the case of the first ligand without methylation, the conformation that can bind at the enzyme active site was in a very low population. This might explain why inhibition was observed only in conditions of 10-fold increase of ligand to protein ratio. Presumably, under these conditions, the ligands possessing the right conformation would be enough to elicit the inhibitory effect against the enzyme whose concentration was reduced to one-tenth compared to the former experiment, where no inhibition was observed.

In conclusion, the competitive inhibition study provided evidence that thiosialosides **1** and **2** were not able to bind to the catalytic site of VCNA, as implied from the lack of any inhibition, at the established ligand concentration (10 mM). However, increasing the

amount of the ligand relative to the enzyme would result in observation of inhibition activity against VCNA, implying the ligand binding to the catalytic site of the enzyme.

### 7.4.3 Implications of inhibition results on ITC interpretation

The competitive inhibition study showed that there was no sign for any significant binding of the thiosialosides **1** and **2** at the catalytic site of VCNA at 10 mM ligand concentration, as implied from the lack of any inhibition activity manifested by the ligands. Obviously, at the ligand concentrations applied in the ITC experiments (0.6 mM for ligand **1**, and 0.5 mM for ligand **2**) there was no binding of the titrated ligands at the catalytic site. Then, it is evident that ligands **1** and **2** interacted only with the sialic acid binding-site in accord with the estimated stoichiometric parameter ( $n = 1$ ), which characterises the 1:1 protein-ligand interaction. Therefore, the thermodynamic parameters drawn from the isothermal titrations of the two ligands were characteristics for the ligand recognition at the N-terminal lectin domain. The affinities of the two ligands towards the sialic acid binding-site were significant ( $K_d$ 's = 37 and 26  $\mu$ M for ligands **1** and **2**, respectively). It is noteworthy to mention that the major contribution to the protein-ligand interaction was due to the sialic acid residues as evident from the STD NMR study of ligand **1** (chapter 8). In fact, such high ligand affinity was in a good correlation with the crystallographic structure, which revealed extensive interactions between the bound  $\alpha$ -sialic acid and the binding site residues (chapter 5). Generally, lectins show weak affinities to monovalent sugars (Dam & Brewer, 2002). Indeed, the micromolar affinity of the thiosialosides studied herein is impressive for such a monomer lectin protein interacting with a monovalent sugar, given that there is no significant role for the second sugar residue of the thiosialosides in binding.

## 7.5 Discussion

The structures of VCNA-complexes revealed the sialic acid binding-site at the N-terminal lectin-like domain (chapter 5). The crystallographic structure displayed the detailed interactions between the bound ligand ( $\alpha$ -sialic acid) and the binding site residues. Nevertheless, crystallographic structure provides only a qualitative description for the ligand affinity towards its target. The quantification of the interaction at the sialic acid binding site was of great importance, as it determined how relevant this particular interaction was under certain conditions. The concern of this chapter was to evaluate the ligand affinity at the newly discovered binding site. Herein, we determined the affinity by applying the ITC technique, which in addition to measuring the affinity characterized the energetics of the interaction under investigation. Ideally, the best ligand to use would be sialic acid itself, either alone or in its natural substrate, e.g., sialyllactose or ganglioside. However, our experience showed that the energetics of the mutarotation equilibrium between the two forms of sialic acid,  $\alpha$ - and  $\beta$ -forms as well as the hydrolysis reaction of the sialyllactose at the catalytic site added complications to the system such that the ITC data became too complicated to interpret. The alternative way was to use a non-hydrolysable substrate such as the thiosialosides, which have been shown to resist the enzymatic hydrolysis of VCNA (Khorlin et al., 1970; Sabesan et al., 1994; Wilson et al., 1999). Thereafter, the two thiosialosides **1** and **2**, both having sialic acid with  $\alpha(2\rightarrow6)$  *S*-ketosidic linkage, were employed in the ITC study.

Based on the analysis of the 3D structure of the sialic acid-binding site with ligand bound, the binding of ligands **1** and **2** at this site was expected. Lower binding of the two thiosialosides was anticipated at the catalytic site, inferred from the early observation that VCNA has more specificity to cleave terminal sialic acid with *O*-ketosidic  $\alpha(2\rightarrow3)$  linkage from its substrate with a rate as twice as that of substrates containing sialic acid

with *O*-ketosidic  $\alpha(2\rightarrow6)$  linkage (Friebolin et al., 1984). Even so, the final word was to be said by the experiments.

The isothermal binding curve obtained for the two thiosialosides revealed a 1:1 protein ligand interaction with binding affinities of 37 and 26  $\mu\text{M}$  for ligands **1** and **2**, respectively. The estimated stoichiometry ( $n = 1$ ) for the interactions indicated that the ligands were recognized at one site rather than the other, which required further investigation. The direct way to check which of the two sites was involved in the interaction was to test the inhibiting power of the studied ligands against VCNA; as the binding of the non-hydrolysable substrate at the catalytic site was expected to manifest itself in the inhibition of the enzyme activity. The inhibition study demonstrated that the two thiosialosides were not able to inhibit at 10 mM ligand concentration. However, increasing the amount of the ligand relative to the enzyme by 10-fold showed weak inhibition activity ( $\approx 20\text{-}25\%$  inhibition). Our interpretation of this result was that thiosialosides **1** and **2** could bind to the catalytic site in certain conformations, which usually exist in a very low population at equilibrium; that is why higher ratio of ligand to enzyme was necessary to observe the inhibition activity. Our view was supported by information from previous study (Sabesan et al., 1994), therein the authors reported that there was no inhibition activity for a thiosialoside, very close to the structure of ligands **2**; however restricting the conformations of the same ligand by a minor change in the structure led to inhibition of the enzyme activity. Of course, better understanding of the thiosialoside interactions with VCNA at the molecular level would be invaluable.

The most important output from the inhibition study was an evidence for the absence of ligand binding at the catalytic domain under the established conditions of ligand concentrations (0.6 and 0.5 mM) in the ITC experiments. It was very unlikely that the

difference of the experimental conditions, pH and the applied protein concentration, between the two experiments would result in a dramatic change of the ligand binding at the catalytic site. Additionally, the results from the inhibition study provided an account for the observed 1:1 protein-ligand interaction in the ITC run.

Evidently, the thermodynamic parameters from the ITC study were characterising the thiosialoside interaction at the lectin domain of the N-terminal. The interaction was driven by the enthalpy change accounting for the major contribution to the free energy change, rather than the unfavourable negative entropy change with its smaller contribution to the overall free energy change. The strength of the interaction was quite significant, and correlated well with the extensive interactions of the bound sialic acid in the X-ray structure. The magnitude of the affinities was determined to be in the medium micromolar range, which is high for such a monomeric lectin protein interacting with a monovalent sugar residue (the NMR results, chapter 8, showed a minor contribution of the N-acetylglucosamine residue of ligand **1** in the binding event). It is more likely that the determined affinity for the thiosialoside ligands could be extrapolated to the real ligand of the sialic acid binding-site. In fact, the micromolar affinity of the sialic acid ligands would be suitable for the function of the lectin domain, which works as a hook to fix the enzyme to the cell surface facilitating the access of the catalytic site to provoke its action. The ligand binding is not too tight, if so the release of the enzyme from its receptor to bind at another site on the cell surface would be hindered, nor too weak, so that it would be able to resist the shear in the small intestine.

## Chapter 8

### NMR studies on VCNA

#### 8.1 Introduction

This chapter is devoted to the NMR spectroscopic study of VCNA. The 1D  $^1\text{H}$  NMR was used to monitor the hydrolysis reaction of the substrate ( $\alpha$ -2,3-sialyllactose) by the action of the enzyme to see whether VCNA itself could form DANA, the enzyme inhibitor, or not. This would support the interpretation of the crystallographic data of the complex of this ligand with the enzyme; see section 5.3.

In addition, the STD (Saturation Transfer Difference NMR spectroscopy) technique was used to map the epitopes of the ligand, Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me, shown to bind specifically at the N-terminal lectin domain as discussed in chapter 7. We also used the STD to screen a small library of common sugars hoping to identify the carbohydrate recognized by the second lectin domain.

An overview of the  $^1\text{H}$  NMR and STD technique is presented at the beginning of the chapter followed by the materials and methods section; the results and discussion are presented in two separate sections.

#### 8.2 Background

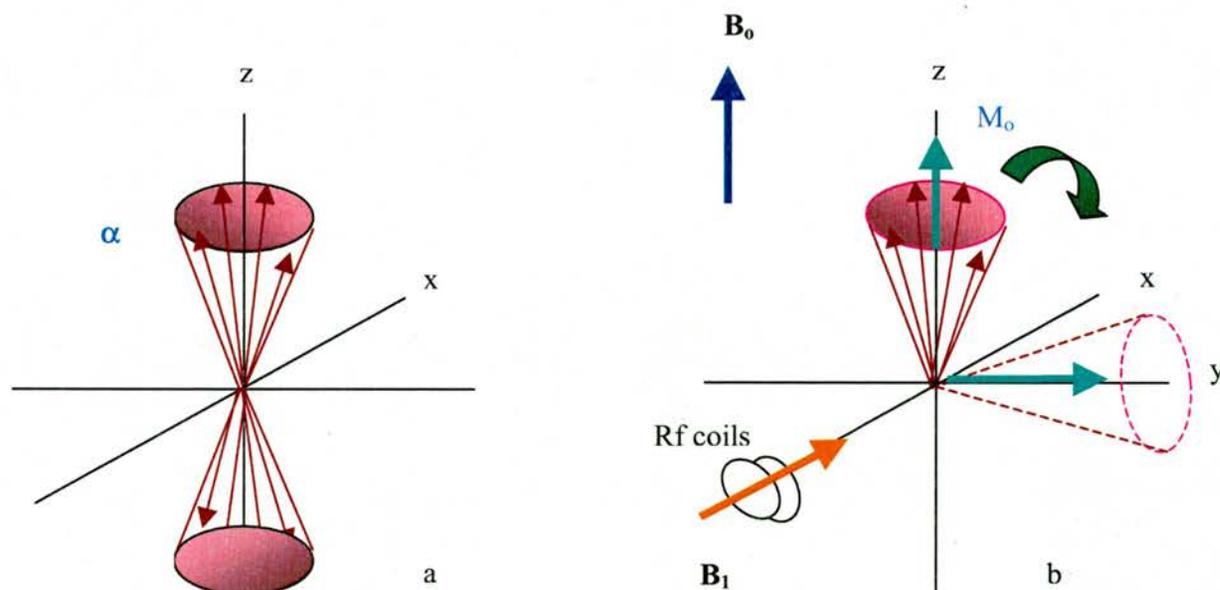
##### 8.2.1 Nuclear magnetic resonance (NMR): Overview

Nuclear magnetic resonance (NMR) is a spectroscopic technique that allows us to detect atomic nuclei and say what sort of environment they are in, within their molecule.

A magnetic field is applied to the sample and those nuclei possessing what is called a nuclear spin interact with the employed field to different extents depending on: the strength of the applied field, the ability of the nuclei to be magnetized, and the local environment of the interacting nuclei (Clayden et al., 2001).

A proton nucleus ( $^1\text{H}$ ) in a strong magnetic field can have two energy levels,  $\alpha$  and  $\beta$ ; the nucleus aligns either with the field (lower energy) or against the field (higher energy). The difference between the two energy states is so small such that the energy from an electromagnetic radiation of radio-wave frequency is sufficient to flip the nucleus from the lower to the higher energy states. In a population of  $^1\text{H}$  nuclei under the applied magnetic field, the excess of nuclei in the ground energy state is very small. Irradiating this population from a radiofrequency source with a frequency matching the energy difference between the two energetic states of the nuclei will disturb their equilibrium by bringing more nuclei from the lower to the higher energy state, see Figure 8.1. Upon switching off the radiofrequency source, the nuclei will return to their equilibrium state and flip from the higher to the lower state emitting a tiny pulse of radiofrequency electromagnetic radiation to be detected by a radio receiver. After acquiring the spectrum, the data are Fourier transformed from the time domain to the frequency domain and the results are displayed in the form of intensity (i.e. number of absorptions) against frequency. The resonance frequencies are expressed in terms of the chemical shift  $\delta$  that is given in ppm (part per million) unit, which is related to the difference between the resonance frequency ( $\nu$ ) of the nucleus in question and that of a reference resonance ( $\nu^o$ ) as:

$$\delta = \frac{\nu - \nu^o}{\nu^o} \times 10^6 \quad (8.1)$$



**Figure 8.1.** Basics of NMR (a) In the absence of an externally applied field, there are equal numbers of  $\alpha$  and  $\beta$  spins at random around z-axis; the net magnetization is zero. (b) An external magnetic field  $B_0$  is applied along z-axis and more nuclear spins aligned parallel to the applied field resulting in a net magnetization  $M_0$ . When radiofrequency coils along the x-axis generate a magnetic field  $B_1$ , the  $M_0$  vector will precess around y-axis.

### 8.2.2 Protein-ligand interaction and STD NMR

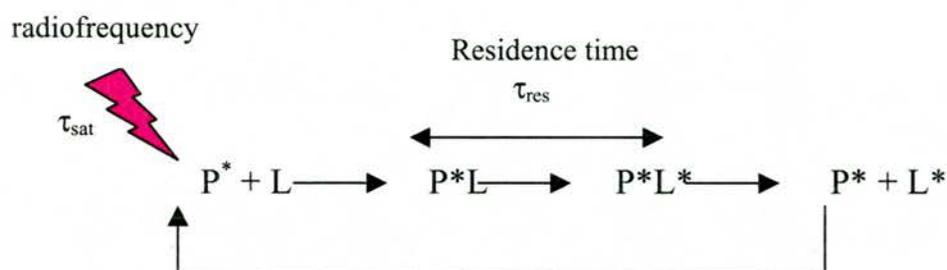
Gaining insight into the protein-ligand interactions become possible by new NMR techniques (Moore, 1999; Wemmer & Williams, 1994; Zartler et al., 2003). The experiments identify the binding events either by looking at the resonance signals from the ligand or the protein (Meyer & Peters, 2003). Ideally, the picture is more complete when both techniques are combined together. Upon binding of the ligand to its protein, several changes can be detected by NMR as a gauge for binding. For example, chemical-shift changes of the target protein, changes in relaxation time of the signals from the ligand, changes of diffusion constants of the ligands, changes of the Nuclear Overhauser

Effects (NOEs), or exchange of saturations (Saturation Difference NMR spectroscopy, STD NMR), which was the method applied in our study.

In STD (Meyer et al., 1997; Mayer & Meyer, 1999), the interaction between a protein and its ligand, which is equilibrating between free and bound states, is studied by selectively saturating a single resonance of the protein, typically for 1 to 2 seconds. No NMR signal is observable for the saturated resonance. The saturation spreads over the entire protein by spin diffusion through the system of protons—tightly coupled by dipole-dipole interactions. The saturation transfers to the protons of the bound ligand as illustrated in (scheme 8.1); the degree of ligand saturation depends on the residence time of the ligand in its binding pocket. For those ligand protons interacting with protons of the protein, a decrease in intensity is observed. It is important to set the irradiation frequency to a value in the region of the (on-resonance) spectrum where there is no resonance of the ligand protons. The saturation frequency may be placed around  $-1$  ppm, or in the aromatic region ( $\delta = 7-8$  ppm), or even further in the region ( $\delta = 11-12$  ppm). For easier identification of the attenuated signals, a second (off-resonance) spectrum is collected by setting the irradiation frequency to a resonance away from any signal of either the protein or the ligand (e.g. at  $\delta = 40$  ppm) and subtracted from the (on-resonance) spectrum. The resulting difference spectrum (STD) will contain only the saturated signals of the ligand protons, which are in close proximity to the protein. The STD experiment requires the ligand to be in excess (up to 100-fold) over the protein, which allows one to work with low protein concentration (Meyer & Peters, 2003).

The STD technique was proved to be a useful tool for mapping the epitopes of a binding ligand (Mayer & Meyer, 2001), as demonstrated in several studies. For example, the binding specificity of Lewis<sup>b</sup>-hexasaccharide to *Aleuria aurantia* agglutinin (AAA)

was studied by STD NMR (Haselhorst et al., 2001). Only the two fucosyl residues were found to bind the AAA. Also, STD was shown to be useful technique in library screening; Mayer & Meyer showed that only *N*-acetylglucosamine (GlcNAc) among a mixture of non-binding sugars was able to bind to the lectin, wheat germ agglutinin (WGA) (Mayer & Meyer, 1999). STD NMR spectroscopy should enable the identification of ligand binding to proteins with dissociation constants  $K_d$  between  $10^{-3}$  and  $10^{-8}$  (Mayer & Meyer, 1999).



**Scheme 8.1.** Principles of STD NMR: the protein  $P$  is saturated by the saturation pulse for duration  $\tau_{\text{sta}}$ , the saturation transfers to the bound ligand  $L$  and the degree of saturation depends on the residence time  $\tau_{\text{res}}$ ; then the ligand is released into the solution and the saturation transfers from the protein to another unsaturated ligand.

## 8.3 Materials and methods

### 8.3.1 Materials

The compounds: DANA,  $\alpha$ -2,3-sialyllactose, Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me, sucrose,  $\alpha$ -D-lactose, D-mannose, D-arabinose, L-arabinose, L-(-)-fucose, D-ribose, D-glucouronic acid, D-xylose, N-acetyl-D-glucosamine, N-acetyl-D-mannoasamine, and hyaluronic acid tetrasaccharide were gifts from Prof. Mark von Itzstein (Institute for Glycomics, Griffith University, Queensland, Australia). Heparin sulfate disaccharide, D<sub>2</sub>O (99.9 % D), deuterated sodium acetate, and deuterated acetic acid, were purchased

from Sigma Chemicals Co. The enzyme VCNA was produced using the new developed protocol as described in chapter 3.

## **8.3.2 Methods**

### **8.3.2.1 General**

**Sample preparations:** Deuterated water (D<sub>2</sub>O, 99.9%) was used in all preparations. All samples were prepared in the (NMR buffer) system composed of: 50 mM deuterated sodium acetate, pH 5.5 (adjusted using the deuterated acetic acid), 150 mM NaCl, and 10 mM CaCl<sub>2</sub>. A stock solution of the pure enzyme (10 mg/ml) in its storage buffer (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, NaN<sub>3</sub>) was dialyzed exhaustively against the NMR buffer and stored as a stock at final concentration of (8.0 mg/ml) in the fridge; the enzyme concentration was determined by Bradford assay (see section 3.2.4). The enzyme sample was diluted to the required concentrations for the different experiments.

**Spectrometer:** All the NMR spectra were collected on the 600 MHz Brüker Avance spectrometer, at 298 K, and data were processed using the software XwinNMR 3.5 on an SGI O2 workstation. Before acquiring any spectrum and upon transferring the NMR tube holding the sample into the magnetic field of the spectrometer, minor adjustments to the shims using a good shim file, and fine tuning of the deuterium-lock signal were made for each new sample placed into the magnet.

### **8.3.2.2 Hydrolysis of $\alpha$ -2,3-sialyllactose by VCNA**

**NMR sample:** A sample solution of the ligand,  $\alpha$ -2,3-sialyllactose, with concentration 30 mM was prepared in the NMR buffer pH 5.5, the same buffer used in the preparation of

the enzyme sample. The ligand solution (200  $\mu$ l) was added to 300 $\mu$ l of the stock solution of the enzyme (8.0 mg/ml) in the NMR tube and the final volume was brought to 600  $\mu$ l using the NMR buffer to give a final concentration of 10 mM of the ligand and 4.0 mg/ml (~48  $\mu$ M) of the enzyme. The sample was then transferred immediately to the magnetic field of the NMR spectrometer. Reference samples of 1mM  $\alpha$ -2,3-sialyllactose and 10 mM DANA solutions were prepared in the same NMR buffer.

**Acquiring the spectrum:** All 1D  $^1\text{H}$  NMR spectra were obtained by collecting 128 scans over a spectral width of 7200 Hz with 32 K data points, and relaxation delay 2s. To suppress the signal of the trace water in the samples, water presaturation was employed in the pulse sequence; spin-lock was used to cancel the background spectrum of the enzyme. For the enzyme/ligand sample, thirteen spectra were acquired: the first spectrum was acquired after 0.5 hr from the mixing time of the enzyme with its substrate followed by two spectra at intervals of 0.5 hr, subsequently ten spectra were acquired at 1 hr intervals. Reference spectra for the  $\alpha$ -2,3-sialyllactose and DANA samples were collected using the same NMR parameters as the enzyme/ligand sample.

### 8.3.2.3 STD NMR of the Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me

**Sample preparation:** The sample containing 27  $\mu$ M VCNA (equivalent to 16 nmol enzyme) and 300  $\mu$ M of the ligand, Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me **5**, scheme 8.4, (equivalent to 180 nmol) in a total volume of 600  $\mu$ l was prepared in the NMR buffer. A reference sample of the ligand **5** alone with concentration 300  $\mu$ M was prepared for the 1D  $^1\text{H}$  NMR and 2D  $^1\text{H}$ — $^1\text{H}$  COSY (homonuclear correlation spectroscopy), used to identify the signals of the ligand in the spectra.

**Acquiring the spectrum:** In the STD experiment, saturation was achieved by a series of equally spaced Gaussian bell-shaped pulses of 50 ms duration separated by 1 ms delay, and a total saturation time of ca. 2s. The frequency of the protein irradiation was set on 4315 Hz ( $\delta = 7.2$  ppm) in the aromatic region; the frequency 20,000 Hz ( $\delta = 33.3$  ppm) was set for the offset irradiation in the off-resonance spectrum. The free induction decays (FIDs) with protein saturation (on-resonance spectra) were recorded alternating with reference FIDs (irradiation in a peak-free region, off-resonance spectra) and directly subtracted to obtain the final STD spectra. Total number of 2k scans was collected for on- and off-resonance spectra each with 32 K data points over a spectral width of 7200 Hz and with a relaxation delay of 1s. A spin-lock pulse (10 ms) was applied to remove protein signals from the background. Water saturation was achieved by using the WATERGATE sequence (Piotto et al., 1992). The off-resonance reference spectrum of the same sample was acquired employing the same parameters of the STD spectrum. A control spectrum with both on- and off-resonance set on an irradiation frequency of 4315 Hz was also obtained; no signals should appear in that spectrum.

The 1D  $^1\text{H}$  NMR spectrum of the ligand sample was obtained by collecting 64 scans with 32 K data points over a spectral width of 7200 Hz, and with a relaxation delay of 2s. To achieve solvent suppression a selective low-power irradiation at frequency 2823.6 Hz ( $\delta = 4.7$  ppm), corresponding to the water signal, was applied during the relaxation delay of 2s. The 2D  $^1\text{H}$ — $^1\text{H}$  COSY spectrum was acquired as 256( $t_1$ ) x 2048( $t_2$ ) data points and 8 scans for each FID were averaged. Water suppression was achieved by presaturation.

#### 8.3.2.4 Screening a small library of Sugars

**Sample preparation:** The stock of 10 mM solutions of the following carbohydrates (scheme 8.5): D-(+)-sucrose **6**, D-lactose **7**, D-(+)-mannose **8**, D-(-)-ribose **9**, D-(+)-

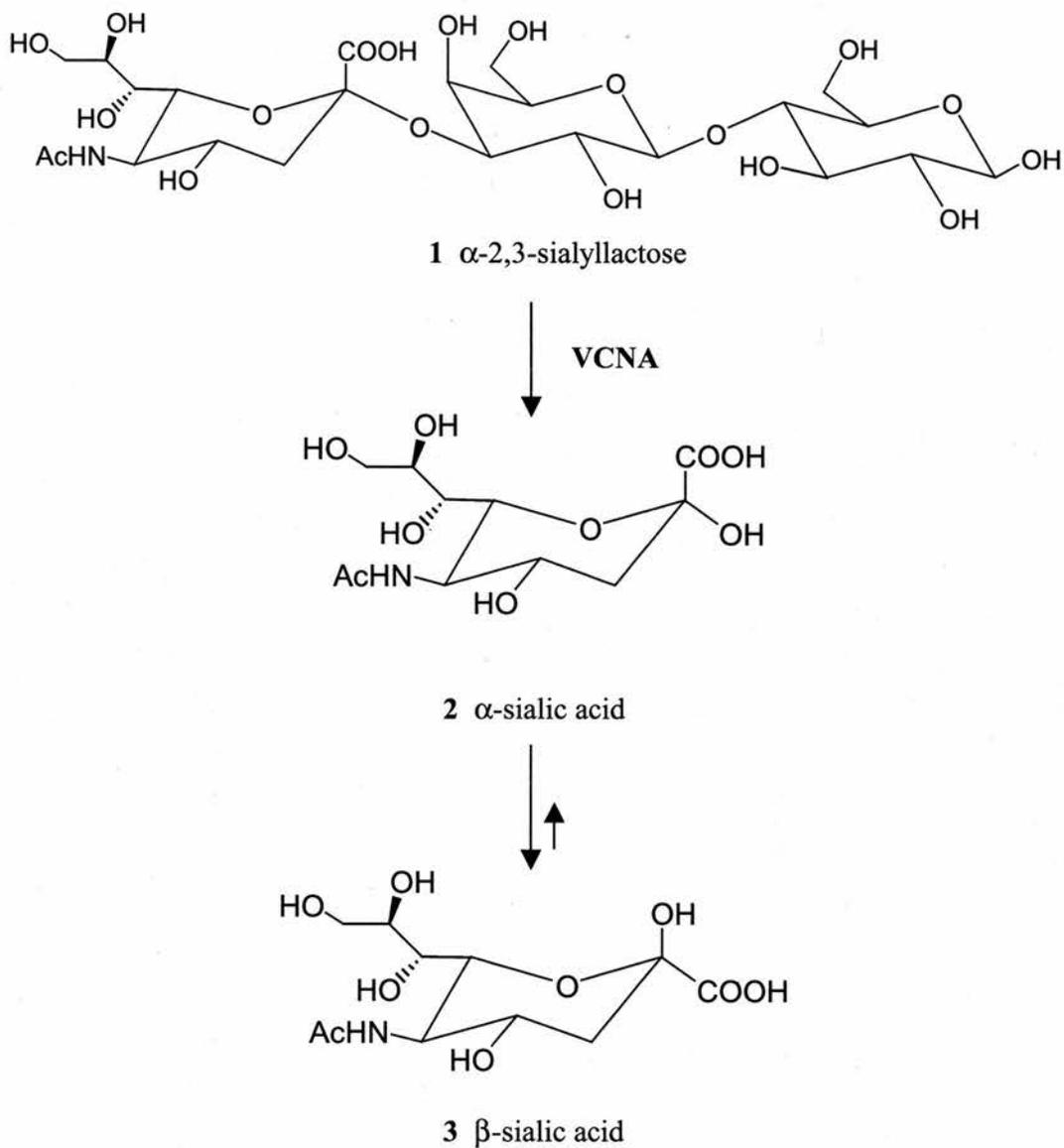
xylose **10**, L-(-)-fucose **11**, D-(-)-arabinose **12**, L-(-)-arabinose **13**, D-glucouronic acid **14**, N-acetyl-D-glucosamine **15**, N-acetyl-D-mannoasamine **16**, hyaluronic acid tetrasaccharide **17**, were prepared in the NMR buffer. A stock solution of 17.4 mM was prepared for heparin sulfate disaccharide **18** in the NMR buffer. The library compounds **6-17** were divided into two mixtures: **6 - 10** and **11 - 17** for STD screening. A sample of each mixture was prepared for an STD NMR run such that it contained 11 nmol enzyme and 20-fold excess of individual carbohydrates in the mixture (ratio of 20:1 ligand to enzyme). This was equivalent to concentrations of 18.6  $\mu\text{M}$  enzyme and 0.37 mM of individual sugars (appropriate volumes added from the stock solutions of the carbohydrate sugars); the final volume of each sample in the NMR tube was brought to 600  $\mu\text{l}$ . Similarly, another sample for each mixture, with the thiosialoside **5** added, was prepared in the same way and applying the same concentrations (ratios 20:1, ligand to enzyme) to be used as a positive control for the STD experiment. The compound **18** was screened in a separate STD experiment applying the ratios of (10:1 and 50:1, ligand to enzyme).

**Acquiring the spectrum:** The STD spectra for the library screening were obtained in the same way as described in section 8.3.2.3 and using the same NMR parameters.

## 8.4 Results

### 8.4.1 Hydrolysis of $\alpha$ -2,3-sialyllactose by VCNA

The hydrolysis of  $\alpha$ -2,3-sialyllactose **1** by the action of VCNA enzyme (scheme 8.2) was investigated in previous studies using  $^1\text{H}$  NMR (Friebolin et al., 1984; Friebolin et al., 1981; Wilson et al., 1999). The enzymatic hydrolysis was found to proceed with a first order rate constant ( $k = 0.62 \times 10^4 \text{ s}^{-1}$ ), and the first product of the hydrolysis was proved

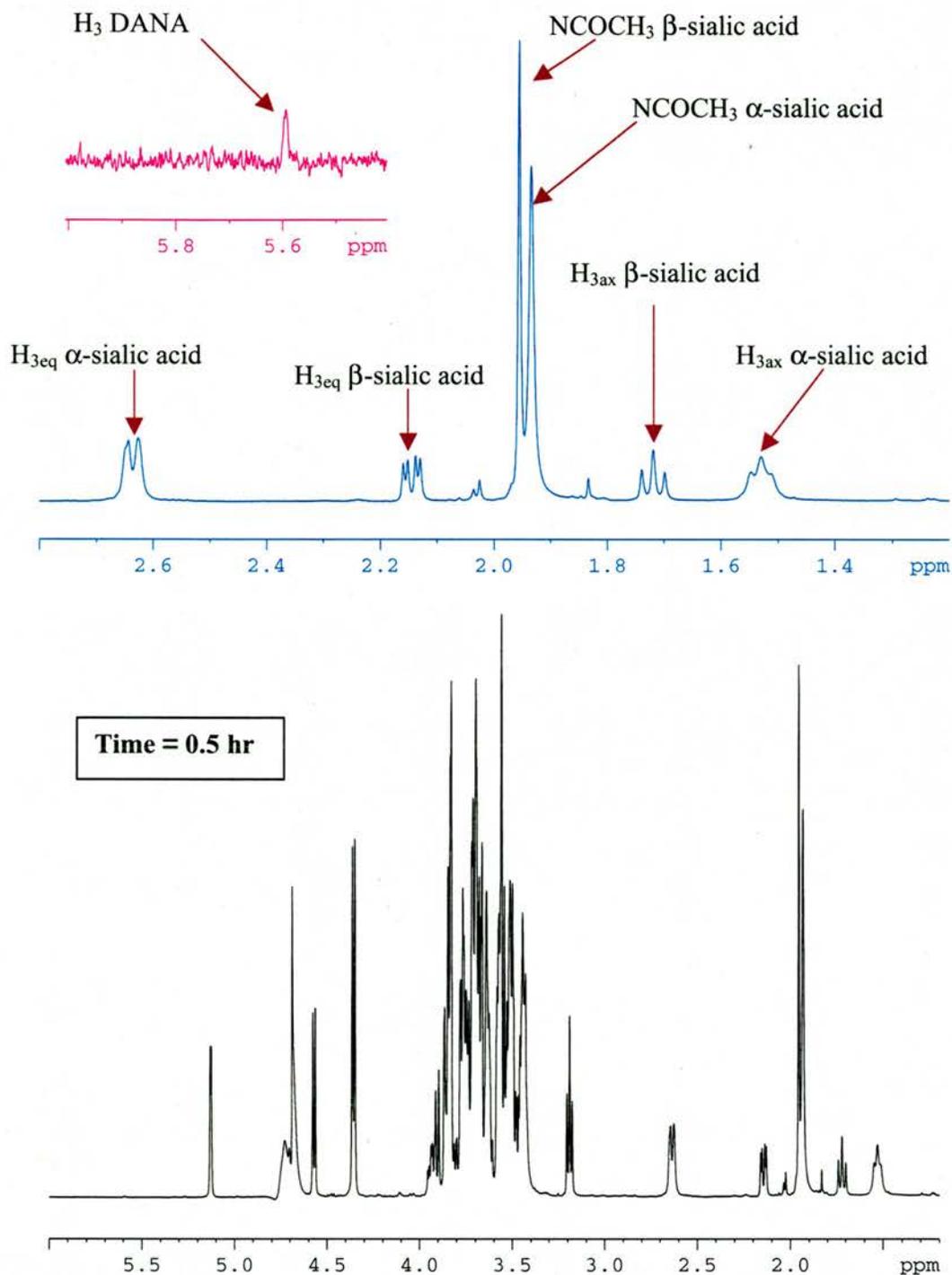


**Scheme 8.2.** The enzymatic hydrolysis of  $\alpha$ -2,3-sialyllactose **1** by the action of the enzyme VCNA; the first product of hydrolysis is  $\alpha$ -sialic **2** acid, which mutarotates to  $\beta$ -sialic acid **3**.

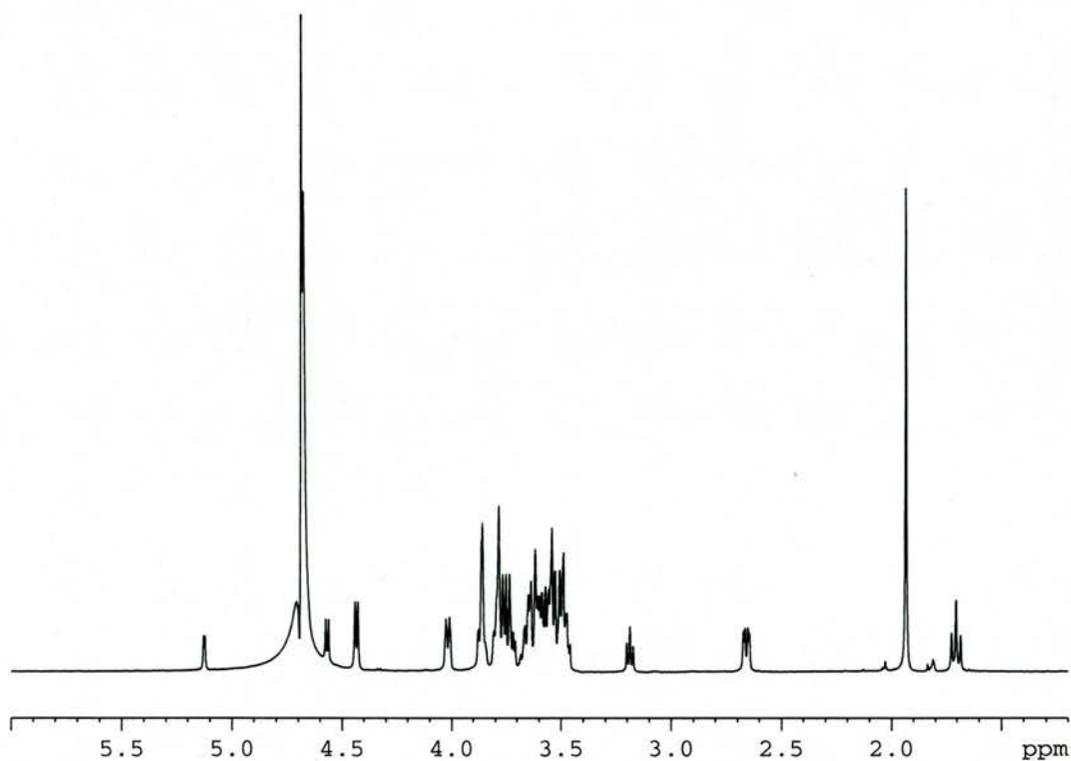
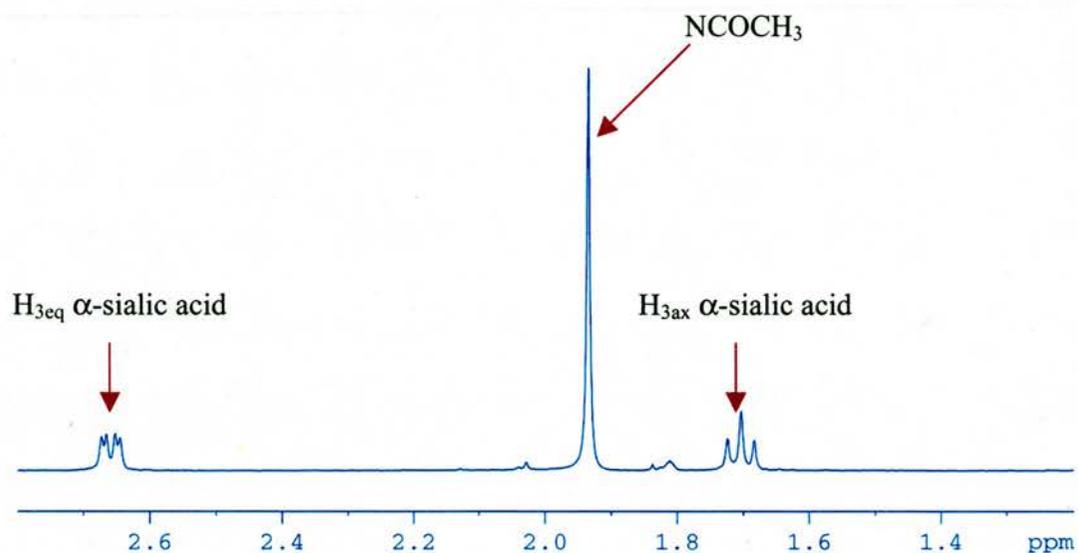
to be  $\alpha$ -sialic acid **2**, which is then mutarotated to the thermodynamically more stable  $\beta$ -sialic acid **3** upon its release from the enzyme active site. In our experiment we monitored the hydrolysis reaction by  $^1\text{H}$  NMR employing high concentration of the enzyme ( $\sim 48$   $\mu\text{M}$ ) and the substrate (10 mM), over a long period of time (11.5 hours). The first spectrum of the hydrolysis reaction collected after 0.5 hr is shown in figure 8.2. The spectrum revealed characteristic signals for the  $\alpha$ - and  $\beta$ -sialic acids as shown in the expanded region ( $\delta = 1.2 - 2.8$  ppm) at the top of figure 8.2, and listed in table 8.1. These resonances characterize: the axial and equatorial protons at C3 ( $\delta[\alpha\text{-H}_{3\text{ax}}] = 1.53$ ,  $\delta[\alpha\text{-H}_{3\text{eq}}] = 2.63$ ;  $\delta[\beta\text{-H}_{3\text{ax}}] = 1.71$ ,  $\delta[\beta\text{-H}_{3\text{ax}}] = 2.14$  ppm), and the three protons of the N-acetyl group at C5 ( $\delta[\alpha\text{-NCOCH}_3] = 1.93$ ;  $\delta[\beta\text{-NCOCH}_3] = 1.96$  ppm). From the reference spectrum of the unhydrolysed substrate  $\alpha$ -2,3-sialyllactose (figure 8.3), it can be shown that the characteristic signals of the  $\alpha$ -sialic acid moiety ( $\delta[\alpha\text{-H}_{3\text{ax}}] = 1.70$ ,  $\delta[\alpha\text{-H}_{3\text{eq}}] = 2.66$ ,  $\delta[\alpha\text{-NCOCH}_3] = 1.93$  ppm, see table 8.1) are impeded underneath the signals of the free  $\alpha/\beta$ -sialic acids in the spectrum displayed in figure 8.2.

**Table 8.1** Chemical shifts of the characteristic signals of  $\alpha/\beta$ -sialic acid

	$\text{H}_{3\text{ax}}$	$\text{H}_{3\text{eq}}$	$\text{NCOCH}_3$
$\alpha$ -sialic acid	1.53	2.63	1.93
$\beta$ -sialic acid	1.71	2.14	1.96
$\alpha$ -2,3-sialyllactose	1.70	2.66	1.93



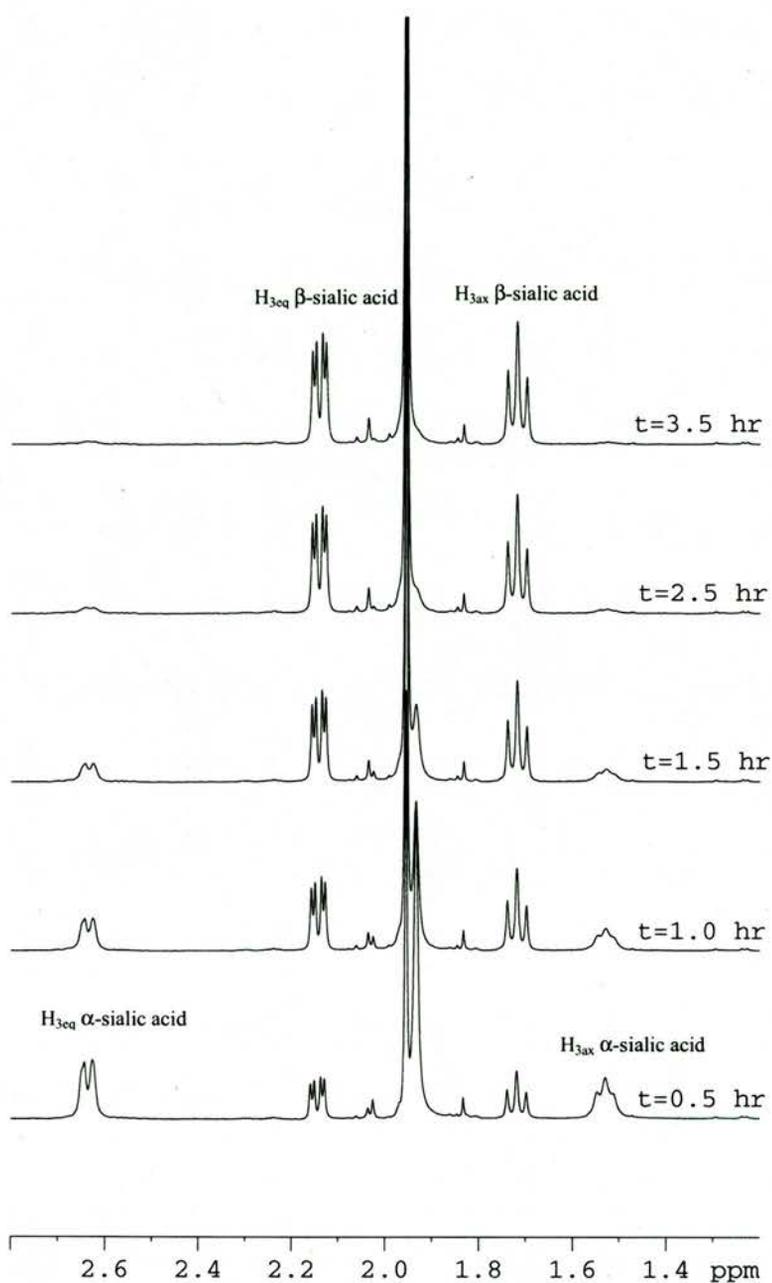
**Figure 8.2.** The spectrum of the hydrolysis of 10 mM  $\alpha$ -2,3-sialyllactose, in presence of 4.0 mg/ml VCNA, collected 0.5 hr after preparing the sample. The expanded region ( $\delta = 1.2$  to 2.8 ppm) is drawn in blue and the characteristic signals of  $\alpha/\beta$ -sialic acid are labelled. The amplified region ( $\delta = 5.40$  to 6.0 ppm) showing the signal at  $\delta = 5.59$  corresponding to the  $H_3$  proton of the double-bonded C3 in DANA is drawn in pink.



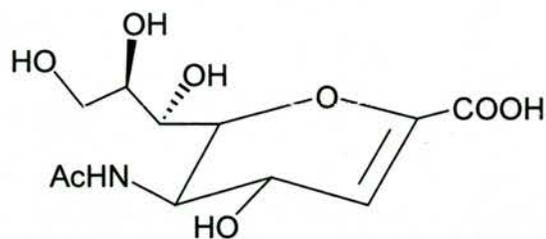
**Figure 8.3.** Reference spectrum (bottom) of 1 mM  $\alpha$ -2,3-sialyllactose in the NMR buffer (50 mM deuterated sodium acetate pD 5.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>). The expanded region ( $\delta = 1.2 - 2.8$  ppm) showing the characteristic signals for the protons on C3 and NCOCH<sub>3</sub> of the sialic acid moiety is drawn in blue colour (top).

The stacked spectra in figure 8.4 represent snapshots of the enzymatic hydrolysis reaction: the first three spectra were collected at 0.5 hr time interval and the next two spectra were obtained at 1 hr interval; the rest of 8 spectra collected at one-hour interval are not shown in the figure for simplicity. As the reaction proceeded, the resonances of  $H_{3ax}$  and  $H_{3eq}$  protons of the  $\alpha$ -conformer ( $\alpha$ -sialic acid:  $\delta[H_{3ax}]=1.53$ ,  $\delta[H_{3eq}]=2.63$ ) decreased in intensity and finally disappeared after 3.5 hr with a concomitant increase in intensity for the resonances of the axial and equatorial protons in the  $\beta$ -form ( $\beta$ -sialic acid:  $\delta[H_{3ax}]=1.71$ ,  $\delta[H_{3eq}]=2.14$  ppm). This is consistent with the product retaining the anomeric configuration of the substrate, followed by mutarotation of the produced  $\alpha$ -sialic acid upon its release as stated in previous studies (Friebolin et al., 1980). Also, the peak with the chemical shift ( $\delta = 1.993$ ) ppm corresponding to  $-NCOCH_3$  group in the free  $\alpha$ -sialic acid and sialic acid moiety of  $\alpha$ -2,3-sialyllactose diminished as the time evolved due to the consumption of the substrate and of the free  $\alpha$ -form of sialic acid until it had almost disappeared after 3.5 hour when all of the substrate was exhausted.

Amplifying the spectral region ( $\delta = 5.4$  to  $\delta = 6.0$  ppm) in figure 8.2, surprisingly, revealed a tiny peak with a chemical shift ( $\delta = 5.59$  ppm) characteristic for the proton on the double bonded C3 of DANA 4 (scheme 8.3) as was confirmed from the reference spectrum of the 10 mM DANA solution in the same NMR buffer (see figure 8.5). The intensity of the tiny peak was clearly above the background of the spectrum indicating it was a real signal rather than a spectral noise. Investigating the same region in the spectra collected over the whole period of the experiment as displayed in the stacked spectra of figure 8.6 (the last two spectra were omitted from the figure for simplicity) showed the same peak at  $\delta = 5.59$  ppm with varying intensity. The intensity of the peak ( $\delta = 5.59$  ppm) increased with time and reached its maximum after about 1.5 hours before it started to decrease till it could not be detected from the background of the spectrum after about

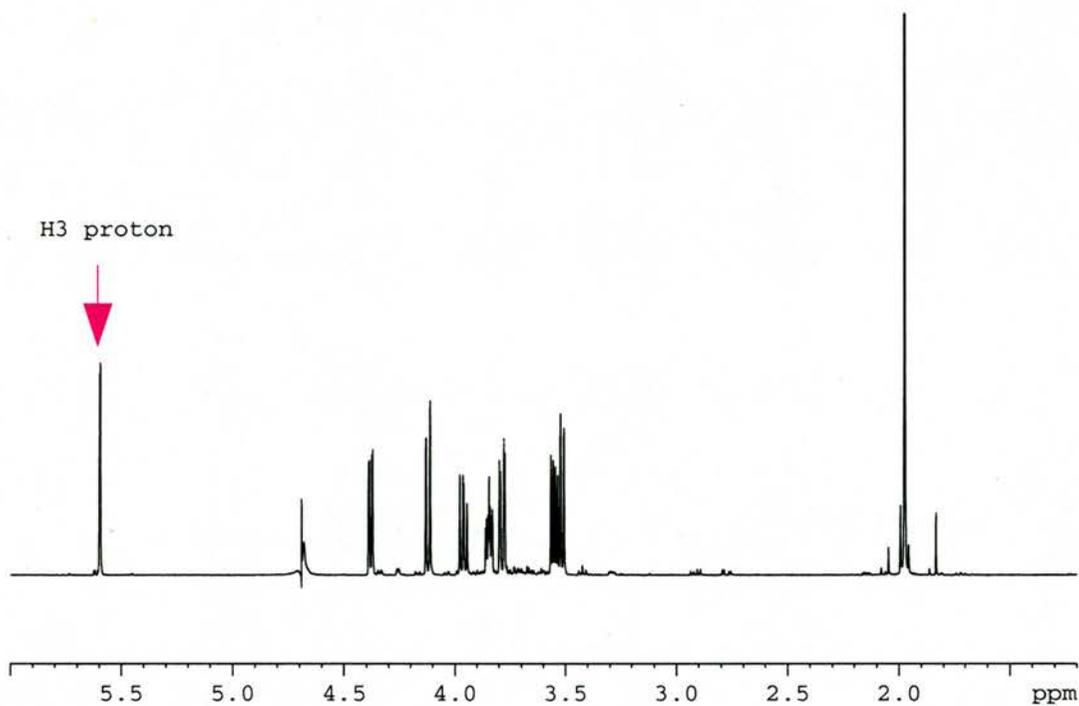


**Figure 8.4.** Stacked spectra of the hydrolysis of  $\alpha$ -2,3-sialyllactose (10 mM) in presence of VCNA (4.0 mg/ml) showing only the region ( $\delta = 1.2 - 2.8$  ppm) with the characteristic resonances for  $\alpha$ - and  $\beta$ -sialic acid are labelled; the spectra collected at the time indicated on each spectrum.

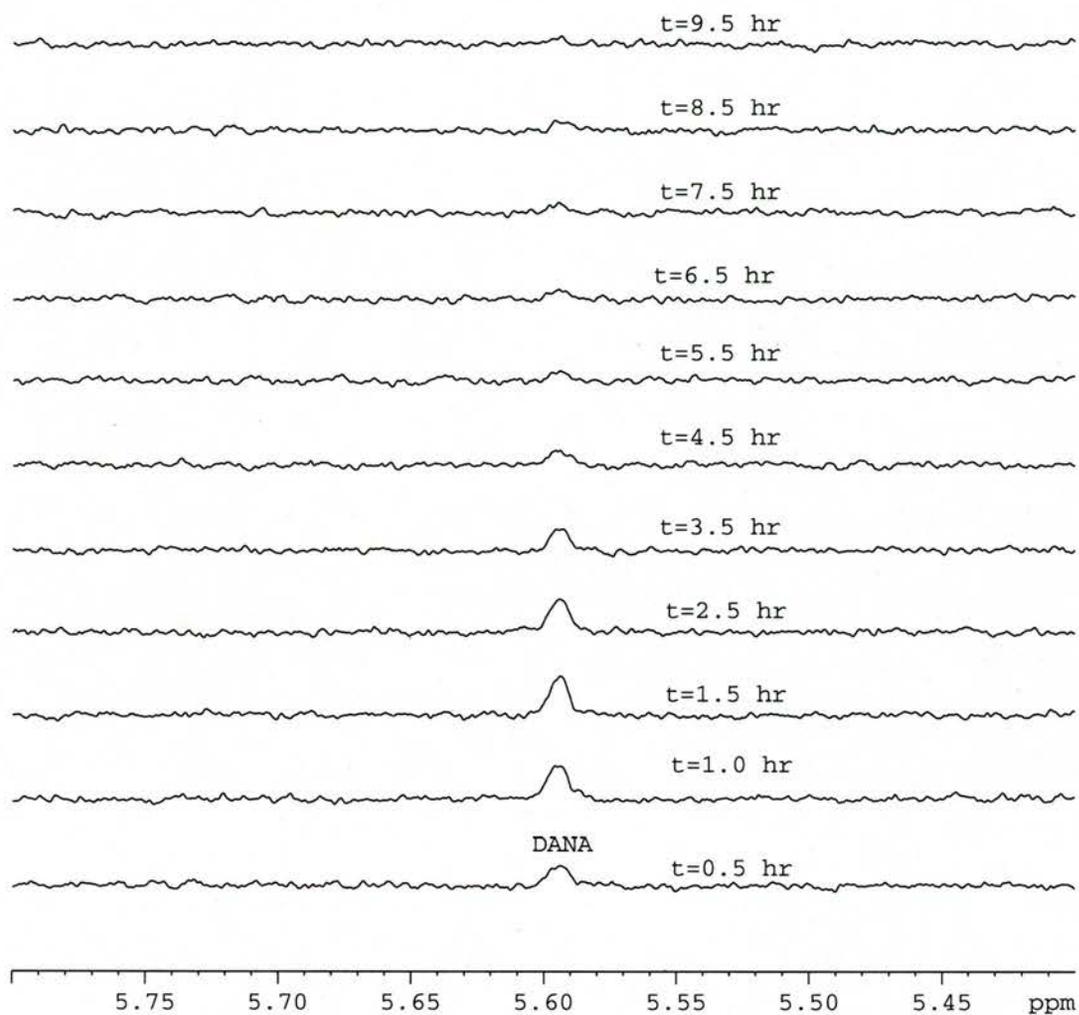


4 DANA

Scheme 8.3



**Figure 8.5.** Reference spectrum of 10 mM DANA solution in the NMR buffer, same buffer used in preparation of samples in the main experiment. The resonance of H3 at chemical shift  $\delta = 5.59$  ppm is labelled in the figure.



**Figure 8.6.** The expanded region ( $\delta = 5.4 - 6.0$  ppm) of the stacked spectra for the enzymatic hydrolysis reaction of 10 mM solution  $\alpha$ -2,3-sialyllactose in presence of VCNA 4.0 mg/ml collected at the time indicated on each spectrum.

9.5 hours. The identification of the signal at  $\delta = 5.59$  ppm corresponding to the proton at the doubly bonded C3 of DANA proved the formation of DANA during the hydrolysis of  $\alpha$ -2,3-sialyllactose by the action of VCNA. The DANA was formed in a very small amount ( $< 1\%$ ) as implied from the size of the signal. The variation of the signal intensity indicated that the DANA was being formed during the hydrolysis reaction and was not pre-existing as a contaminant. The decrease of the signal after reaching its maximum could be interpreted as the rehydration of the formed DANA by the reverse action of the enzyme itself.

It has been reported before that the influenza neuraminidase can form its own inhibitor (DANA) during the course of the hydrolytic reaction of the substrate  $\alpha$ -2,3-sialyllactose; DANA was shown to be rehydrated by the action of the enzyme (Burmeister et al., 1993). This is the first example of a bacterial neuraminidase shown to be able to form its own inhibitor as in the case of influenza virus neuraminidase.

The formation of DANA by the action of VCNA on its substrate  $\alpha$ -2,3-sialyllactose was a supporting evidence for the interpretation of the crystallographic data of the 3'-sialyllactose-complex structure (see section 5.3.5), where the DANA molecule was fitted to the electron density at the catalytic site. Also, the formation of DANA during the hydrolysis reaction can give a rationalization for the inhibition activity of sialic acid when applied at high concentration as reported in an earlier study (Mohr, 1960). This observed inhibition was thought to be due to the binding of  $\beta$ -sialic acid, which has a lower affinity for the enzyme. According to our results, the sialic acid inhibition could be explained by the formation of the enzyme inhibitor DANA upon the enzymatic action on sialic acid.

The exact mechanism used by the enzyme to form DANA during the action on its substrate is unknown. However, in the case of influenza neuraminidase, it was suggested that DANA could be formed by subtracting a proton from C3 of the sialosyl cation

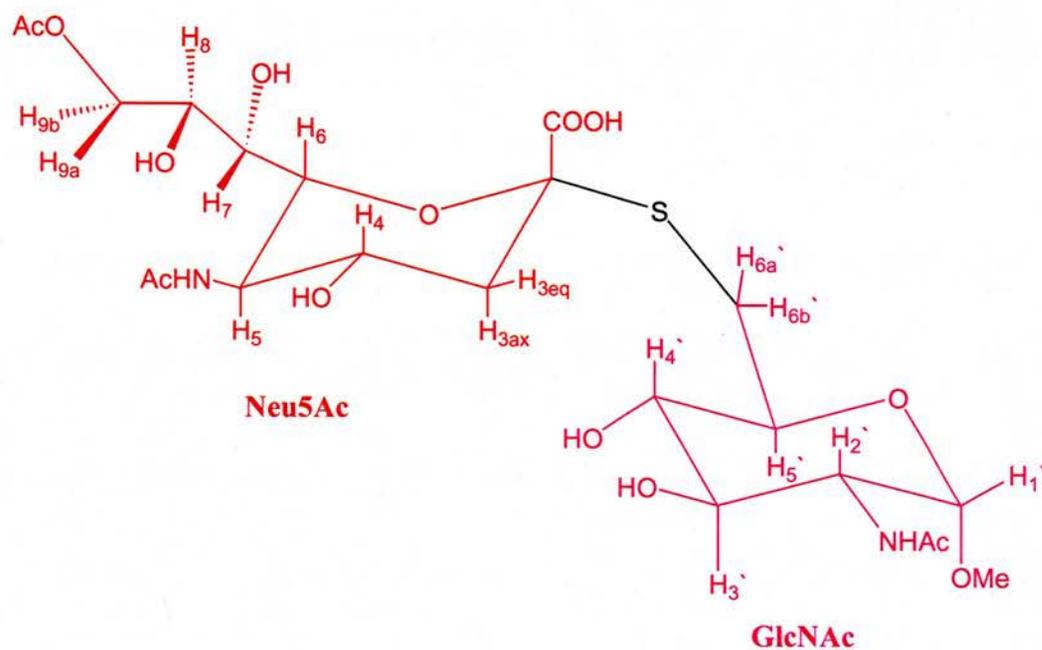
species (thought to be the transition state), formed during the hydrolysis reaction (Chong et al., 1992). Another interesting question to be addressed is why does the enzyme need to form its own inhibitor? This addressed question will be discussed further in section 8.5.

#### 8.4.2 STD NMR of the Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me

It was shown in the previous chapter that the thiosialoside Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me **5** (scheme 8.4) binds specifically to the N-terminal lectin like domain in the VCNA enzyme with a dissociation constant in the micromolar range. The epitope mapping of the binding ligand was obtained from the STD NMR experiment. It is of interest to determine whether the N-acetylglucosamine part of the disaccharide substrate has any contribution in the binding process or not.

##### 8.4.2.1 1D <sup>1</sup>H NMR spectrum of the ligand

In the STD NMR experiment we monitored the change in the spectrum of the interacting ligand; therefore, it was a prerequisite to assign the reference spectrum of the ligand in the NMR buffer and to identify the different protons of the ligand, shown in the scheme 8.4. The figure 8.7a displays the 1D <sup>1</sup>H-NMR spectrum of **5**. For easier identification of the ligand-proton signals, the spectral regions: ( $\delta = 1.2 - 2.2$ ;  $\delta = 2.2 - 3.4$ ; and  $\delta = 3.4 - 4.8$  ppm) are expanded in the figures 8.7b-d, in the same order. The assigned signals are listed in the table 8.2. The assignment from the 1D spectrum was confirmed by the 2D <sup>1</sup>H—<sup>1</sup>H COSY spectrum of the ligand **5**, see figure 8.8. Chemical shifts of **5** were in accordance with values reported previously by the compound provider

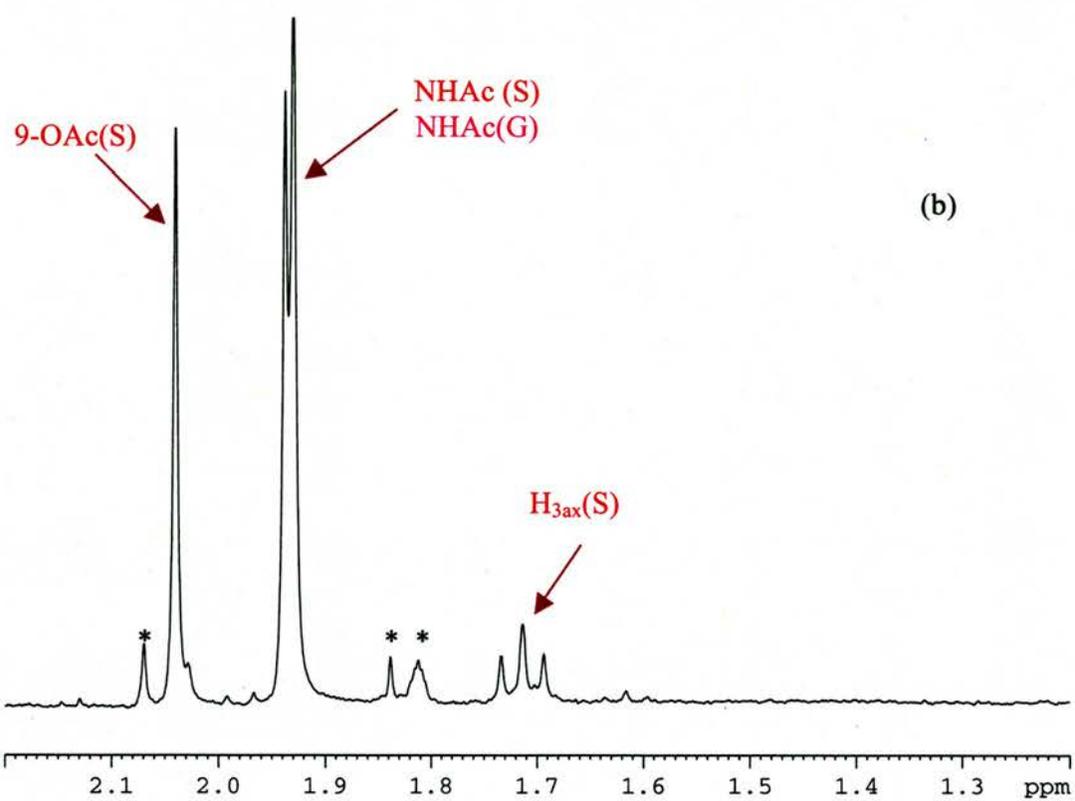
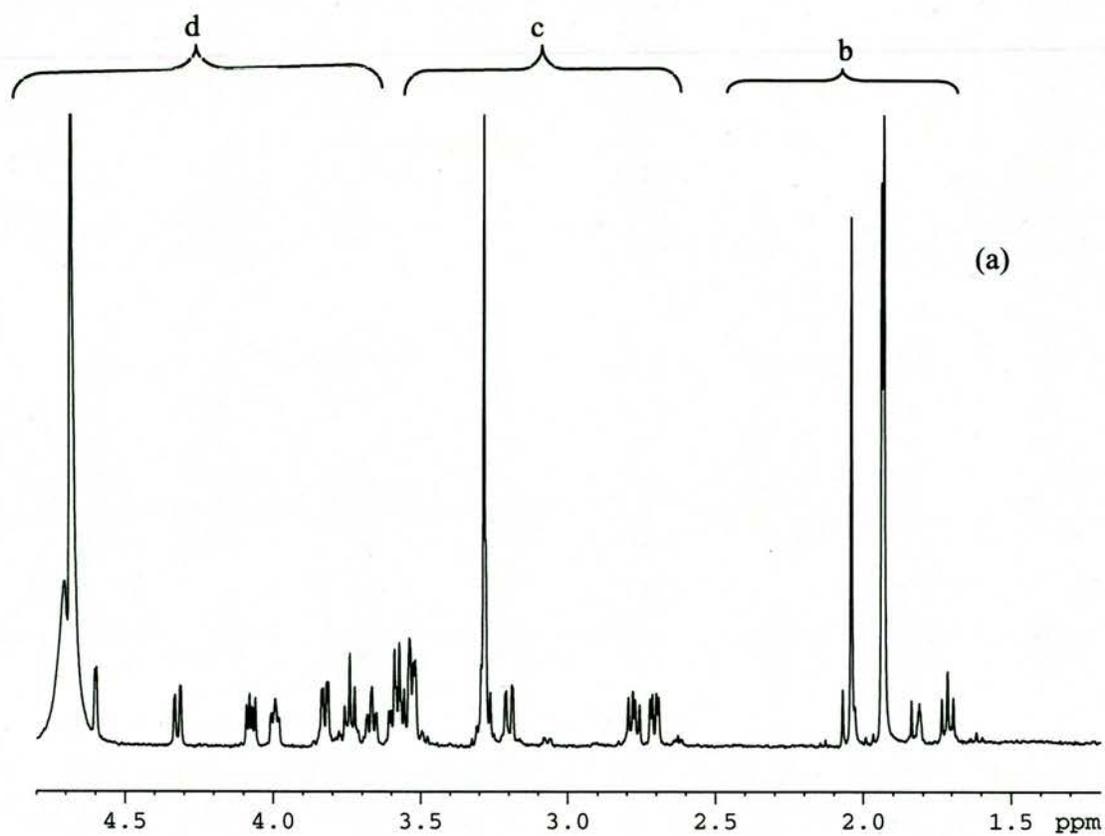


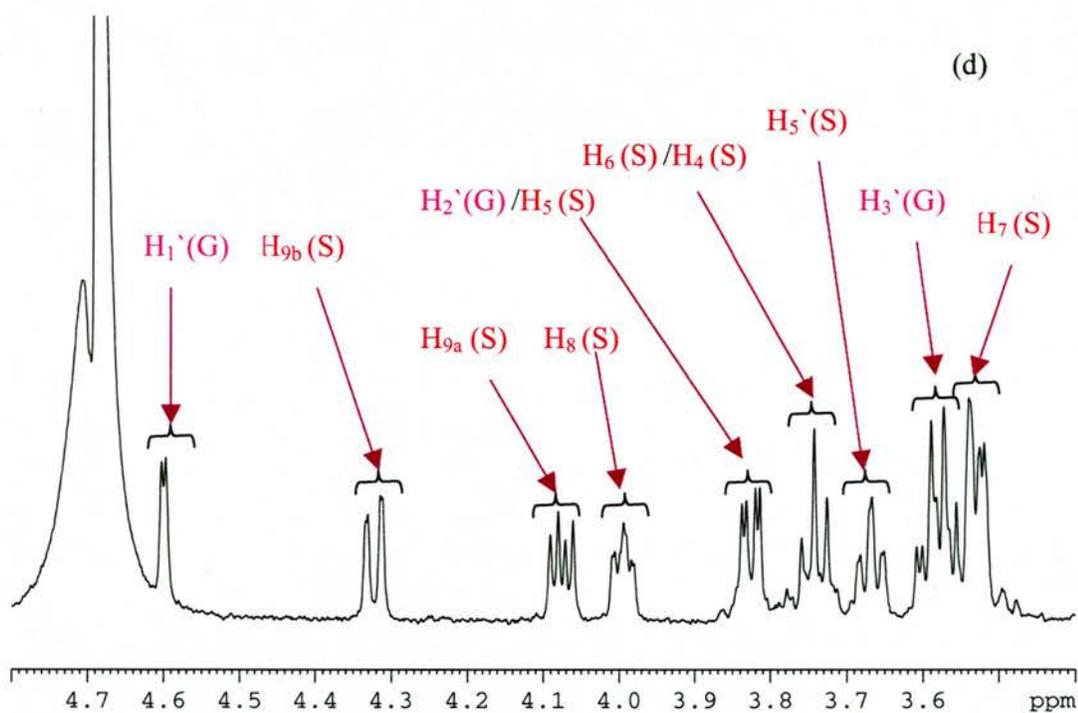
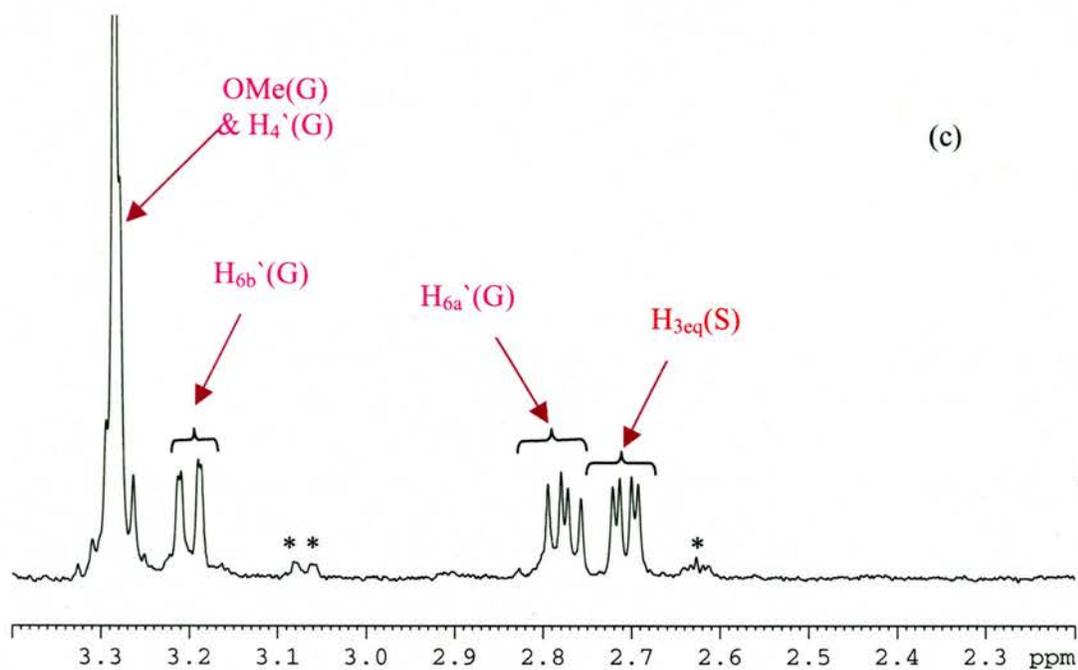
5 Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me

Scheme 8.4

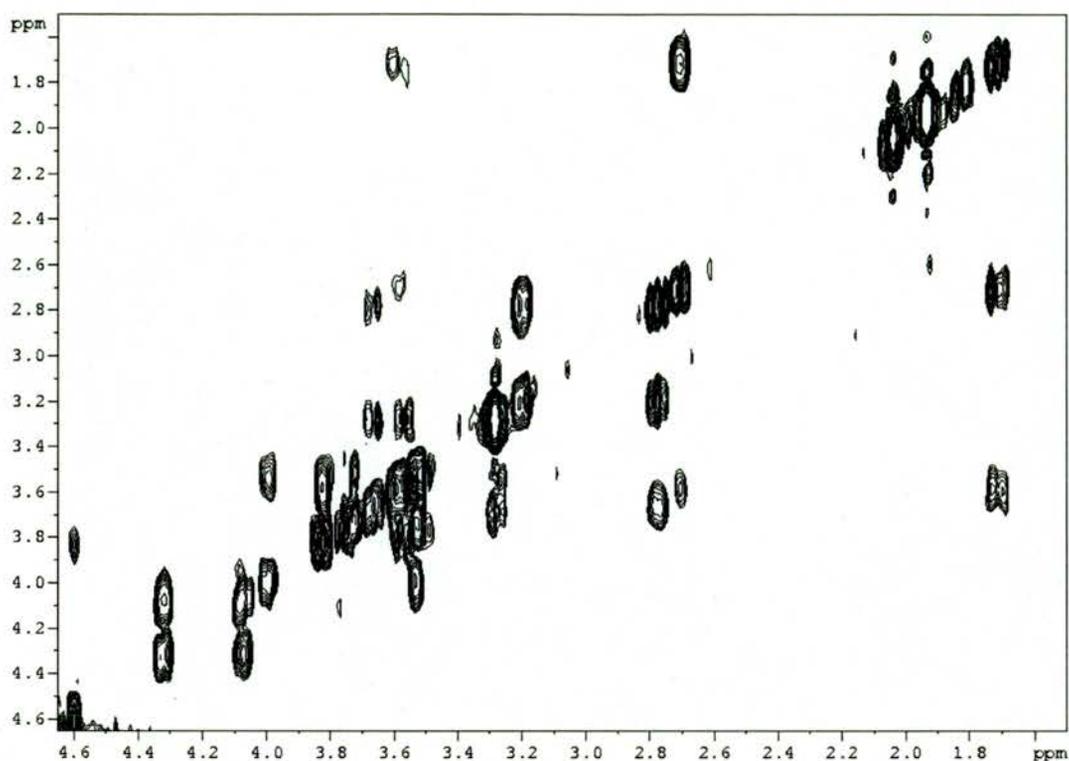
**Table 8.2** Chemical shifts of the protons of **5** as appeared in its 1D <sup>1</sup>H NMR spectrum. The numbering scheme of the protons are shown in scheme 8.4.

Neu5Ac		GlcNAc	
Proton	$\delta$ (ppm)	Proton	$\delta$ (ppm)
H <sub>3ax</sub>	1.71	H <sub>1'</sub>	4.60
H <sub>3eq</sub>	2.70	H <sub>2'</sub>	3.82
H <sub>4</sub> /H <sub>6</sub>	3.74	H <sub>3</sub>	3.58
H <sub>5</sub>	3.83	H <sub>4'</sub> /OMe	3.28
H <sub>7</sub>	3.52	H <sub>5'</sub>	3.67
H <sub>8</sub>	3.99	H <sub>6a'</sub>	3.20
H <sub>9a</sub>	4.32	H <sub>6b'</sub>	2.77
H <sub>9b</sub>	4.07	NHAc (G)	1.90-1.95
9-OAc	2.03		
NHAc (S)	1.90-1.95		





**Figure 8.7.** (a) 1D  $^1\text{H}$  NMR spectrum of **5** in the NMR buffer collected at 600 MHz, (b), (c) and (d), expanded regions ( $\delta=1.2\text{--}2.2$ ,  $\delta=2.2\text{--}3.4$ ,  $\delta=3.4\text{--}4.8$  ppm, respectively). The protons of **5** with the numbering in scheme 8.4 are assigned to the corresponding signal in the spectrum; protons of Neu5Ac are labelled (S) and of GlcNAc labelled (G). Peaks labelled (\*) are contaminant.



**Figure 8.8.** 2D  $^1\text{H}$ — $^1\text{H}$  COSY spectrum of the ligand **5**

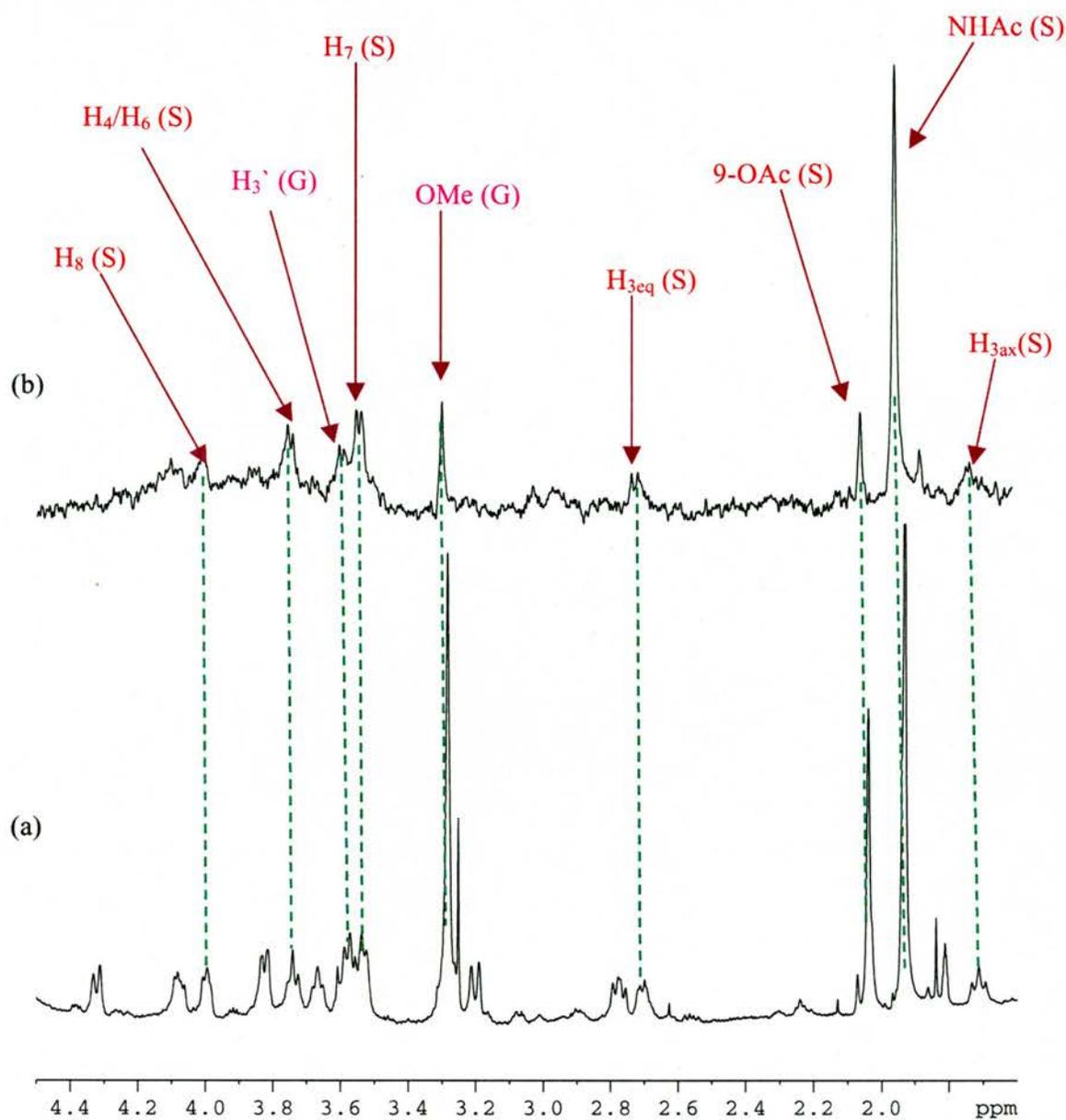
(Dr. K. Melton, Griffith university, personal communication).

#### 8.4.2.2 The STD NMR spectrum

STD NMR relies on the transfer of saturation from the protein to the ligand and it is strongest for the ligand parts that are in closest contact with the protein. It is advantageous to have the ligand in large excess in order to emphasize the differentiation between the protons in close contact to the protein from those pointing away but still receiving indirect saturation from the protein, which is required for efficient epitope mapping. In our case we used only an 11-fold excess of **5** because of the limited amount of the ligand; however this excess was sufficient to map the ligand epitopes as shown in the original work of

applying STD in epitope mapping (Mayer & Meyer, 2001). Figure 8.9 displays the normal 1D  $^1\text{H-NMR}$  spectrum of the thiosialoside **5** with VCNA (ratio 11:1, ligand to enzyme) in comparison with the corresponding STD NMR spectrum of the ligand/enzyme mixture. The STD spectrum shows that, not all protons of the ligand were saturated; only those protons which received saturation, because of their direct contact with the protein protons, exhibited STD signals in the spectrum. The normal 1D spectrum of the protein/ligand mixture enabled us to assign the STD signals. Most of the STD signals originated from the sialic acid residue of the disaccharide ligand; which means that sialic acid was the chief residue responsible for the ligand binding.

The N-acetyl group exhibited the strongest STD signal. Although, the resonances of the two N-acetyl groups, from the sialic acid residue at C5 and from the N-acetylglucosamine at C2 were overlapping in the reference 1D spectrum of the ligand and appeared too close to each other, the normal 1D spectrum of the ligand in presence of the enzyme showed only one peak at the position of the N-acetyl group. This was due to the line broadening effect of the peak arising from the N-acetyl group due to its interaction with the protein. We believe that the N-acetyl group that gave the largest STD signal was that at C5 of the sialic acid residue not the one belonging to the N-acetylglucosamine. This was inferred from the results of the STD library screening experiment (section 8.4.3). Where, in the STD spectrum of a sugar mixture containing the GlcNAc monosaccharide, no STD signals were observed that could be assigned to the N-acetyl group of the N-acetylglucosamine. On the other hand, adding ligand **5** to the monosaccharide mixture resulted in an STD spectrum similar to the one in figure 8.9, with the N-acetyl STD signal at the same position. Hence, the N-acetyl group at C5 of the sialic acid moiety played the major role in the binding of the thiosialoside ligand to its receptor.



**Figure 8.9.** One-dimensional STD spectrum (b) in comparison to a normal proton NMR spectrum (a) of **5** in the presence of VCNA (molar ratio of approximately 11:1, ligand to protein). Resonances in the STD spectrum correspond to the ligand protons interacting with the enzyme; the identified peaks are labelled using the same numbering scheme for **5** as in scheme 8.4. Protons of sialic acid are assigned (S) and those of GlcNAc assigned (G).

Signals of the sialic acid protons: 9-OAc, H<sub>3ax</sub>, H<sub>3eq</sub>, H<sub>4</sub>/H<sub>6</sub>, H<sub>7</sub> and H<sub>8</sub> were displayed in the STD spectrum. The STD signal of the 9-OAc was much less intense than the N-acetyl signal, though both have three protons. Also, no clear STD signal was observed for H<sub>9a</sub> and H<sub>9b</sub> protons of sialic acid residue. This indicated that the 9-OAc was not in close proximity to the protein, and therefore had less contribution to the binding process. The intensities of the signals of protons: H<sub>3ax</sub>, H<sub>3eq</sub>, H<sub>4</sub>/H<sub>6</sub>, H<sub>7</sub> and H<sub>8</sub> of sialic acid were much less than the intensity of the N-acetyl group but relatively comparable to their corresponding signals in the normal 1D spectrum. It was hard to differentiate between the proximity of these protons to the protein; the signals of these protons showed line broadening indicating their interaction with the protein.

Protons of the  $\beta$ -OMe group at the anomeric carbon of GlcNAc showed a low intensity STD signal relative to the corresponding signal in the normal 1D spectrum. Also, a small signal that was identified as H<sub>3'</sub> was observed in the STD spectrum. Certainly, no STD signals were observed for the GlcNAc protons: H<sub>1'</sub>, H<sub>2'</sub>, H<sub>4'</sub>, H<sub>5'</sub> and H<sub>6a'</sub> and H<sub>6b'</sub>. Thus, the residue GlcNAc in ligand **5** was not in an intimate contact with the protein. Presumably, binding of the sialic acid residue, brought the protons:  $\beta$ -OMe and H<sub>3'</sub> of GlcNAc in contact to the protein surface.

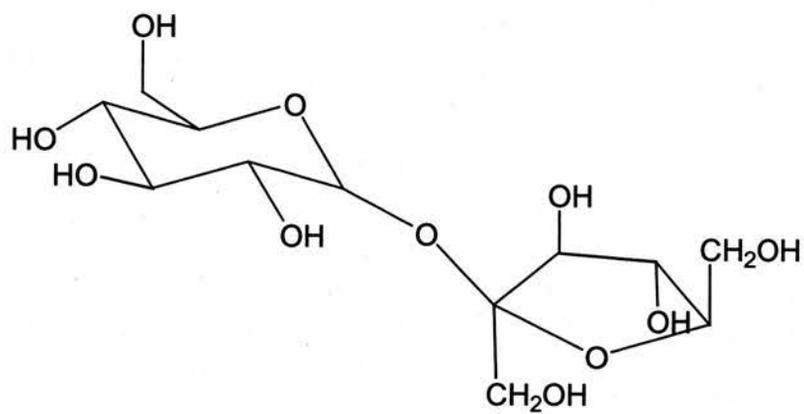
The interpretation of the results from the STD epitope mapping experiment was in agreement with the crystallographic data, which showed sialic acid bound at the N-terminal lectin-like domain (see chapter 5). The N-acetyl group of sialic acid (exhibited the strongest STD signal) showed the largest number of contact with binding site residues compared to the other substituents of the sialic acid sugar ring. The N-acetyl group showed extensive interaction with residues: Gly196, Arg118, and Gln188 through H-bonding; and Thr109, Tyr111, and Leu120 through hydrophobic interaction. The 9-OAc

group, which exhibited a small STD signal, was expected not to have a strong interaction with the binding site, as the 9-OH group was pointing away from the protein surface in the crystallographic model. The region of sialic acid comprising the protons: H<sub>7</sub> and H<sub>8</sub> of the glycerol side chain and the sugar ring protons H<sub>3ax</sub>, H<sub>3eq</sub>, H<sub>4</sub>, and H<sub>6</sub>, was exposed to the protein environment, which was in accordance with the appearance of the corresponding STD signals in the NMR spectrum.

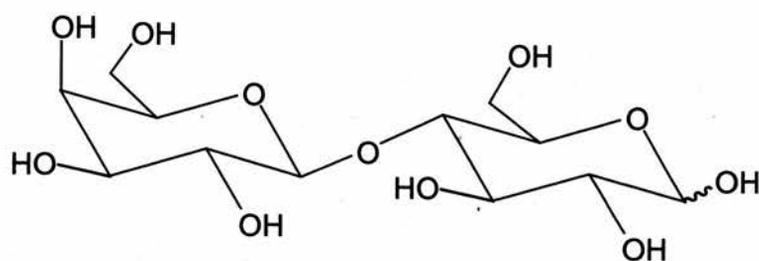
### 8.4.3 Screening of carbohydrate library by STD NMR

It is well known that NMR spectroscopy is able to detect the binding of ligands to receptors. Recently, many NMR methods have been developed to screen collections of compounds, including carbohydrate libraries, for binding to protein receptors (reviewed in (Coles et al., 2003; Fejzo et al., 2003; Meyer & Peters, 2003)). STD NMR has been applied to NMR screening with good results (Klein et al., 1999; Mayer & Meyer, 1999). STD experiments were setup for screening a small carbohydrate library composed of compounds **6 - 18** (chemical structures of the compounds are drawn in scheme 8.5). The library of compounds **6 - 17** were divided into two mixtures: the first mixture composed of compounds **6 - 10** and the second mixture composed of compounds **11 - 17**.

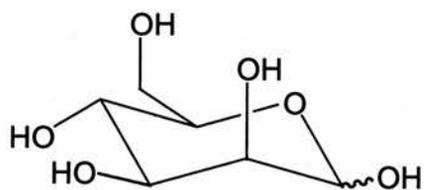
Figure 8.10 displays (a) the normal 1D <sup>1</sup>H-NMR spectrum of the mixture of compounds **6 - 10** with VCNA in a (20:1) ratio, and (b) the corresponding STD NMR spectrum. The normal 1D <sup>1</sup>H NMR spectrum and the corresponding STD spectrum obtained for the second mixture composed of **11 - 17** are shown in figure 8.11. As can be seen from the figures, no signals appeared in the STD spectra of the two mixtures that could suggest a specific binding of any of the components in the mixtures to the enzyme. When repeating the same STD experiments for the two mixtures in presence of the



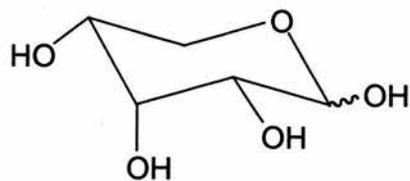
6 D-(+)-Sucrose



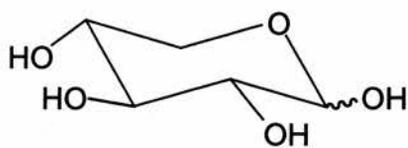
7 D-Lactose



8 D-(+)-Mannose

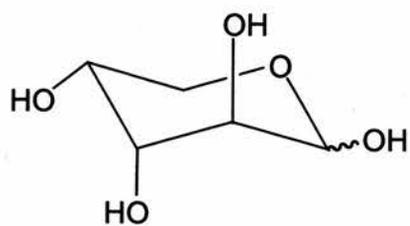


9 D-(-)-Ribose

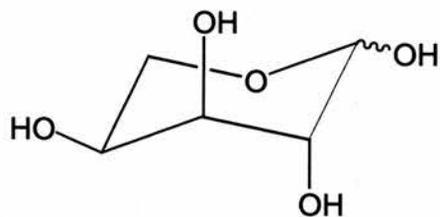


10 D-(+)-Xylose

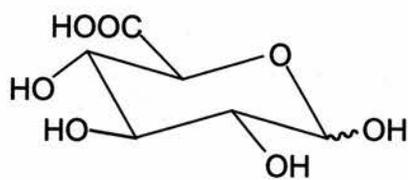
Scheme 8.5



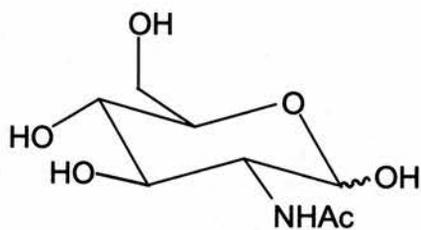
11 <sup>4</sup>C<sub>1</sub>-D-(-)-Arabinose



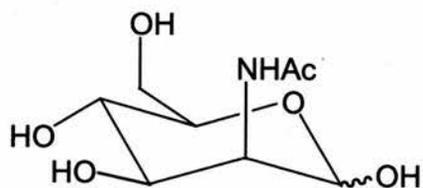
12 <sup>1</sup>C<sub>4</sub>-L-(-)-arabinose



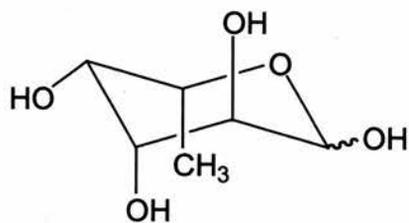
13 D-Glucouronic acid



14 N-acetylglucosamine

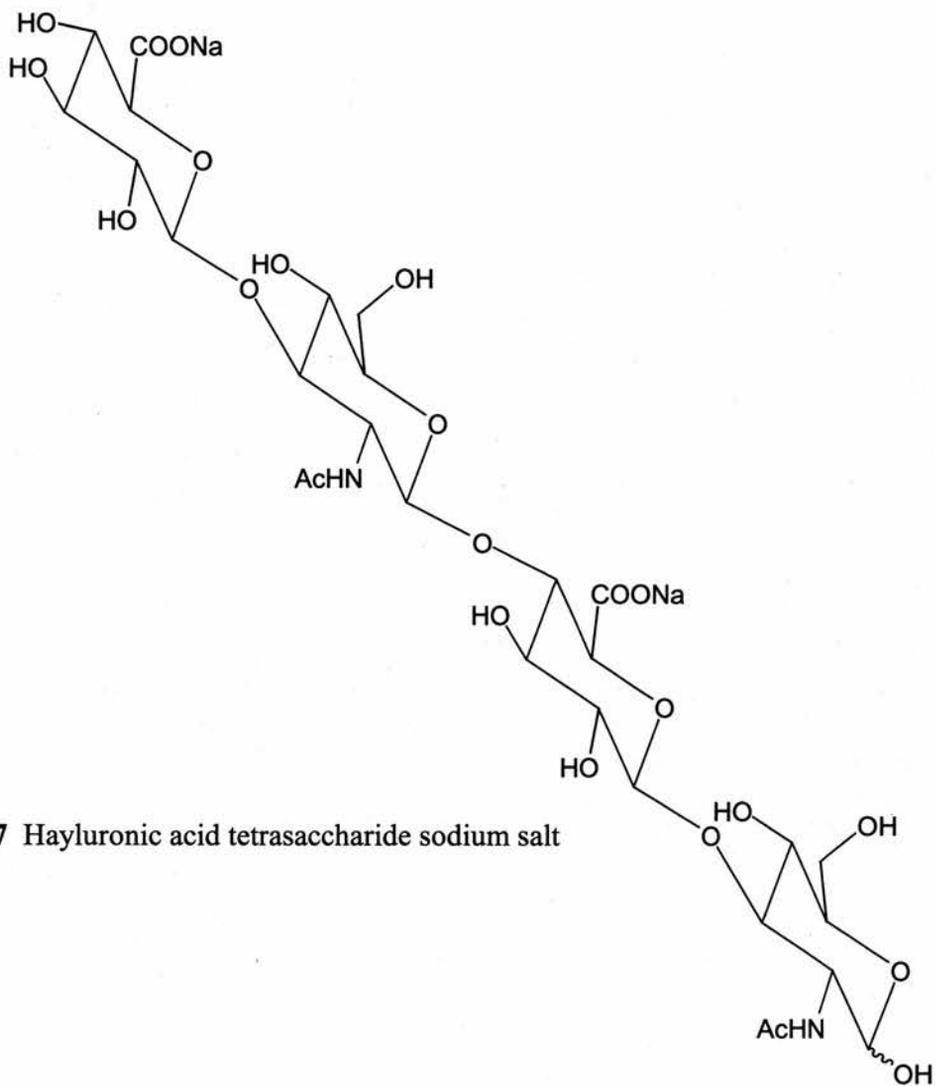


15 N-acetylmannosamine

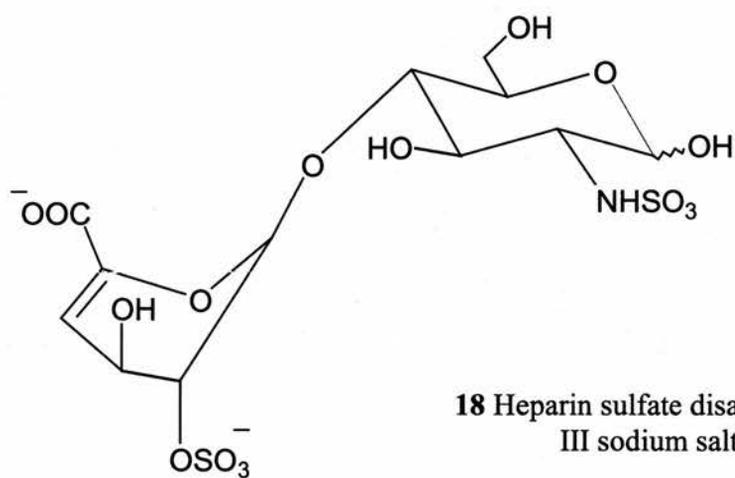


16 L-(-)-Fucose

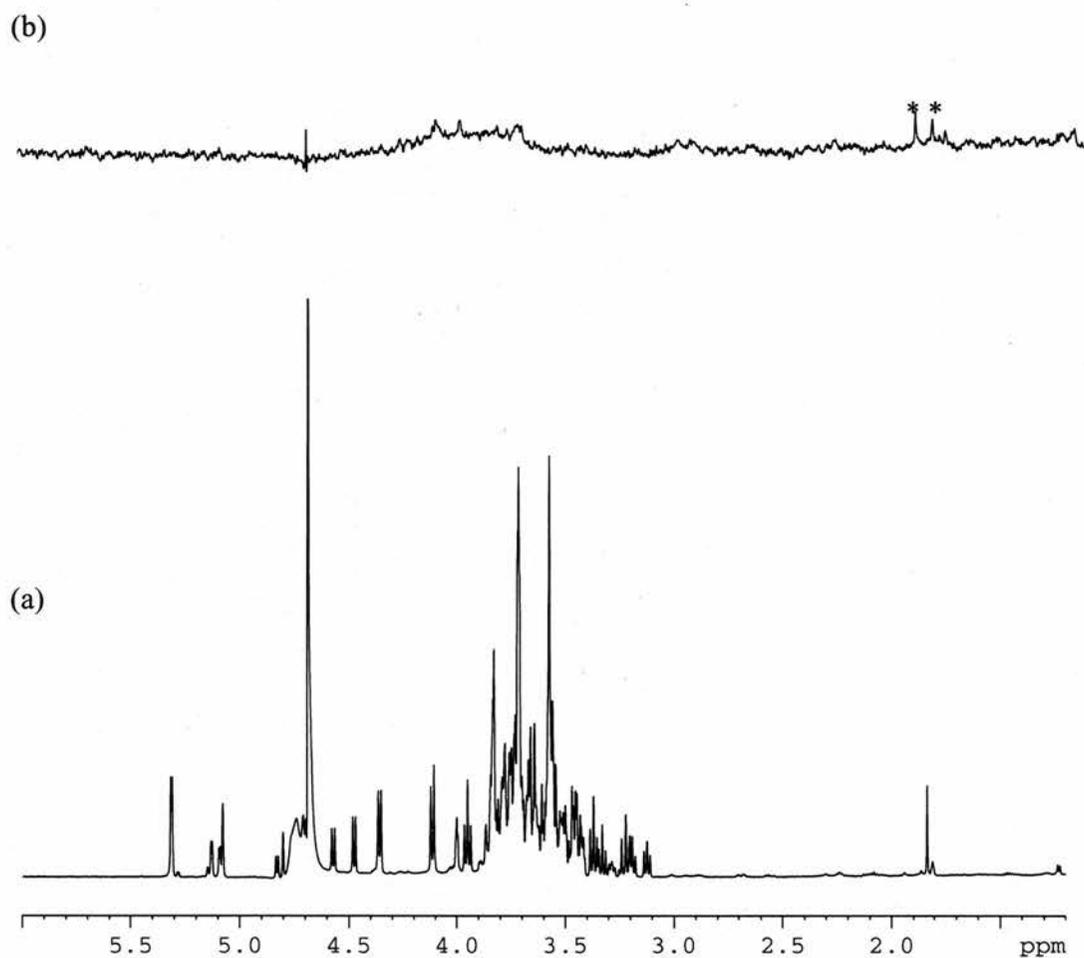
Scheme 8.5 (continued)



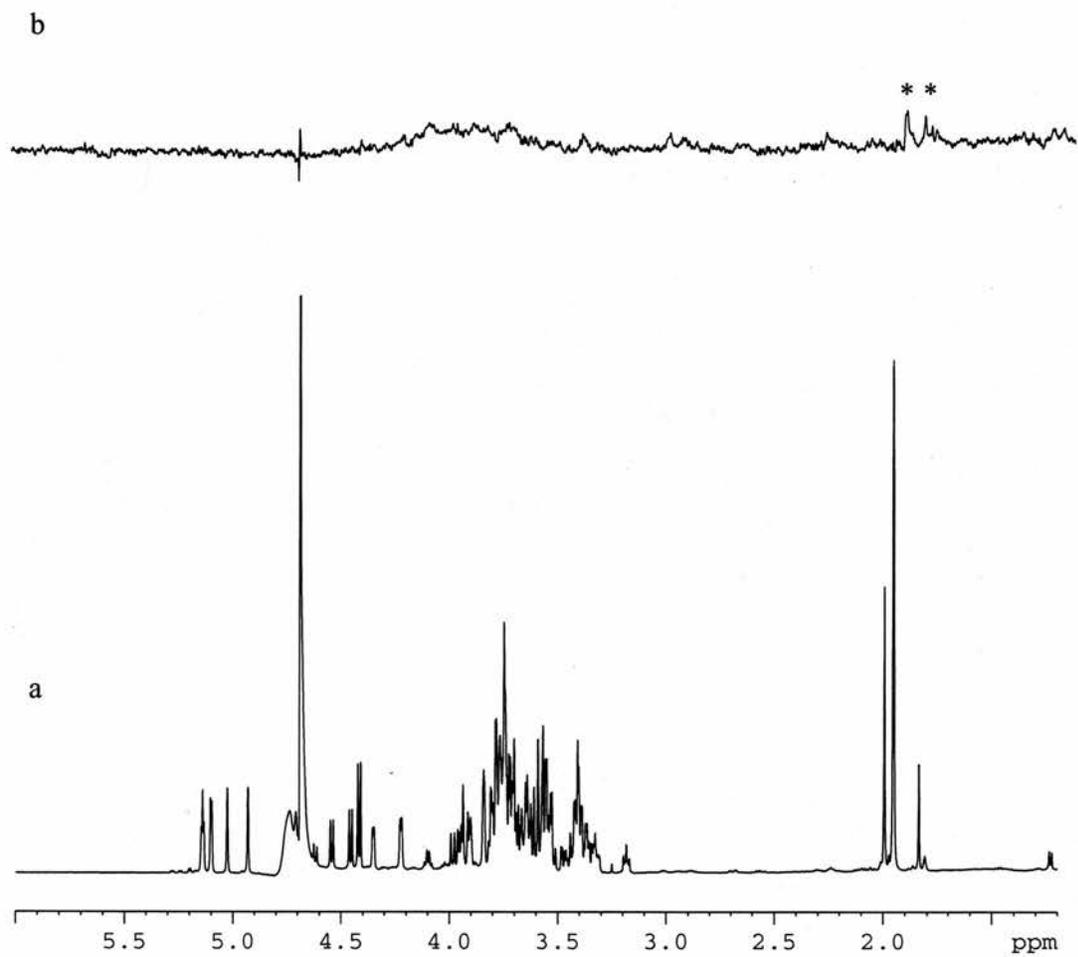
17 Heparin sulfate disaccharide III sodium salt



Scheme 8.5 (continued). Carbohydrates used in the STD NMR screening.



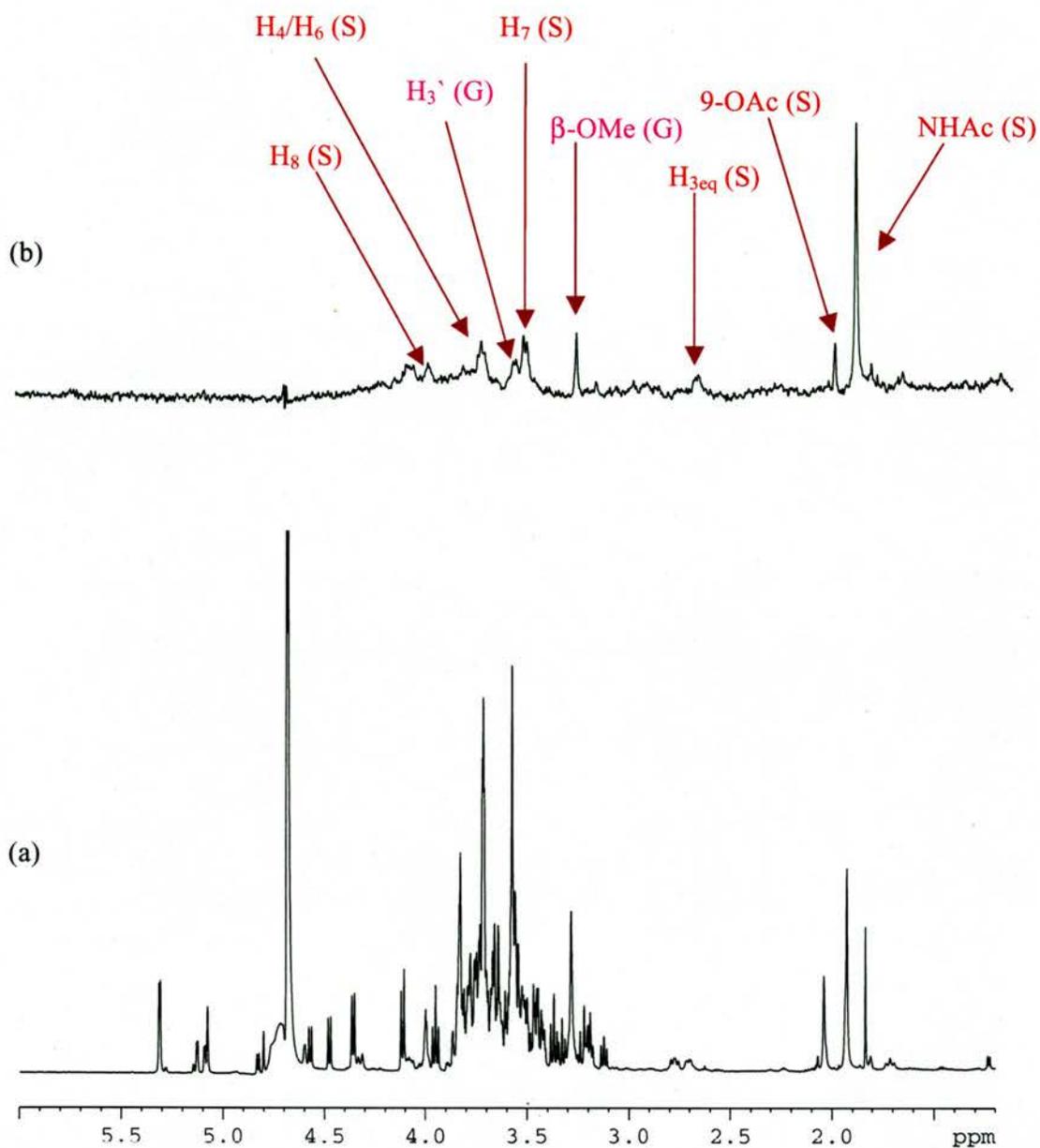
**Figure 8.10.** One-dimensional STD spectrum (b) in comparison to a normal proton NMR spectrum (a) of the library compounds **6 - 10** in the presence of VCNA (20-fold individual ligand excess over protein). No real STD signals could be identified; peaks labelled with (\*) in (b) are impurities.



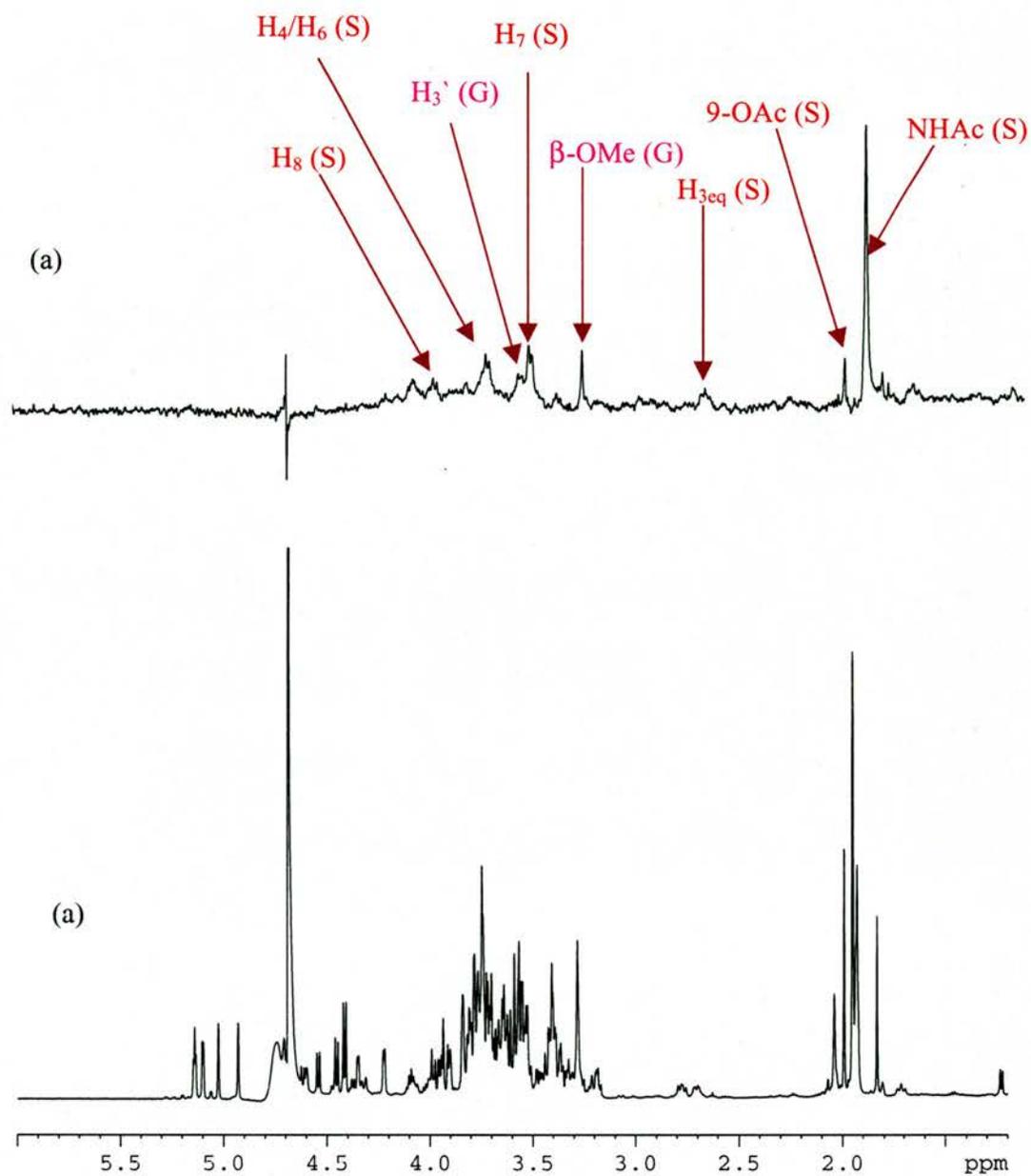
**Figure 8.11.** One-dimensional STD spectrum (b) in comparison to a normal proton NMR spectrum (a) of the library compounds **11** - **17** in the presence of VCNA (20-fold individual ligand excess over protein). No real STD signals could be identified; peaks labelled with (\*) in (b) are impurities.

thiosialoside **5**, added to each mixture as a 20-fold excess over the enzyme (same as the other components), STD spectra similar to the one collected previously during the epitope mapping of ligand **5** were obtained (see figures 8.12 and 8.13). Clearly, the STD signals in the spectra in figures 8.12 & 8.13 were originating from the saturation of the thiosialoside ligand **5** as a result of its binding to the enzyme. No STD signals were obtained for the second mixture, which contained the monosaccharide N-acetylglucosamine, suggesting that the N-acetylglucosamine residue has no specific binding to VCNA. The observed STD signals of the GlcNAc residue of ligand **5** in the epitope mapping experiment arise from the indirect interaction of that residue with the enzyme upon binding of the adjacent sialic acid residue to the N-terminal domain.

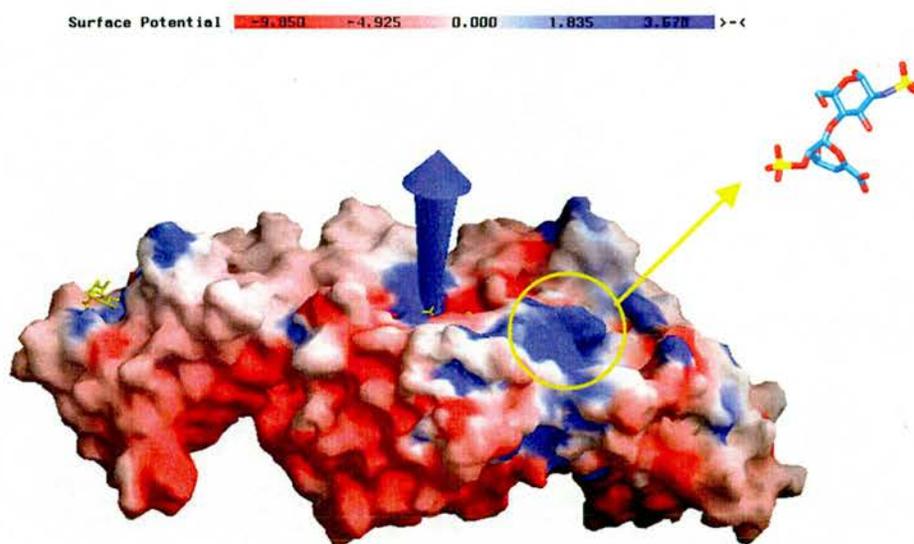
Investigation of the electrostatic potential of the surface of VCNA revealed a positively charged pocket rich in basic residues (Lys520, Lys524, Lys529, Lys531, Lys323, and Arg577) at the interface between the central and the C-terminal domains; the pocket also includes hydrophobic residues (Tyr368, Tyr539, Phe357, and Trp325). Figure 8.14 displays the surface electrostatic potential with the positive pocket bounded by the circle. The combination of positive and hydrophobic residues is characteristic for heparin sulfate-binding sites (Margalit et al., 1993). Therefore, we carried out a separate STD run for **18** to investigate if there is any binding of heparin sulfate to VCNA. The STD spectra of heparin sulfate **18** in presence of the enzyme in ratios (10:1 & 50:1, ligand to enzyme) compared to the normal 1D <sup>1</sup>H NMR spectra are shown in figure 8.15. The STD spectra exhibited no signals, even though 50-fold excess of **18** was used, indicating there was no specific interaction between **18** and the enzyme. Unfortunately, the guess that heparin sulfate could be a binding ligand for the VCNA domain was incorrect.



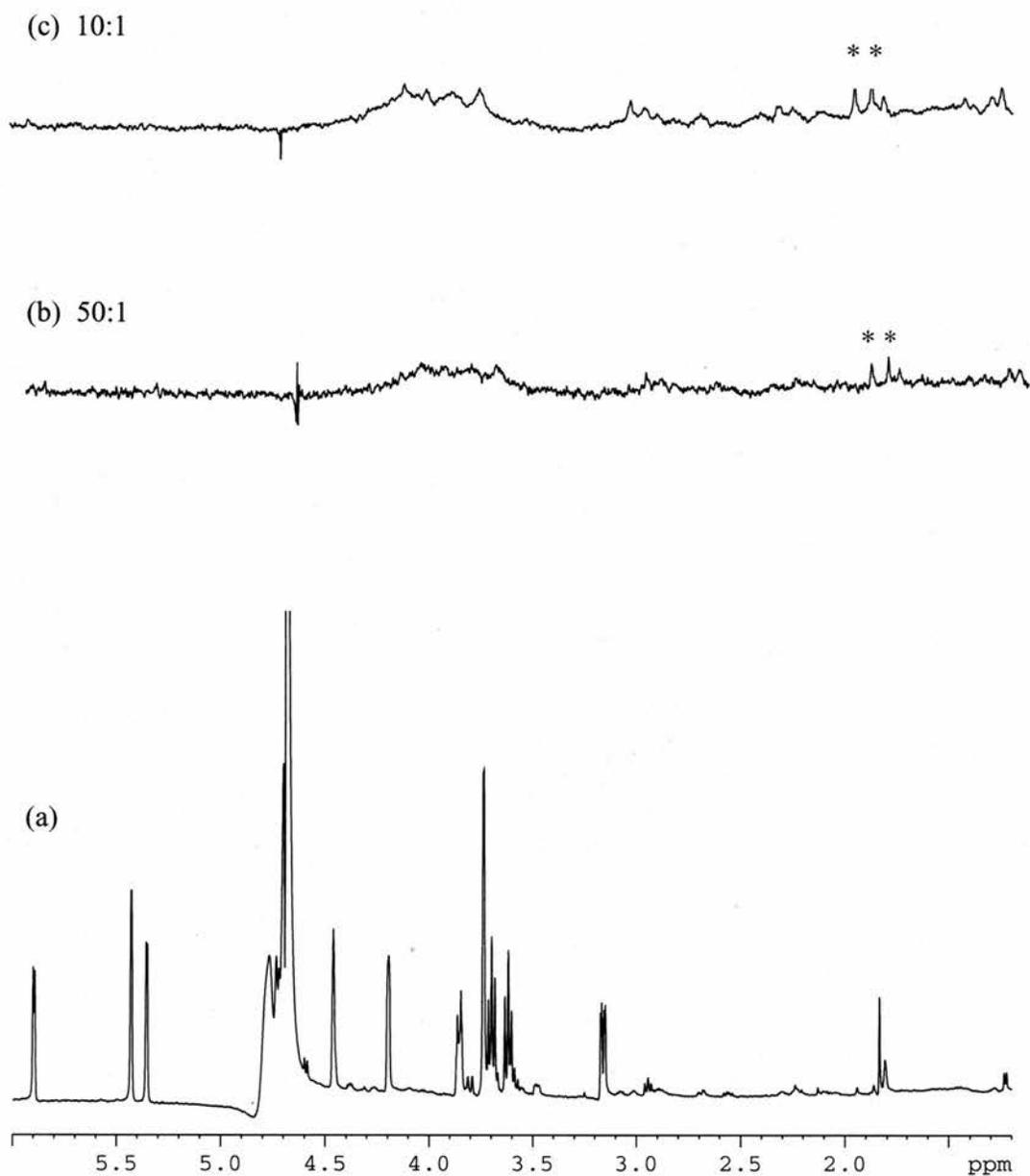
**Figure 8.12.** (a) Normal 1D  $^1\text{H}$  NMR spectrum of the mixture of compounds **6** - **10** with the thiosialoside **5** added in presence of VCNA (ratio 20:1, ligand to enzyme), in comparison with (b) the STD spectrum of the same sample.



**Figure 8.13.** (a) Normal 1D  $^1\text{H}$  NMR spectrum of the mixture of compounds **11** - **17** with the thiosialoside **5** added in presence of VCNA (ratio 20:1, ligand to enzyme), in comparison with (b) the STD spectrum of the same sample.



**Figure 8.14.** Putative binding site, enclosed within the circle, for the ligand heparin sulfate disaccharide **18**, drawn as a ball-and-stick model. The protein surface is colour coded according to the electrostatic potential values between: -9.85 KT (red), and 3.67 KT (blue). The blue arrow represents the calculated dipole moment of the protein.



**Figure 8.15.** The normal 1D  $^1\text{H}$  NMR spectrum of heparin sulfate **18** in presence of VCNA in ratio 10:1, ligand to enzyme (a), in comparison with the STD spectra of the same sample with ratios 50:1 (b) and 10:1 (c), ligand to enzyme, respectively. Peaks labelled with (\*) in (b) & (c) are impurities.

The results from the library screening of compounds **6 - 18** applying STD technique showed no clear evidence for specific binding of any of the carbohydrate components in the screened library. Therefore, the function of the C-terminal lectin domain and the ligand(s) recognized by this domain are still to be established.

## 8.5 Discussion

Significant insight into ligand-protein interaction can be obtained from NMR, including the identification of the ligands that bind to proteins, the dissection of the ligand epitopes, and the monitoring the enzyme reactions. These three addressed applications were applied to VCNA in order to understand this multifunctional protein and to complement the X-ray studies.

Three questions were addressed; the first question was does VCNA form its own inhibitor DANA when acting on the substrate  $\alpha$ -2,3-sialyllactose? The enzymatic cleavage of the glycosidic bond in  $\alpha$ -2,3-sialyllactose by the action of VCNA to produce a free sialic acid has been studied previously, and was proved to proceed with full retention of configuration for the cleaved  $\alpha$ -sialic acid, which in turn mutarotates to  $\beta$ -sialic acid. The crystallographic study of the complex structure of VCNA with the substrate  $\alpha$ -2,3-sialyllactose showed the enzyme inhibitor (DANA) bound to the catalytic site rather than the substrate or the hydrolytic product. In another neuraminidase from the influenza virus, it has been shown that the enzyme was able to form the inhibitor (DANA) during its action on the substrate  $\alpha$ -2,3-sialyllactose, and was also able to rehydrate DANA to form sialic acid. To obtain a conclusive evidence for the ability of VCNA to form its inhibitor DANA from  $\alpha$ -2,3-sialyllactose, we monitored the hydrolysis reaction by  $^1\text{H-NMR}$  over a

long period of time (about 11 hours) under conditions of high concentrations of enzyme and substrate, in the same concentration range applied in the crystallographic experiment. The NMR results showed that, as expected,  $\alpha$ -sialic acid was the first product of the hydrolysis reaction followed by mutarotation to the more stable  $\beta$ -form. The NMR spectra also revealed that a trace of DANA (< 1%) was formed transiently; the characteristic signal of H<sub>3</sub> proton of DANA appeared progressively and disappeared in about 9 hours.

The transient formation of DANA means that the inhibition action produced by the enzyme itself was not permanent. This could be speculated of as a kind of regulating mechanism for the enzyme under the physiological conditions. The bacteria *V. cholerae* would not get any benefits from VCNA being all the time processing the higher gangliosides, as the host would suffer severely and may eventually get lost. This would have a negative effect on the colonizing bacteria, which needs the host to sustain their life. Therefore, at high concentration of the cleaved sialic acid, a self-inhibitor could be formed by the enzyme, but in a small amount such that it would not block the enzyme activity for long time. To keep their momentum going, in absence of the substrate, the enzyme uses the inhibitor DANA as a substrate.

The second question addressed was does the N-acetylglucosamine moiety of the ligand, Neu5,9Ac<sub>2</sub>-2-S- $\alpha$ -(2,6)-GlcNAc $\beta$ Me **5**, play any major role in the ligand binding to the sialic acid-binding site at the N-terminal lectin-like domain? The affinity of the thiosialoside ligand was determined by ITC ( $K_d \approx 30 \mu\text{M}$ ), which is significant for carbohydrate-recognition by a protein, especially when it is acting as a monomer. The competitive inhibition study showed that **5** has no binding affinity towards the catalytic site; the ITC results fitted best to the one-site-model, which means that **5** binds only at the N-terminal lectin like domain. However, the ITC experiment did not show whether the

ligand affinity was mainly due to the binding of the sialic acid residue or there was some contribution from the GlcNAc residue. Such information could be obtained from the STD NMR experiment, which proved to be efficient in epitope mapping of the ligand binding to proteins. The results from the STD experiment clearly demonstrated that sialic acid of **5** was the chief residue responsible for binding to the sialic acid site. The interpretation of the STD spectrum was in accord with the crystallographic information of the sialic acid recognition by the N-terminal lectin domain. The STD spectrum showed the significant role of the N-acetyl group at C5 of the sialic acid; the N-acetyl group received the largest saturation from its contact with the nearby protein residues. It is noteworthy to mention that the 9-OAc group received some saturation from the protein, though it was not expected to participate in the ligand interaction based on the crystallographic data. This apparent interaction of the 9-OAc may suggest that the acetylated sialic acid at C9 could be the true ligand for the N-terminal domain, which needs further investigation.

The STD spectrum of the thiosialoside ligand showed that there could be indirect contact of the GlcNAc residue with the protein, which might be brought about by the binding of the sialic acid residue to the binding site. But, it is unlikely that the non-specific contact of the GlcNAc would have a large contribution to the ligand affinity.

The third question was which carbohydrate sugar could be recognized by the second lectin domain of VCNA? The lectin-topology of the C-terminal domain in VCNA suggested that this domain might also be used for sugar recognition. Recently, STD NMR was applied successfully in screening a library of sugar compounds and good results were obtained. We setup the STD protocol in order to screen a small library of the available sugars hoping to identify the correct ligand for the C-terminal lectin domain. Unfortunately, no sharp evidence could be obtained from the STD spectra of the

enzyme/library mixtures to indicate any specific affinity of the components towards the enzyme. Even the idea that a negatively charged carbohydrate, like heparin sulfate or hayluronic acid, could bind at the positive pocket at the interface between the central and C-terminal domains turned to be untrue. Further investigation is required to establish the exact function of the second lectin domain.

## Appendix 1

### Determination of the absorption coefficient ( $\epsilon_{280}$ ) of VCNA

- 1- The UV-absorbance at  $\lambda_{205}$  ( $A_{205}$ ) and at  $\lambda_{280}$  ( $A_{280}$ ) of the VCNA sample was measured.
- 2- The extinction coefficient ( $\epsilon_{205}$ ) was obtained applying the following equation,

$$\epsilon_{205} = 27 + 120 \left( \frac{A_{280}}{A_{205}} \right) \quad (4)$$

- 3- From the determined  $\epsilon_{205}$  and  $A_{205}$ , the protein concentration ( $C$ ) was obtained using Beer-Lambert law (Equation 5),

$$C = \frac{A_{205}}{\epsilon_{205}} \quad (5)$$

- 4- The extinction coefficient ( $\epsilon_{280}$ ) was then determined using the measured concentration ( $C$ ) and the absorbance ( $A_{280}$ ) from the relation in Equation (5),

$$\epsilon_{280} = \frac{A_{280}}{C}$$

## References

- Ada, G. L., French, E. L., Lind, P. E. (1961). Purification and properties of neuraminidase from *Vibrio cholerae*. *J. Gen. Microbiol.* 24, 409-421.
- Amaya, M. F., Buschiazzo, A., Nguyen, T. & 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.
- Angata, T. & Varki, A. (2002). Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem. Rev.* 102, 439-469.
- Babal, P., Pindak, F. F., Russell, L. C. & Gardner Jr., W. A. (1999). Sialic acid-specific lectin from *Tritrichomonas foetus*. *Biochim. Biophys. Acta* 1428, 106-116.
- Barcelo, F., Ortiz-Lombardia, M. & Portugal, J. (2001). Heterogeneous DNA binding modes of berenil. *Biochim. Biophys. Acta* 1519, 175-184.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 253-242.
- Blandamer, M. J. (1998). Thermodynamic Background to Isothermal Titration Calorimetry. In *Biocalorimetry: Applications of calorimetry in the biological sciences* (Ladbury, J. E. & Chowdhry, B. Z., ed.), pp. 5-25. Wiley & Sons Ltd, West Sussex.
- Bonten, E., van der Spoel, A., Fornerod, M., Grosveld, G. & d'Azzo, A. (1996). Characterization of human lysosomal neuraminidase defines the molecular basis of the metabolic storage disorder sialidosis. *Genes Dev.* 10, 3156-3169.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248-254.

- Bratosin, D., Mazurier, J., Debray, H., Lecocq, M., Boily, B., Alonso, C., Moisei, M., Motas, C. & Montreuil, J. (1995). Flow cytofluorometric analysis of young and senescent human erythrocytes probed with lectins-evidence that sialic acids control their life-span. *Glycoconj. J.* 12, 258-267.
- Brook, S. A., Dwek, M. V. & Schumacher, U. (2002). Sialic acids. In *Functional & Molecular Glycobiology* (Dines, P., ed.), pp. 203-220. Bios Scientific Publishers Ltd., Oxford.
- Brossmer, R., Wagner, M. & Fischer, E. (1992). Specificity of the sialic acid-binding lectin from the snail *Cepaea hortensis*. *J. Biol. Chem.* 267, 8752-876.
- Brunger, A. T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-474.
- Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nigels, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., & Warren, G. L. (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Cryst.* D54, 905-921.
- Brunger, A. T., Adams, P. D. & Rice, L. M. (2001). Enhanced macromolecular refinement by simulated annealing. In *International Tables for Crystallography: Crystallography of Biological macromolecules* (Rossmann, M. G. & Eddy, A., ed.), Vol. F, pp. 375-380. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Burmeister, W. P., Henrissat, B., Bosso, C., Cusack, S. & Ruigrok, R. W. (1993). Influenza B virus neuraminidase can synthesize its own inhibitor. *Structure* 1, 19-26.
- Burnet, F. M. & Stone, J. D. (1947). The receptor destroying enzyme of *V. cholerae*. *Aust. J. exp. Biol. med. Sci.* 25, 227-233.

- Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C., Alzari, P. M. (2002). The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma Cruzi* pathogenesis. *Mol. Cell* 10, 757-768.
- Charron, C., Sauter, C., Zhu, D.W., Ng, J.D., Kern, D., Lorber, B. & Giege, R. (2001). Packing contacts in orthorhombic and monoclinic crystals of a thermophilic aspartyl-tRNA synthetase favor the hydrophobic regions of the protein. *J. Cryst. Growth* 232, 376-386.
- Chan, T.-H., Xin, Y.-C. & von Itzstein, M. (1997). Synthesis of phosphonic acid analogues of sialic acids (Neu5Ac and KDN) as potential sialidase inhibitors. *J. Org. Chem.* 62, 3500-3504.
- Chong, A. K., Pegg, M. S., Taylor, N. R. & von Itzstein, M. (1992). Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur. J. Biochem.* 207, 335-343.
- Clayden, J., Greeves, N., Warren, S. & Wothers, P. (2001). *Organic Chemistry*. first edition, Oxford University Press, Oxford.
- Coles, M., Heller, M. & Kessler, H. (2003). NMR-based screening technologies. *Drug Discov. Today* 8, 803-810.
- Collaborative Computational Project, number 4. (1994). The CCP4 suite: Programs for protein crystallography. *Acta Cryst.* D50, 760-763.
- Colman, P. M. (2002). Neuraminidase inhibitors as antivirals. *Vaccine* 20 Suppl 2, S55-S58.
- Connors, K. A. (1987). *Binding constants: The measurement of molecular complex stability*, Wiley & Sons.
- Cooper, A., Jhonson, C. M., Lakey, J. H. & Nollmann, M. (2001). Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum

- confinement, pressure perturbation calorimetry, solvation and multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.* 93, 215-230.
- Copley, R. R., Russel, R. B. & Ponting, C. B. (2001). Sialidase-like Asp-boxes: Sequence-similar structures within different protein folds. *Protein Sci.* 10, 285-292.
- Corfield, A. P., Higa, H., Paulson, J. C. & Schauer, R. (1983). The specificity of viral and bacterial sialidases for  $\alpha(2-3)$ - and  $\alpha(2-6)$ -linked sialic acids in glycoproteins. *Biochim. Biophys. Acta* 744, 121-126.
- Corfield, T. (1992). Bacterial sialidases: roles in pathogenicity and nutrition. *Glycobiology* 2, 509-521.
- Crennell, J. S., Garman, E. F., Laver, W. G., Vimr, E. R. & Taylor, G. L. (1994). Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure* 2, 535-544.
- Crennell, J. S., Garman, E. F., Laver, W. G., Vimr, E. R. & 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. USA* 90, 9852-9856.
- Crennell, S. J., Garman, E. F., Philippon, C., Vasella, A., Laver, W. G., Vimr, E. R. & Taylor, G. L. (1996). The structures of *Salmonella typhimurium* LT2 neuraminidase and its complexes with three inhibitors at high resolution. *J. Mol. Biol.* 259, 264-280.
- Crennell, J. S., Takimoto, T., Portner, A. & Taylor, G. L. (2000). Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. *Nat. Struct. Biol.* 7, 1068-1074.
- Dabrowski, U., Friebolin, H., Brossmer, R. & Supp, M. (1979). 1H-NMR studies at N-acetyl-D-neuraminic acid ketosides for the determination of the anomeric configuration II. *Tetrahedron Lett.* 20, 4637-4640.

- Dam, T. K. & Brewer, C. F. (2002). Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* 102, 387-429.
- Dauter, Z. & Wilson, K. S. (2001). Monochromatic data collection: Principles of monochromatic data collection. In *International Tables for Crystallography: Crystallography of Biological macromolecules* (Rossmann, M. G. & Eddy, A., ed.), Vol. F, pp. 177-195. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Delano, W. L. (2002). *The PyMOL molecular graphics system*. Delano Scientific, San Carlos, CA, USA.
- Dirita, V. J. (2001). Molecular basis of *Vibrio Cholerae* pathogenesis. In *Principles of Bacterial Pathogenesis* (Groisman, E. A., ed.), pp. 457-508. Academic Press.
- Dormitzer, P. R., Sun, Z. Y., Blixt, O., Paulson, J. C., Wagner, G. & Harrison, S. C. (2002a). Specificity and affinity of sialic acid binding by the rhesus rotavirus VP8\* core. *J. Virol.* 76, 10512-10517.
- Dormitzer, P. R., Sun, Z. Y., Wagner, G. & Harrison, S. C. (2002b). The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* 21, 885-897.
- Drenth, J. (1999). *Principles of Protein X-ray Crystallography*. second edition, Springer-Verlag.
- Drenth, J. (2001). Basic crystallography. In *International Tables For Crystallography: Crystallography of Biological macromolecules* (Rossmann, M. G. & Eddy, A., ed.), Vol. F, pp. 45-63. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Ducruix, A. & Giege, R. (1999). *Crystallization of Nucleic Acids and Proteins*. second edition. *The Practical Approach Series* (Hames, B. D., ed), Oxford University Press, Oxford.

- Emsley, P., Fotinou, C., Black, I., Fairweather, N. F., Charles, I. G., Watts, C., Hewitt, E. & Isaacs, N. W. (2000). The structures of the H(C) fragment of tetanus toxin with carbohydrate subunit complexes provide insight into ganglioside binding. *J. Biol. Chem.* 275, 8889-8894.
- Engel, P. C. (1977). *Enzyme Kinetics: The Steady-State Approach*. first edition. *The Outlines studies in biology series*. (Ashworth, J. M., ed), Chapman & Hall.
- Esnouf, R. M. (1997). An extremely modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* 15, 132-134.
- Evans, P. R. (1993). Data collection and processing. In *Proceedings of the CCP4 study weekend*. Daresbury Laboratory, Warrington, UK.
- Eyke, L. F. & Watenpaugh, K. D. (2001). Introduction to refinement. In *International Tables for Crystallography: Crystallography of Biological macromolecules* (Rossmann, M. G. & Eddy, R., ed.), Vol. F, pp. 369-374. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Fejzo, J., Lepre, C. & Xie, X. (2003). Application of NMR screening in drug discovery. *Curr. Top. Med. Chem.* 3, 81-97.
- Finkelstein, R. A., Boesman, F. M., Chang, Y & Häse, C. C. (1992). *Vibrio cholerae* hemagglutinin/protease, colonial variation, virulence, and detachment. *Infect. Immun.* 60, 472-479.
- Freire, E., Mayorga, O. L. & Straume, M. (1990). Isothermal titration calorimetry. *Anal. Chem.* 62, 950A-959A.
- Friebolin, H., Baumann, W., Hauck, M., Kurz, D., Wajda, R., Weisshaar, G., Keilich, G., Ziegler, D., Brossmer, L. & von Nicolai, H. (1984). H-NMR spectroscopy: specificity of microbial sialidases against complex substrates. *Hoppe Seylers Z. Physiol. Chem.* 365, 1309-1321.

- Friebolin, H., Baumann, W., Keilich, G., Ziegler, D., Brossmer, R. & von Nicolai, H. (1981). <sup>1</sup>H-NMR spectroscopy: a potent method for the determination of substrate specificity of sialidases. *Hoppe Seylers Z. Physiol. Chem.* 362, 1455-1463.
- Friebolin, H., Brossmer, R., Keilich, G., Ziegler, D. & Supp, M. (1980). <sup>1</sup>H-NMR-spectroscopic evidence for the release of N-acetyl- $\alpha$ -D-neuraminic acid as the first product of neuraminidase action. *Hoppe Seylers Z. Physiol. Chem.* 361, 697-702.
- Gaisford, S. & Buckton, G. (2001). Potential applications of microcalorimetry for the study of physical processes in pharmaceuticals. *Thermochim. Acta* 380, 185-198.
- Galen, J. E., Ketley, J. M., Fasano, A., Richardson, S. H., Wasserman, S. S. & Kaper, J. B. (1992). Role of *Vibrio cholerae* neuraminidase in the function of cholera toxin. *Infect. Immun.* 60, 406-415.
- Gallivan, J. P. & Dougherty, D. A. (1999). Cation- $\pi$  interactions in structural biology. *Proc. Natl. Acad. Sci. USA* 96, 9459-9464.
- Gascoyne, N. & van Heyningen, W. E. (1979). Unmasking of actual and potential receptor sites for cholera toxin in intestinal mucosal homogenates. *J. Infect. Dis.* 139, 235-236.
- Gaskell, A., Crennell, J. S. & Taylor, G. L. (1995). The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* 3, 1197-205.
- Gottschalk, A. (1957). Neuraminidase: the specific enzyme of influenza virus and *Vibrio cholerae*. *Biochem. Biophys. Acta* 23, 645-646.
- Gottschalk, A. & Lind, P. E. (1949). Product of interaction between influenza virus enzyme and ovomucin. *Nature* 164, 232.
- Guo, X. & Sinnott, M. L. (1993). A kinetic-isotope-effect study of catalysis by *Vibrio cholerae* neuraminidase. *Biochem. J.* 294, 653-656.

- Haseley, S. R., Talaga, P., Kamerling, J. P. & Vliegthart, J. F. (1999). Characterization of the carbohydrate binding specificity and kinetic parameters of lectins by using Surface Plasmon Resonance. *Anal. Biochem.* 274, 203-210.
- Haselhorst, T., Weimar, T. & Peters, T. (2001). Molecular recognition of sialyl Lewis(x) and related saccharides by two lectins. *J. Am. Chem. Soc.* 123, 10705-10714.
- Heimer, R. & Meyer, K. (1956). Studies on sialic acids of submaxillary mucoid. *Proc. Natl. Acad. Sci. USA* 42, 728.
- Hendrickson, W. A. (1985). Stereochemically restrained refinement of macromolecular structures. In *Methods of enzymology* (Wyckoff, H. W., Hirs, C. H. & Timasheff, S. N., ed.), Vol. 115, pp. 252-270. Academic Press Inc., London.
- Hirst, G. K. (1941). The agglutination of red cells by allantoic fluid of chick embryos infected with influenza virus. *Science* 94, 22-23.
- Holmquist, L. (1975). Activation of *Vibrio cholerae* neuraminidase by divalent cations. *FEBS Lett.* 50, 269-271.
- Holzer, C. T., von Itzstein, M., Jin, B., Pegg, M. S., Stewart, W. P. & Wen-Yang, W. U. (1993). Inhibition of sialidases from viral, bacterial and mammalian sources by analogues of 2-deoxy-2,3-didehydro-N-acetylneuraminic acid modified at the C-4 position. *Glycoconj. J.* 10, 40-44.
- Hoyer, L. L., Roggentin, P., Schauer, R. & Vimr, E. R. (1991). Purification and properties of cloned *Salmonella typhimurium* LT2 sialidase with virus-typical kinetic preference for sialyl  $\alpha$ -2-3- linkages. *J. Biochem. (Tokyo)* 110, 462-467.
- Imberty, A., Gautier, C., Lescar, J., Perez, S., Wyns, L. & Loris, R. (2000). An unusual carbohydrate binding site revealed by the structures of two *Maackia amurensis* lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* 275, 17541-17548.

- Iwamori, M., Ohta, Y., Uchida, Y. & Tsukada, Y. (1997). *Arthobacter ureafaciens* sialidase isoenzymes, L, M1 and M2, cleave fucosyl GM1. *Glycoconj. J.* 14, 67-73.
- Janakiraman, M. N., White, C. L., Laver, W. G., Air, G. M. & Luo, M. (1994). Structure of influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dehydro analog at 1.8-Å resolution: implications for the catalytic mechanism. *Biochemistry* 33, 8172-8179.
- Jermyn, W. S. & 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.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density map and the location of errors in these models. *Acta Cryst.* A47, 110-119.
- Karaolis, D. K., Somara, S., Maneval-Jr, D. R., Johnson, J. A. & Kaper, J. B. (1999). A bacteriophage encoding a pathogenicity island, type-IV pilus and a phage receptor in cholera bacteria. *Nature* 399, 375-379.
- Khorlin, A. Y., Privalova, I. M., Zakstelskaya, L. Y., Molibog, E. V. & Evstigneeva, N. A. (1970). Synthetic inhibitors of *Vibrio cholerae* neuraminidase and neuraminidases of some influenza virus strains. *FEBS Lett.* 8, 17-19.
- Kim, C. U., Lew, W., Williams, M. A., Liu, H., Zhang, L., Swaminathan, S., Bischofberger, N., Chen, M. S., Mendel, D. B., Tai, C. Y., Laver, W. G. & Stevens, R. C. (1997). Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* 119, 681-690.

- Klein, J., Meinecke, R., Mayer, M. & Mayer, B. (1999). Detecting binding affinity to immobilized receptor proteins in compound libraries by HR-MAS-STD NMR. *J. Am. Chem. Soc.* 121, 5336-5337.
- Koch, R. (1894). An address on cholera and its bacillus. *Br. Med. J.* 2, 453-459.
- Krell, T., Greco, F., Nicolai, M. C., Dubayle, J., Renauld-Mongenie, G., Poisson, N. & Bernard, I. (2003). The use of microcalorimetry to characterize tetanus neurotoxin, pertussis toxin and filamentous haemagglutinin. *Biotechnol. Appl. Biochem.* 38, 241-251.
- Kronis, K. A. & Carver, J. P. (1982). Specificity of isolectins of wheat germ agglutinin for sialyloligosaccharides: a 360-MHz proton nuclear magnetic resonance binding study. *Biochemistry* 21, 3050-3057.
- Ladbury, J. E. (2001). Isothermal titration calorimetry: application to structure-based drug design. *Thermochim. Acta* 380, 209-215.
- Ladbury, J. E. & 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., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283-291.
- Lawrence, M. C., Borg, N. A., Streltsov, V. A., Pilling, P. A., Epa, V. C., Varghese, J. N., McKimm-Breschkin, J. L. & Colman, P. M. (2004). Structure of the haemagglutinin-neuraminidase from human parainfluenza virus type III\*1. *J. Mol. Biol.* 335, 1343-1357.
- Leavitt, S. & Freire, E. (2001). Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* 11, 560-566.

- Leslie, A. G. (1992). *Crystallographic Computing 5: From chemistry to biology*. (Moras, D., Pojarny, A. D. & Thierry, J. C., ed). Oxford University Press, Oxford.
- Lis, H. & Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* 98, 637-674.
- Luo, Y., Li, S. C., Chou, M. Y., Li, Y. T. & Luo, M. (1998). The crystal structure of an intramolecular trans-sialidase with a NeuAc  $\alpha$ -2-3-Gal specificity. *Structure* 6, 521-530.
- Malisan, F. & Testi, R. (2002). GD3 ganglioside and apoptosis. *Biochim. Biophys. Acta* 1585, 179-187.
- Margalit, H., Foscher, N. & Ben-Sasson, S. A. (1993). Comparative analysis of structurally defined heparin binding sequences reveals a distinct spatial distribution of basic residues. *J. Biol. Chem.* 268, 19228-19231.
- Maru, I., Ohnishi, J., Ohta, Y. & Tsukada, Y. (2002). Why is sialic acid attracting interest now? complete enzymatic synthesis of sialic acid with N-acylglucosamine 2-epimerase. *J. Biosci. Bioeng.* 93, 258-265.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* 33, 491-497.
- May, A. P., Robinson, R. C., Vinson, M., Crocker, P. R. & Jones, E. Y. (1998). Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 Å resolution. *Mol. Cell* 1, 719-728.
- Mayer, M. & Meyer, B. (1999). Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew Chem. Int. Ed. Engl.* 38, 1784-1788.
- Mayer, M. & Meyer, B. (2001). Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* 123, 6108-6117.

- Mepheron, A. & Geige, R. (2001). Crystallization. In *International Tables for Crystallography: Crystallography of Biological macromolecules* (Rossmann, M. G. & Eddy, A., ed.), Vol. F, pp. 81-93. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Meindl, P., Bodo, G., Palese, P., Schulman, J. & Tuppy, H. (1974). Inhibition of neuraminidase activity by derivatives of 2-deoxy-2,3-dehydro-N-acetylneuraminic acid. *Virology* 58, 457-463.
- Meindl, P. & Tuppy, H. (1969). 2-Deoxy-2,3-dehydrosialic acids. II. Competitive inhibition of *Vibrio cholerae* neuraminidase by 2-deoxy-2,3-dehydro-N-acetylneuraminic acids. *Hoppe Seylers Z. Physiol. Chem.* 350, 1088-1092.
- Merritt, E. A., Sarfaty, S., Jobling, M. G., Chang, T., Holmes, R. K., Hirst, T. R. & Hol, W. G. (1997). Structural studies of receptor binding by cholera toxin mutants. *Protein Sci.* 6, 1516-1528.
- Meyer, B. & Peters, T. (2003). NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem. Int. Ed. Engl.* 42, 864-890.
- Meyer, B., Weimar, T. & Peters, T. (1997). Screening mixtures for biological activity by NMR. *Eur. J. Biochem.* 246, 705-709.
- Miller, R. L., Collawn, J. F. Jr & Fish, W. W. (1982). Purification and macromolecular properties of a sialic acid-specific lectin from the slug *Limax flavus*. *J. Biol. Chem.* 257, 7574-7580.
- Milner, C. M., Smith, S. V., Carrillo, M. B., Taylor, G. L., Hollinshead, M. & Campbell, R. D. (1997). Identification of a sialidase encoded in the human major histocompatibility complex. *J. Biol. Chem.* 272, 4549-4558.

- Miyagi, T., Wada, T., Iwamatsu, A., Hata, K., Yoshikawa, Y., Tokuyama, S., Sawada, M., Milner, C. M., Smith, S. V., Carrillo, M. B., Taylor, G. L., Hollinshead, M., Campbell, R. D., Bonten, E., van der Spoel, A., Fornerod, M., Grosveld, G. & d'Azzo, A. (1999). Molecular cloning and characterization of a plasma membrane-associated sialidase specific for gangliosides. *J. Biol. Chem.* 274, 5004-5011.
- Mohr, E. (1960). Hemmung der Neuraminidase aus *Vibrio cholerae*. *Z. Naturforschg.* 15b, 575-577.
- Moore, J. M. (1999). NMR techniques for characterization of ligand binding: utility for lead generation and optimization in drug discovery. *Biopolymers* 51, 221-243.
- Muraki, M., Ishimura, M. & Harata, K. (2002). Interactions of wheat-germ agglutinin with GlcNAc- $\beta$ -1,6-Gal sequence. *Biochim. Biophys. Acta* 1569, 10-20.
- Myers, R. W., Lee, R. T., Lee, Y. C., Thomas, G. H., Reynolds, L. W. & Uchida, Y. (1980). The synthesis of 4-methylumbelliferyl alpha-ketoside of N-acetylneuraminic acid and its use in a fluorometric assay for neuraminidase. *Anal. Biochem.* 101, 166-74.
- Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Cryst.* A50, 157-163.
- Nayal, M. & Di Cera, E. (1994). Predicting Ca(2+)-binding sites in proteins. *Proc. Natl. Acad. Sci. USA* 91, 817-821.
- Neu, H. C. & Heppel, L. A. (1965). The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. *J. Biol. Chem.* 240, 3685-3692.
- Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281-296

- Nohle, U., Shukla, A., Schroder, C., Reuter, G., Schauer, R., Kamerling, J. & Vliegthart, J. (1985). Structural parameters and natural occurrence of 2-deoxy-2,3-didehydro-N- glycoloylneuraminic acid. *Eur. J. Biochem.* 152, 459-463.
- O'Brien, R., Ladbury, J. E. & Chowdhry, B. Z. (2001). Isothermal titration calorimetry of biomolecules. In *Protein-ligand interactions: hydrodynamic and calorimetry. The practical approach series* (Harding, S. E. & Chowdhry, B. Z., ed.), pp. 263-286. Oxford University Press, Oxford.
- Otwinoski, Z. & Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Macromolecular Crystallography, part A* (Carter, C. W. & Sweet, R. M., ed.), Vol. 276, pp. 307-326. Academic Press.
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 4, 2411-2423.
- Patterson, A. L. (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Phys. Rev.* 46, 372-376.
- Pickens, J. C., Merritt, E. A., Ahn, M., Verlinde, C. L., Hol, W. G. & Fan, E. (2002). Anchor-based design of improved cholera toxin and *E. coli* heat-labile enterotoxin receptor binding antagonists that display multiple binding modes. *Chem. Biol.* 9, 215-224.
- Pilatte, Y., Bignon, J. & Lambre, C. R. (1993). Sialic acids as important molecules in the regulation of the immune system: pathophysiological implications of sialidases in immunity. *Glycobiology* 3, 201-218.
- Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661-665.

- Poppe, L., Brown, G. S., Philo, J. S., Nikrad, P. V. & Shah, B. H. (1997). Conformation of sLex tetrasaccharide, free in solution and bound to E-, P-, and L-selectin. *J. Am. chem. soc.* 119, 1727-1736.
- Potier, M., Mameli, L., Belisle, M., Dallaire, L. & Melancon, S. B. (1979). Fluorometric assay of neuraminidase with a sodium (4-methylumbelliferyl-alpha-D-N-acetylneuraminate) substrate. *Anal. Biochem.* 94, 287-296.
- Pye, J. & Curtain, C. C. (1961). Electrophoretic, sedimentation and diffusion behaviour of crystalline neuraminidase from *Vibrio cholerae*. *J. Gen. Microbiol.* 24, 423-425.
- Read, R. J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Cryst.* D57, 1373-1382.
- Reidl, J. & Klose, K. E. (2002). *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiol. Rev.* 26, 125-139.
- Reuter, G. & Gabuis, H. -J. (1996). Sialic acids: structure-analysis-metabolism-occurrence-recognition. *Biol. Chem. Hoppe-Seyler* 377, 325-342.
- Rhodes, G. (2000). *Crystallography made crystal clear: A guide for users of Macromolecular models*. second edition, Academic Press, London.
- Roggentin, P., Rothe, B., Kaper, J. B., Galen, J., Lawrisuk, L., Vimr, E. R. & Schauer, R. (1989). Conserved sequences in bacterial and viral sialidases. *Glycoconj. J.* 6, 349-353.
- Roggentin, P., Schauer, R., Hoyer, L. L. & Vimr, E. R. (1993). The sialidase superfamily and its spread by horizontal gene transfer. *Mol. Microbiol.* 9, 915-921.
- Rosenberg, A. (2004). Pneumococcus virulence factor sialidase: A new direction in neuro-AIDS research. *J. Neuroimmunol.* 147, 33-34.
- Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* 15, 24-31.

- Sabesan, S., Neira, S., Davidson, F., Duus, J. O. & Bock, K. (1994). Synthesis, enzymic, and NMR studies of novel sialoside probes: unprecedented, selective neuraminidase hydrolysis of and inhibition by C-6-(methyl)-Gal Sialosides. *J. Am. Chem. Soc.* 116, 1616-1634.
- Sack, D. A., Lyke, C., McLaughlin, C. & Suwanvanichkij, V. (2001). Antimicrobial resistance in shigellosis, cholera and campylobacteriosis. World Health Organization.
- Sack, D. A., Sack, R. B., Nair, G. B. & Siddique, A. K. (2004). Cholera. *Lancet* 363, 223-233.
- Sakurada, K., Ohta, T. & Hasegawa, M. (1992). Cloning, expression, and characterization of the *Micromonospora viridifaciens* neuraminidase gene in *Streptomyces lividans*. *J. Bacteriol.* 174, 6896-6903.
- Sauter, N. K., Bednarski, M. D., Wurzburg, B. A., Hanson, J. E., Whitesides, G. M., Skehel, J. J. & Wiley, D. C. (1989). Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. *Biochemistry* 28, 8388-8396.
- Schauer, R. (1985). Sialic acids and their role as biological masks. *Trends Biochem. Sci.* 10, 357-360.
- Schauer, R. (1995). Biochemistry and role of sialic acids. In *Biology of the sialic acids* (Rosenberg, A., ed.), pp. 7-67. Plenum Press, New York.
- Schneider, D. R., Parker, C. D. (1982). Purification and characterization of the mucinase of *Vibrio cholerae*. *J. Infect. Dis.* 145, 474-482.
- Schon, A. & Freire, E. (1989). Thermodynamics of intersubunit interactions in cholera toxin upon binding to the oligosaccharide portion of its cell surface receptor, ganglioside GM1. *Biochemistry* 28, 5019-5024.

- Schramm, G. & Mohr, E. (1959). Purification of Neuraminidase from *Vibrio cholerae*. *Nature* 183, 1677-1678.
- Schreiner, E., Zbiral, E., Kleineidam, R. G. & Schauer, R. (1991a). 2,3-Didehydro-2-deoxysialic acids structurally varied at C-4 and their behaviour towards the sialidase from *Vibrio cholerae*. *Carbohydr. Res.* 216, 129-134.
- Schreiner, E., Zbiral, E., Kleineidam, R. G. & Schauer, R. (1991b). 2,3-Didehydro-2-deoxysialic acids structurally varied at C-5 and their behaviour towards the sialidase from *Vibrio cholerae*. *Carbohydr. Res.* 216, 61-66.
- Scopes, R. K. (1974). Measurement of protein by spectrophotometry at 205 nm. *Anal. Biochem.* 59, 277-282.
- Scopes, R. K. (1994). *Protein Purification: Principles and Practice*. Third edition. Springer Advanced Texts in Chemistry (Cantor, C. R., Ed.), Springer-Verlag, New York.
- Sedlacek, H. H. & Seiler, F. R. (1974). Demonstration of *Vibrio-Cholerae* Neuraminidase (VCN) on the surface of VCN-treated Cells. *Behring Inst. Mitt.* 55, 254-257.
- Sherwood, D. (1976). *Crystals, X-rays and Proteins*, Longman Group Limited, London.
- Snyder, J. D. & Walker, W. A. (1987). Structure and function of intestinal mucin: developmental aspects. *Int. Arch. Allergy Appl. Immunol.* 82, 351-356.
- Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. (2000). Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell* 103, 467-479.
- Stark, J. R., H. J., Wiegand, H. (1974). Neuraminidase, a virulence factor in *Vibrio cholerae* infection. *Behring Inst. Mitt.* 55, 145-146.

- Stehle, T. & Harrison, S. C. (1996). Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure* 4, 183-194.
- Stein, P. E., Boodhoo, A., Armstrong, G. D., Heerze, L. D., Cockle, S. A., Klein, M. H. & Read, R. J. (1994). Structure of a pertussis toxin-sugar complex as a model for receptor binding. *Nat. Struct. Biol.* 1, 591-596.
- Stewart-Tull, D. E., Ollar, R. A. & Scobie, T. S. (1986). Studies on the *Vibrio cholerae* mucinase complex. I. Enzymic activities associated with the complex. *J. Med. Microbiol.* 22, 325-333.
- Stewart-Tull, D. E., Lucas, C. & Bleakley, C. R. (2004). Experimental immunisation and protection of guinea pigs with *Vibrio cholerae* toxoid and mucinases, neuraminidase and proteinase. *Vaccine* In Press.
- Tame, J. R., O'Brien, R., Ladbury, J. E. (1998). Isothermal Titration Calorimetry of Biomolecules. In *Biocalorimetry: Applications of calorimetry in the biological sciences* (Ladbury, J. E. & Chowdhry, B. Z., ed.), pp. 27-38. Wiley & Sons Ltd, West Sussex.
- Taylor, G. L. (1996). Sialidases: structures, biological significance and therapeutic potential. *Curr. Opin. Struct. Biol.* 6, 830-837.
- Taylor, G., Vimr, E. R., Garman, E. F. & Laver, G. W. (1992). Purification, crystallization and preliminary crystallographic study of neuraminidase from *Vibrio cholerae* and *Salmonella typhimurium* LT2. *J. Mol. Biol.* 226, 1287-1290.
- Taylor, N. R. & 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.

- Taylor, R. K., Miller, V. L., Furlong, D. B. & Mekalanos, J. J. (1987). Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc. Natl. Acad. Sci. USA* 84, 2833-2837.
- Thobhani, S., Ember, B., Siriwardena, A., Boons, G. J., Parr, I. B., Horenstein, B. A., Streicher, H. & Bohner, C. (2003). Multivalency and the Mode of Action of Bacterial Sialidases. *J. Am. Chem. Soc.* 125, 7154-7155.
- Troncoso, M. F., Iglesias, M. M., Isecke, R., Todel, C. W., Brossmer, R., Powell, L. D., Varki, A., Wagner, M. & Fischer, E. (2000). Specificity of the binding site of the sialic acid-binding lectin from ovine placenta, deduced from interactions with synthetic analogues. *Glycoconj. J.* 17, 705-711.
- Varghese, J. N., McKimm-Breschkin, J. L., Caldwell, J. B., Kortt, A. A. & Colman, P. M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* 14, 327-32.
- Varki, A. (1992). Diversity in the sialic acids. *Glycobiology* 2, 25-40.
- Varki, A. (1997). Sialic acids as ligands in recognition phenomena. *FASEB J.* 11, 248-255.
- Venerando, B., Cestaro, B., Fiorilli, A., Ghidoni, R., Preti, A. & Tettamanti, G. (1982). Kinetics of *Vibrio cholerae* sialidase action on gangliosidic substrates at different supramolecular-organizational levels. *Biochem. J.* 203, 735-742.
- Vijayan, M. & Chandra, N. (1999). Lectins. *Curr. Opin. Struct. Biol.* 9, 707-714.
- Vimr, E. R., Lawrisuk, L., Galen, J. & Kaper, J. B. (1988). Cloning and expression of the *Vibrio cholerae* neuraminidase gene nanH in *Escherichia coli*. *J. Bacteriol.* 170, 1495-1504.
- 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., Oliver, S. W., Colman, P. M., Varghese, J. N., Ryan, D.

- M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M. & Penn, R. C. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 363, 418-423.
- Vorwerk, S. & Vasella, A. (1998). Carbocyclic analogues of N-Acetyl-2,3- didehydro-2-deoxy-D-neuraminic acid (Neu5Ac2en, DANA): synthesis and inhibition of viral and bacterial neuraminidases. *Angew. Chem. Int. Ed. Engl.* 37, 1732-1733.
- Wadso, I. (2001). Isothermal microcalorimetry: current problems and prospects. *J. Therm. Anal. Calorim.* 64, 75-84.
- Waldor, M. K & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910-1914.
- Weber, P. C. & Salemme, F. R. (2003). Applications of calorimetric methods to drug discovery and the study of protein interactions. *Curr. Opin. Struct. Biol.* 13, 115-121.
- Weis, W. I., Brown, J. H., Cusack, S., Paulson, J. C., Skehel, J. J. & Wiley, D. C. (1988). Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature* 333, 426-431.
- Weis, W. I., Brunger, A. T., Skehel, J. J. & Wiley, D. C. (1990). Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* 212, 737-761.
- Weis, W. I. & Drickmer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* 65, 441-473.
- Wemmer, D. E. & Williams, P. G. (1994). Use of nuclear magnetic resonance in probing ligand-macromolecule interactions. *Methods Enzymol.* 239, 739-767.
- Wilson, A. J. C. & Prince, E. (1999). *International Tables for Crystallography: Mathematical, Physical and Chemical Tables*. Vol. C. second edition. Kluwer Academic Publishers, Dordrecht/Boston/London.

- Wilson, J. C., Kiefel, M. J., Angus, D. I. & von Itzstein, M. (1999). Investigation of the stability of thiosialosides toward hydrolysis by sialidases using NMR spectroscopy. *Org. Lett.* 1, 443-446.
- Wilson, J. C., Thomson, R. J., Dyason, J. C., Florio, P., Quelch, K. J., Abo, S. & von Itzstein, M. (2000). The design, synthesis and biological evaluation of neuraminic acid-based probes of *Vibrio cholerae* sialidase. *Tetrahedron: Asymmetry.* 11, 53-73.
- Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* 179, 131-137.
- World Health Organization. (2000). WHO report on global surveillance of epidemic-prone infectious diseases. World Health Organization.
- Wright, C. S. (1990). 2.2 Å resolution structure analysis of two refined N-acetylneuraminyl-lactose-wheat germ agglutinin isolectin complexes. *J Mol. Biol.* 215, 635-651.
- Yu, K. (1995). Biochemistry and Function of Sialidases. In *Biology of the sialic acids* (Rosenberg, A., ed.), pp. 261-313. Plenum Press, New York.
- Zaitsev, V., von Itzstein, M., Groves, D., Kiefel, M., Takimoto, T., Portner, A. & Taylor, G. L. (2004). A second sialic acid binding site in Newcastle disease virus hemagglutinin-neuraminidase: implications for fusion. *J. Virology* In press.
- Zartler, E. R., Yan, J., Mo, H., Kline, A. D. & Shapiro, M. J. (2003). ID NMR Methods in ligand-receptor interactions. *Curr. Top. Med. Chem.* 3, 25-37.
- Zbiral, E., Schreiner, E., Christian, R., Kleineidam, R. G. & Schauer, R. (1989). Synthesis of 2,7-, 2,8-, and 2,9-dideoxy- and 2,4,7-trideoxy-2,3-didehydro-N-acetylneuraminic acids and their behavior towards sialidase. *Liebigs Ann. Chem.*, 159-165.

Zhang, R. G., Scott, D. L., Westbrook, M. L., Nance, S., Spangler, B. D., Shipley, G. G.  
& Westbrook, E. M. (1995). The three-dimensional crystal structure of cholera toxin. *J. Mol. Biol.* 251, 563-573.