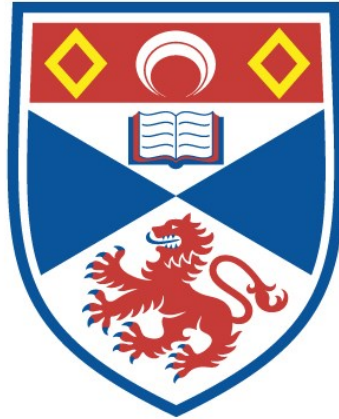


PROTEIN-CARBOHYDRATE INTERACTIONS

Davina Noelle Moothoo

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



1998

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

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

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/14528>

This item is protected by original copyright

**PROTEIN - CARBOHYDRATE
INTERACTIONS**

Davina Noelle Moothoo

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



26th August 1998

ProQuest Number: 10171019

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10171019

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

TL
D 178

Contents.

Contents	2
List of Figures	5
List of Tables	9
List of Abbreviations	12
Declaration	13
Abstract	15
Acknowledgements	16
Chapter 1 : Protein - Carbohydrate Interactions	17
1.1. Summary	18
1.2. Introduction	18
1.3. Energetics of protein - carbohydrate interactions	28
1.4. Hydrogen bonding and van der Waals interactions	32
1.5. Carbohydrate conformation	37
1.6. Monosaccharide binding site geometry and specificity	39
1.7. Oligosaccharide binding	49
1.8. Charged carbohydrate groups	56
1.9. Ligand multivalency	58
1.10. Concluding remarks	62
Chapter 2 : The 1.75Å structure of concanavalin A complexed with α 1-2 mannobiose	63
2.1. Summary	64
2.2. Introduction	65
2.3. Experimental	69

2.4. Analysis of the final model	75
2.5. Results and discussion	79
2.6. Conclusions	97
2.7. Future work	98
Chapter 3 : The 2.75Å structure of concanavalin A complexed with methyl α 1-2 mannobioside	99
3.1. Summary	100
3.2. Introduction	101
3.3. Experimental	103
3.4. Analysis of the final model	111
3.5. Results and discussion	114
3.6. Conclusions	128
3.7. Future work	128
Chapter 4 : The 2.7Å structure of concanavalin A complexed with a pentasaccharide	129
4.1. Summary	130
4.2. Introduction	130
4.3. Experimental	133
4.4. Analysis of the final model	138
4.5. Results and discussion	143
4.6. Conclusions	162
4.7. Future work	162
Chapter 5 : Crystal structures of concanavalin A complexed with fructose	164
5.1. Summary	165

5.2. Introduction	165
5.3. Experimental	167
5.4. Analysis of the final model	178
5.5. Results and discussion	184
5.6. Conclusions	198
5.7. Future work	199
Appendix 1	200
Bibliography	203
Publications	219

List of Figures.

1.1.	The con A tetramer	20
1.2.	The extended binding site of con A	21
1.3.	Numbering scheme for mannose	22
1.4.	Monosaccharides; Structures and abbreviations	23
1.5.	Oligosaccharides; Structures and abbreviations	25
1.6.	Enthalpy - entropy compensation plot for con A data	31
1.7.	Cooperative hydrogen bond	33
1.8.	Bidentate hydrogen bond	33
1.9.	Conserved water molecule in an antibody fragment - trisaccharide complex	36
1.10.	Group I binding	39
1.11.	Group II binding	40
1.12.	Extensive hydrogen bonding in the ABP - galactose complex	41
1.13.	The D loop of the legume lectins	44
1.14.	Binding of Man and Gal by the legume lectins	45
1.15.	Monosaccharide binding by GNA	46
1.16.	The binding site of WGA	47
1.17.	Monosaccharide binding site of Jacalin	48
1.18.	Involvement of metals in carbohydrate binding	49
1.19.	LOL II complexed with a fucosylated glycan	51
1.20.	The GS4 - Lewis b binding site	53
1.21.	Distortion of a sugar ring by lysozyme	55
1.22.	Extended binding site of glycogen phosphorylase	56
1.23.	The heparin linked dimer of aFGF	58

1.24.	Cross-linked bovine heart Gal-1	59
1.25.	The two modes of binding of oligomannose binding by GNA	60
1.26.	Cross-linked con A monomers	61
2.1.	α 1-2 mannoside	66
2.2.	Ramachandran plot for the 1.75Å α 1-2 mannoside - con A structure	77
2.3.	Backbone temperature factor plot for the α 1-2 mannoside - con A structure	78
2.4.	F_o-F_c omit map contoured at 2.8σ	82
2.5.	Hydrogen bond diagram of the α 1-2 mannoside - con A complex	85
2.6.	Comparison of native and α 1-2 mannoside bound monosaccharide binding sites	89
2.7.	Comparison of the sugar positions in the α 1-2 mannoside and trimannoside bound concanavalin A structures	91
2.8.	Displacement of water from the binding site	93
2.9.	Differences in the monosaccharide binding site	95
3.1.	Methyl α 1-2 mannoside	103
3.2.	Ramachandran plot of the con A - methyl α 1-2 mannoside complex	113
3.3.	Temperature factor plots along the backbone of each subunit	114
3.4.	F_o-F_c electron density observed in the binding sites	117
3.5.	Schematic diagram of the hydrogen bonds in subunit A	118
3.6.	Methyl α -12 mannoside and α 1-2 mannoside	121
3.7.	Schematic diagram of the hydrogen bonds in subunit D	123
3.8.	The binding site of con A displaying the two distinct binding modes	125
4.1.	The core oligosaccharide of the N-linked glycan found on mammalian cell	131

	surfaces	
4.2.	Ramachandran plot for the final model of the con A - pentasaccharide complex	140
4.3.	Temperature factor plot along the backbone of each subunit	142
4.4.	Packing of the two tetramers found in the asymmetric unit	143
4.5.	Differences between the ABCD tetramer and the EFGH tetramer	145
4.6.	$F_o - F_c$ electron density seen in the binding site	147
4.7.	Schematic representation of the hydrogen bonds between con A and the pentasaccharide	150
4.8.	Stereoview of the displacement of waters by the pentasaccharide	152
4.9.	Stereoview of the pentasaccharide binding site superimposed onto the native structure	155
4.10.	Crystal packing in the native binding site	156
4.11.	The con A - pentasaccharide complex	157
4.12.	Prediction of the mode of binding of high mannose oligosaccharides	163
5.1.	α -D-fructofuranose	166
5.2.	Superposition of α -D-fructofuranose and methyl α -D-mannopyranoside	167
5.3.	Section of one 0.8° oscillation of the C222 ₁ con A - fructose complex	173
5.4.	Ramachandran plot for the final model of the P2 ₁ 2 ₁ 2 ₁ con A - fructose complex	179
5.5.	Temperature factor plots along the backbone of each subunit in the P2 ₁ 2 ₁ 2 ₁ con A - fructose complex	181
5.6.	Ramachandran plot of the final model of the C222 ₁ con A - fructose complex	182

5.7.	Temperature factor plots along the protein backbone for each subunit of the C222 ₁ con A - fructose complex	184
5.8.	F _o -F _c electron density contoured at 2σ seen in the binding site of subunit A	187
5.9.	F _o -F _c electron density contoured at 2σ seen in the binding site of subunit B	188
5.10.	F _o -F _c electron density contoured at 2σ seen in the binding site of subunit C prior to inclusion of fructose	188
5.11.	F _o -F _c electron density contoured at 2σ seen in the binding site of subunit C calculated after inclusion of α-D-fructose	189
5.12.	2F _o -F _c density of subunit C binding site	190
5.13	Comparison of bound fructose and bound mannose	192
5.14	F _o -F _c electron density seen in the binding site of subunit D	192
5.15.	Organisation of the two dimers present in the asymmetric unit	194
5.16.	F _o -F _c electron density contoured at 2σ observed in the binding site of subunit A	196
5.17.	F _o -F _c electron density contoured at 2σ observed in the binding site of subunit B	197
5.18.	F _o -F _c electron density contoured at 2σ observed in the binding site of subunit C	197
5.19.	F _o -F _c electron density contoured at 2σ observed in the binding site of subunit D	198

List of Tables.

1.1.	Protein structures discussed in Chapter 1	27
1.2.	Thermodynamics of con A - carbohydrate binding	30
2.1.	Thermodynamics for binding of mannobioses to con A	67
2.2.	Quality of the 2.0Å α 1-2 mannobiose data	70
2.3.	Quality of the 1.75Å α -12 mannobiose data	71
2.4.	Progress of refinement of the 2.0Å α 1-2 mannobiose - con A structure	73
2.5.	Quality of the 2.0Å α 1-2 mannobiose - con A structure	74
2.6.	Quality of the 1.75Å α 1-2 mannobiose - con A structure	76
2.7.	Metal to ligand distances	81
2.8.	Hydrogen bonds and polar contacts (<3.5Å) between α 1-2 mannobiose and con A	83
2.9.	Van der Waals interactions (<4.0Å) between α 1-2 mannobiose and con A	84
2.10.	R.m.s. deviations for the C α superposition of con A complexes	87
2.11.	Temperature factor ratios of the saccharide binding loops	92
3.1.	Thermodynamics of con A binding mannobioses and their methylated counterparts	102
3.2.	Cell output from DENZO during the first 40 oscillations of data collection	105
3.3.	Quality of data for the con A - methyl α 1-2 mannobioside complex	106
3.4.	Rotation and translation solutions	107
3.5.	Progress of the refinement of the con A - methyl α 1-2 mannobioside complex	110
3.6.	NCS restraints used in the refinement of the con A - methyl α 1-2 mannobioside complex	111

3.7.	Final model statistics of the con A - methyl α 1-2 mannoside complex	112
3.8.	Average temperature factors for each subunit	114
3.9.	Metal to ligand distances and ligand temperature factors	116
3.10.	Hydrogen bonding <u>d</u> istances between methyl α 1-2 mannoside and con A	119
3.11.	Van der Waals interactions between methyl α 1-2 mannoside and con A	120
3.12.	Dihedral angles around the inter-sugar glycosidic linkage of the mannoside sugars	121
4.1.	Thermodynamic data for con A binding to oligosaccharides at the core of the <i>N</i> -linked glycan	132
4.2.	Data quality for the 2.7Å complex of the pentasaccharide - con A complex	134
4.3.	NCS restraints used in positional and temperature factor refinement	136
4.4.	Progress of refinement of the con A - pentasaccharide complex	137
4.5.	Final model statistics of the 2.7Å con A pentasaccharide complex	139
4.6.	Average temperature factors for each subunit in the con A - pentasaccharide complex	141
4.7.	Metal to ligand distances and temperature factors in the con A - pentasaccharide complex	146
4.8.	Hydrogen bonding and polar contact distances between con A and the pentasaccharide	148
4.9.	Van der Waals contacts between con A and the pentasaccharide	149
4.10.	Average temperature factor of the binding loops in the con A - pentasaccharide complex	151
4.11.	Torsion angles for the various oligosaccharides	153

4.12.	Differences amongst the con A complexes in the contacts at the carbohydrate binding site	158
5.1.	Data quality for the P2 ₁ 2 ₁ 2 ₁ fructose - con A complex	168
5.2.	Rotation and translation solutions for the P2 ₁ 2 ₁ 2 ₁ crystal form	169
5.3.	Progress of refinement of the P2 ₁ 2 ₁ 2 ₁ con A - fructose complex	171
5.4.	NCS restraints used in positional and temperature factor refinement	172
5.5.	Quality of the C222 ₁ con A - fructose structure	174
5.6.	Rotation and translation solutions for the C222 ₁ crystal form	175
5.7.	Progress of refinement of the C222 ₁ con A - fructose complex	177
5.8.	NCS restraints used in positional and temperature factor refinement	178
5.9.	Quality of the final model of the P2 ₁ 2 ₁ 2 ₁ con A - fructose complex	180
5.10.	Temperature factors for each subunit of the P2 ₁ 2 ₁ 2 ₁ fructose - con A complex	181
5.11.	Statistics on the final model of the C222 ₁ fructose - con A complex	183
5.12.	Temperature factors for each subunit of the C222 ₁ fructose - con A complex	183
5.13.	Metal to ligand distances and temperature factors in the P2 ₁ 2 ₁ 2 ₁ con A complex	186
5.14.	Hydrogen bonds and polar contacts (<3.5Å) between fructose and con A	190
5.15.	Van der Waals interactions (<4.0Å) between fructose and con A	191
5.16.	Metal to ligand distances and temperature factors in the C222 ₁ fructose - con A complex	195
A1	Statistics on the con A - methyl α-D-mannopyranoside structure before and after CNS refinement	201

List of abbreviations.

PDB	Protein Data Bank
K_a	Association constant
NMR	Nuclear Magnetic Resonance
PEG	Polyethylene Glycol
r.m.s.	root mean square
$C\alpha$	alpha carbon of amino acid
con A	concanavalin A
OW	water

Declaration.

I, Davina Moothoo, hereby certify that this thesis, which is approximately 35,000 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 26.8.98

Signature of candidate *D Moothoo*

I was admitted as a research student in September 1995 and as a candidate for the degree of Ph.D. in September 1996; the higher study for which this is a record was carried out in the University of St. Andrews between 1995 and 1998 .

Date 26.8.98

Signature of candidate *D Moothoo*

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 26.8.98

[Handwritten Signature]
Signature of supervisor

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being 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 26.8.98

Signature of candidate *D. M. M. M.*

Abstract.

Carbohydrates are ubiquitous in nature and have become the focus of much scientific investigation. The proteins which recognise carbohydrates have become widely used in the areas of cell and molecular biology. Protein - carbohydrate interactions have been probed by theoretical, structural and thermodynamic techniques. The lectins are a class of carbohydrate binding proteins which bind carbohydrates through non covalent interactions such as hydrogen bonds and van der Waals interactions. In addition to these interactions, other factors play an important role in determining affinity such as carbohydrate conformation, solvent reorganisation and changes in the protein binding site.

The legume lectin concanavalin A specifically recognises mannose and glucose terminal residues. The thermodynamics of concanavalin A binding to carbohydrates has been well documented. Concanavalin A binds the core trimannoside and pentasaccharide of the biantennary glycan found on mammalian cell surfaces with a high affinity. This thesis describes the structural basis of carbohydrate binding by con A, through the interpretation of crystal structures of concanavalin A bound with α 1-2 mannoside, methyl α 1-2mannoside, the core pentasaccharide of the biantennary glycan and fructose. The structural information obtained from these structures is related to thermodynamic information available and unravels the importance of the role played by carbohydrate conformation, solvent reorganisation and statistical population of ligand in determining affinity. This work helps to develop an understanding of the physical basis of carbohydrate recognition.

Acknowledgements.

Firstly, thanks go to my supervisor Jim Naismith for his constant enthusiasm and encouragement over the past three years, and for helping me to achieve my goals. Thanks also go to Steve Homans, Trevor Rutherford, Charlie Weller and Tony Chiovotti for useful discussions. A special thanks goes to everyone in the Naismith group, especially Stephen for making the lab a good place to work.

Thanks are not enough for my parents, they are truly amazing.

Finally, a special thanks goes to Alex, who achieved the impossible and made the sun shine every day in St. Andrews.

Chapter 1

Protein - Carbohydrate Interactions.

1.1. Summary

Carbohydrates are ubiquitous in biological systems. A sophisticated array of these molecules are found on cell surfaces and they are recognised with varying degrees of specificities by carbohydrate binding proteins such as lectins, anti-carbohydrate antibodies, bacterial periplasmic binding proteins and some enzymes. The interactions between proteins and carbohydrates are attractive therapeutic targets. However, the ubiquity of carbohydrates presents a huge challenge for rational design of therapeutics. The physical basis of protein - carbohydrate interactions is still too poorly understood to permit such design. Improving our understanding needs an atomic level understanding of the recognition of a carbohydrate by a protein. In recent years, X-ray crystallography together with isothermal titration microcalorimetry has begun to provide a powerful insight into the molecular processes governing protein - carbohydrate interactions. This chapter will summarise the current information available from X-ray structures and thermodynamics studies, identifying common features of protein - carbohydrate interactions.

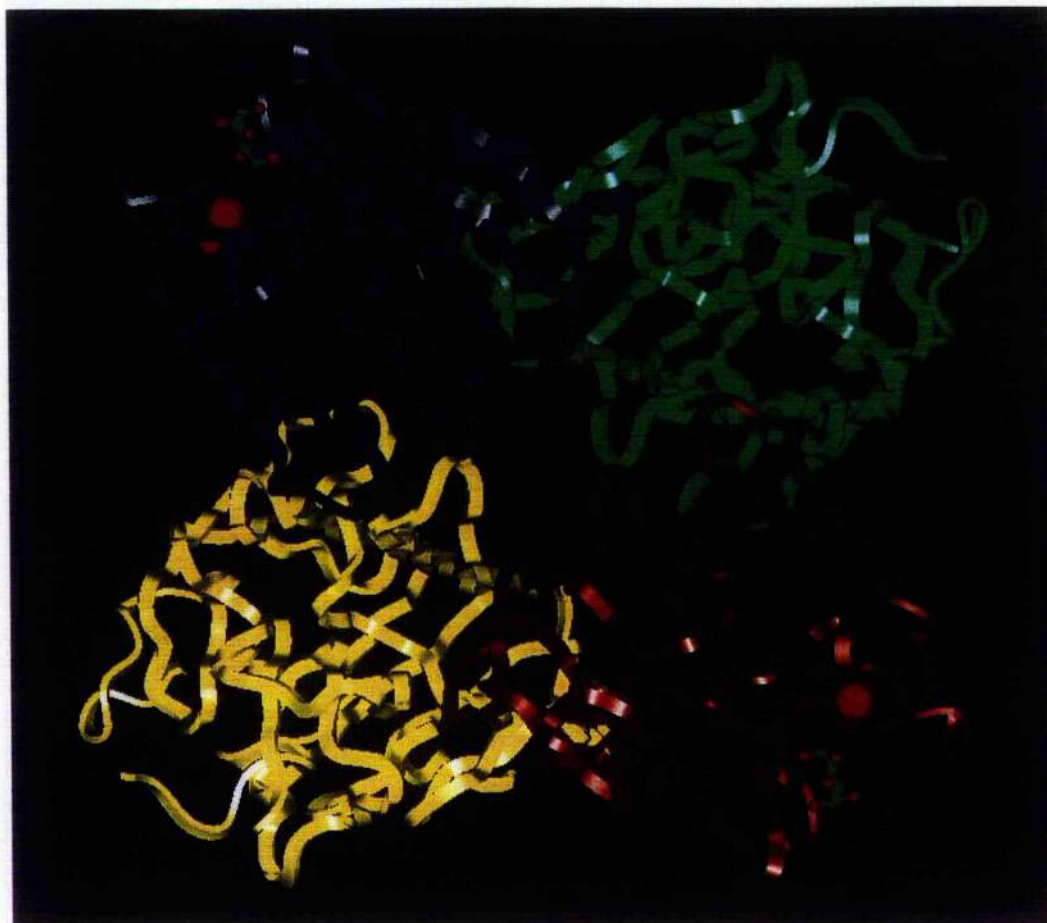
Concanavalin A from the Jack bean (*Canavalia ensiformis*), is the most well studied lectin. It was the first of the lectins to have its three dimensional structure solved by X-ray crystallography, the first to have a carbohydrate bound structure solved and the first lectin to be solved to atomic resolution. The thermodynamics of its binding to both natural and modified carbohydrates have been extensively analysed and are discussed.

1.2. Introduction.

As the wealth of information regarding protein structure increases, common features such as sequence, fold and function have been identified enabling proteins to be grouped

accordingly. Proteins which recognise carbohydrates are ubiquitous and their functions are extremely diverse, from defense to metabolism (Sharon and Lis, 1989). They recognise carbohydrates with varying degrees of specificity, and bind them with varying degrees of affinity. As more structural and thermodynamic information becomes available, the deconvolution of molecular processes which control specificity and affinity becomes possible. The most well studied group of carbohydrate binding proteins are the lectins. The lectins bind carbohydrates reversibly and with a high degree of specificity but are devoid a catalytic activity (Goldstein and Portez, 1986). Classically, lectins have been grouped according to their source, for example legume, cereal, bacterial etc. The number of lectin crystal structures available is now approaching 150, the majority of which are with bound carbohydrate (for a comprehensive listing see <http://www.cermav.cnrs.fr/databank/>). Lectins were first characterised early in this century, when concanavalin A (con A) was isolated from the Jack Bean (Sumner, 1919). Since that time plant, animal, bacterial and viral lectins as well as the other carbohydrate binding proteins have been isolated and characterised. To date, con A is the most well studied lectin. The thermodynamics of con A - carbohydrate binding have been extensively studied (discussed in section 1.3) and its structure has been solved to atomic resolution (Deacon *et al.*, 1997). The monosaccharide binding site of con A was characterised in the first lectin carbohydrate complex studied by crystallography (Derewenda *et al.*, 1989). The 2.0Å structures of con A complexed with methyl α -D-mannopyranoside (Figure 1.1) (Naismith *et al.*, 1994) and methyl α -D-glucopyranoside (Harrop *et al.*, 1996) have allowed a detailed description of monosaccharide binding. The sugar is anchored to the protein by several direct hydrogen bonds, polar contacts and van der Waals interactions burying some 75% of its surface accessible area.

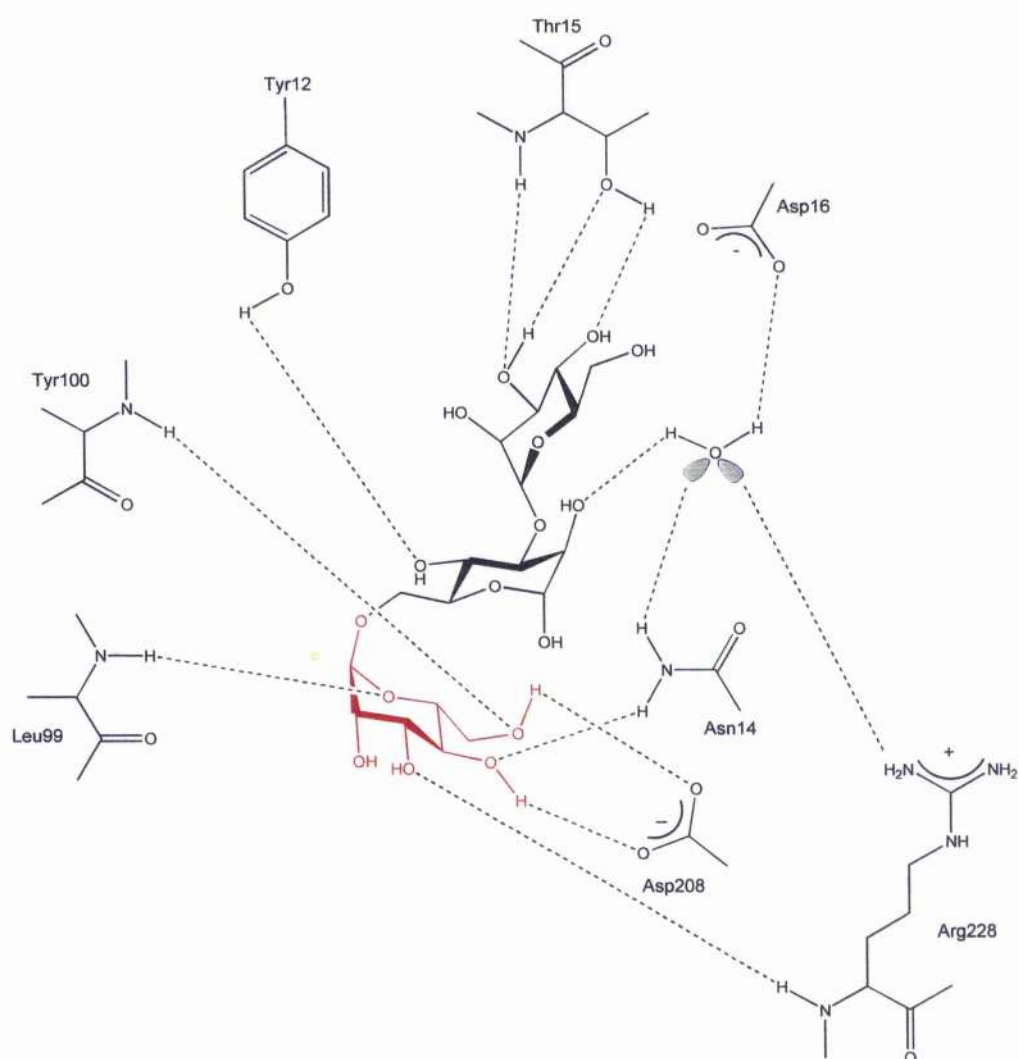
Figure 1.1. The con A tetramer. Bound methyl α -D-mannopyranoside is shown as ball and stick representation, the metal ions are shown as red spheres (Naismith *et al.*, 1994).



Con A shows an enhanced affinity for the trimannoside core of the biantennary glycan found on mammalian cell surfaces (Figure 1.5.d) (Brewer, 1993; Toone, 1994) and the structural basis for this became clear with the 2.3Å resolution structure of this trimannoside complexed with con A (Naismith and Field, 1996). This structure showed that all three residues were involved in direct interactions with the protein (Figure 1.2.), in contrast with other lectin - oligosaccharide structures such as the pea lectin trimannoside structure (Rini *et al.*, 1993) in which only the sugar at the monosaccharide site makes direct interactions with the protein. The 1-6 arm terminal mannose is recognised at the monosaccharide binding site in a very similar way to Me α Man. A water

molecule was identified which plays an important role in anchoring the reducing Man to the protein. A second report of the con A trimannoside complex found that in one of the four subunits the trimannoside adopts a different conformation, although it is unclear whether or not this is an artefact of crystallisation (Loris *et al.*, 1996).

Figure 1.2. The extended binding site of con A. The 1-6 arm terminal mannose residue is bound at the monosaccharide binding site (shown in red). All three mannose residues make direct contact with the protein (Naismith and Field, 1996).



As well as the lectins, other proteins whose role is to bind carbohydrates have come under investigation in recent years including carbohydrate processing enzymes,

periplasmic binding proteins and antibodies directed against carbohydrates. Isothermal titration microcalorimetry has enabled the direct measurement of the association constant K_a and the stoichiometry of binding interactions. Changes in free energy, specific heat capacities entropy can then be derived (Toone, 1994).

Figure 1.3 shows an example of the numbering scheme for carbohydrates, Figures 1.4 and 1.5 illustrate the carbohydrates discussed in this chapter. Table 1.1 lists the protein - carbohydrate structures discussed in this chapter, their PDB accession codes and their abbreviations used.

Figure 1.3. Numbering scheme for mannose. The dotted bonds show the N-Acetyl group in GlcNAc.

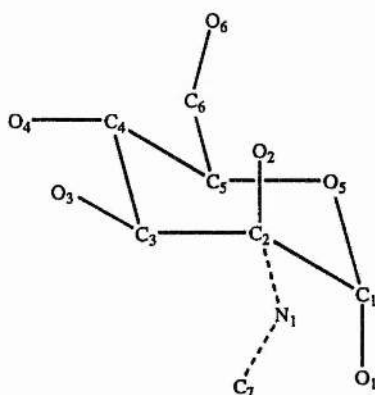
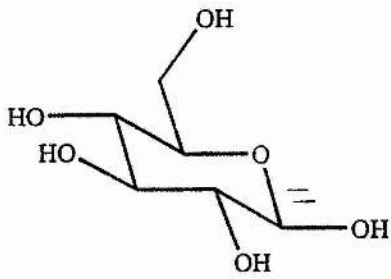
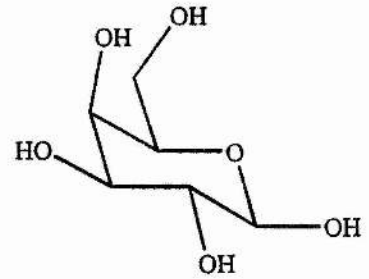


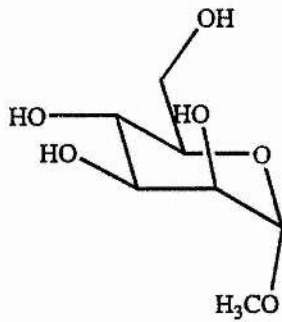
Figure 1.4. Monosaccharides; Structures and abbreviations.



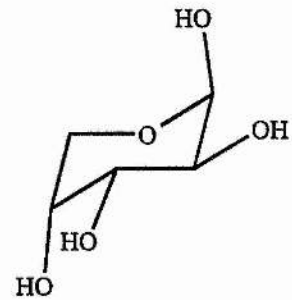
β -D-glucopyranose (Glc)



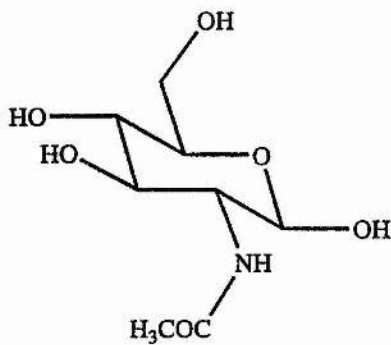
β -D-galactopyranose (Gal)



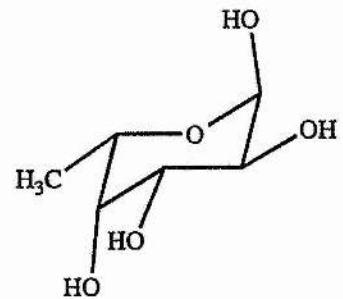
Methyl α -D-mannopyranose (Mec α Man)



β -D-arabinopyranose (Ara)

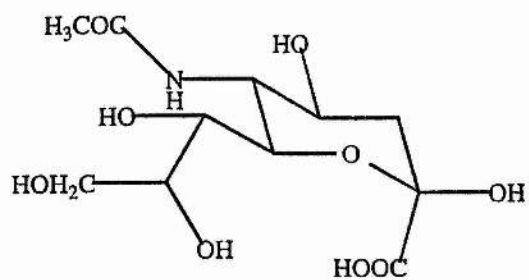


N-Acetyl glucosamine (GlcNAc)

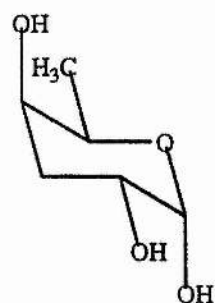


α -L-fucopyranose (Fuc)

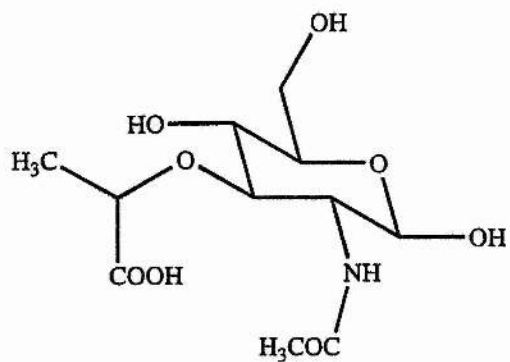
Figure 1.4. Continued.



N-Acetyl neuraminic acid (NeuNAc)



abequose (Abe)



N-Acetyl muramic acid (NAM)

Figure 1.5. Oligosaccharides; Structures and abbreviations.

a)

Gal β 1-3GalNAc

T-antigenic disaccharide

b)

Gal β 1-4Glc

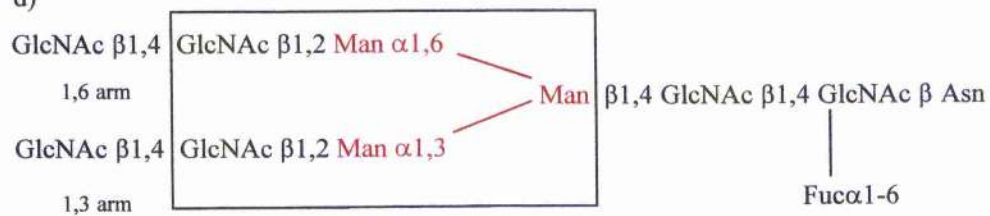
Lactose (Lac)

c)

Gal β 1-4GlcNAc

N-acetyl lactosamine (LacNAc)

d)



Biantennary Glycan. In red is the trimannoside core, boxed is the pentasaccharide core.

e)

Gal α 1-2Abe α 1-3Man

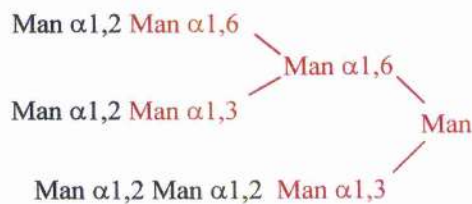
Salmonella O-antigen trisaccharide

f)

Fuc α 1-2Gal β 1-4(Fuc α 1-4)GlcNAc

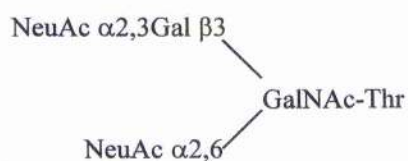
Lewis b antigen

g)



Branched high mannose oligosaccharide. In red is the branched pentamannose.

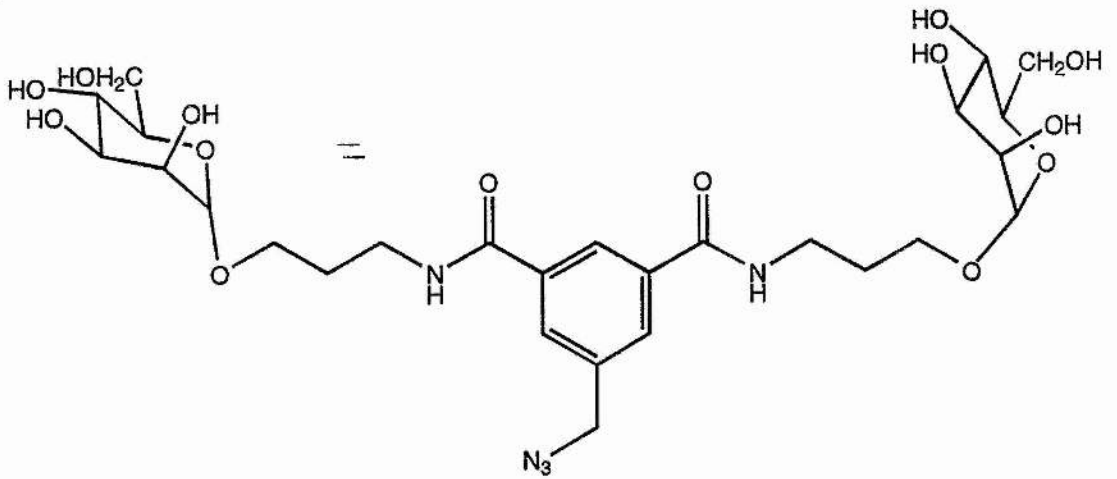
h)



Sialyl Glycopeptide

Figure 1.5. Continued.

i)



Synthetic cross-linker

j)

Glc α 1-4Glc

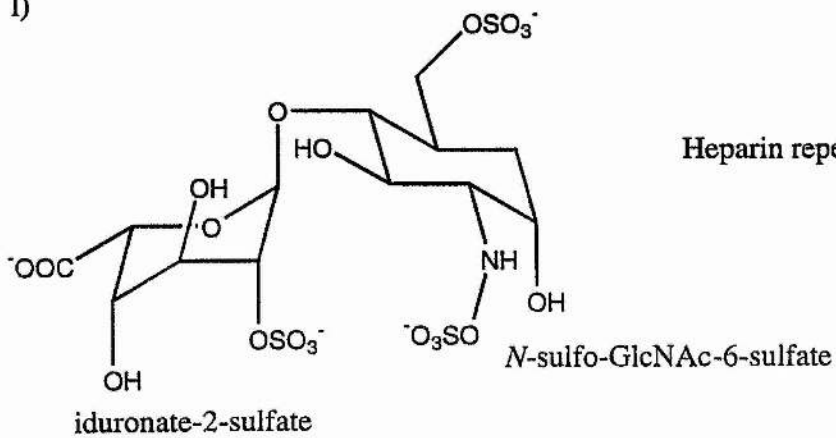
Maltose

k)

NAM β 1-4GlcNAc β 1-4NAM

Bacterial cell wall trisaccharide

l)



Heparin repeating unit

Table 1.1. Protein structures discussed in Chapter 1.

Protein	abbreviation	Ligand	PDB code
Concanavalin A	con A	Me α Man	5CAN
Concanavalin A	con A	core trimannoside	1CVN/ 1ONA
Concanavalin A (demetallised)	con A	none	1APN
<i>Erythrina corallodendron</i>	Ecor L	Gal	1AXZ
<i>Erythrina corallodendron</i>	Ecor L	Lac	1LTE/ 1AX1
<i>Erythrina corallodendron</i>	Ecor L	LacNAc	1AX2
Mannose binding protein- A	MBP-A	branched oligomannose	2MSB
Wheat germ agglutinin	WGA	sialyl lactose	1WGC
Wheat germ agglutinin	WGA	sialyl glycopeptide	2CWG
<i>Lathyrus ochrus isolectin I</i>	LOL I	Me α Man	1LOB
<i>Lathyrus ochrus isolectin I</i>	LOL I	linear trisaccharide	1LOG
<i>Lathyrus ochrus isolectin I</i>	LOL I	biantennary octasaccharide	1LOF
<i>Lathyrus ochrus isolectin II</i>	LOL II	biantennary glycopeptide	1LGC
<i>Winged bean agglutinin</i>	WBA	Gal	1WBL
<i>Galanthus nivalis</i>	GNA	pentamannose	1MSA
Soybean agglutinin	SBA	biantennary glycan	1SBA
peanut lectin	PNA	T-antigenic disaccharide	1TEP
Jacalin	Jacalin	Me α Gal	1JAC
<i>Griffonia simplicifolia</i>	GS4	Lewisb	1LED
Galectin 1	Gal-1	biantennary octasaccharide	1SLB, 1SLC, 1SLT
Galectin 2	Gal-2	Lac	1HLC
L-arabinose binding protein	ABP	Gal	5ABP
Glc/ Gal binding protein	GGBP	Gal	1GLG
		Glc	2GBP
<i>Staphylococcus aureus</i>	ET	Lac	1SE3
<i>Vibrio cholera</i>	CT	GM1 oligosaccharide	1CHP

Table 1.1. (Continued).

Protein	abbreviation	Ligand	PDB code
<i>Bordetella pertussis</i>	PT	sialyl galactose	1PTO
Polyoma virus	none	sialyl lactose	1SID
acidic fibroblast growth factor	aFGF	Heparin octasaccharide	2AXM
Se155-4 antibody fragment	Se155-4	<i>Salmonella</i> O-antigen	1MFD
Lysozyme	Lysozyme	bacterial cell wall trisaccharide	9LYS
Lysozyme	Lysozyme	Tri-N-Acetyl-Chitotriose	1LZB
Glycogen phosphorylase	GP	maltoheptaose	6GPB

This chapter discusses protein - carbohydrate interactions with reference to the common themes in energetics that have emerged over the past few years.

1.3. Energetics of protein - carbohydrate interactions.

Prior to isothermal titration microcalorimetry, the principal method for investigating protein - carbohydrate interactions was inhibition of hemagglutination or oligosaccharide precipitation (Sharon and Lis, 1989). These methods were useful in identifying specificity but they did not measure thermodynamic data. These techniques measure K_i which can be dominated by kinetic parameters. Direct measurement of thermodynamic data is now possible by isothermal titration calorimetry. Contributions to the overall free energy by each hydroxyl can be estimated by utilising saccharide analogues such as methoxy, fluorodeoxy and deoxy derivatives (Quioco, 1993; Schwarz *et al.*, 1996, Brewer *et al.*, 1997). These are discussed more fully in section 1.4.

Specific recognition of oligosaccharides is indicated by a more negative enthalpy (valence corrected where necessary) of binding than that seen for the cognate

monosaccharide. This is seen for many of the carbohydrate complexes of con A (Chervenak *et al.*, 1992; Mandal *et al.*, 1994b), the *Erythrina corallodendron* lectin (Ecor L) (Gupta *et al.*, 1996; Surolia *et al.*, 1996), mannose binding protein C (MBP-C) (Quesenberry *et al.*, 1997) and wheat germ agglutinin (WGA) (Bains *et al.*, 1992). The data shows that almost all lectin - carbohydrate interactions are entropically unfavourable in water (Chervenak and Toone, 1994). Calorimetric studies have shown that in general, a significant proportion of the enthalpy arises from solvent reorganisation (Chervenak and Toone, 1994) (the transfer of solvent molecules between the bulk of the solvent and the solvation shell of the protein and ligand). This was shown experimentally by evaluating the thermodynamics of several binding systems in H₂O and D₂O (Chervenak and Toone, 1994).

Enthalpy - entropy compensation is a general feature in many biological systems, especially those occurring in aqueous solution and involving changes in hydrogen bonding (Grunwald and Steel, 1995). For the lectins, a strong trend of enthalpy - entropy compensation can be seen in oligosaccharide binding (Dunitz, 1995). The binding of larger oligosaccharides has a more favourable enthalpy term which is offset by an increasingly unfavourable entropic term. This has been interpreted both in terms of changes in rotational degrees of freedom as glycosidic torsion angles are frozen and in terms of solvent reorganisation upon binding (Toone, 1994). Two extensive thermodynamic investigations involving con A - carbohydrate binding are available (Mandal *et al.*, 1994) (Chervenak and Toone, 1995). Although the binding constants and hence free energies are in agreement between the two studies, the magnitudes of the entropy change is uniformly smaller in the work by Toone and co-workers. Discrepancies between calorimetry experiments are due to the dependence of the results on pH, ligand concentration experimental design and temperature. Further derivation of ΔH° and ΔS°

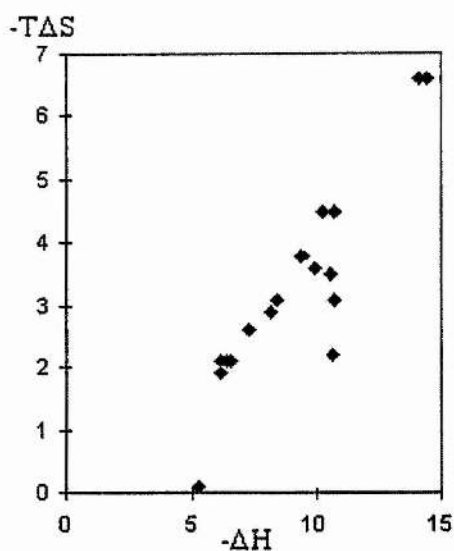
from raw data is problematic. The results of Mandal and co-workers (Mandal *et al.*, 1994b) are quoted in Table 1.2 as it encompasses all of the oligosaccharides used throughout this thesis. However, it should be noted that the results of the other investigations are very different. The con A studies can be used to illustrate entropy - enthalpy compensation. A plot of enthalpy versus entropy for the con A complexes is shown in Figure 1.6.

Table 1.2. Thermodynamics of con A - carbohydrate binding. (Mandal et al., 1994b). Enclosed in [] are shown values from (Chervenak and Toone, 1995) to illustrate the discrepancies between the studies.

	K_a ($\times 10^4 M^{-1}$)	$-\Delta H$ (kcal mol ⁻¹)	$-T\Delta S$ (kcal mol ⁻¹)
Me- α -Man	0.82 [0.76]	8.2 [6.8]	2.9 [1.5]
α 1-2 manno- biose	4.17	9.9	3.6
α 1-6 manno- biose	1.34	9.4	3.8
α 1-3 manno- biose	1.41	10.2	4.5
methyl α 1-2 manno- bioside	14.1	10.5	3.5
Trimannoside ^a	33.7	14.1	6.6
biantennary pentasaccharide ^b	140	10.6	2.2

^a Trimannoside core of the biantennary glycan shown in red in Figure 1.4.d ^bPentasaccharide core (boxed) from the biantennary glycan in Figure 1.4.d.

Figure 1.6. Enthalpy - Entropy compensation plot for con A data.



The interactions between lectins and monosaccharides are relatively weak and non specific. Affinities and specificities increase for cognate oligosaccharides. Thus multiple protein - carbohydrate interactions co-operate to give high specificity and affinity.

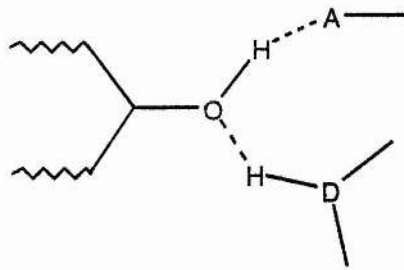
Monosaccharide binding sites are very varied amongst carbohydrate binding proteins, and they have been divided into two major groups termed groups I and II (Rini, 1995). Members of group I completely envelope their ligand and include bacterial periplasmic transport proteins and enzymes. Group II carbohydrate binding proteins include lectins, viral proteins, toxins, anti-carbohydrate antibodies and pentraxins; they bind their ligands in shallow pockets on their surfaces. Binding by group I proteins is high affinity, typically with sub-micromolar dissociation constants. This is illustrated for the periplasmic binding proteins, with K_a values in the micromolar range (Miller *et al.*, 1983). These proteins are promiscuous in their specificity for monosaccharides, for example, arabinose binding protein (ABP) recognises fucose ($K_a = 0.26\mu M^{-1}$) and galactose ($K_a = 4.35\mu M^{-1}$) as well as arabinose ($K_a = 10.20\mu M^{-1}$) (Quioco, 1988). These associations are also

characterised by a favourable entropic contribution (Sigurskjold and Bundle, 1992) which has been interpreted in terms of a dominant role of the hydrophobic effect in carbohydrate binding. The favourable entropic contribution is explained by the expulsion of water from the binding site and complete exclusion of the ligand from bulk solvent. Favourable entropies have been seen for the group II proteins, the ScFv antibody - O-antigen trisaccharide complex (Cygler *et al.*, 1991) and the snowdrop lectin (GNA). Antibody recognition falls between Group I and Group II as a single Abe sugar is almost completely enveloped. GNA recognises oligomannosides with much weaker affinity than the legume lectins (in fact the monosaccharide binding falls below the range of sensitivity for calorimetry (Shibuya *et al.*, 1988; Kaku and Goldstein, 1992)). The tightest GNA complex is with the trimannoside ($K_a = 3.0 \times 10^3 \text{ M}^{-1}$), some 100 times weaker than seen for con A. The only other oligosaccharides for which binding to GNA is detectable by calorimetry are the α 1-6 and α 1-3 mannobioses. The trimannoside, and the α 1-3 and α 1-6 mannobioses all bind GNA with a positive entropic contribution. However, as this binding is weak, GNA may be a special case.

1.4. Hydrogen bonding and van der Waals interactions.

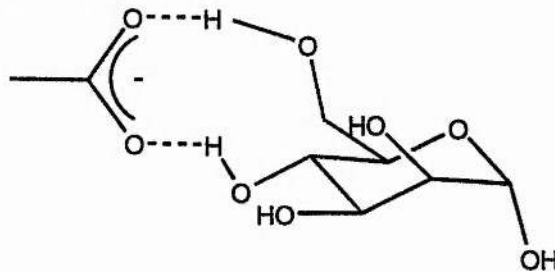
The most significant contribution to specificity is made by polar interactions. In the case of lectins, generally, one acidic side chain acts as a hydrogen bond acceptor from one or two of the carbohydrate hydroxyls. The functional groups (mainly hydroxyls) on carbohydrates are exposed and therefore heavily involved in recognition. The hydroxyl oxygen is sp^3 hybridised, resulting in a tetrahedral arrangement of two lone pairs and a proton. Thus the hydroxyl group has the capacity to act simultaneously as an acceptor and donor of hydrogen bonds (Figure 1.7.).

Figure 1.7. Cooperative hydrogen bond (A is acceptor, D is donor).



When two adjacent hydroxyl groups interact with different atoms of the same amino acid, a bidentate hydrogen bond is formed (Figure 1.8). This is common in protein - carbohydrate complexes. An example of this is in the mannose binding legume lectins, where the 4- and 6- hydroxyls form a bidentate hydrogen bond with an aspartic acid. Other bidentate ligands are glutamic acid, glutamine, asparagine and arginine.

Figure 1.8. Bidentate hydrogen bond.



A bidentate hydrogen bond is more stable than two monodentate interactions. The chelating interaction is more stable as two hydrogen bonds are formed whilst only one side chain is immobilised. The entropic penalty is therefore much less. This is the chelate effect and is well known in biology (Page and Jencks, 1971).

Main chain amide groups, the asparagine side chain amide group and sometimes a glutamine serve as hydrogen bond donors. In addition, charged side chains such as arginine often act as hydrogen bond donors, and in some cases, -OH containing amino acids such as tyrosine or serine can act as either hydrogen bond donors or acceptors. The torsional freedom of the proton and lone pairs of a hydroxyl group on the carbohydrate

presumably allows optimisation of hydrogen bonds with the protein with some entropic cost due to fixing the rotamer. The ring oxygen of the sugar is sp^3 hybridised, with the two bound carbons and the lone pairs of electrons of the ring oxygen arranged tetrahedrally allowing it to act as an acceptor of hydrogen bonds.

Chemically designed ligands have been used to assess the contribution made by individual hydrogen bonds (Withers *et al.*, 1986; Bhattacharyya and Brewer, 1988; Quioco, 1993). By using deoxy- and fluorodeoxy- saccharides, estimates for the strength of a hydrogen bond ranged from 0.5 to 3kcal mol⁻¹. The experiments are useful for determining which hydroxyl group on the ligand is involved in hydrogen bonding. However, the strength of an individual hydrogen bond remains ambiguous as the elimination of a hydroxyl group also abolishes van der Waals interactions, and there are differences in the free energies of solvation between the neutral and analogue ligands. Thus these compounds have not accurately measured hydrogen bonding energies as pointed out by Quioco and co-workers. In their studies of deoxy- and fluorodeoxy-analogues of galactose binding to ABP, they highlight the neglect of the contribution of van der Waals interactions in complex formation, and found that the compounds proved to be inadequate probes due to the binding of new water molecules and structural changes arising from the electronegative fluoro- group (Vermersch *et al.*, 1992).

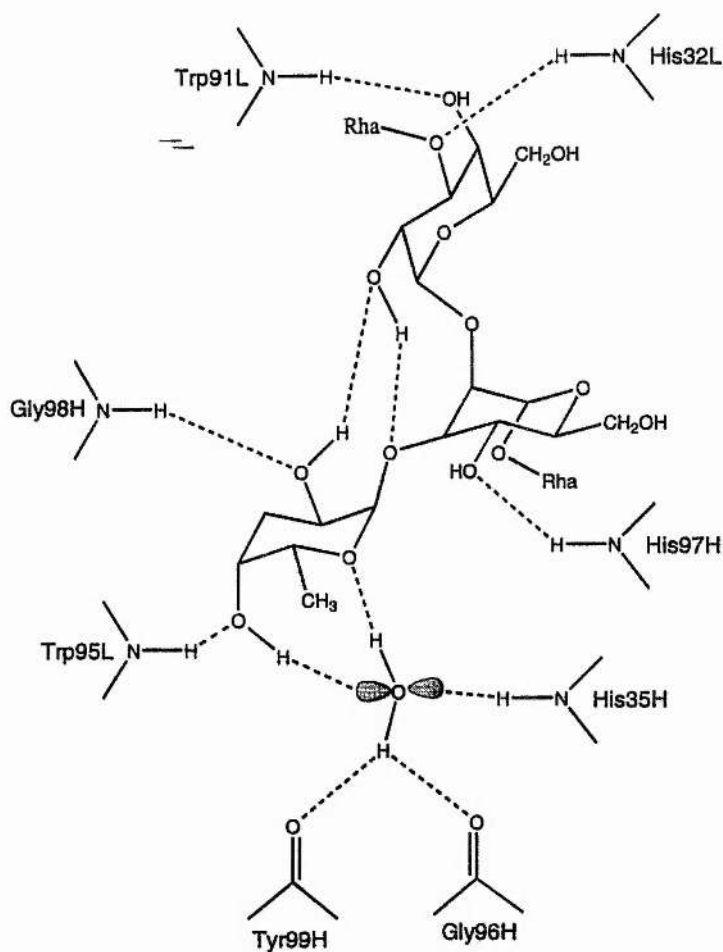
Hydrogen bonds between a carbohydrate and a protein which are mediated by a single water molecule can be as strong as a direct protein - carbohydrate hydrogen bond (Toone, 1994). A water molecule can make four hydrogen bonds. Recruitment of water to the binding site to mediate carbohydrate binding has been observed, and proposed to play an important role. In the structure of the isolectin I from *Lathyrus ochrus* (LOL I) bound with the trisaccharide Man α 1-3Man β 1-4GlcNAc (Bourne *et al.*, 1990), direct contact with the protein is only seen with the α 1-3 terminal mannose sitting in the

monosaccharide binding site. The two remaining sugars contact the protein only through long chains of water molecules. A total of 20 water molecules are involved in the binding of the trisaccharide to the protein, the majority of which are not found in a similar position in the native structure. Interestingly, in the structure of peanut lectin (PNA) complexed with the T-antigen disaccharide (Figure 1.5.a) (Ravishankar *et al.*, 1997), PNA binds the disaccharide with a 20 fold higher affinity than with lactose (Figure 1.5.b), although the structure showed that overall there are no additional direct interactions with the protein. The *N*-acetyl group of the T-antigenic disaccharide makes water mediated contact with the protein. This is the only example so far where the increase in affinity seems to be due entirely to water mediated protein-carbohydrate interactions.

Recently, the complexes of con A (Naismith and Field, 1996), the Se155-4 antibody (Figure 1.9) (Cyglar *et al.*, 1991) and enterotoxin (Sixma *et al.*, 1992), identified water molecules in the binding site of the native structure which are seen in similar positions in the carbohydrate bound structures ligating the carbohydrate. These structural water molecules can be considered part of the protein binding site.

Figure 1.9. Conserved water molecule in an antibody fragment - trisaccharide complex.

H and L refer to the heavy and light chains respectively (Cygler et al., 1991).



The hydrophobic face of the saccharide, for example in mannose - the face of the sugar with the 2- and 6- hydroxyls, commonly interacts with the partial negative charge on the π -electrons of the aromatic ring in a stacking arrangement. This stacking arrangement has been observed in all of the legume lectin structures and almost all other protein - carbohydrate structures. In the Se155-4 antibody - O-antigen trisaccharide complex (Zdanov *et al.*, 1994) the only protein side chains which interact with the ligand are aromatic. Carbohydrates have significant nonpolar substituents such as the glycerol moiety of neuraminic acid. In the lectin structures, the nonpolar groups interact with the hydrophobic residues present in the binding site.

1.5. Carbohydrate conformation

It has been demonstrated by theoretical models and NMR calculations that oligosaccharides are relatively flexible molecules, and in solution exist in a number of low energy conformations (Homans, 1995). When a carbohydrate binds to a protein, most of the rotatable bonds of the carbohydrate are frozen, creating an entropy penalty to binding.

Carbohydrate conformation has been found in some cases, to have a global effect on organisation of molecules in the crystallisation process. For example, in the cross-linked arrays of galectin-1 (Gal-1) dimers, the three different crystal forms are brought about by different low energy conformations of the cross-linking oligosaccharide, for example in the triclinic crystal form the oligosaccharides are bent (Bourne *et al.*, 1994) (discussed in section 1.9). Most carbohydrates do appear to be bound in one of the conformational energetic minima, thus no additional entropy penalty is paid.

In the O- antigen trisaccharide (Figure 1.5.e) - Se155-4 antibody complex structure (Cygler *et al.*, 1991) the Gal α 1-2Man inter-glycosidic torsion angle is shifted about 40° from the calculated minimum energy conformation, $\sim 4 \text{ kcal mol}^{-1}$ from the calculated global energy minimum. This conformation is essential for a network of inter-molecular hydrogen bonds between the galactose and the mannose and is the result of a trade off. In the structure of the Se155-4 variable domain fragment complexed with the same trisaccharide (Zdanov *et al.*, 1994), the Gal α 1-2Man inter-glycosidic linkages are different from those seen in the parent Fab - trisaccharide complex. As a result, the intermolecular hydrogen bond between the galactose and the mannose does not exist in the scFv - trisaccharide complex and the two hydroxyls are bridged by a water molecule. The calculated minimum energy structure lies between the two conformations. There is no obvious reason as to why the conformations of the saccharide should differ. There are

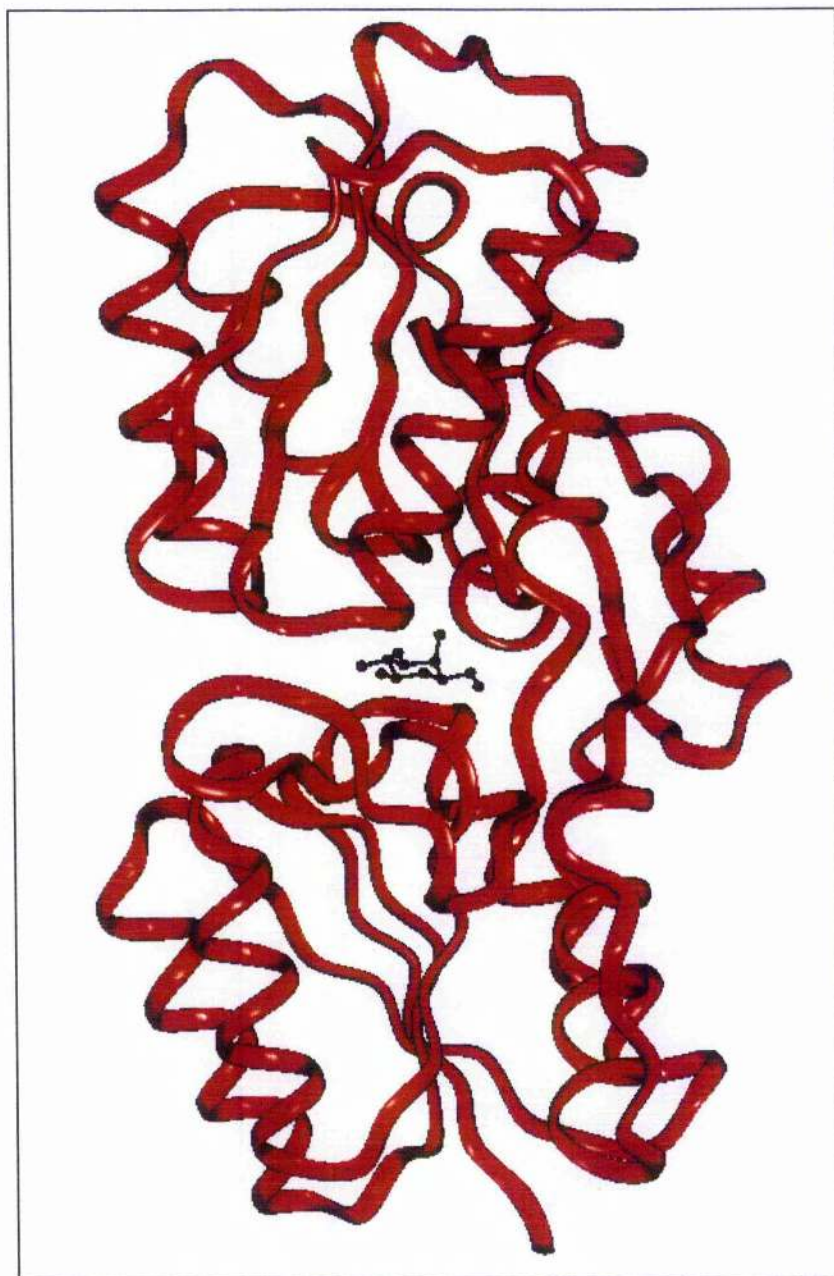
more ordered water molecules around the trisaccharide in the scFv structure and the authors suggest that reordering of water molecules near the galactose could have triggered the rotation observed.

In the case of LOL II, which requires a fucose moiety for high affinity binding of the biantennary glycan (Figure 1.5.d) different carbohydrate conformations are observed in structures of the protein bound to the glycan with and without the fucose (Bourne *et al.*, 1992; Bourne *et al.*, 1994). When the fucose moiety is present, the Man α 1-3Man inter-sugar glycosidic torsion angle adopts a low energy conformation which is $\sim 3\text{kcalmol}^{-1}$ above the global energy minima. In the complex where the fucose moiety is absent, the inter-sugar glycosidic torsion angle is remote from any minima (Imberty *et al.*, 1990). This points to the role of the fucose in attaining a more favourable conformation for binding. Distortion of carbohydrate conformation is seen in con A complexes and is discussed in Chapters 2 and 4 of this thesis.

1.6. Monosaccharide binding site geometry and specificity.

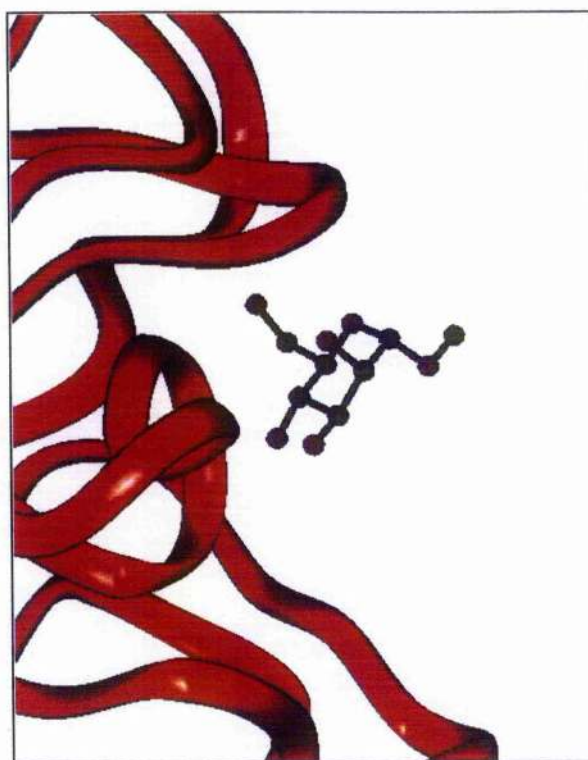
Members of group I binding proteins, such as the bacterial periplasmic binding proteins, undergo a large conformational change upon binding that buries the entire saccharide from solvent (Figure 1.10)

Figure 1.10. Group I binding. Arabinose binding protein complexed with Galactose (Quioco, 1993).



In contrast, the binding sites of group II proteins seem to be preformed undergoing no substantial main chain conformational changes upon binding, and only one or two edges of the carbohydrate ligand generally bind to the protein. A significant proportion of the sugar is exposed to solvent. This is illustrated in the lectin crystal structures, one example being con A which is shown in Figure 1.11.

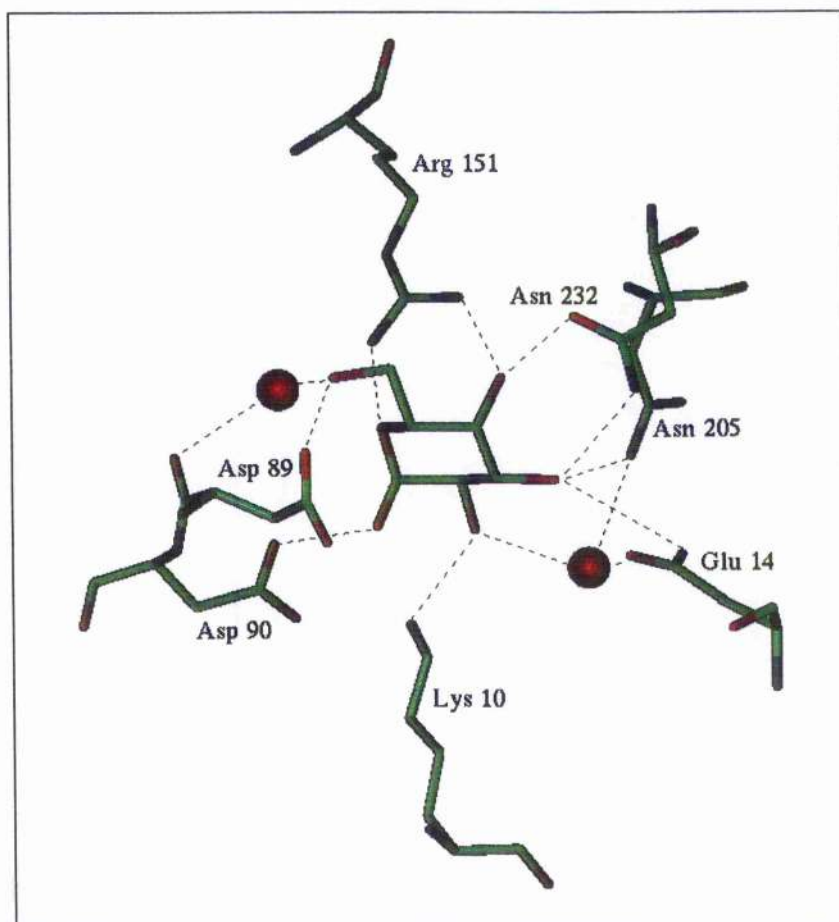
Figure 1.11. Group II binding. Concanavalin A complexed with Me α Man (Naismith et al., 1994).



In general in group I carbohydrate binding proteins such as the bacterial periplasmic binding proteins, virtually all of the hydrogen bonding potential of the saccharide is used. Members of group II such as the lectins seem not to harness the full binding potential of the saccharide. Structural studies of the bacterial periplasmic protein ABP (Quioco, 1993) bound to arabinose, galactose and its derivatives showed the pyranose rings embedded deeply between two domains of the protein (Figure 1.10) making 11 hydrogen

bonds compared with 6 made between con A and methyl α -D-mannopyranoside (Figure 1.12). Noticeably, there are two bidentate hydrogen bonds and one bidentate water mediated bond.

Figure 1.12. Extensive hydrogen bonding in the ABP - galactose complex (Quioco, 1993).



Another bacterial periplasmic protein studied is the glucose/galactose binding protein (GGBP) (Vyas *et al.*, 1994). The tight binding of this protein to its ligands is due to extensive hydrogen bonding, again with every oxygen making hydrogen bonds with the protein. The higher affinity for glucose than galactose by the GGBP is explained by more favourable hydrogen bonding and van der Waals interaction for the equatorial O4 of glucose. In the structure of the O-antigen trisaccharide fragment bound to the antibody

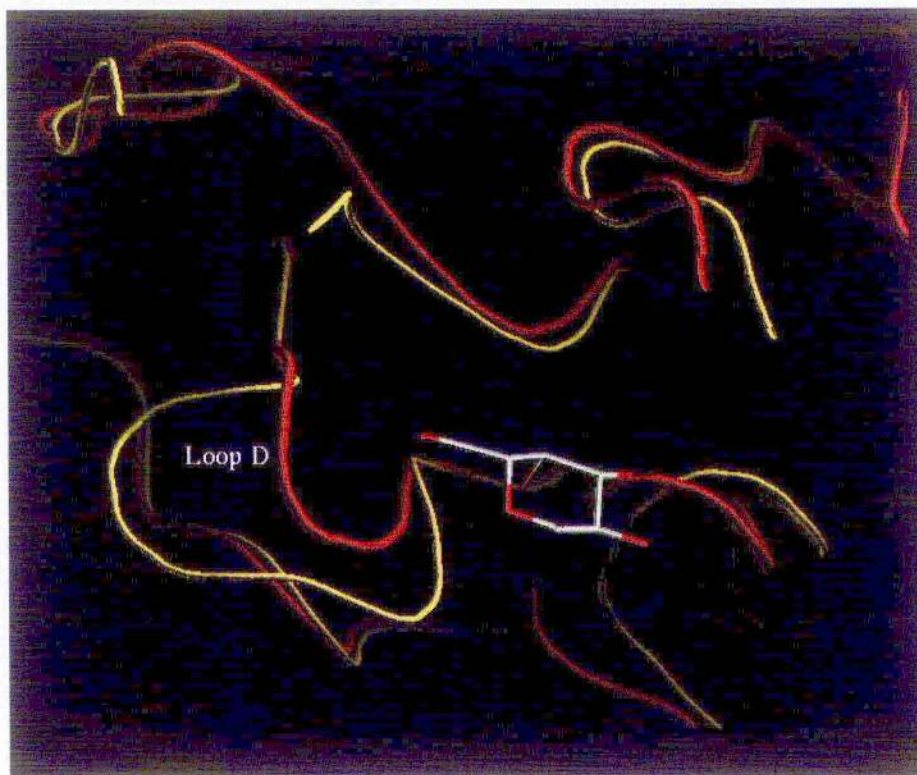
Se155-4 (Zdanov *et al.*, 1994), every hydroxyl group of the sugar is observed to make hydrogen bonds with the protein with the exception of the 6- hydroxyls on the galactose and mannose (Figure 1.9).

Legume lectins are highly conserved (Sharon and Lis, 1990) and irrespective of their specificity bind ligands through the side chains of an aspartic acid, an asparagine and an aromatic amino acid or a leucine. A single bidentate hydrogen bond is conserved in the monosaccharide binding site of the mannose and galactose specific legume lectins. The aspartic acid and asparagine, as well as interacting with the ligand also coordinate with a Ca^{2+} ion. The legume lectins also all have a non-proline *cis* peptide between the conserved asparagine and the preceding amino acid, which is typically an alanine. The *cis* peptide is necessary for the correct orientation of the asparagine. In the absence of the metals ions this peptide linkage is in the normal *trans* conformation (Bouckaert *et al.*, 1990), thus directing the asparagine away from the binding site and abolishing sugar binding. The legume lectins can be split into two main groups based on their monosaccharide specificity; those which are specific for galactose such as EcorL and those which are specific for glucose and mannose such as con A. In addition there are a small number which are specific for fucose and for GlcNAc. Most of the monosaccharide interactions are with backbone atoms suggesting that the length and position of the binding loops are more important than the precise sequence of the side chains in these loops. The shape of the monosaccharide binding site is well conserved. The three conserved residues, the aspartic acid, the asparagine and the hydrophobic residue belong to three of four binding loops. The main difference between the galactose specific and mannose/ glucose specific lectins is seen in a fourth binding loop termed the D loop (Sharma and Surolia, 1997) (Thr-97 to Glu-102 in con A, Thr-216 to Glu-224 in Ecor L (Sharon and Lis, 1990)) which makes additional interactions with the monosaccharide.

This loop is the same length in all mannose/ glucose specific lectins and is two or more residues longer in almost all galactose specific lectins (Figure 1.13). It confers specificity between the mannose/ glucose and galactose specific lectins. In the galactose specific lectins, the extended D loop interacts with the axial 4- OH of galactose, but not an equatorial 4- OH. The shorter loop in the mannose specific lectins would clash with an axial 4- OH.

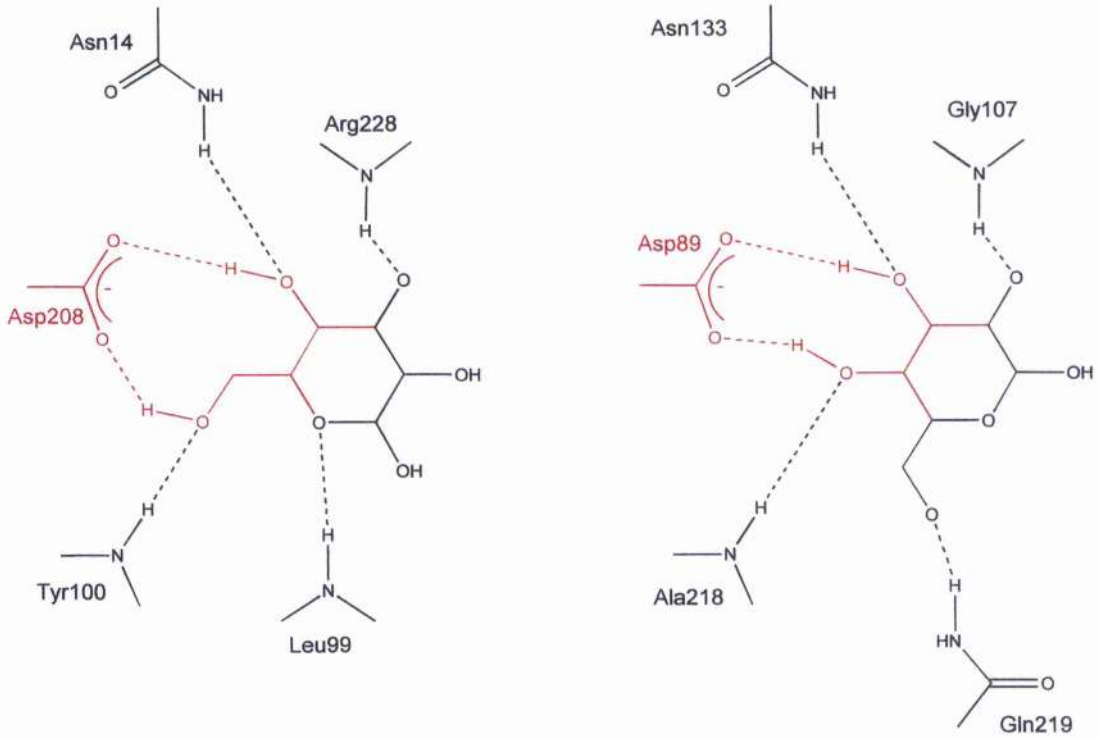
Interestingly, the recent structure of winged bean lectin (WBA) in complex with galactose (Prabu *et al.*, 1998), the D loop is longer than seen in any of the other lectins structures, it is seven residues longer than any of the mannose/ glucose binding lectins. Unlike the other galactose specific lectins, WBA shows specificity for galactose moieties with α - linkages. The structure shows that a β - linkage (such as lactose) would make a steric clash with this long D loop. This loop not only confers specificity between the glucose/ mannose and galactose specific lectins, but also recognises glycosidic linkages in the galactose specific lectins.

Figure 1.13. The D Loop of the legume lectins. Superposition of ECor L and con A binding sites (Shaanan *et al.*, 1991).



Given the similarities between the mannose/ glucose and galactose binding sites it is instructive to compare the bound sugar complexes. In con A (Derewenda *et al.*, 1989; Naismith *et al.*, 1994), lentil lectin (Loris *et al.*, 1994) and LOL I (Bourne *et al.*, 1990), the 4- and 6- hydroxyls form a bidentate hydrogen bond with an aspartic acid. In the structures of the galactose specific lectins EcorL (Shaanan *et al.*, 1991), PNA (Ravishankar *et al.*, 1997), SBA (Dessen *et al.*, 1995), the same protein residues form bidentate hydrogen bonds with the 3- and 4- hydroxyls of the saccharide. This is illustrated in Figure 1.14.

Figure 1.14. Binding of Man and Gal by the legume lectins. The con A monosaccharide binding site is shown on the left, and that of Ecor L on the right (Shaanan *et al.*, 1991; Naismith *et al.*, 1994).

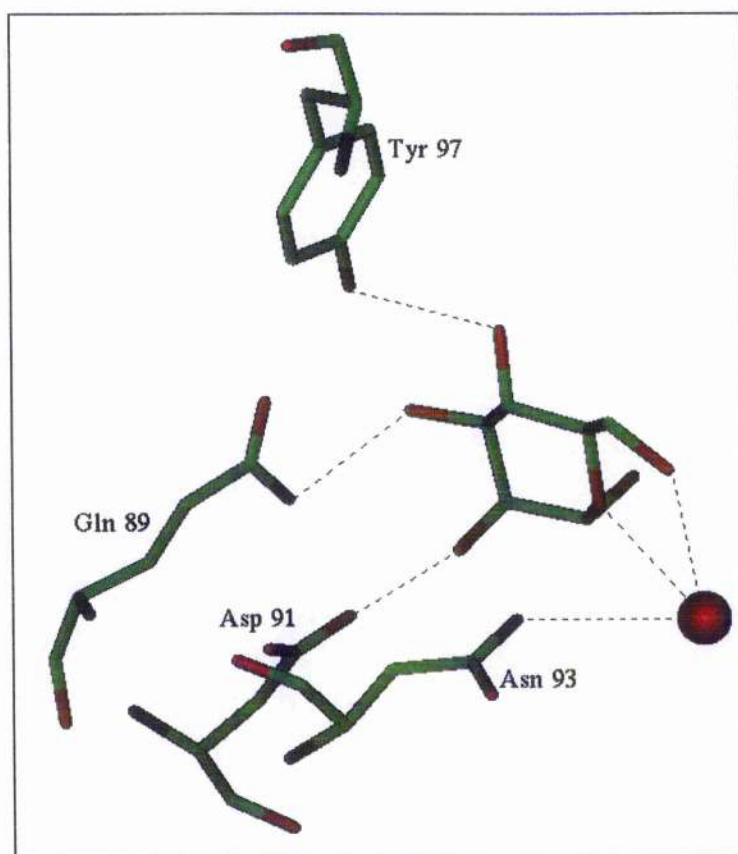


The C-types lectins belong to a different family, however similar to the legume lectins they can be grouped into those which are specific for galactose and those which are specific for mannose. Two glutamic acid residues and two asparagine residues coordinate the Ca^{2+} ion in all of the C-type lectins. However, in the galactose specific C-type lectins, one glutamic acid is replaced by a glutamine and one asparagine by an aspartic acid. In both cases, the 4- and 3- hydroxyls coordinate the bound calcium. Mutation of Glu-185 and Asn-187 in MPB-A gave a mutant specific for galactose (Drickamer, 1992).

In contrast to the legume lectins, the bulbs of plants of the amaryllis, orchid and garlic families contain lectins that bind mannose and not glucose (Kaku and Goldstein, 1992). The residues making up the binding sites are invariant and are glutamine, aspartic acid, asparagine, valine and tyrosine. Unlike the mannose/ glucose binding legume lectins

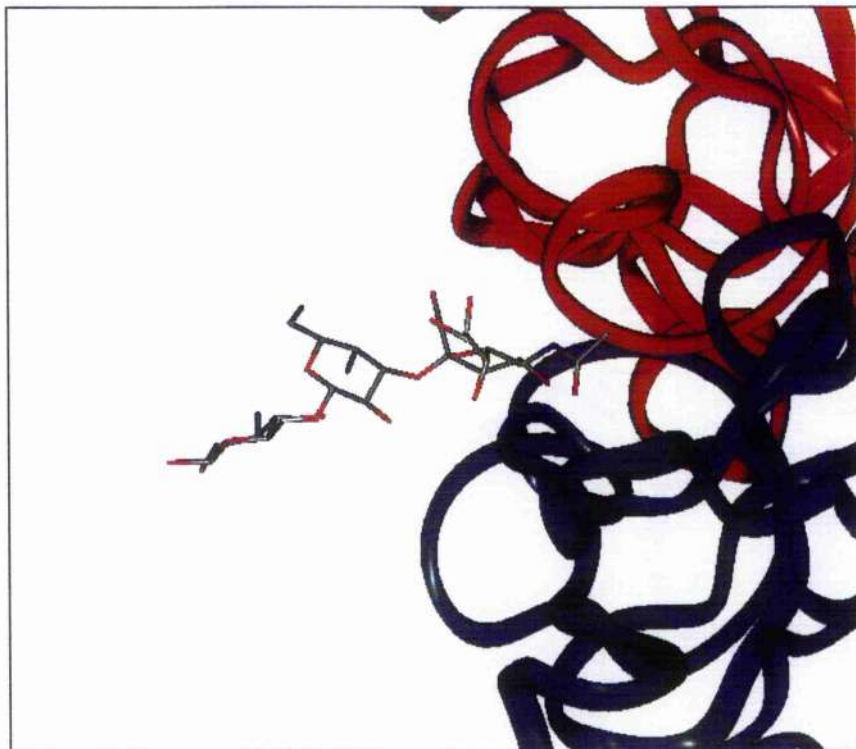
which recognise the monosaccharide principally via the 3-, 4- and 6- hydroxyls, the bulb lectins specifically recognise the axial 2- hydroxyl (Hester and Wright, 1996), thus differentiating glucose from mannose. The binding is like that seen for the legume lectins in that the saccharide makes contacts with an aspartic acid and an asparagine (Figure 1.15). However, the interactions are less extensive, for example there are no bidentate interactions, and no stacking interactions with aromatic residues which would explain why the binding is so much weaker - monosaccharide binding by GNA falls below the range of sensitivity for calorimetry and it binds the trimannoside core of the *N*-linked glycan (Figure 1.5.d) with a K_a of $3.0 \times 10^3 \text{ M}^{-1}$, some 100 times weaker than measured for con A (Chervenak and Toone, 1995).

Figure 1.15. Monosaccharide binding by GNA. A bound water molecule is shown as a red sphere.



In some proteins, residues from different subunits form the binding sites. WGA contains four binding sites per monomer, in only two of which is binding detected by calorimetry (Wright and Kellog, 1996). This is due to variability between the binding sites as the residues which form each binding site originate from different subunits (Figure 1.15). The similarities between the binding sites lie in three tyrosine residues and a serine from one subunit while variability between the binding sites is seen in interaction involving residues from the other subunit (Wright, 1990; Nagahora *et al.*, 1995).

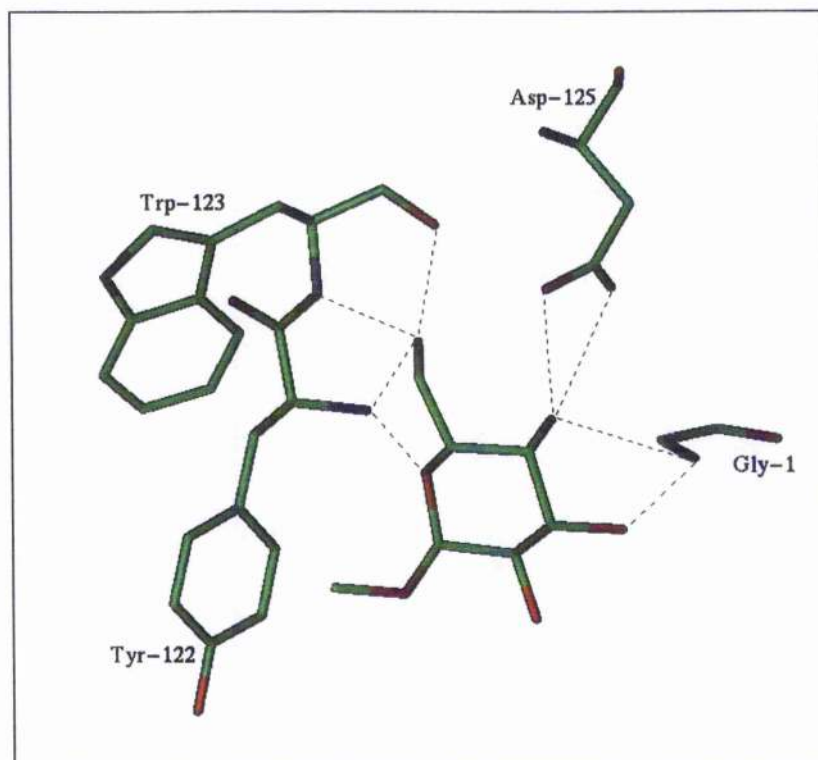
Figure 1.16. The binding site of WGA. The residues forming the binding site originate from different monomers.



A different binding site was observed in the crystal structure of the galactose specific lectin Jacalin complexed with Me α Gal (Sankaranarayanan *et al.*, 1996). This protein binds the monosaccharide about 10 to 20 times more tightly than other lectins (Gupta *et al.*, 1992) and this was reflected in the crystal structure in which the ligand was

substantially enclosed by the protein, burying 80% of the surface accessible area of the sugar compared with 75% of Me α Man buried by con A (Figure 1.17).

Figure 1.17. Monosaccharide binding site of jacalin.



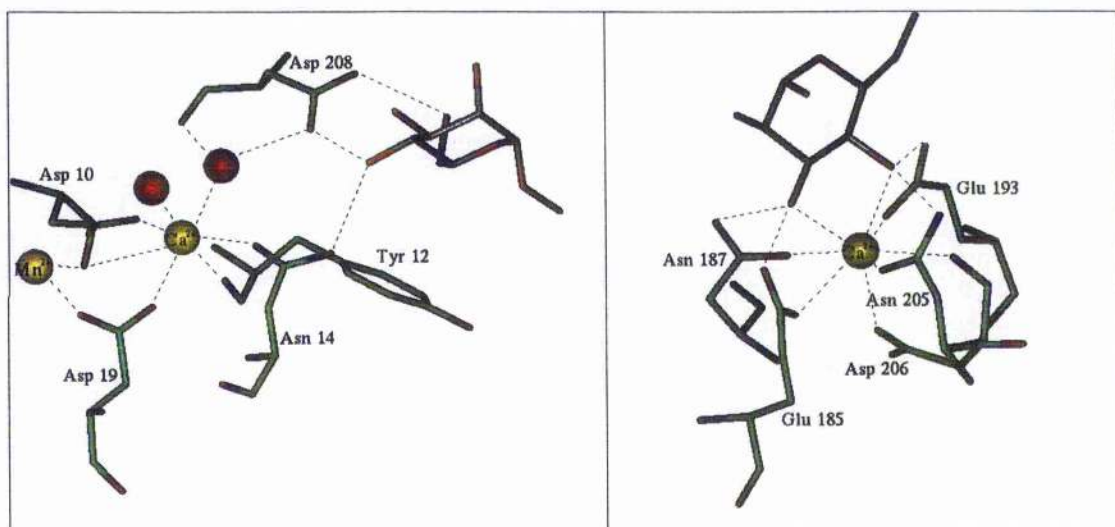
Modelling glucose and mannose into the jacalin binding site (equatorial 4- hydroxyl) revealed that although Asp-125 can still hydrogen bond the 4- hydroxyl, Gly-1 cannot, thus demonstrating that the specificity of jacalin for galactose over glucose and mannose is chiefly determined by the terminal amino group.

Many of the lectins require metals for carbohydrate binding. These can play either a direct role in contacting the carbohydrate as in the case of the C-type lectins, or an indirect role, in forming the shape of the binding site, as in the case of the legume lectins. The legume lectins require Mn^{2+} and Ca^{2+} for carbohydrate binding. The Ca^{2+} ion ligates the side chain carboxyl oxygen of the conserved asparagine while the side chain of the same residue donates a hydrogen bond to the monosaccharide. One carboxyl oxygen of the acidic amino acid is bound to a water molecule which is in turn ligated to the Ca^{2+}

ion. In this way the Ca^{2+} ion fixes side chains for optimal carbohydrate binding. The Mn^{2+} ion holds the Ca^{2+} ion in place (Figure 1.18).

In the C-type lectins, the Ca^{2+} ion ligates directly with the carbohydrate. Two lone pairs of electrons each from adjacent carbohydrate hydroxyls coordinate the Ca^{2+} ion in a bidentate fashion. These hydroxyls accept a hydrogen bond from side-chain amide groups (Figure 1.18). The only other direct sugar - metal interaction known is xylose isomerase in which the sugar ligates two Mg^{2+} ions (Allen *et al.*, 1994; Lavie *et al.*, 1994).

*Figure 1.18. Involvement of metals in carbohydrate binding. On the left is the indirect role played by con A and on the right is the direct role played by MBP-A (Naismith *et al.*, 1994; Allen *et al.*, 1994).*



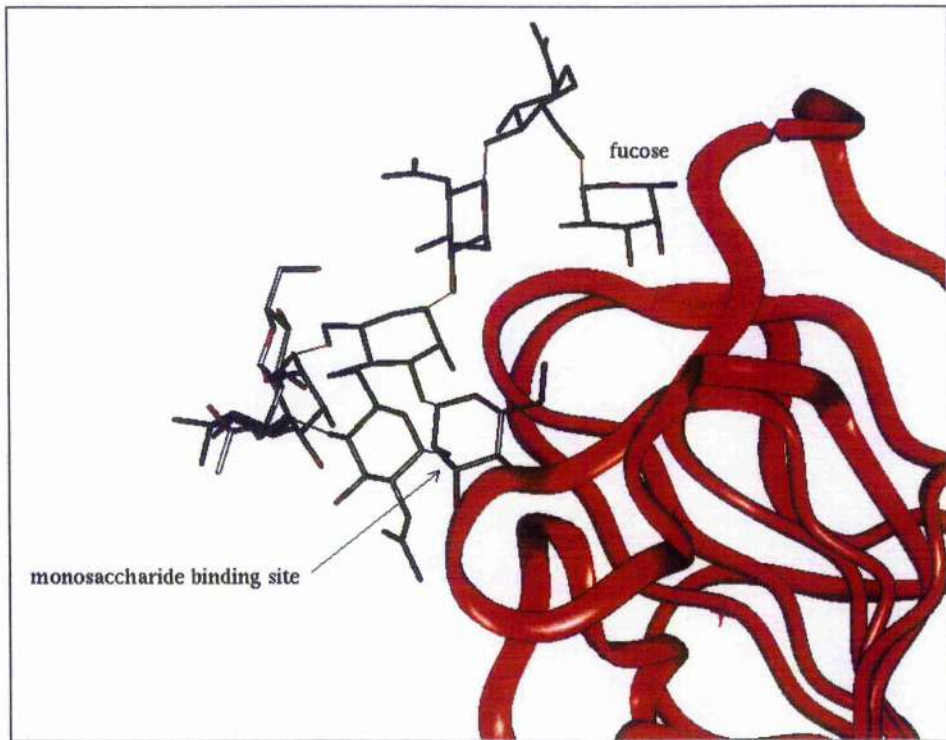
1.7. Oligosaccharide binding.

In many cases, enhancements in affinity and specificity are observed for oligosaccharides relative to monosaccharides. This has been visualised extensively in the lectins, for which there are many crystal structures of oligosaccharide bound complexes.

Some extended binding sites use water molecules extensively to mediate interactions between the oligosaccharide and the protein. Water molecules also serve to stabilise

sugar conformation. Both of these roles have been observed in oligosaccharide bound structures of LOL II (Bourne *et al.*, 1994), pea lectin (Rini *et al.*, 1993), WGA (Wright, 1992), MBP (Weis *et al.*, 1992) and the bluebell lectin (SCA) (<http://plab.ku.dk/tcbh/Lectins12/Wright/paper.htm>). These complexes, which show no significant enhancement in affinity over the monosaccharide are not specific and the relevance to understanding recognition has been questioned (Naismith and Field, 1996). In LOL II and pea lectin, an 8 to 16 fold increase in affinity is observed for a fucosylated glycan (Figure 1.5.d) (Goldstein and Portez, 1986) even though they do not bind free fucose. This increase is not seen for other mannose specific legume lectins such as con A. The structure of LOL II complexed with a fucosylated glycan (Bourne *et al.*, 1990) reveals that LOL II possess a spatially separate binding site for fucose (Figure 1.19). Superposition of con A and LOL II reveals that the oligosaccharide interaction is conserved but that fucose would clash with His-205 in the con A structure.

Figure 1.19. LOL II complexed with a fucosylated biantennary glycan (Bourne et al., 1990).

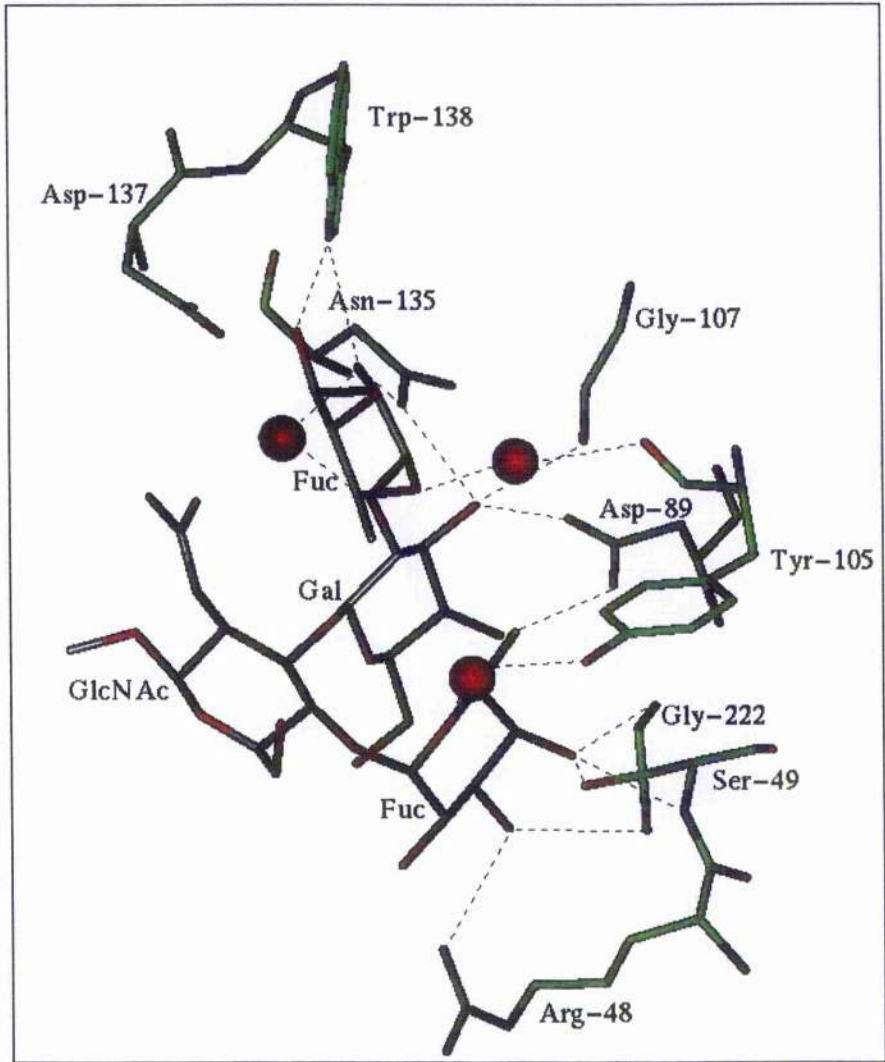


For some proteins, binding can only be detected at the oligosaccharide level, with no measurable affinity for monosaccharides. This includes the selectins which require both fucose and a charged group for high affinity binding, the *Griffonia simplicifolia* lectin (GS4) which binds the Lewis b antigen (Figure 1.5.f) and the galectins which have a low affinity for galactose, but bind lactose (Figure 1.5.b) and LacNAc (Figure 1.5.c). The galectins specificity appears to be due to the formation of four additional hydrogen bonds upon binding the disaccharide over the monosaccharide . Two of the residues (an arginine and a glutamic acid) which are involved in binding the glucose of lactose are also involved in monosaccharide binding, thus the entropy penalty for freezing the protein side chains is already paid on monosaccharide binding, but makes twice the number of hydrogen bonds upon disaccharide binding. In addition the glutamic acid

bidentate ligand upon disaccharide binding. These factors explain the 6-fold increase in binding affinity for the galectins in binding LacNAc over galactose.

The GS4 - Lewis b complex (Delbaere *et al.*, 1993) shows that the galactose residue is recognised by the monosaccharide binding site in a similar way to monosaccharide binding by EcorL. The main difference between the binding sites of GS4 and EcorL is in the D loop which is truncated in GS4. In the EcorL complex (Shaanan *et al.*, 1991), the D loop forms hydrogen bonds with the galactose. In the GS4 complex the loop forms hydrogen bonds with the α 1-4 linked fucose (Figure 1.20). The result is that the interactions between the galactose in the monosaccharide binding site and the D loop have been removed weakening the monosaccharide binding. Additional interactions are generated between the protein and other residues of the tetrasaccharide. The lectins which do recognise monosaccharides do so weakly, so it is not surprising that the removal of one or two hydrogen bonds abolishes monosaccharide binding, which is what seems to have happened with GS4. This is further evidence that lectins do not bind monosaccharides in nature.

Figure 1.20. The GS4 - Lewis b binding site. The galactose residue sits in the monosaccharide binding site and forms hydrogen bonds with the protein with its 3- and 4- hydroxyls. Due to the truncated D loop, 2 hydrogen bonds, 1 from the 3- hydroxyl and 1 from the 6- hydroxyl are lost compared with monosaccharide binding by EcorL (Delbaere *et al.*, 1993).

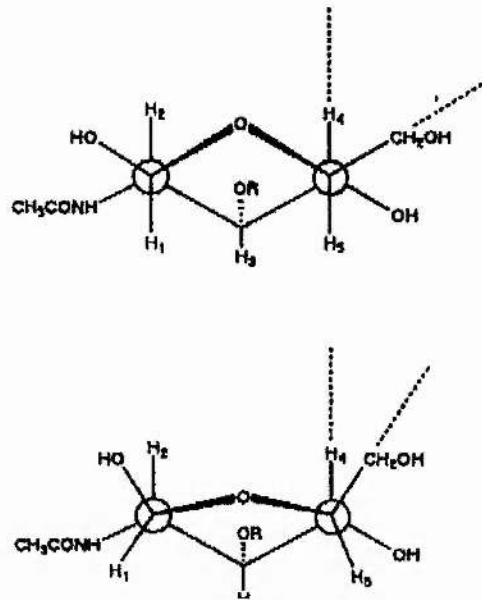


The binding sites of the mammalian galectins are made up from four adjacent β -strands which are conserved among the members of the family. The side chains of three invariant amino acids, a histidine, an asparagine and an aspartic acid are hydrogen bonded to the 4-

OH of the galactose. A conserved tryptophan stacks against the face of the saccharide and the 6-OH is also hydrogen bonded with the protein (Lobsanov *et al.*, 1993). Structures of Gal-1 with branched oligosaccharides containing LacNAc at the non-reducing end of the branches (Bourne *et al.*, 1994) leads to cross-linked arrays which are discussed in section 1.9.

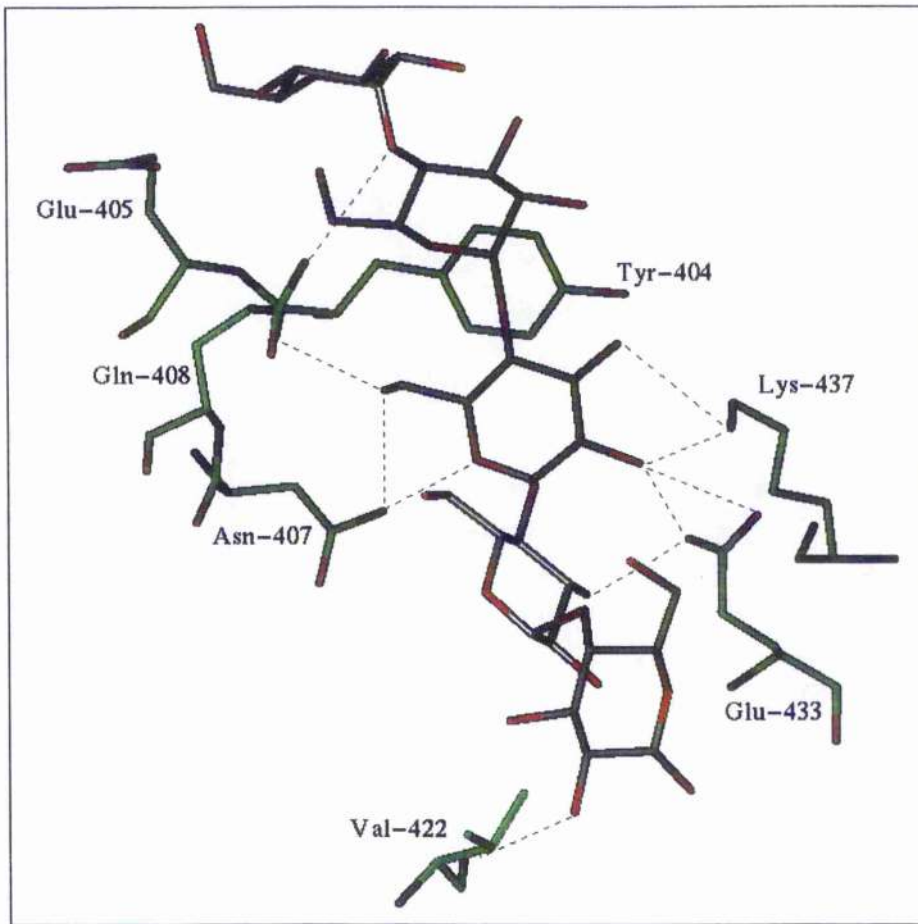
Extended binding sites have been seen in several enzyme - substrate complexes. One very well studied example is lysozyme (Cheetham *et al.*, 1992). Enzyme - carbohydrate recognition is characterised by similar features to those described for other protein - carbohydrate interactions in this chapter, stacking interactions between an aromatic ring and the carbohydrate, hydrogen bonding and participation of water molecules. In the structure of the lysozyme - tri-*N*-Acetylchitotriose ((GlcNAc)₃) complex (Cheetham *et al.*, 1992), all three residues of the carbohydrate make hydrogen bonds with the active site. One GlcNAc residue is buried in a 6Å deep pocket and makes 4 hydrogen bonds and 25 van der Waals interactions. In addition this GlcNAc forms hydrogen bonds with 2 conserved water molecules. The remaining 2 GlcNAcs of the trisaccharide lie on the surface of the protein, each making a hydrogen bond with the protein via their 6-hydroxyls, and a total of 18 van der Waals interactions. In the structure of lysozyme complexed with the bacterial cell wall trisaccharide (Strynadka and James, 1991) (Figure 1.5.k) again all three sugar residues form hydrogen bonds with the protein. Interestingly, in this structure, one of the *N*-acetyl muramic acid (NAM) residues adopts a conformation which is distorted from the expected ⁴C₁ chair conformation. The C-2, C-1, O-5, C-5 and C-4 atoms are approximately coplanar. The result is that the CH₂OH group on C5 is pseudo axial (Figure 1.21). A hydrogen bond between the 6-hydroxyl and Val-109 would not be favourably oriented in the ⁴C₁ chair conformation.

Figure 1.21. Distortion of a sugar ring by lysozyme; schematic representation of the NAM sugar in the lysozyme - bacterial cell wall trisaccharide complex. On top is the expected 4C_1 chair conformation. Below is the more planar sofa conformation seen in the complex (Cheetham *et al.*, 1992).



A maltoheptose bound structure of the inactive form of glycogen phosphorylase (T state phosphorylase b) (Johnson *et al.*, 1990) showed the oligosaccharide sitting in the glycogen binding site of the enzyme, remote from the active site. All five saccharide residues located in electron density are seen making direct van der Waals contact with the protein, four of them form direct hydrogen bonds. (Figure 1.22). One sugar makes extensive hydrogen bonds with the protein, including one bidentate hydrogen bond. Two of the sugars form a groove in which a tyrosine ring fits. Hydrophilic interactions play an important role with five non-polar side chains involved in the binding site.

Figure 1.22. The extended oligosaccharide binding site of glycogen phosphorylase (Johnson *et al.*, 1990).



1.8. Charged carbohydrate groups.

Sialic binding proteins include cholera toxin (Merritt *et al.*, 1994), influenza hemagglutinin (Weis *et al.*, 1988; Sauter *et al.*, 1992; Watowich *et al.*, 1994), pertussis toxin (Stein *et al.*, 1994) and WGA (Wright, 1990). With the exception of polyoma virus (Stehle *et al.*, 1994) the carboxylate moiety of NeuNAc interacts with main chain amide groups, polar side chains and ordered water molecules rather than forming salt bridges with positively charged amino acids as one might expect. However, formation of salt bridges is a common feature of neuraminidases (Colman *et al.*, 1983; Burmeister *et al.*,

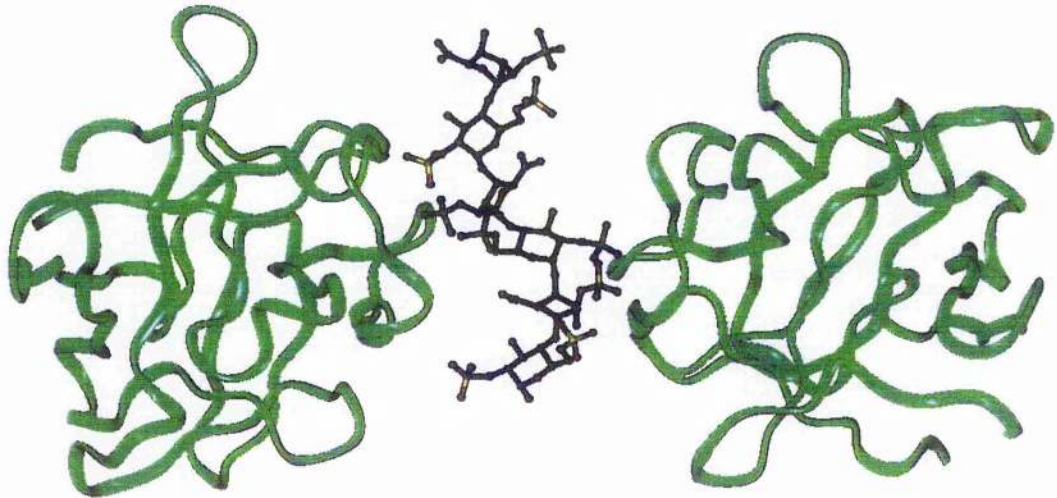
1992; Crennell *et al.*, 1994). And in the structure of polyoma virus complexed with sialyllactose, the NeuNAc carboxylate interacts with an asparagine side chain (Stehle *et al.*, 1994). Based on this structure and modelling, it was suggested that the inability of one particular polyoma strain to bind branched ligands containing both α 2-3 and α 2-6 linked NeuNAc is due to charge repulsion by a glutamic acid at a putative second site. Charge repulsion has been seen in WGA structures where GlcNAc and NeuNAc both bind in the primary binding site but only GlcNAc binds in the secondary binding site. Acidic residues near the secondary site are thought to cause this and is consistent with the observation that the neutral methyl ester of NeuNAc will bind at the secondary site (Wright, 1984). In the case of the selectins, sialic acid is required for binding.

Charged carbohydrate moieties have been seen in legume lectin - carbohydrate complexes; a structure of LOL I complexed with muramic acid (Figure 1.4) showed that the acid moiety is solvent exposed and interacts with the protein only through water molecules (Bourne *et al.*, 1994).

The structures of acidic fibroblast growth factor (aFGF) highlight the importance of charged carbohydrates for activity (Zhu *et al.*, 1993; DiGabriele *et al.*, 1998). The first structure, that of aFGF complexed with the anti-cancer drug sucrose octasulfate (SOS) demonstrated how SOS stabilises the aFGF dimer by neutralising several positively charged residues that would otherwise destabilise the dimer by electrostatic repulsion. Two lysine residues interact with at least two sulfates from each sugar ring, and in addition two arginine side chains interact with sulfate groups. Other heparin like molecules (Figure 1.5.1) - linear anionic polysaccharide chains, typically heterogenously sulfated on alternating L-iduronic and glucosamino sugars have been seen to function in a similar manner; the recent structure of aFGF bound with a heparin decasaccharide was the first of an active complex. The resulting dimer (Figure 1.23) has no protein - protein

interface, instead the monomers are bridged by heparin. The monomers bind sulfate groups of 5 or 6 monosaccharide units on opposite sides of the heparin helix axis. Most of the contacts are ionic contacts between basic residues such as lysine and the sulfate and carboxylate groups of heparin.

Figure 1.23. The heparin linked dimer of aFGF (DiGabriele *et al.*, 1998).

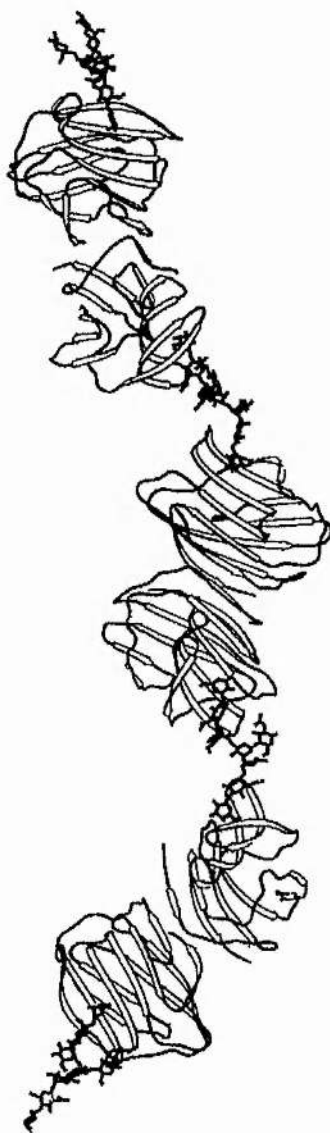


1.9. Ligand Multivalency.

The binding of monosaccharides by some proteins is too weak to be of therapeutic value, and oligosaccharides are too polar for satisfactory uptake. Apparently high affinity for monosaccharide binding can be achieved by multivalent ligands. A synthetic polymer containing multiple mannose residues was recognised by con A with a 10^5 - fold higher K_i than Me α Man (Mortell *et al.*, 1996). Indeed, in nature, many carbohydrate binding proteins appear to function as oligomers. Multivalent oligosaccharides are capable of binding at spatially separate sites on the protein. Linear arrays of protein - carbohydrate molecules can be formed with divalent oligosaccharides and dimeric lectins. This has been seen in the oligosaccharide bound structures of the bovine heart muscle lectin Gal-1

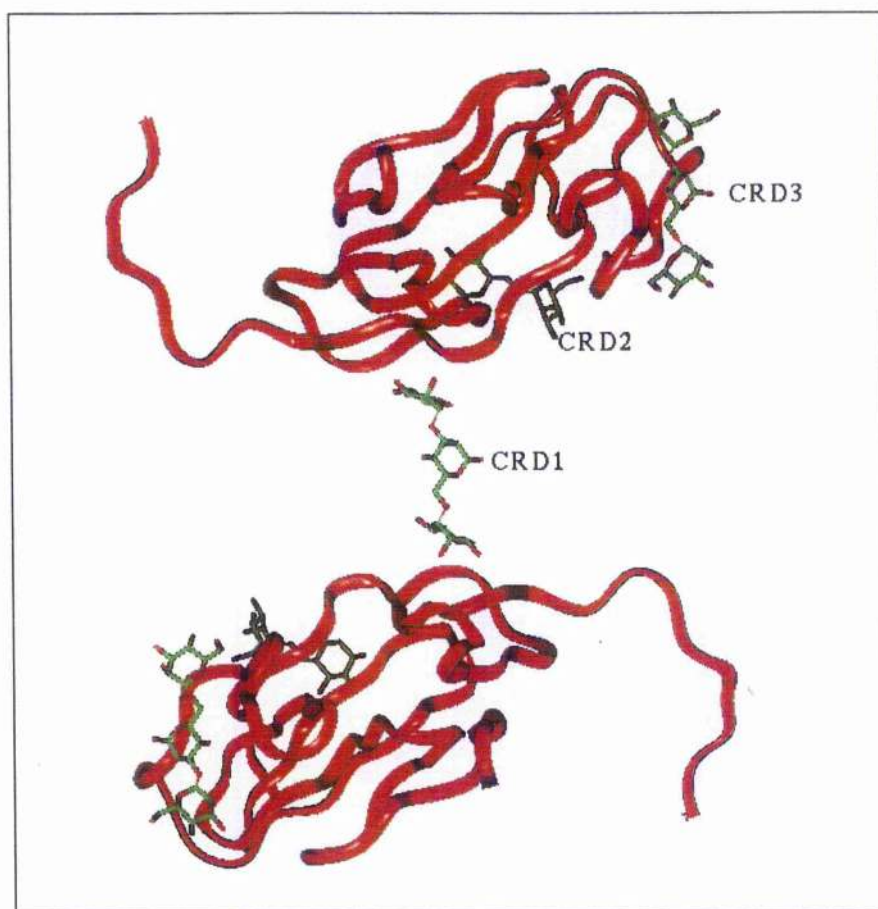
(Bourne *et al.*, 1994). For each of the oligosaccharide complexes (an octasaccharide, an asparaginyl-nonasaccharide, both branched and containing LacNAc at the non-reducing end of the branches) three crystal forms are obtained - hexagonal, trigonal and triclinic. The structures are infinite chains of lectin dimers cross-linked through the LacNAc moieties (Figure 1.24).

Figure 1.24. Cross-linked bovine heart Gal-1. The dimers are cross-linked by a biantennary octasaccharide, the arrangement shown is found in the hexagonal crystal form (Bourne et al., 1994).



The monosaccharide binding site of the bulb lectins is made up from 3-fold internal repeats within the monomer; GNA is a tetrameric dodecavalent molecule. Each monomer contains 3 carbohydrate recognition domains; CRD1, CRD2 and CRD 3. In the structure of GNA complexed with a branched pentamannose (Figure 1.5.g) (Wright and Hester, 1996) two modes of binding are observed. At CRD1, the highest affinity monosaccharide binding site, the pentamannose cross-links GNA dimers. At CRD3, the outer arm interacts with a single subunit with three of the mannose residues contacting the protein. These binding modes are illustrated in Figure 1.24. None of the legume lectins have yet been complexed with this sugar, but one would expect to see a similar cross linking effect.

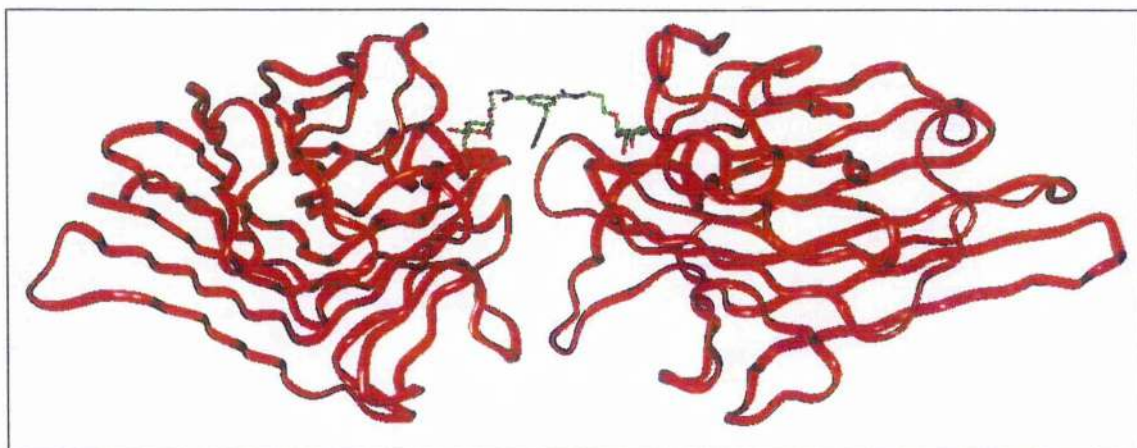
Figure 1.25. The two modes of oligomannose binding by GNA (Wright and Hester, 1996).



This kind of multivalency is also seen for WGA which has multiple binding sites due to the 4- fold internal repeats. The binding sites in this case are located at the interface between the subunits which form the molecular dimer. There are eight sites per dimer, four of which are unique. In the complex of WGA with a sialylglycopeptide (Wright, 1992) (Figure 1.5.h) the carbohydrate binds to the protein so that the α 2-6 linked sugar occupies the binding site in one domain of one dimer and the α 2-3 linked sugar binds a different domain of a crystallographically related dimer.

The first structure of a lectin with a synthetic polyvalent ligand has recently been reported (Moothoo and Naismith, 1998b; Dimick *et al.*, 1998). The ligand is a bidentate ligand (Figure 1.5.i) which cross-links con A monomers (Figure 1.25) leading to an infinite cross-linked array of con A molecules.

Figure 1.26. Cross linked con A monomers.



Monosaccharide binding affinities by the C-type lectins MBP-A and MBP-C are comparable with con A (Quesenberry *et al.*, 1997). However, MBP-C binds branched oligomannosides more strongly than MBP-A. Binding studies using mono- and oligosaccharides and synthetic cluster glycosides, led to the conclusion that MBP-C has

two binding sites per subunit, one for Man, the other for both Man and GlcNAc, and that MBP-A has only one binding site per subunit.

1.10. Concluding remarks. _

The widespread occurrence of protein - carbohydrate interactions in nature has led to them becoming investigated by a number of techniques. In order to be able to rationally exploit protein - carbohydrate interactions for therapeutic purposes, they must become well understood at the molecular level. This points to a need for accurate parameterisation of the factors which control the interactions. This Chapter has highlighted the important aspects of protein - carbohydrate interactions which are known through structural and thermodynamic studies. Con A is a carbohydrate binding protein which has been well characterised. Its carbohydrate binding properties have been described and the structure is known at an atomic level. In addition, the protein is readily available and crystallises with ease. Therefore, con A provides a useful model in this thesis for exploring protein - carbohydrate interactions. This thesis attempts to delineate and quantify the components of protein - carbohydrate recognition.

Chapter 2

**The 1.75Å structure of concanavalin A complexed with
α1-2 manno­biose.**

2.1. Summary

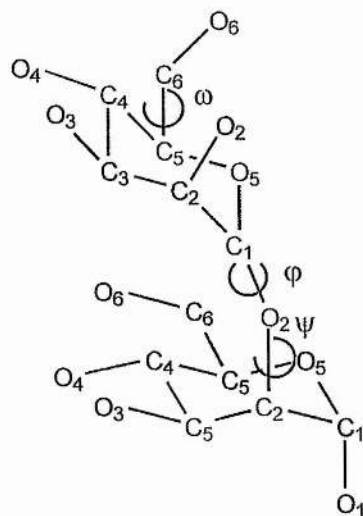
The three-dimensional structure of the complex between α 1-2 manno-biose and concanavalin A has been determined to 1.75Å resolution. Diffraction data were first recorded from a single crystal to a resolution of 2.0Å. The structure was determined by molecular replacement and refined to 2.0Å. Subsequently, further data were collected from a second crystal which diffracted to a resolution of 1.75Å. The final model, refined to 1.75Å has good geometry, an R factor of 18.5% for 27201 reflections and a free R factor of 21.8%. The asymmetric unit contains one monomer which binds one disaccharide molecule via hydrogen bonds and van der Waals interactions. The non reducing terminal mannose is recognised at the so called monosaccharide binding site. The reducing terminal mannose is bound by residues Tyr-12, Asp-16 and Arg-228 and a single water molecule which interacts with the protein. The interactions for this sugar ring are less extensive than observed for the reducing sugar in the trimannose bound and pentasaccharide bound complexes, but are more than predicted by molecular modelling studies. However, buried surface area analysis shows that the reducing sugar of α 1-2 manno-biose buries more apolar surface accessible area than the reducing sugar of α 1-6 manno-biose. This observation and the relative rigidity of α 1-2 manno-biose may explain its high affinity for con A.

This complex crystallises in the same space group and has similar cell dimensions to the native con A structure. This allows a detailed description of the changes in protein structure, bound solvent and crystal packing which occurs upon sugar binding.

2.2. Introduction.

Several studies have addressed con A binding to α 1-6, α 1-3 and α 1-2 mannobiosides. Studies by Goldstein and co-workers (Goldstein *et al.*, 1965; So and Goldstein, 1968) grouped linear oligosaccharides into two classes based upon their affinities for con A. The first group, including α 1-3, α 1-4 and α 1-6 linked oligosaccharides with non-reducing terminal glucose or mannose residues, display weak affinity for the protein, similar to monosaccharide binding. The second group includes the α 1-2 linked oligomannosides and show 5 to 20 fold higher affinities for con A. The authors speculated that the higher affinity displayed by the α 1-2 linked oligomannosides was due to the presence of an extended binding site present which could accommodate them. Studies investigating con A - dextran precipitation by oligosaccharides with varying number of mannose or glucose residues were carried out (Goldstein, 1975). The enhanced affinities of the α 1-2 linked oligomannosides with respect to methyl α -D-mannopyranoside led to the postulation that con A possesses an extended binding site consisting of sub-sites, each of which is complementary to α 1-2 mannobiose (Figure 2.1).

Figure 2.1. α 1-2 mannobiose. The inter-sugar glycosidic torsion angles are defined as: $\varphi = O5-C1-OX-CX$, $\psi = C1-OX-CX-C(X+1)$, $\omega = C4-C5-C6-O6$ and are shown in red.



Further biophysical techniques have been employed to address the binding behaviour of con A, such as the solvent proton nuclear magnetic relaxation dispersion technique, near ultraviolet circular dichroism and rapid flow kinetic analysis using fluorescent labelled ligands (Goldstein, 1975; Reddy and Rao, 1992; Mandal *et al.*, 1994; Mandal *et al.*, 1994) (Brewer and Brown, 1979). Interestingly, another oligosaccharide, melezitose (Glc β 1-3Fru α 2-1Glc) has a similar enhanced affinity for con A to α 1-2 mannobiose. Although melezitose is structurally different to α 1-2 mannobiose, both sugars have more than one residue with free hydroxyls at the 3-, 4- and 6- positions, i.e. they both have more than one residue able to bind in the monosaccharide binding site. These studies led to the postulation that the higher affinities displayed by con A for the α 1-2mannobiosides were not due to the extended binding site of con A as previously postulated, rather the forward rate constant for complex formation being increased through an enhanced probability of binding and recapture of the ligand.

Table 2.1. Thermodynamic data for binding of mannobioses to con A

	K_a ($\times 10^4 M^{-1}$)	$-\Delta H$ (kcal mol $^{-1}$)	$-T\Delta S$ (kcal mol $^{-1}$)
Me- α -Man	0.82	8.2	2.9
α 1-2 mannobiose	4.17	9.9	3.6
α 1-6 mannobiose	1.34	9.4	3.8
α 1-3 mannobiose	1.41	10.2	4.5

Table 2.1 shows thermodynamic data for con A binding to α 1-3, α 1-6 and α 1-2 mannobiose (Mandal *et al.*, 1994). Another study also investigated thermodynamics of con A binding (Chervenak and Toone, 1995). Although the relative magnitudes of the binding constants and enthalpies of binding are in agreement between the two studies, the enthalpies of the work by Chervenak and co-workers are uniformly smaller. Discrepancies between calorimetry experiments are due to the dependence of the results on pH, ligand concentration and temperature. The thermodynamic results of Brewer and co-workers (Mandal *et al.*, 1994) are quoted here as it encompasses all of the oligosaccharide used throughout this thesis. The interactions are characterised by an unfavourable entropic term, as are most lectin - carbohydrate interactions. α 1-2 Mannobiose stands apart from the other two mannobiosides as having an enhanced association constant.

Prediction of the modes of binding of α 1-2 mannobiose to con A using energy minimisation methods (Reddy and Rao, 1992) docked the disaccharide to the protein with the non-reducing terminal sugar in the previously designated primary binding site with a more favourable energy than the reducing terminal sugar. This is in agreement with the mode of binding observed in the crystal structure described in this chapter. However, the

structure presented here allows observation of detail not seen in modelling studies such as reorganisation of bound solvent, precise position of the sugar and changes in the protein structure - the native structures was used for the ligand docking in the modelling study. This highlights the need for increased parameterisation of protein - carbohydrate interactions to enable meaningful modelling studies.

This chapter describes the 1.75Å resolution structure of con A complexed with α 1-2 mannobiose. The basis for high affinity of con A for α 1-2mannobiose over its α 1-3 and α 1-6 counterparts is discussed and comparisons made with the high resolution native structure.

2.3. Experimental.

Crystallisation and Data Collection.

Co-crystals of the con A - α 1-2 manno- β -D-glucopyranosyl complex were obtained after an initial screen consisting of 10, 15, 20 and 25% Polyethylene Glycol (PEG) 6K, pH 4.0, 5.0, 6.0 and 7.0. Optimisation of pH and PEG concentration then followed (Moothoo *et al.*, 1998b). Large crystals were obtained from 10 μ l of a solution containing 18mM α 1-2 manno- β -D-glucopyranosyl (Dextra Laboratories, Reading, UK), 0.6mM con A (Sigma, Poole, UK), 1mM CaCl₂, 1mM MnCl₂, 20mM TRIS pH 7.0 and 0.1M NaCl, mixed with 10 μ l of a reservoir containing 10% PEG 6K, 0.1M citric acid pH 5.0 in a sitting drop tray (Charles Supper Co.) and left to equilibrate against 1ml of the reservoir. Some crystals grew within one week but were prone to cracking. Slower growing crystals which took two weeks to reach optimum size showed no signs of cracking and these were used for data collection.

Data from the largest crystal (0.8 x 0.8 x 0.6 mm³) mounted in a glass capillary were collected to 2.0Å as 200 1.0° 25 minute non - overlapping oscillations using a Nonius DIP2000 dual image plate at 20.5°C. The data had a mosaic spread of 0.7°. Data were indexed and merged using DENZO and SCALEPACK (Otwinowski, 1993). Data were indexed in lattice groups I centred tetragonal, I centred orthorhombic and F centred orthorhombic. A stable unit cell matrix could not be obtained in I centred tetragonal or F centred orthorhombic lattices. Data were integrated and merged in space groups I222 and I2₁2₁2₁ to give an R-merge of 7.2%, with unit cell parameters a = 91.75Å, b = 86.68Å, c = 66.64Å. 0.8% Of observations were rejected. A monomer in the asymmetric unit gives a Matthew's number of 2.6 DaÅ⁻³, corresponding to a solvent content of 52%. Data quality is summarised in Table 2.2.

Table 2.2. Quality of data of the 2.0Å α 1-2 mannobiose - con A data.

Resolution	No. Reflections	% Complete	R-Merge/ %	Redundancy
25.00-4.30	1897	97.2	5.3	7.5
4.30-3.42	1847	100.0	6.2	7.8
3.42-2.99	1839	99.9	7.1	8.0
2.99-2.71	1837	99.9	8.0	8.3
2.71-2.52	1815	100.0	9.2	8.1
2.52-2.37	1795	99.9	10.0	8.1
2.37-2.25	1822	100.0	11.0	8.1
2.25-2.15	1796	100.0	12.1	8.0
2.15-2.07	1815	100.0	13.2	7.9
2.07-2.00	1670	93.3	13.5	5.6
25.00-2.00	18133	99.0	7.2	7.8

Several months later, inspection of the crystallisation plates revealed several large crystals. Data were collected from the largest crystal (1.2 x 1.0 x 0.8 mm³) which diffracted to a resolution of 1.75Å. Data were recorded as 141 non-overlapping 25 minute 0.75° oscillations. A second data set was collected with the crystal to detector distance set to 150mm. This enabled low resolution data to 3.0Å which was obscured by the backstop in the high resolution data collection to be collected. The data had a mosaic spread of 0.6°. The observations from both passes were merged together. Data quality is shown in Table 2.3.

Table 2.3. Quality of data of the 1.75Å α 1-2 mannobiose - con A data.

Resolution	No. Reflections	% Complete	R-Merge/ %	Redundancy
25.00-3.77	2608	91.1	4.8	6.3
3.77-2.99	2761	99.9	6.0	8.0
2.99-2.61	2648	97.1	6.6	4.0
2.61-2.37	2605	96.9	7.2	3.1
2.37-2.20	2650	97.2	7.9	3.0
2.20-2.07	2610	96.9	8.9	3.1
2.07-1.97	2587	96.2	10.0	3.0
1.97-1.89	2596	97.0	13.5	3.0
1.89-1.81	2613	96.4	19.4	3.0
1.81-1.75	2127	80.3	23.9	2.8
25.00-1.75	25805	94.9	6.1	3.9

Structure solution.

The structure was determined using the molecular replacement method as implemented in the CCP4 (CCP4, 1994) program AMORE (Navaza, 1994) with data from 12.0Å to 3.5Å. One monomer of the methyl α -D-mannopyranoside - con A complex (PDB code 5CNA) was used as the search model, with all metal ions, sugar molecules and water molecules removed, and with all atoms set to full occupancy. This was carried out for both I222 and I₂12₁2₁ space groups. A cross rotation function gave only one solution $\alpha = 60.48$, $\beta = 14.96$, $\gamma = 34.92$ with a correlation coefficient of 0.33, 11 standard deviations above noise. The translation search found one solution in each of the space groups. For space group I222 the translation solution was $x = 0.3489$, $y = 0.1270$, $z = 0.2859$, correlation coefficient 73.8, R-factor 32.3%, 52 standard deviations above noise and this solution rigid body refined to give a correlation coefficient of 0.79 and an R factor of 28.9%, for values $\alpha = 59.83$, $\beta = 15.53$, $\gamma = 34.64$, $x = 0.34927$, $y = 0.12700$, $z = 0.28705$. For space

group $I2_12_12_1$, the translation solution was $x = 0.1001$, $y = 0.0849$, $z = 0.2869$ with a correlation coefficient of 0.40, an R-factor of 46.4%, 29 standard deviations above noise. This solution rigid body refined to give a correlation coefficient of 0.45 and an R-factor of 45.0%, for values $\alpha = 58.89$, $\beta = 15.54$, $\gamma = 34.36$, $x = 0.09878$, $y = 0.08490$, $z = 0.28799$. At this point the space group was assigned as $I222$.

Refinement.

The solution in $I222$ was imported into X-PLOR (Brunger, 1992) and refinement carried out using the free R-factor (R_{free}) as a guide. R_{free} was calculated on 10% of data which were excluded from all refinement calculations. The first round of X-PLOR (rigid body refinement) reduced the R-factor to 23.2% and gave an R_{free} of 27.7%. $F_o - F_c$ and $2F_o - F_c$ electron density maps were calculated and strong density was observed for both metal ions and the two mannose residues. A number of changes in protein structure were made manually at this stage (Table 2.4) using 'O' (Jones *et al.*, 1991). The metal ions were included in the model with zero electrostatic charge and the sugar molecules included. Further refinement proceeded smoothly by alternating cycles of automated refinement (X-PLOR restrained positional and temperature factor refinement) and manual intervention using 'O'. The protein was refined with the Engh & Huber stereochemical parameter dictionary (Engh and Huber, 1991). Individual temperature factors were reset to the overall average for the structure after each manual intervention. Apart from the 10% of measured data excluded to monitor refinement no cut-offs were applied to the data. A bulk solvent correction was included in the X-PLOR refinement with parameters for solvent density $0.33\text{e}\text{\AA}^{-3}$, solvent radius 0.25\AA and B factor 50\AA^2 . Electron density maps were calculated using SIGMAA (Read, 1986) modified coefficients with all data to 2.0\AA . Water molecules were added to the model in batches if they satisfied the following

criteria: (1) they corresponded to a peak $> 3.5\sigma$ in the F_o-F_c map, (2) they made hydrogen bonds with reasonable stereochemistry, (3) they reappeared in at least 1σ in subsequently calculated $2F_o-F_c$ maps and (4) a drop in the R_{free} was observed. The progress of the refinement is summarised in Table 2.4. The quality of the final 2.0\AA model is shown in Table 2.5.

*Table 2.4. Progress of refinement for the 2.0\AA α 1-2 manno-*bio*se - con A model.*

Refinement	R factor/ R_{free}
X-Plor: Rigid body fitting (30 cycles).	23.2/ 27.7
O: Mn ²⁺ , Ca ²⁺ , α 1-2 manno- <i>bio</i> se and 66 waters added. 51 Atoms set to zero occupancy from residues 39, 60, 69, 118 - 122, 132, 136, 158, 162, 185, 217, and 228. >20 side chains moved into electron density.	
X-Plor: 70 cycles of positional refinement. 45 cycles of temperature factor refinement.	17.4/ 21.1
O: 11 waters deleted, 39 waters added. Residues 69, 132, 136, 185, 217 and 228 were modelled into density.	
X-Plor: 80 cycles of positional refinement. 50 cycles of temperature factor refinement.	17.4/ 21.7
O: 27 Waters added. Atoms from residues 13, 59, 60, 71, 76, 101, 156, 158, 162, 164, 200, 202, 223 and 102 modelled into density.	
X-Plor: 40 cycles of positional refinement. 40 cycles of thermal refinement.	17.3/ 21.6

Table 2.5. Quality of the final model of the 2.0Å α 1-2 manno-*biose* - con A complex.

Resolution range (Å)	∞ - 2.00
R _{free} (%)	21.6
R factor (%)	17.3
Bond r.m.s. deviation (Å) ^a —	0.009
Angle r.m.s. deviation (°) ^a	1.887
B-factor bonded atoms r.m.s. deviation (Å ²) ^d	2.543
Ramachandran core/additional (%) ^b	85.1/ 14.9
Protein mean B-factor (Å ²) ^c (all)	20.8
Protein mean B-factor (Å ²) ^c (main chain)	19.4
Protein mean B-factor (Å ²) ^c (side chain)	22.3
Sugar mean B-factor (Å ²)	32.8
Solvent mean B (Å ²)	36.0

^a r.m.s. deviation from Engh and Huber ideal values (Engh and Huber, 1991). ^bCore and additionally allowed regions as defined by PROCHECK (Laskowski, MacArthur et al., 1993). ^cCalculated with MOLEMAN (G.J. Kleywegt, unpublished program). All stereochemically modelled atoms were removed prior to B-factor analysis, all bonded atoms including those in the sugars are included in the calculation of r.m.s. ^dB-factor deviation for bonded atoms.

The refined 2.0Å α 1-2 manno-*biose* - con A complex co-ordinates were then used as the starting model against which to refine the 1.75Å data. The UNIQUEFY procedure from the CCP4 suite was implemented in order to flag the same reflections for the R_{free} set as were used in the refinement of the 2.0Å structure. Refinement proceeded using CNS (Crystallography and NMR System) (Brunger and Adams, 1998). This programme is the most recent software from the developers of X-PLOR. It includes as an option the maximum likelihood method of minimisation rather than the least squares method utilised in X-PLOR. In addition, the bulk solvent parameters are calculated automatically during refinement and the data is corrected for anisotropy. Addition of 5 waters, 100

cycles of restrained positional refinement and 65 cycles of restrained temperature factor refinement gave a final R factor of 18.5% and an Rfree of 21.8%.

2.4. Analysis of the final model.

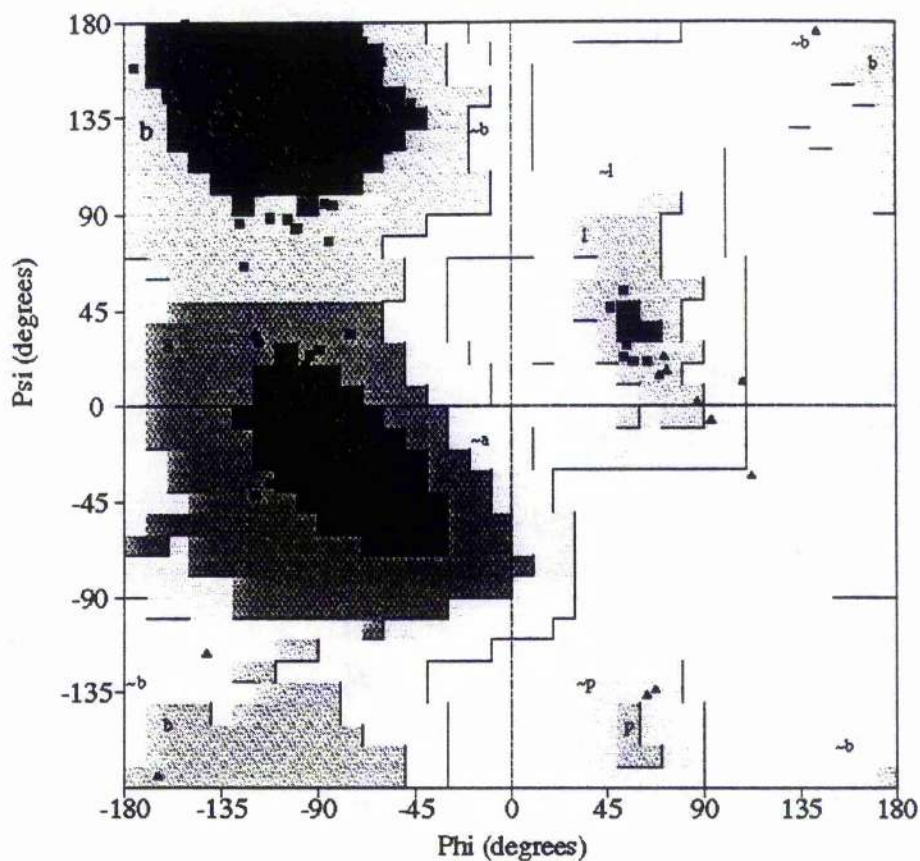
The final structure contains one monomer of 237 residues (1089 protein atoms), 23 sugar atoms, one Ca²⁺ ion, one Mn²⁺ ion and 126 water molecules. The N terminal residue and surface loops at residues 120, 150, 160 and 202 were poorly defined by electron density and the side chains of residues 39, 46, 60, 135 and 223 were stereochemically modelled. The quality of the final model is shown in Table 2.6 and the Ramachandran plot (Ramakrishnan and Ramachandran, 1965) for the final model is shown in Figure 2.2. The Ramachandran plot was calculated using PROCHECK (Laskowski *et al.*, 1993). Glycine residues (not restricted to any particular region of the plot) are represented by a ▲, Non glycine residues by a ■. The darker the shaded region, the more favourable is the ϕ , ψ combination. No residues lie outside the additionally allowed regions. Regions are labelled as follows: A = core alpha, a = allowed alpha, ~a = generous alpha, B = core beta, b = allowed beta, ~b = generous beta, L - core left handed alpha, l = allowed left handed alpha, ~l = generous left handed alpha, p = allowed epsilon, ~p = generous epsilon.

Table 2.6. Quality of the 1.75Å α 1-2 mannobiose - con A model.

Resolution range (Å)	∞ - 1.75
R _{free} (%)	21.8
R factor (%)	18.5
Bond r.m.s. deviation (Å) ^a	0.007
Angle r.m.s. deviation (°) ^a	1.70
B-factor bonded atoms r.m.s. deviation (Å ²) ^c	1.832
Ramachandran core/additional (%) ^b	86.6/13.4
Protein mean B-factor (Å ²) ^c all/ core (all atoms)	20.7/ 17.3
Protein mean B-factor (Å ²) ^c all/ core (main chain)	19.4/ 15.7
Protein mean B-factor (Å ²) ^c all/ core (side chain)	22.2/ 18.9
Sugar mean B-factor (Å ²)	28.0
Solvent mean B (Å ²)	37.9

^a r.m.s. deviation from Engh and Huber ideal values (Engh and Huber, 1991). ^bCore and additionally allowed regions as defined by PROCHECK (Laskowski *et al.*, 1993). ^cCalculated with MOLEMAN (G.J. Kleywegt, unpublished program). All stereochemically modelled atoms were removed prior to B-factor analysis, all bonded atoms including those in the sugars are included in the calculation of r.m.s. B-factor deviation for bonded atoms.

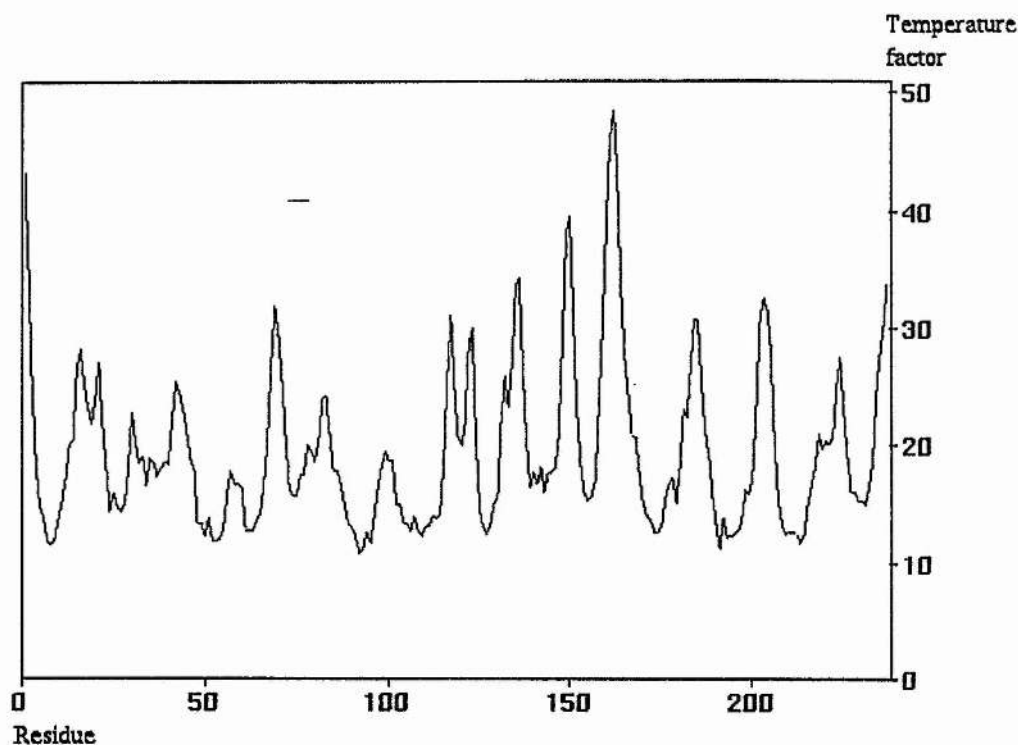
Figure 2.2. Ramachandran plot for the 1.75Å con A - α 1-2 mannobiose model.*



* See section 2.4 for an explanation of the Ramachandran plot.

The temperature factors along the residues of the protein for the subunit are shown in Figure 2.3. The highest peaks correspond with regions found to be consistently disordered in previous crystal structures of con A (Naismith *et al.*, 1993; Naismith *et al.*, 1994; Naismith and Field, 1996; Moothoo and Naismith, 1998).

Figure 2.3. Backbone temperature factor plot for the con A - α 1-2 mannobiose model.



To investigate the effect of sugar binding on the tryptophans, in particular Trp-182 which adopts an unusual conformation in the complexed structures, fluorescence experiments were carried out on a Perkin Elmer luminescence spectrophotometer (LS50B). A solution of 1M glucose was titrated up to a concentration of 170mM into a solution containing 1.5 mgml⁻¹ con A (0.06mM), 1mM CaCl₂, 1mM MnCl₂ and 20mM TRIS pH 7.0. No change in the A₂₈₀ absorbance was observed indicating that no change in the hydrophobic environment of tryptophan in the protein occurs upon binding glucose in solution.

2.5. Results and Discussion.

Tetramer organisation.

The dimer of con A is formed by applying the transformation $(-x+1, -y, z)$ (a two fold rotation axis) to the asymmetric unit. The tetramer is generated by a further two operations ($(-x+1, y, -z+1)$ and $(x, -y, -z+1)$). The four monomers are labelled A, B, C and D in agreement with previous structures. Monomers AB and CD are the two characteristic dimers.

The dimer interaction AB buries 2600\AA^2 of protein surface accessible area and comprises 18 hydrogen bonds, 14 bridging interactions via water molecules and 158 van der Waals interactions. This is comparable with other con A complexes except for an increase in the water mediated contacts (probably because more waters are observed in this higher resolution structure) but less extensive than observed in the native (sugar free) structure. The interactions between monomers A and C, (and B and D) stabilise the tetramer. The interaction consists of 6 hydrogen bonds, 10 hydrogen bond interactions through single water molecules, 82 van der Waals interactions and 6 polar contacts. These interactions are also comparable to those in other con A complex crystal structures but are less than those seen in the native crystal structure. Formation of the complex tetramer buries 9169\AA^2 of protein surface accessible area, almost 1000\AA^2 less accessible surface area than the native.

Metal sites

Table 2.7 gives metal to ligand distances and the temperature factors relative to the core (β -sheet) residues of the protein for both the α 1-2 mannanose structure and the 0.94\AA native structure. The β -sheets are made up from residues 4 to 10, 24 to 29, 36 to 39, 47 to 55, 60 to 66, 73 to 78, 88 to 96, 103 to 117, 123 to 130, 140 to 144, 147 to 148, 154 to

155, 170 to 175, 186 to 200 and 209 to 216. A value close to 1.0 represents a temperature factor close to the average for the core of the protein. A value below 1.0 represents a lower temperature factor than the average for the core of the protein. These ratios allow one to see that this is a well ordered part of the protein, but less well ordered than the metal ligands in the native structure. As this is seen in all of the con A - saccharide bound structures (Chapters 3 to 5), this suggests that this region of the protein is destabilised upon binding sugars or by crystal contacts. The distance between the Ca^{2+} and Mn^{2+} ions is 4.2\AA

Table 2.7. Metal - Ligand distances in the α 1-2 mannoiose - con A complex.

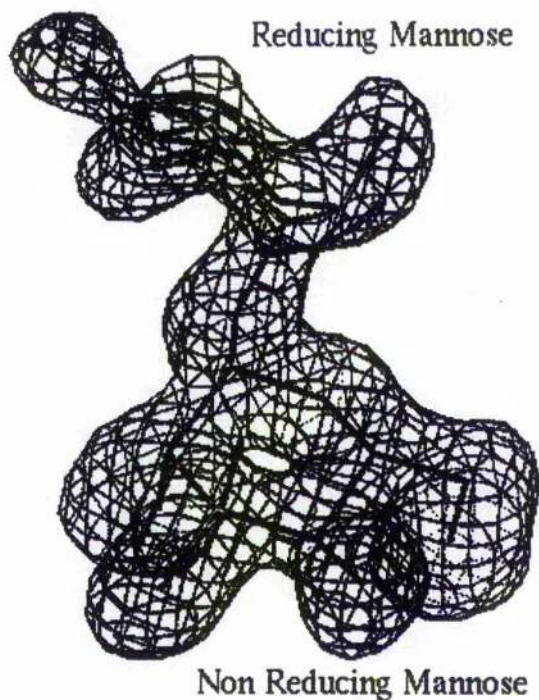
Metal	Ligand	Distance /Å ^a	Temperature factor Ratio ^b	Temperature factor Ratio ^b for native
Ca ²⁺	Asp10 OD1	2.48 (2.51)	0.83	0.67
	Asp10 OD2	2.48 (2.47)	0.67	0.62
	Tyr12 O	2.38 (2.38)	1.01	0.75
	Asn14OD1	2.46 (2.35)	0.99	0.75
	Asp19 OD2	2.32 (2.41)	0.83	0.69
	WA O	2.47 (2.43)	0.98	0.78
	WB O	2.29 (2.36)	0.90	0.73
Average		2.41 (2.41)	0.93	0.71
Mn ²⁺	Glu8 OE2	2.25 (2.16)	0.79	0.65
	Asp10 OD2	2.18 (2.15)	0.83	0.62
	Asp19 OD1	2.28 (2.19)	0.89	0.70
	His24 NE2	2.27 (2.23)	0.73	0.64
	WC O	2.17 (2.18)	0.87	0.82
	WD O	2.28 (2.26)	0.74	0.70
Average		2.24 (2.20)	0.82	0.69

^aValues for 0.94Å native structure shown in parentheses. ^bTemperature factor/ average temperature factor of β -sheet residues.

Sugar Binding

Clear difference electron density is observed in the binding site for the disaccharide (Figure 2.4).

Figure 2.4. F_o-F_c omit map contoured at 2.8σ .



The monosaccharide binding site is formed by residues 12-14, 98-100, 207-208 and 227-228 as described previously (Naismith *et al.*, 1994). The non-reducing mannose binds in the monosaccharide site making a total of a total of six hydrogen bonds (Figure 2.5), 36 van der Waals interactions and 4 polar contacts with the protein (Tables 2.8 and 2.9). The reducing sugar sits in an extended binding site formed by Tyr-12, Asp-16 and Arg-228. The binding site is surrounded by crystal contacts, the closest being 2.6\AA from O1 of the reducing sugar. Ser-185 from a symmetry related molecules occupies the site of the 1,3 linked mannose in the trimannoside - con A structure (Naismith and Field, 1996).

Table 2.8. Hydrogen bonds and polar contacts ($< 3.5 \text{ \AA}$) between α 1-2 mannobiose and con A.

Sugar atom	Protein atom	Distance/ \AA
Non-reducing mannose		
O3	Arg-228 N	3.0
O4	Asn-14 ND2	3.0
O4	Asp-208 OD2	2.5
O4	Arg-228 N	3.1 ^a
O5	Leu-99 N	3.3
O6	Gly-98 N	3.1 ^a
O6	Leu-99 N	3.1 ^a
O6	Tyr-100 N	3.1
O6	Tyr-100 O	3.4 ^a
O6	Asp-208 OD1	2.8
Reducing mannose		
O1	Tyr-12 OH	3.5
O5	OW	2.8
O6	Asp-16 OD2	3.4 ^a

^aThese are polar contacts; the distance is $< 3.5 \text{ \AA}$ but geometry deviates from the linearity expected for a hydrogen bond.

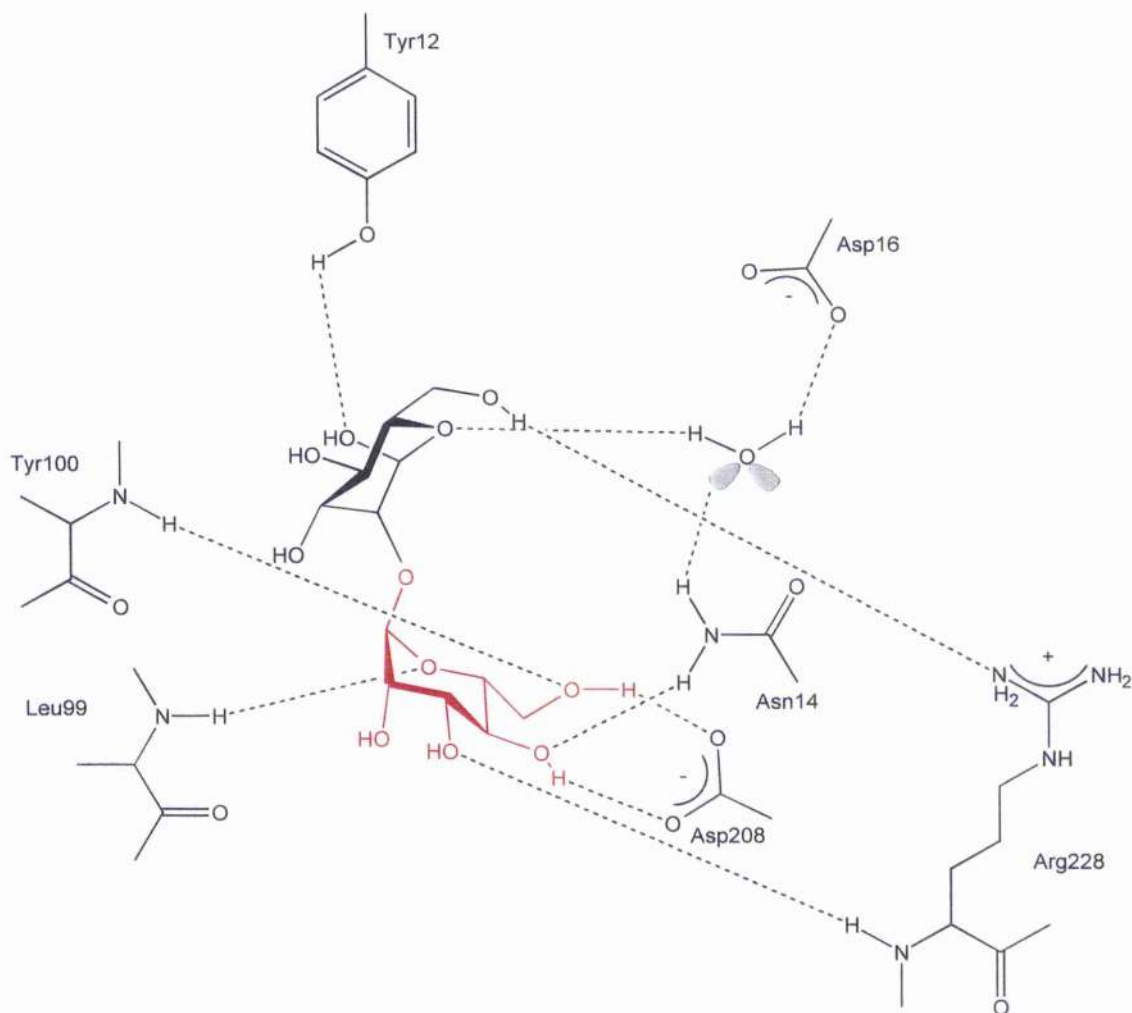
Table 2.9. Van der Waals interactions (< 4.0 Å) between α 1-2 manno-*bio*se and con A.

Sugar atom	Residues	Number of contacts
The non-reducing sugar		
C1	Leu-99	1
O2	Gly-98, Leu-99	2
C3	Arg-228	1
O3	Gly-227, Arg-228	5
C4	Asp-208, Arg-228	4
O4	Asn-14, Asp-208, Gly-227, Arg-228	6
O5	Leu-99	2
C6	Tyr-12, Leu-99, Tyr-100, Ala-207, Asp-208	7
O6	Gly-98, Leu-99, Tyr-100, Ala-207, Asp-208	8
The reducing sugar		
C1	Tyr-12	1
O1	Tyr-12	1
C6	Asp-16	1
O6	Arg-228	1

In this complex, the reducing mannose makes a hydrogen bond with Tyr-12, a polar contact with Asp-16 and in addition a water bridged hydrogen bond to the protein (Table 2.8). This is the 'structural' water molecule (Naismith and Field, 1996) and plays an

important role in mediating con A - oligosaccharide contacts. In addition the reducing sugar makes 4 van der Waals interactions with the protein (Table 2.9).

Figure 2.5. Hydrogen bond diagram of the α 1-2mannobiose - con A complex. The sugar residue coloured red is bound at the monosaccharide binding site.



Formation of the α 1-2 mannobiose - con A complex buries 179\AA^2 of protein surface area and 289\AA^2 of sugar surface area. For the sugar this comprises 206\AA^2 of the non-reducing sugar buried and 83\AA^2 of the reducing sugar. Buried protein surface area is increased by 59\AA^2 over the methyl α -D-mannopyranoside - con A structure (see appendix 1) and buried surface area of the monosaccharide binding sugar is reduced by 15\AA^2 . Asp-16

which was not buried by the carbohydrate in the methyl α -D-mannopyranoside - con A structure now has 19\AA^2 of surface accessible area buried and is newly recruited into binding. Tyr-12 and Arg-228 are less accessible due to interactions with the reducing sugar. Breaking this down further in to buried polar and apolar surface accessible area, a total of 55% of the total protein buried surface accessible area is polar and 45% is apolar. 70% of the sugar buried surface accessible area is polar, 30% apolar. The ϕ, ψ glycosidic linkage conformation angles of the disaccharide in this structure are close to those observed in solution by NMR and lie in an energy minima (Imberty *et al.*, 1990) (NMR structure: $\phi = 59^\circ$, $\psi = -140^\circ$, crystal structure: $\phi = 65^\circ$, $\psi = -141^\circ$, for definitions of ϕ and ψ see Figure 2.1).

Prediction of the modes of binding of α 1-2 mannobiose to con A using energy minimisation methods (Reddy and Rao, 1992) docked the disaccharide to the protein with the non-reducing sugar in the monosaccharide binding site with a more favourable energy than the reducing sugar. The reducing sugar makes a hydrogen bond as well as several van der Waals interactions with the protein. The crystal structure of the complex confirms that it is the non-reducing sugar which is recognised by the monosaccharide site. However, the reducing sugar makes interactions with the protein which are different from those proposed by the minimisation study. This reflects the insufficient parameters available for protein - saccharide interactions. The importance of the so-called conserved structural water first observed in the trimannose bound structure (Naismith and Field, 1996) was unknown and not included in modelling. In addition, the native con A structure was used in the modelling studies and changes in the protein structure upon carbohydrate binding were not taken into account.

Effect of sugar binding on the protein structure

The dimannose bound structure superimposes onto the 0.94Å native structure with a root mean square deviation (r.m.s.d.) of 0.50Å for all C α backbone atoms and 0.25Å for the C α atoms of the β -sheet residues. Table 2.10 shows average r.m.s.d.'s when superimposing the C α atoms of the various con A structures onto each other. The native structure has a consistently higher r.m.s.d. for superposition to any of the other complex structures. This suggests that sugar binding brings about substantial changes in the protein structure, even in regions remote from the binding site itself.

Table 2.10. R.m.s. deviations for the C α superposition of con A complexes.

	r.m.s. deviation /Å	r.m.s. deviation /Å β -sheet C α only
Dimannose \rightarrow Native	0.50	0.25
Dimannose \rightarrow Man OMe ^a	0.36	0.18
Dimannose \rightarrow Trisaccharide ^b	0.35	0.17
Dimannose \rightarrow Pentasaccharide ^c	0.28	0.17
Me- α -Man \rightarrow Native ^a	0.50	0.27
Trisaccharide \rightarrow Native ^b	0.57	0.30
Pentasaccharide \rightarrow Native ^c	0.51	0.27

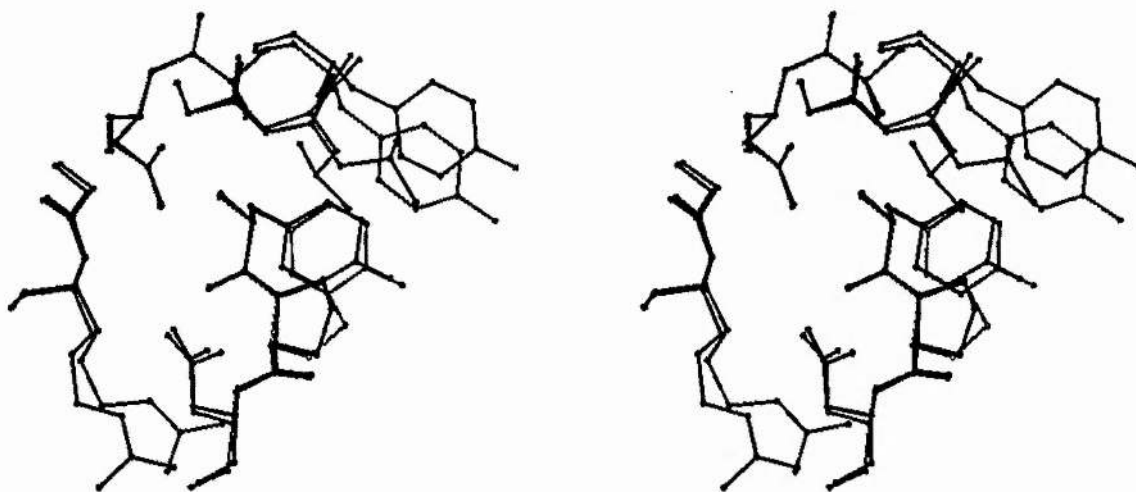
^aAveraged over three of the four subunits present in the asymmetric unit. ^bAveraged over the 4 subunits present in the asymmetric unit. ^cAveraged over the 6 well ordered subunits present in the asymmetric unit.

The sugar binding site is remote from the core of the protein and involves only loop regions. Thus it seems unlikely that the sugar could exert such effects throughout the

protein, nor is there any evidence for transmission of conformation. From the con A - pentasaccharide complex (Moothoo and Naismith, 1998)(Chapter 4) with two tetramers in the asymmetric unit, the details of the tetramer organisation depends upon the precise environment of the protein in the crystal. In addition, from the fluorescence measurements of sugar binding, there is no change upon binding of glucose in solution. On a residue by residue comparison of the native structure with any con A complex the most obvious change is centred around Trp-182, this change has been discussed in detail (Naismith *et al.*, 1994). Trp-182 is remote from the sugar binding site but sits close to the dimer interface and interacts with the β sheet. The change in its conformation thus has a significant effect on global comparison of the structure. Fluorescence measurements show that this Trp movement does not occur upon sugar binding in solution, therefore it is almost certainly due to crystal packing. No changes in the metal co-ordination number result from binding the dimannose molecule, however there are some subtle changes in co-ordination distances, particularly the Ca^{2+} ligands Asn-14 and Asp-19. These appear to be due to the movement of the protein to accommodate the disaccharide.

Effect of sugar binding at the sugar binding sites

Figure 2.6. Comparison of native and α 1-2 mannobiose-bound monosaccharide binding sites. Superposition of the α 1-2 mannobiose binding site (black) onto the binding region in the native structure (blue) was carried out via the protein C α atoms.



Conformations of Tyr-100 and Leu-99 are different with respect to the native structure and are consistent with the other saccharide bound con A structures, the only exception being in the pentasaccharide bound structure where there is a 0.5Å movement in the position of the Tyr-100 side chain (Chapter 4) (Moothoo and Naismith, 1998). The loop from Pro-202 to Pro-206 varies almost 2Å from the native structure. Most pronounced is the conformation of the side chain of His-205. In this structure, the residue is in the same conformation as that observed in the native, trimannose and pentasaccharide structure. NE2 of the histidine is 3.11Å from a symmetry related water molecule, 2.98Å from a symmetry related Asp-82 O and is less than 4.0Å away from Trp-182 of the same symmetry related molecule. Tyr-12 makes a weak hydrogen bond with the sugar. At 3.5Å it is at the limit for the length of a hydrogen bond (Baker and Hubbard, 1984). In the trimannose and pentasaccharide bound structures, this interaction is much stronger. This

is due to the conformation of the 1-2 linkage of the disaccharide placing the hydroxyls of the sugar in a different position (Figure 2.7). The loop from Thr-15 to Ile-17 is pulled in towards the sugar significantly, 1.3Å at Asp-16, the residue which interacts with the structural water molecule. This loop is more removed from the native structure than in the trimannose and pentasaccharide bound structures. The side chain of Arg-228 is moved up to 1.9Å from its position in the native structure as observed in the other saccharide bound structures, allowing a weak hydrogen bond with the sugar. The side chain of this residue was not observed to previously interact significantly with the sugars, except for a 3.7Å van der Waals interaction in the trimannose bound structure.

The difference in the conformation of the 1-2 linkage compared with the longer 1-6 linkage (Figure 2.7) which is present in the other oligosaccharide bound structures results in the loss of the hydrogen bond between Arg-228 and the structural water.

Figure 2.7. Comparison of the sugar positions in the α 1-2mannobiose-bound and trimannose-bound structures. The reducing and non reducing sugar of the α 1-2 mannobiose (Di) and trimannoside complex (Tri) have been superimposed through the protein backbone illustrating the effect on sugar position of the α 1-2 and α 1-6 linkage.

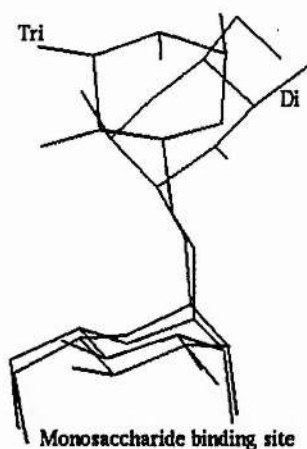


Table 2.11 shows the ratios of the average temperature factor of each of the binding loops to that of the protein core residues. Also shown are the values for the native structure. These figures suggest that these loops are somewhat better ordered in the native structure relative to the core. The loop 207 - 208 is still the better ordered of these loops, and loop 227-228 is still the most disordered. This is seen in all of the con A - carbohydrate structures discussed in this thesis. Destabilisation of the binding loops upon binding is somewhat unexpected as the extensive interaction between the ligand and the carbohydrate mean that the loops should become tied down relative to the native structure. The temperature factor of the carbohydrate is above the core of the protein (ratio of 1.6). It is possible that mobility of the ligand results in some disordering of the

binding loops relative to the native structure. However it may reflect the unusual tightness of crystal contacts.

Table 2.11. Temperature factor ratios of saccharide binding loops

Binding loop	Temperature factor ratio ^a	Temperature factor ratio for native ^a
Tyr-12 - Asn-14	1.11	0.80
Gly-98 - Tyr-100	1.09	0.77
Ala-207 - Asp-208	0.83	0.67
Gly-227 - Arg-228	0.93	0.81

^aTemperature factor of binding loop/ average b factor of β -sheet residues.

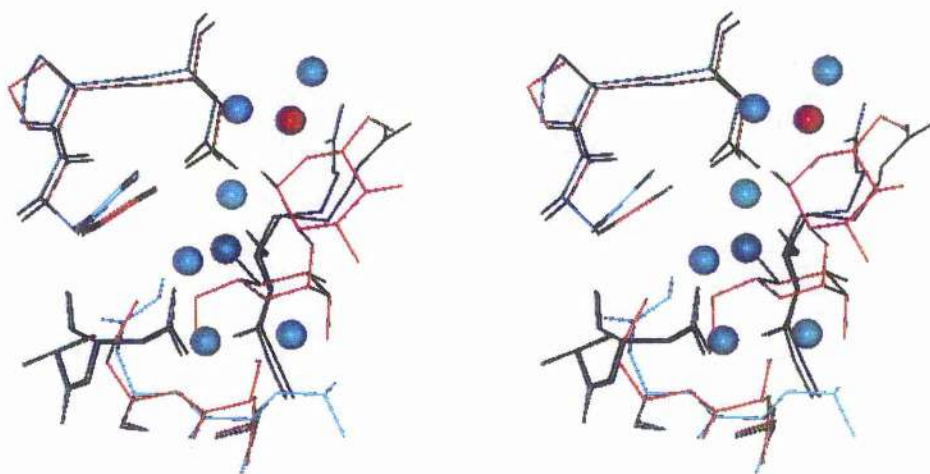
Effect of sugar binding: water structure

When making a comparison of the water structure between the con A structures, it should be noted that the number of waters observed is highly dependant on the resolution of the structure. For instance, 319 waters were included in the final model of the 0.94Å native structure (Deacon *et al.*, 1997), 175 more than were included in the earlier 2.0Å native structure and the 1.75Å refined α 1-2 mannanose - con A structure contains 5 extra water molecules than the 2.0Å refined structure.

The disaccharide molecule displaces a total of 5 water molecules from its binding site (Figure 2.8). The 'structural' water (Naismith and Field, 1996) which is conserved in all con A structures is present but undergoes a 2.1Å movement from its position in the native structure (Deacon *et al.*, 1997). In the native structure, instead of co-ordinating Asn-14 it co-ordinates an additional water which in turn co-ordinates with Asn-14. Three of the five displaced waters correspond with the positions of O4, O5 and O6 of the non-

reducing sugar and were observed in the lower resolution studies (Naismith *et al.*, 1993). An additional water molecule sits close to the position of C6 of the non-reducing sugar was not present in the 2.0Å native structure. The fifth displaced water sits < 1.5Å from the position of C1 of the reducing sugar. There are four water molecules occupying the site of the 1,6 arm terminal GlcNAc in the pentasaccharide structure (Chapter 4) (Moothoo and Naismith, 1998) one of which bridges the non reducing sugar and the protein.

Figure 2.8. Displacement of water from the binding site. Superimposition of the α 1-2 mannobiose binding site (red) onto the native binding site (blue) through the backbone of the binding site residues. The red sphere is the 'structural' water.



Structural basis of affinity for the α 1-2 linked disaccharide

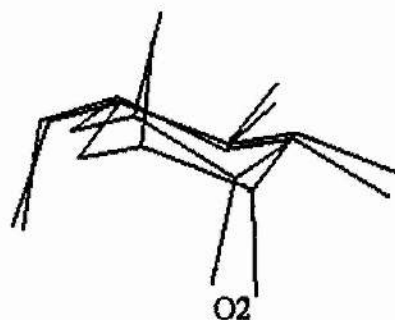
Table 2.1 shows thermodynamic data for con A binding to α 1-3, α 1-6 and α 1-2 mannobiose (Mandal *et al.*, 1994). The K_a observed for the association between con A and α 1-2 mannobiose has a five fold increase over methyl α -D-mannopyranoside and a three fold increase over α 1-3 and α 1-6 linked disaccharides. The comparison with both

α 1-3 and α 1-6 disaccharides is pertinent as they are constituent fragments of the trisaccharide Man α 1-3(Man α 1-6) Man which is recognised with high affinity. It may be thought that they should bind in this manner to the protein. From studies of the monosaccharide complex of con A, only one mannose was observed bound to the complex at the monosaccharide site, despite being crystallised from 30 fold excess of sugar to monomer. This would indicate that for both α 1-3 and α 1-6 manno-
bi-ose the non-reducing terminal mannose is bound at the monosaccharide site. Modelling α 1-3 and α 1-6 manno-
bi-ose into the binding site with the non-reducing terminal mannose superimposed onto that in the α 1-2 manno-
bi-ose structure, there is no apparent steric hindrance to the binding of either the α 1-3 or α 1-6 manno-
bi-ose.

Comparison with the methyl α -D-mannopyranoside - con A complex

In the methyl α -D mannopyranoside complex (see appendix 1) (Naismith *et al.*, 1994), the saccharide made an average of 46 van der Waals interactions with the protein, two of which are made by the methyl group. In the dimannose bound structure the corresponding atom (C2 of the reducing sugar) is more than 4.0Å from the nearest protein atom. The plane of the non-reducing mannose ring is rotated slightly from its position in the monosaccharide bound complex (Figure 2.9).

Figure 2.9. Differences in the monosaccharide binding site. α 1-2 Mannobiose and methyl α -D-mannopyranoside superimposed through the protein backbone, illustrating the slight rotation of the ring resulting in a shift in the position of O2.



A similar but more pronounced disturbance at the monosaccharide site was seen in the pentasaccharide bound structure (Chapter 4; Moothoo and Naismith, 1998). The movement observed here is sufficient that O2 is displaced by 0.6Å from its position in the monosaccharide complex structure. Buried surface analysis detects a slight reduction in the amount of surface area buried at the monosaccharide site. Thus it appears that the monosaccharide site is marginally destabilised with respect to that seen in the methyl α -D-mannopyranoside - con A complex. Assigning values to this destabilisation is not possible without more accurate modelling technologies than are currently available. Mobility of the monosaccharide in the binding site was investigated recently by Helliwell and co-workers (Bradbrook *et al.*, 1998). This work encompassed structural information along with static and dynamic calculations to investigate the contributions which give rise to the tighter binding of mannoside than glucoside by con A (Bradbrook *et al.*, 1998). This work highlighted the need for dynamic models to help relate thermodynamic and structural information in order to provide a more complete picture of ligand recognition.

It has been suggested that both mannose residues in α 1-2 manno-*bio*se, with free hydroxyl groups at 3-, 4- and 6- positions, can independently bind at the monosaccharide site (Mandal *et al.*, 1994). We see no evidence of disorder in this complex, however, crystallisation is a kinetic process and an equilibrium between two binding modes may exist in solution. In the structure of con A complexed with methyl α 1-2 manno-*bio*side, two binding modes for the disaccharide are observed to exist in identical binding sites (Chapter 3).

The α 1-6 linkage contains three rotatable bonds rather than the two in the α 1-2 manno-*bio*se. Further, in the α 1-6 disaccharide, the linkages are known from solution studies to be extremely flexible. Presumably, freezing three bonds will be more entropically unfavourable than freezing two lowering the overall affinity for the sugar. Precise values are unavailable for this although a figure of 0.5 kcalmol⁻¹ per rotatable bond is often quoted as guide.

The different conformation and length of the α 1-2 linkage which links the reducing sugar to the monosaccharide binding sugar compared with the α 1-6 linkage found in the trimannoside, forces the ring of the reducing sugar to sit in a different position relative to the protein (Figure 2.7). The result is that the protein uses the same residues to recognise a different spatial distribution of oxygen and carbon atoms of the carbohydrate. This effect is reminiscent of the way similar sites in con A and Ecor L are able to accommodate mannose and galactose, respectively, illustrated in Figure 1.14. In comparison to the trisaccharide complex the reducing sugar contacts are less extensive for the α 1-2 manno-*bio*se. In particular, the hydrogen bonds between the sugar and residues Tyr-12 and Asp-16 are lost. However, buried surface area analysis of the dimannose bound complex shows that the reducing sugar buries an additional 2Å² of

polar surface area and 105\AA^2 of apolar surface area. Assuming that the α 1-6 manno-
biose from the trisaccharide structure provides a reasonable approximation to the binding of
 α 1-6 manno-
biose, buried surface area calculations can be carried out for this complex. In
the α 1-6 manno-
biose complex then, an additional 23\AA^2 of polar surface area and 7\AA^2 less
apolar surface area is buried relative to α 1-2 manno-
biose. Despite making fewer
hydrogen bonds, the reducing sugar in the α 1-2 dimannose complex buries more apolar
surface area. Superimposing α 1-3 manno-
biose into the binding site via the non-reducing
terminal mannose reveals that the reducing sugar makes no hydrogen bonds or polar
contacts with the protein. Furthermore, 35\AA^2 less polar and 1\AA^2 less apolar surface
accessible area is buried by the con A - α 1-3 manno-
biose complex than the con A - α 1-6
manno-
biose complex. It appears that the origin of the increased affinity of the α 1-2
manno-
biose lies in a complex mixture of factors. Amongst the disaccharides it alone
seems able to enhance its affinity by burying sufficient additional apolar surface area
without paying entropic or enthalpic penalties.

2.6. Conclusions

Con A specifically recognises both the α 1-6 and α 1-2 linkages within the same binding
site. The non-reducing sugar of each of these disaccharides is bound at the
monosaccharide site in an almost identical manner. However, the reducing sugar rings are
rotated and translated relative to each other and the details of the interactions are quite
different. From the three fold increase in affinity displayed by α 1-2 manno-
biose over its
 α 1-6 linked counterpart one might expect to see strong interactions between the reducing
sugar and the protein. However, modelling the α 1-6 manno-
biose complex from the
trisaccharide complex, it can be seen that the reducing mannose of 1-6 linked dimannose

makes more numerous contacts with the protein than it does in the 1-2 linked sugar. It appears that a combination of the entropy penalty for immobilising the α 1-6 linkage and the increase in apolar surface area buried by the α 1-2 manno-*bio*se may explain the difference in affinity. A comparison to the native structure, which crystallises with a very similar packing arrangement, identifies clear differences in protein conformation affected by the carbohydrate. This high resolution study increases our knowledge about the forces governing protein saccharide interactions, providing an insight to the balance of forces involved in the affinity of protein for a carbohydrate.

2.7. Future Work

In this Chapter, it has been assumed that the 1-6 arm in the trimannoside and pentasaccharide bound con A structures provides a reasonable approximation to the binding of α 1-6 manno-*bio*se to con A. Experimental evidence to support this would be valuable and therefore an important next step would be to obtain the crystal structure of α 1-6 manno-*bio*se with con A. In addition, further valuable information would be provided by the crystal structure of α 1-3 manno-*bio*se with con A and extending the resolution of this structure, particularly using neutron diffraction.

Chapter 3

**The 2.75Å structure of concanavalin A complexed with
methyl α 1-2 mannoside.**

3.1. Summary.

The free energies of binding α 1-3 and α 1-6 mannobioses and their methylated counterparts by con A are comparable with monosaccharide binding. Con A has an enhanced affinity for α 1-2mannobiose corresponding to both more favourable enthalpic and entropic contributions to the binding energy. The methyl glycoside of α 1-2mannobiose (methyl α 1-2 mannobioside) binds to con A with an increased affinity over α 1-2mannobiose.

Methyl α 1-2mannobioside was synthesised and its crystal structure in complex with con A determined to 2.75 Å. The crystals of the methyl α 1-2 mannobioside - con A complex form in the space group P2₁2₁2₁ with cell dimensions a = 119.7Å, b = 119.7Å, c = 68.9Å and diffract to 2.75Å. The final model has good geometry and an R factor of 19.6% (R_{free} = 22.8%). One tetramer is present in the asymmetric unit in which the disaccharide clearly displays two different modes of binding. In one mode, the non-reducing terminal mannose is recognised at the monosaccharide binding site. The oligosaccharide adopts an unfavourable conformation about the psi inter-sugar dihedral angle in order to maximise the interaction of the anomeric methyl group with the protein. In the other binding mode the reducing terminal mannose is recognised at the monosaccharide binding site. No distortion of the inter-sugar glycosidic linkage is seen, but the hydrophobic interactions with the secondary sugar are much less extensive. This provides structural evidence to support the hypotheses that the α 1-2 linked disaccharide has 2 distinct binding modes in that either mannose residue can be accommodated in the monosaccharide binding site. Thus a balance between the strain of oligosaccharide conformation and the difference in the interactions made between the sugar and protein on the different binding modes means that the energy barrier between the different binding sites is relatively low, so that

either may be adopted, giving rise to a more favourable binding energy than the α 1-3 and α 1-6 linked mannobioses.

3.2. Introduction.

In chapter 2, the 1.75Å structure of α 1-2 mannobiose complexed to con A was described, in which the non-reducing terminal mannose occupies the monosaccharide binding site and the reducing mannose occupies the region of that in the trimannoside complex. Methylation at the anomeric position of α 1-2 mannobiose leads to an enhancement in affinity. Table 3.1 shows thermodynamic data for binding of the mannobioses and their methylated counterparts by con A. The values for α 1-6 mannobiose are very similar with binding methyl α -D-mannopyranoside. The values for α 1-3 mannobiose show a relative favourable enthalpy term, more favourable than α 1-2 mannobiose, however the overall free energy is still similar with monosaccharide binding due to an unfavourable entropic term. α 1-2 Mannobiose is unusual amongst these disaccharides in its enhanced affinity for con A due to a favourable relative entropic term. The effect of methylation varies in all cases. In α 1-3 Mannobiose, methylation leads to a slightly enhanced affinity due to an increased enthalpic term which suggests that extra interactions are made over α 1-3 mannobiose. For α 1-6 mannobiose, methylation leads to a decrease in affinity. In the case of α 1-2 mannobiose however, methylation leads to a 3 fold increase in affinity.

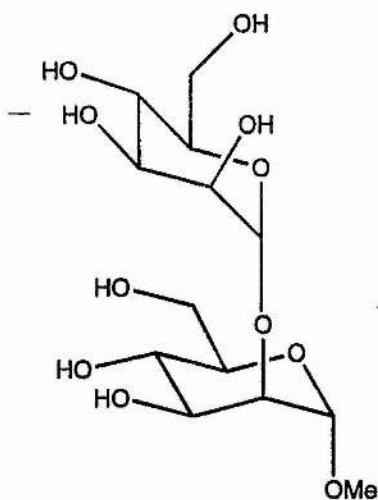
Table 3.1. Thermodynamics of con A binding mannobioses and their methylated counterparts. (Mandal *et al.*, 1994b)

	Ka ($\times 10^4 \text{ M}^{-1}$)	$-\Delta H^\circ$ (kcal mol $^{-1}$)	$-T\Delta S^\circ$ (kcal mol $^{-1}$)	ΔG°
Me- α Man	0.82	8.2	2.9	-5.3
α 1-2mannobiose	4.17	9.9	3.6	-6.3
methyl- α 1-2mannobioside	14.1	10.5	3.5	-7.0
α 1-6mannobiose	1.34	9.4	3.8	-5.6
methyl- α 1-6mannobioside	0.81	8.4	3.1	-5.3
α 1-3mannobiose	1.41	10.2	4.5	-5.7
methyl- α 1-3mannobioside	3.35	10.7	4.5	-6.2

Goldstein's rules (Goldstein *et al.*, 1974) describe the requirements for monosaccharide binding to con A as being free equatorial hydroxyl groups at the 3- and 4- and 6-positions. Both sugar rings in α 1-2 mannobiose can potentially be recognised by the monosaccharide binding site. It has been proposed that this enhances the probability of binding the carbohydrate and leads to a more favourable entropy contribution to the binding energy (Mandal *et al.*, 1994b). On the basis of thermodynamic data, another group proposed that the increased affinity is due to the presence of additional sites for binding the disaccharide (Goldstein, 1975). Neither hypothesis has been confirmed experimentally.

In order to continue these studies and to further answer the questions posed by calorimetry, α 1-2 mannobiose has been methylated at its anomeric position (Figure 3.1) and the structure in complex with con A has been determined to 2.75Å resolution.

Figure 3.1. methyl α 1-2 mannobioside.



3.3. Experimental.

Crystallisation and Data Collection.

Co-crystals of the con A - methyl α 1-2 mannobioside complex were obtained after an initial screen consisting of 10, 15, 20 and 25% Polyethylene Glycol (PEG) 6K, pH 4.0, 5.0, 6.0 and 7.0. Optimisation of pH and PEG concentration then followed. Crystals were obtained from 10 μ l of a solution containing 0.6mM con A (sigma, Poole, UK), 18mM methyl α 1-2 mannobioside (synthesised by Booma Canan and Robert A. Field at the University of St. Andrews), 1mM MnCl₂, 1mM CaCl₂, 20mM TRIS pH 7.0 and 0.1M NaCl was mixed with 10 μ l of a reservoir which contained 13.5% PEG 6K, 1.0M LiCl, 0.1M TRIS pH 7.0 and left to equilibrate (Moothoo *et al.*, 1998b). The crystals were elongated rods and varied considerably in size. Some crystals grew as hollow rods. The crystals took 2 to 4 weeks to grow and a hollow crystal with dimensions 0.5 x 0.2 x 0.2 mm³ mounted in a glass capillary was used for X-ray analysis. Data to 2.75Å were recorded at room temperature as 190 non-overlapping 32 minute 0.8° oscillations on an

Enraf-Nonius/ MacScience DIP2000 dual image plate. Auto indexing with DENZO (Otwinowski, 1993) suggested two different lattice types: primitive tetragonal and primitive orthorhombic. A stable unit cell matrix was obtained in both lattice types. Data were integrated in both lattices. Frames 1 to 30 showed a high R factor and low $\langle I/\sigma \rangle$ compared with the remaining data. Inspection of the images showed that the cell contracted $\sim 1.5\text{\AA}$ along the c dimension during the first 30° of data collection after which the cell became more stable. This is shown in Table 3.2.

Table 3.2. Cell output from DENZO during the first 40 frames of data collection.

	a	b	c	cell volume/ Å ³	I/σ
Frame 1	68.71	119.86	121.46	1000294	3.1
Frame 2	68.75	119.98	121.47	1001960	3.3
Frame 3	68.79	120.06	121.55	1003873	3.5
Frame 4	68.76	120.03	121.43	1002194	3.8
Frame 5	68.77	120.03	121.55	1003330	3.4
Frame 6	68.75	120.05	121.52	1002958	3.2
Frame 7	68.81	120.15	121.61	1005413	3.2
Frame 8	68.75	120.19	121.42	1003301	3.1
Frame 9	68.75	119.76	121.60	1001194	3.3
Frame 10	68.67	119.93	121.35	999389	3.5
Frame 11	68.69	119.88	121.39	999593	3.3
Frame 12	68.73	120.21	121.33	1002433	3.6
Frame 13	68.64	119.81	121.39	998282	3.7
Frame 14	68.67	120.11	121.29	1000394	3.9
Frame 15	68.58	119.74	121.31	996170	4.2
Frame 16	68.56	119.96	121.04	995488	3.7
Frame 17	68.55	119.84	120.80	992376	3.8
Frame 18	68.71	120.18	120.72	996854	4.3
Frame 19	68.56	119.70	120.27	987012	4.4
Frame 20	68.71	119.91	120.24	990659	4.2
Frame 21	68.72	119.56	120.27	988158	4.4
Frame 22	68.80	120.12	120.04	992041	4.6
Frame 23	68.89	119.69	120.26	991597	4.9
Frame 24	68.87	119.74	120.14	990734	4.7
Frame 25	68.89	119.88	120.18	992511	5.3
Frame 26	68.93	119.90	120.00	991765	5.2
Frame 27	68.88	119.68	120.13	990299	4.8
Frame 28	68.89	119.82	119.84	989207	5.6
Frame 29	68.90	119.57	120.00	988605	5.4
Frame 30	68.94	119.92	119.80	990421	5.5
Frame 31	68.96	119.49	120.16	990122	5.6
Frame 32	68.97	119.90	119.99	992258	5.4
Frame 33	68.96	119.67	120.27	992521	5.7
Frame 34	68.95	119.87	120.09	992548	6.3
Frame 35	68.91	119.78	120.16	991805	5.9
Frame 36	68.93	119.97	119.93	991765	5.7
Frame 37	68.91	119.86	120.19	992716	6.0
Frame 38	68.93	119.93	120.10	992840	5.7
Frame 39	68.89	119.74	120.27	992094	6.1
Frame 40	68.88	119.94	120.13	992450	6.0

Contraction of the cell could possibly be due to drying out of the crystal. The diffraction intensity clearly increases as the cell decreases. Frames 1 to 30 were excluded from all further analysis. Merging the data in Laue group 4/m gave an R-merge of 47.4% and 50.8% for Laue group 4/mmm. Merging the data using SCALEPACK (Otwinowski, 1993) in Laue group 2/mmm gave an R-merge of 10.1%. The space group was assigned as P2₁2₁2₁ (on the basis of systematic absences) with unit cell parameters a = 119.7Å, b = 119.7Å, c = 68.9Å to give an R_{merge} of 10.1%, an average I/σ of 8.9, a multiplicity of 3.29, and 99.7% completeness. The data had a mosaic spread of 0.8°. The asymmetric unit contains one tetramer, approximately 49% solvent. Data collection and reduction statistics are shown in Table 3.3.

Table 3.3. Quality of the data for the con A - methyl α1-2 mannoside complex.

Resolution	No.Reflections	% Complete	R-Merge/ %	Redundancy
20.00-5.88	2772	98.0	6.2	4.9
5.88-4.69	2680	100.0	7.4	5.1
4.69-4.10	2672	100.0	8.4	5.1
4.10-3.73	2620	100.0	9.8	5.2
3.73-3.46	2609	100.0	10.8	5.2
3.46-3.26	2636	99.9	13.2	5.1
3.26-3.10	2625	99.9	15.8	5.1
3.10-2.96	2603	100.0	19.4	5.1
2.96-2.85	2590	99.9	24.6	5.0
2.85-2.75	2585	99.5	29.0	4.9
20.00-2.75	26392	99.7	10.5	5.0

Structure Solution.

The structure was determined by the molecular replacement method as implemented in the CCP4 (CCP4, 1994) program AMORE (Navaza, 1994) using all data from 10Å to 3.8Å. The trimannoside - con A complex (Naismith and Field, 1996) was used as the search model, with all metal ions, sugar molecules and water molecules removed and all atoms given full occupancy. A cross rotation function gave seven solutions, 10 to 11 standard deviations above noise (Table 3.4). Rotation solution 3 gave the best translation correlation coefficient 0.71, R-factor 33.3% which refined to give a correlation coefficient of 0.76 and R factor of 31.4%.

Table 3.4. Rotation and translation solutions.

Sol ⁿ	Rotation			Translation			cc ^a	R ^b
	α	β	γ	x	y	z		
R1 ^c	77.43	7.14	82.49	-	-	-	0.52	-
R2	113.74	7.12	82.41	-	-	-	0.52	-
R3	519.71	0.00	0.00	-	-	-	0.47	-
R4	159.73	0.00	0.00	-	-	-	0.47	-
R5	339.67	0.00	0.00	-	-	-	0.47	-
R6	376.27	0.00	0.00	-	-	-	0.46	-
R7	196.20	0.00	0.00	-	-	-	0.46	-
T ^d	519.71	0.00	0.00	0.1470	0.2582	0.4512	0.71	33.3
R ^e	476.24	0.69	42.99	0.1469	0.2587	0.4520	0.76	31.4

^a correlation coefficient. ^b R factor. ^c R = Rotation function solution. ^d T = Translation function solution from R solution 3. ^e Refined solution which was imported into CNS.

Refinement.

The solution was imported into CNS (Brunger and Adams, 1998) and refinement carried out using the R_{free} as a guide. R_{free} was calculated on 10% of data which were excluded from all refinement calculations. The raw molecular replacement solution had an initial R factor in CNS of 44%. Rigid body refinement of the tetramer followed by each subunit lowered this to 28.0%. F_o-F_c and $2 F_o-F_c$ electron density maps were calculated from this model. Strong density was observed for a single sugar sitting in the monosaccharide binding site in each of the subunits. Weak density was observed for the second sugar in some subunits and only methyl α -D-mannopyranoside was included in the model at this stage. A number of changes in the protein structure were made manually at this stage using 'O' (Jones *et al.*, 1991) (summarised in Table 3.5). The metal ions were included in the model with zero electrostatic charge and the sugar molecules included when the F_o-F_c map showed density stronger than 3σ for all atoms in the sugar residue. Further refinement proceeded smoothly by alternating cycles of automated CNS (Brunger and Adams, 1998) (restrained positional and temperature factor) refinement and manual intervention using 'O'. The protein was refined with the Engh & Huber (Engh and Huber, 1991) stereochemical parameter dictionary. A bulk solvent and anisotropic temperature factor correction are calculated automatically during CNS refinement. In the final round of refinement, the parameters for the bulk solvent density were $0.36 \text{ e}\text{\AA}^{-3}$, solvent radius 0.67\AA and B factor 39.4\AA^2 . The initial anisotropic B factor correction applied to the data had parameters $B_{11} = 2.30$, $B_{22} = -1.97$, $B_{33} = -0.33$. The non-diagonal terms were 0. Non-crystallographic symmetry (NCS) restraints were applied throughout for both positional and temperature factor refinement (Table 3.6). All individual temperature factors were reset to the overall average after each manual intervention. Apart from the 10% of measured data excluded to monitor refinement no cut-offs were applied to the

data. Electron density maps were calculated using SIGMAA (Read, 1986) modified coefficients with all data to 2.75\AA . Water molecules were added to the model in batches if they satisfied four criteria: they corresponded to a peak $> 3.5\sigma$ in the F_o-F_c map; they made hydrogen bonds with reasonable stereochemistry; they reappeared in at least 1σ in subsequently calculated $2F_o-F_c$ maps and that a drop in the R_{free} was observed.

Table 3.5. Progress of refinement of the con A - methyl α 1-2 mannoside complex.

Refinement	R factor/ R _{free}
CNS: Rigid body refinement of tetramer (30 cycles) followed by each monomer (11 cycles).	28.0/ 28.2
O: Two Mn ²⁺ ions, two Ca ²⁺ ions, 19 waters and 4 methyl α -D-mannopyranoside molecules added. ~40 residues modelled into electron density.	
CNS: 80 Cycles of positional refinement, 75 cycles of temperature factor refinement.	21.7/ 26.9
O: Second sugar included in subunits A and D (minus the methyl group (C6) in subunit D). 4 Waters deleted, 46 waters added. Residues 101, 135, 204 and 237 modelled into electron density.	
CNS: 80 Cycles of positional refinement, 50 cycles of temperature factor refinement.	21.6/ 26.1
O: Saccharide methyl group added in subunit D, 7 waters added.	
CNS: 40 Cycles of positional refinement, 40 cycles of temperature factor refinement.	21.6/ 26.1
O: 5 waters added. Residues 46, 59, 160, 161, 200 modelled into density.	
CNS: 100 cycles of positional refinement, 60 cycles of temperature factor refinement.	19.6/ 22.8

Table 3.6. NCS restraints used in positional and temperature factor refinement.

	Weight ^a	σ_{nCS} ^b
Group 1. β -sheet backbone. =	150	1.8
Group 2. β -sheet side chains.	80	2.8
Group 3. Non β -sheet backbone.	80	2.5
Group 4. Non β -sheet side chains.	60	3.5
Group 5. Sugars.	40	9.0

^a Energy constant for positional restraints/ kcal mol⁻¹ Å⁻².

^b Target deviation for b factor restraints/ Å².

3.4. Analysis of the final model.

The final structure contains 4 subunits of 237 residues (7236 atoms), 4 Ca²⁺ ions, 4 Mn²⁺ ions, 75 water molecules and 74 sugar atoms. Statistics on the final model are shown in Table 3.7. Poorly ordered regions occur at residues 118 to 122 and 204. These are consistent with poorly ordered regions found in previous con A complex structures (Naismith *et al.*, 1994; Naismith and Field, 1996; Moothoo and Naismith, 1998). In the final model, 202 atoms were stereochemically modelled and set to zero occupancy as they could not be located in the electron density. The Ramachandran plot for the final model is given in Figure 3.2. Four residues lie in the generously allowed region, these are residues Asp-203 from subunit A, and Ser-204 from subunits B, C and D. These regions are in poorly defined regions of the structure. The structure superimposes onto the native structure with an average r.m.s. deviation for C α atoms of 0.34Å² (0.21Å² for β -sheet

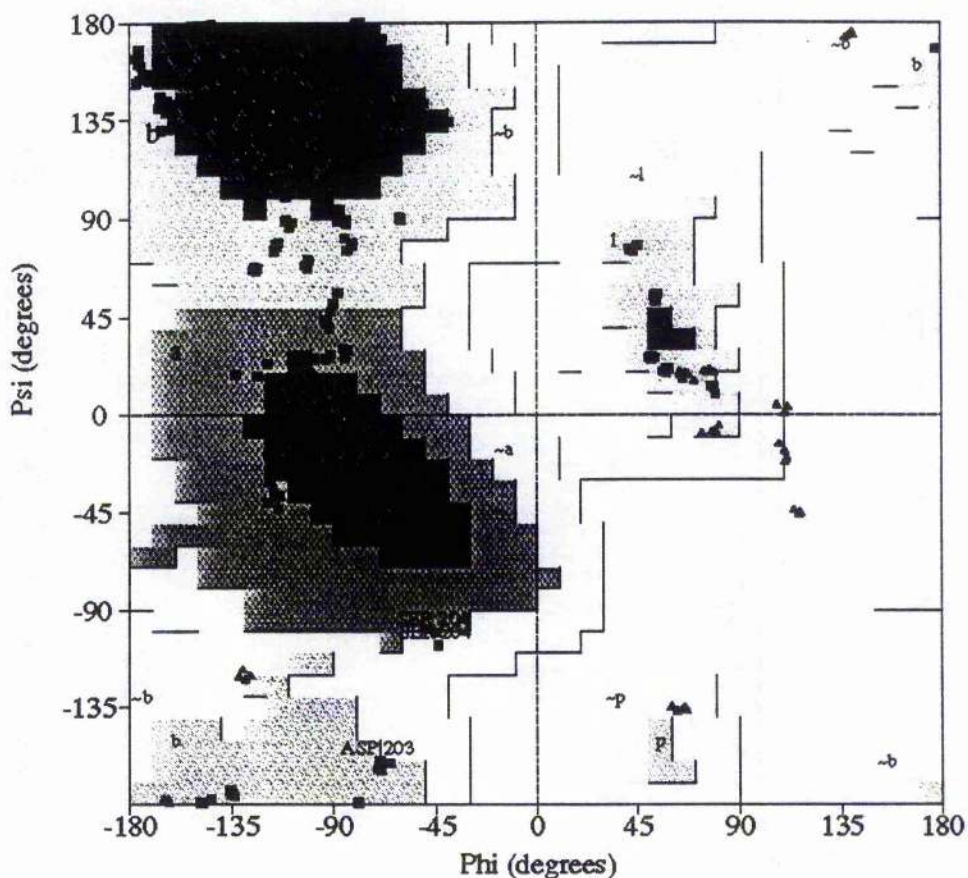
residues only) and onto the α 1-2 mannobiose structure with an r.m.s. of 0.19\AA^2 (0.12\AA^2 for β -sheet residues only).

Table 3.7. Quality of the final model of the con A - methyl α 1-2 mannobioside complex.

Resolution range (\AA)	∞ -2.7
R_{free} (%)	26.09
R factor (%)	21.51
Luzzati co-ordinate error (\AA) (5.0 - 2.75 \AA)	0.30
Bond r.m.s. deviation (\AA) ^a	0.015
Angle r.m.s. deviation ($^\circ$) ^a	1.913
Non-crystallographic symmetry r.m.s. deviation ($\text{C}\alpha$ atoms) (\AA)	0.04
B-factor bonded atoms r.m.s. deviation (\AA^2) ^b (main chain)	1.403
B-factor bonded atoms r.m.s. deviation (\AA^2) ^b (side chain)	1.833
Ramachandran core/additional/generously allowed (%) ^c	83.2/16.3/0.5
Protein mean B-factor (\AA^2) ^b all/ core (all atoms)	27.21/ 26.5
Protein mean B-factor (\AA^2) ^b all/ core (main chain)	27.02/ 26.4
Protein mean B-factor (\AA^2) ^b all/ core (side chain)	27.42/ 26.7
Sugar mean B-factor (\AA^2)	35.42
Solvent mean B (\AA^2)	23.63

^a r.m.s. deviation from Engh and Huber ideal values (Engh and Huber, 1991). ^b Calculated with MOLEMAN (G.J. Kleywegt, unpublished program). All stereochemically modelled atoms were removed prior to B-factor analysis, all bonded atoms including those in the sugars are included in the calculation of r.m.s. B-factor deviation for bonded atoms. ^c core and additionally allowed regions as defined by PROCHECK (Laskowski *et al.*, 1993).

Figure 3.2. Ramachandran plot for the con A - methyl α 1-2 mannoside complex.^a



^a See section 2.4 for explanation of the Ramachandran plot.

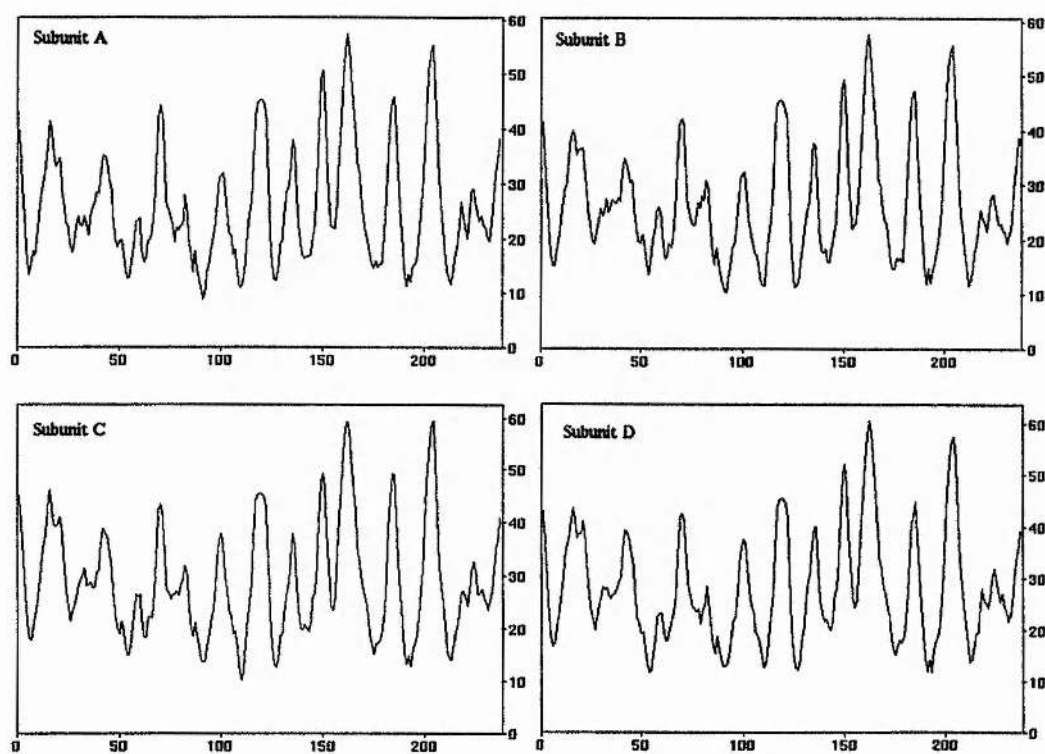
Temperature factors for the atoms in each of the subunits are shown in Table 3.8. Atoms which were stereochemically modelled were excluded from the calculations. The overall temperature factors for subunits C and D are higher than subunits A and B and this is not simply due to loop regions; the temperature factors for the core (β -sheet residues) are also higher in subunits C and D. Figure 3.5. shows the temperature factor plot along the backbone of the protein.

Table 3.8. Average temperature factors for each subunit.

Subunit	Main chain (Å)/ core ^a	Side chain (Å)/ core ^a	All atoms (Å)/ core ^a
A	25.6/ 20.2	26.2/ 20.4	25.8/ 20.3
B	26.2/ 20.8	26.5/ 21.0	26.3/ 20.9
C	28.5/ 22.7	28.9/ 23.0	28.7/ 22.9
D	27.9/ 22.2	28.1/ 22.3	28.0/ 22.3

^aTemperature factor calculated for β -sheet residues only, listed in Chapter 2.

Figure 3.3. Temperature factor plots along the backbone of each subunit. The y-axis represents the average temperature factor along the backbone of each residue, the x-axis represents residue number.



3.5 Results and discussion.

Dimer Formation. Formation of the AB dimer buries 1157\AA^2 of subunit A and 1148\AA^2 of subunit B surface accessible area. Formation of the CD dimer buries 1145\AA^2 of subunit C and 1152\AA^2 of subunit D surface accessible area. This is very similar with the

con A - α 1-2 manno- β -D-glucopyranoside - con A structure described in chapter 1. The interaction involves 16 hydrogen bonds and 138 van der Waals interactions between A and B and 16 hydrogen bonds and 139 van der Waals interactions between C and D.

Tetramer formation.

Formation of the tetramer buries 8953 Å² of surface accessible area comprising 2245 Å² of subunit A, 2242 Å² of subunit B, 2251 Å² of subunit C and 2215 Å² of subunit D. This is slightly less than seen in the con A - α 1-2 manno- β -D-glucopyranoside structure and much less than the native tetramer and supports the observation that binding of the sugar results in a more loosely packed tetramer. The tetramer is stabilised by interactions between subunits A and C (2 hydrogen bonds, 1 polar contact and 38 van der Waals interactions) and subunits B and D (1 hydrogen bond, 1 polar contact and 33 van der Waals interactions).

Metal sites.

Metal co-ordination is unchanged in this structure from the previous con A structures. Metal to ligand distances and ligand temperature factors are given in Table 3.9. This shows that Asp-19 and Tyr-12 ligands are above the average temperature factors. The average Mn²⁺ to Ca²⁺ distance is 4.2 Å.

Table 3.9. Metal to ligand distances and ligand temperature factors.

Ligand		Average distance/ Å	Average temperature factor ratio ^a
Mn ²⁺	Glu-8 OE2	2.18	0.99
	Asp-10 OD2	2.42	0.90
	Asp-19 OD1	2.24	1.4
	His-24 NE2	2.29	0.98
	OW	1.92	0.60
	OW	2.33	0.48
	Average	2.23	0.89
Ca ²⁺	Asp-10 OD1	2.56	0.88
	Asp-10 OD2	2.53	0.90
	Tyr-12 O	2.24	1.21
	Asn-14 OD1	2.55	0.94
	Asp-19 OD2	2.35	1.34
	OW	2.73	0.71
	OW	2.29	0.69
	Average	2.46	1.01

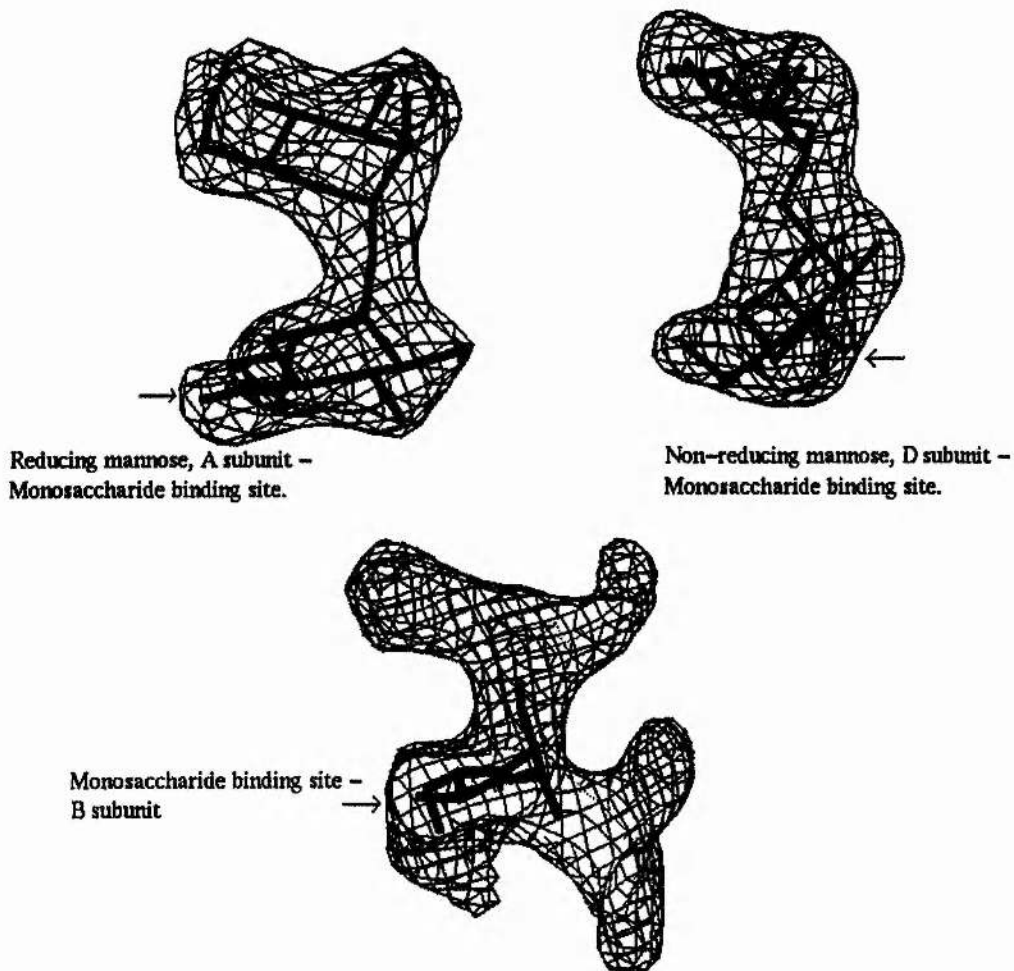
^atemperature factor/ average temperature factor of β -sheet residues (defined in Chapter 2).

Sugar Binding.

The structure shows well defined difference density for both sugars of the disaccharide in two of the four subunits denoted A and D (Figure 3.4). In these two subunits the mode of binding is clearly different (Figure 3.5). It was not possible to unambiguously model a second sugar into residual electron density in the remaining two subunits (B and C). In subunit B difference density is visible for a second sugar sitting in an extended binding

region framed by residues 12-14, 98-100, 207-208 and 227-228 as seen in subunit A. Surprisingly however, electron density is **also** visible for the binding region framed by residues Gly-98, Ser-168 and the loop Thr-226 to Leu-229 as seen in subunit D. In subunit C no density for a second saccharide is seen.

Figure 3.4. F_o-F_c electron density observed in the binding sites. Subunits A and D are shown above, (contoured at 3σ) and subunit B is shown below (contoured at 1σ). Maps were calculated after CNS simulated annealing with all sugars omitted.



Subunit A: Figure 3.5 shows a schematic diagram of the hydrogen bonding network between the sugar and the protein. The non-reducing sugar is recognised by the

monosaccharide binding site via a combination of hydrogen bonds, a polar contact and van der Waals interactions (Tables 3.10 and 3.11). The interaction is similar to that observed for methyl α -D-mannopyranoside.

Figure 3.5. Schematic diagram of the hydrogen bonds in the binding site of subunit A.

The sugar ring bound at the monosaccharide binding site is shown in red.

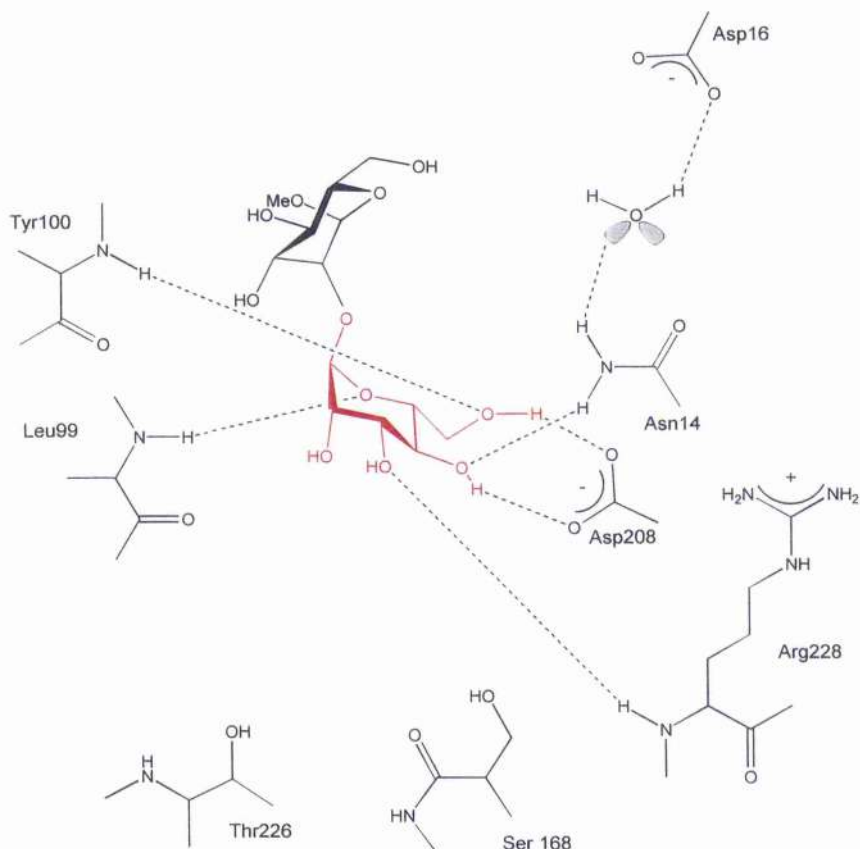


Table 3.10. Hydrogen bonding and polar contacts between methyl α 1-2 mannoside and con A.

Sugar	Protein	Distance ^a
monosaccharide binding sugar ^b		
O3	Arg-228 N	3.1
O4	Asn-14 ND2	3.0
O4	Asp-208 OD1	2.7
O4	Arg-228 N	3.2 ^c
O5	Leu-99 N	3.2
O6	Gly-98 N	3.3 ^c
O6	Leu-99 N	3.0 ^c
O6	Tyr-100 N	3.1
O6	Asp-208 OD2	3.0
Second sugar		
O6 (A subunit)	Tyr-12 OH	3.0 ^c
O3 (D subunit)	Thr-226 OG1	3.3
O4 (D subunit)	Ser-168 OG	3.3 ^c

^a Averaged across the four subunits, typically $\pm 0.1\text{\AA}$. ^b In subunit A this is the non-reducing sugar, in subunit D this is the reducing sugar. ^c Distance is less than 3.5\AA but the donor-H-acceptor angle deviates substantially from the linearity expected for a hydrogen bond.

Table 3.11. Van der Waals interaction ($< 4.0\text{\AA}$) between methyl α 1-2 mannoside and con A.

Sugar	Contacts	Residue
Primary binding sugar ^{a, b}	= 42	Tyr-12, Asn-14, Gly-98, Leu-99, Gly-207, Asp-208, Gly-277, Arg-228
Secondary sugar (A subunit)	15	Tyr-12, Leu-99, Tyr-100
Secondary sugar (D subunit)	10	Gly-98, Ser-168, Thr-226

^a Averaged across the four subunits. ^b In subunit A this is the non-reducing sugar, in subunit D this is the reducing sugar.

There is a shift in the position of the sugar ring of 0.3\AA out of the monosaccharide binding site with respect to the methyl α -D-mannopyranoside complex. Such a distortion has been seen in the con A - pentasaccharide complex discussed in Chapter 4 (Moothoo and Naismith, 1998a). The reducing sugar is recognised via 1 polar contact and 15 van der Waals interactions. The van der Waals interaction is considerably more extensive than that observed for the reducing sugar in the α 1-2 mannoside complex described in Chapter 2. This is due to a rotation ($\sim 50^\circ$) in the ψ dihedral angle of the inter-sugar glycosidic linkage (Table 3.12) with respect to the solution minima which was (observed in the α 1-2 mannoside crystal structure (Chapter 2). This results in the sugar ring being rotated about a pseudo six - fold axis through the sugar ring (Figure 3.6) placing the anomeric methyl group in a hydrophobic pocket formed by Tyr-12, Leu-99 and Tyr-100. This maximises its interaction with these residues and also maximises the interaction of the C6 atom of the sugar with the protein. This conformation lies 1 kcal mol^{-1} above the

energy minima (Imberty *et al.*, 1991). A distortion of similar size in a protein bound oligosaccharide was also observed in the pentasaccharide - con A complex which placed the glycosidic linkage ~ 3 kcalmol⁻¹ above the global minimum (Moothoo and Naismith, 1998). In this case a distortion in the interaction between the sugar and the monosaccharide binding site was also observed.

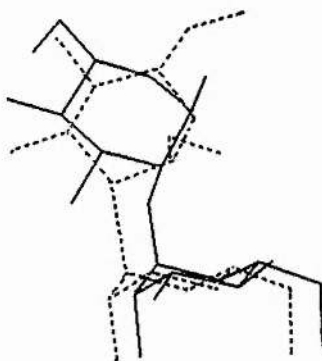
Table 3.12. Dihedral angles around the inter-sugar glycosidic linkage of the mannoside sugars.

	ϕ^a	ψ^a
Methyl $\alpha 1-2$ mannoside: A subunit	78.3	-91.9
Methyl $\alpha 1-2$ mannoside: D subunit	69.6	-146.6
$\alpha 1-2$ mannoside ^b	64.6	-140.8

^a See Figure 1.2.

^b from 1.75Å con A - $\alpha 1-2$ mannoside structure (Chapter 2).

Figure 3.6. Methyl $\alpha 1-2$ mannoside and $\alpha 1-2$ mannoside. The two disaccharide have been superimposed via the C α atoms of the protein.



The structurally conserved water molecule observed in all of the con A structures is observed here making contacts with the protein and the sugar. This water, however, no

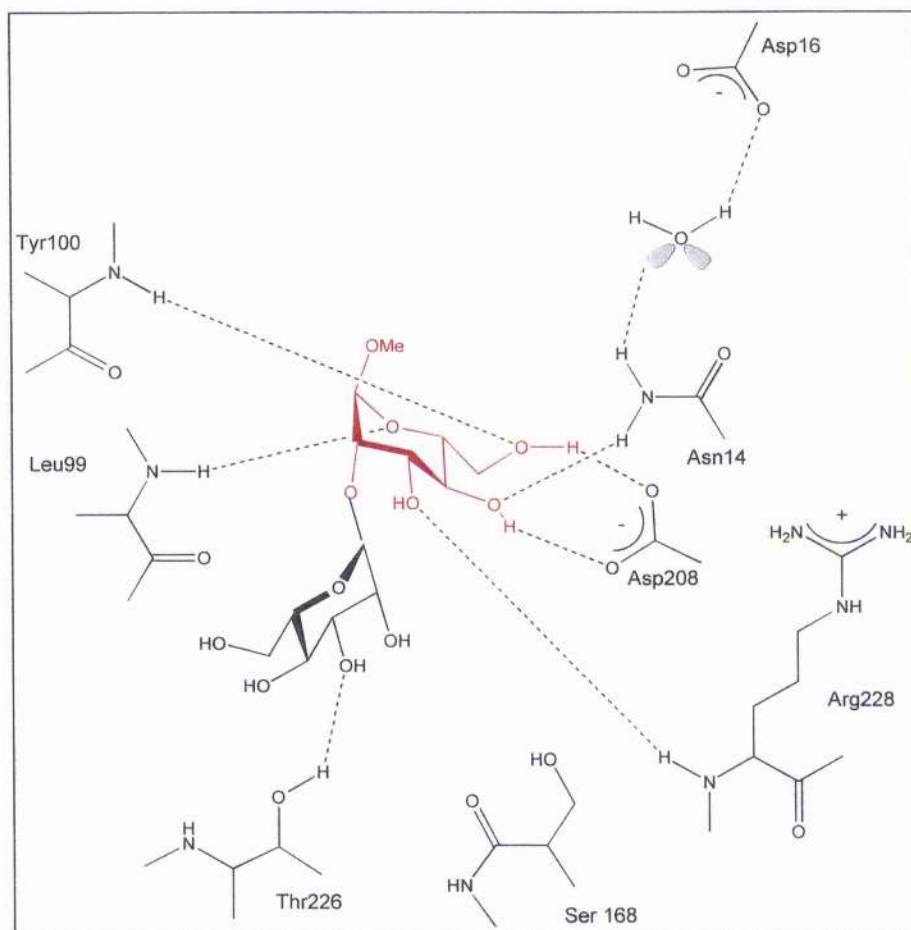
longer makes a hydrogen bond with the sugar as seen in the α 1-2 mannoside complex, but it does make van der Waals contact with C6 of the reducing terminal sugar. The hydrogen bond between Arg-228 NH₂ and the reducing sugar seen in the α 1-2 mannoside structure is also lost as a result of the rotation of the ψ angle.

Methyl α 1-2 mannoside buries a total of 61 Å² of additional surface area over α 1-2 mannoside upon binding to con A. Significantly, this increase is mainly due to burying of apolar surface accessible area; the anomeric methyl group and C6 of the sugar and Tyr-12 and Asp-16 of the protein. Burying apolar surface area is energetically favourable and therefore contributes favourably to the binding affinity.

Subunit D: In this subunit the disaccharide is recognised by an alternative binding site to that seen in the A subunit (Figure 3.7). It is the reducing terminal mannose residue which is recognised by the monosaccharide binding site via hydrogen bonds, van der Waals interactions and a polar contact (Tables 3.10, 3.11).

Figure 3.7. Schematic diagram of the hydrogen bonds in the binding site of subunit D.

The sugar bound at the monosaccharide binding site is shown in red.



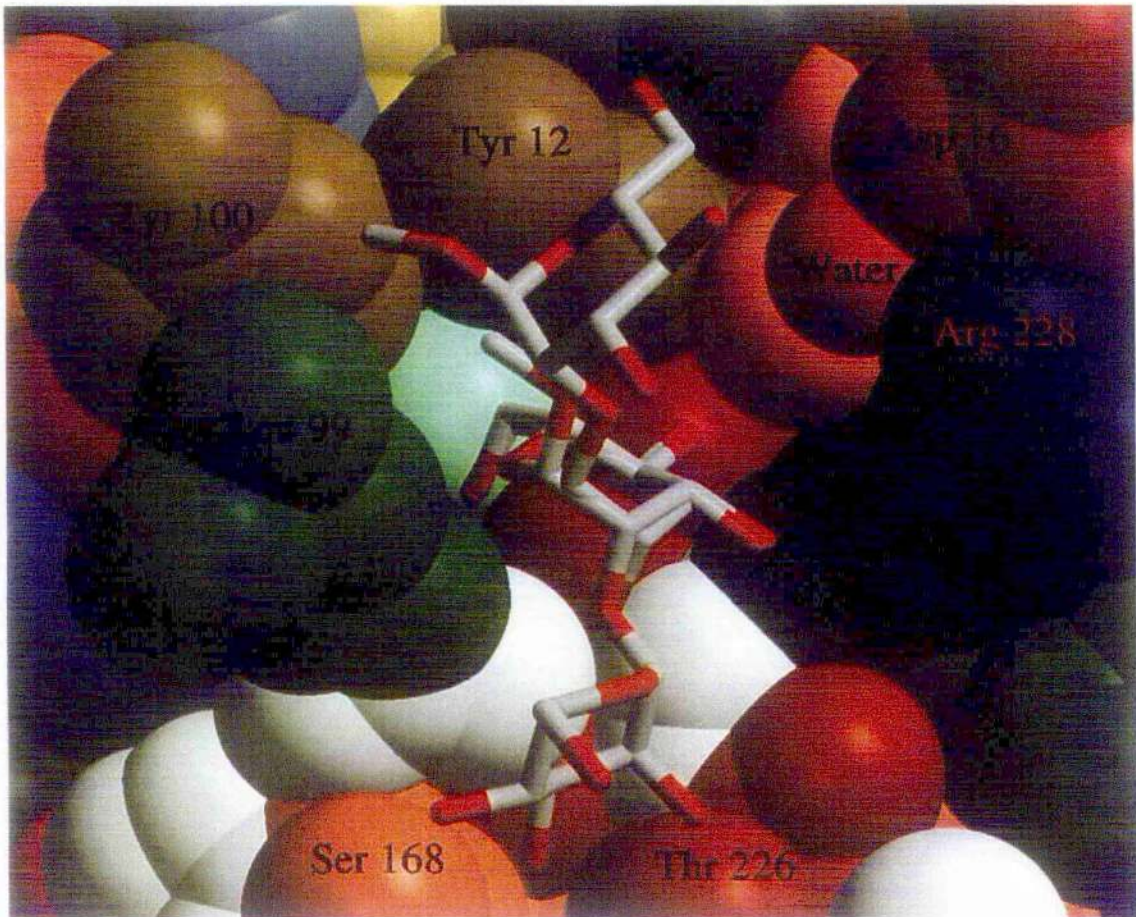
The non-reducing terminal sugar is recognised by the site occupied by the 1-6 arm GlcNAc in the pentasaccharide bound structure (Moothoo and Naismith, 1998) and interacts with Gly-98, Ser-168 and Thr-226. The binding mode seen in subunit A is unavailable in this subunit due to crystal contacts and likewise this mode would be prohibited by crystal contacts in subunit A. The φ and ψ inter-glycosidic linkage angles are close to those observed in the α 1-2 mannoside structure which lies in an energetic minimum (Imberty *et al.*, 1991). Although the non-reducing terminal sugar of methyl α 1-2 mannoside occupies the binding site of the 1-6 arm GlcNAc of the pentasaccharide,

the rings of the two sugars are in different positions. The number of interactions made by the non-reducing terminal sugar with the protein are more than halved in this complex compared with the GlcNAc sugar. This is due to this different position of the sugar rings and the absence of the *N*-acetyl group. A total of 307\AA^2 of polar surface accessible area is buried upon sugar binding, and 205\AA^2 of apolar surface area. This is 11\AA^2 less buried apolar surface area than subunit A. The side chain of Asp-16 is markedly different than the other subunits with its chi1 angle rotated about 90° and there is a 10° rotation in the chi-1 angle of Thr-226 which is involved in carbohydrate binding in this subunit.

Subunit B: There is difference density for the second sugar in two different places, analogous to the binding modes seen in subunits A and D (Figure 3.4). As a crystal structure is a static average of many molecules, some proportion of molecules apparently bind the saccharide in a manner analogous to the seen in subunit A, the remainder analogous to subunit D.

The structure of the methyl α 1-2 mannoside - con A complex (Chapter 2) presented here demonstrates that a carbohydrate can be bound specifically in two distinct environments by the same lectin binding site (Figure 3.8).

Figure 3.8. The binding site of con A displaying the two distinct binding modes displayed by the lectin.



In this complex one binding mode involves the secondary sugar making extensive contacts that are particularly 'hydrophobic' in nature. However, in order to maximise these interactions a substantial energy penalty is paid in conformational strain and destabilisation of the primary site of attachment. In the other mode, the secondary sugar avoids the distortion penalty but seems not to harness the full binding potential in terms of hydrogen bonds and van der Waals contacts with the protein. The obvious question arises, why is this phenomenon not observed in the α 1-2 mannobiose complex? The crystal packing arrangement in that complex would preclude the binding arrangement seen in subunit D in this complex. Thus, either methyl glycoside formation results in

equalisation of the energy of interaction at both sites or crystallisation kinetically selects one mode in preference to the other. Modelling protocols are not sufficiently advanced enough to distinguish between these two possibilities.

Statistical thermodynamics gives the relative occupancy for the two freely inter-converting but mutually exclusive states as $1:e^{-\Delta\Delta G^\circ/RT}$, where $\Delta\Delta G^\circ$ is the difference in free energy between the states. In our case we define $\Delta\Delta G^\circ = \Delta G^\circ_{\text{weak}} - \Delta G^\circ_{\text{strong}}$ where $\Delta G^\circ_{\text{strong}}$ is the intrinsic free energy of binding of the higher affinity mode and $\Delta G^\circ_{\text{weak}}$ is the intrinsic free energy of binding of the lower affinity mode. This results in the ratio of binding modes ($[\text{weak mode}] / [\text{strong mode}]$) as given in equation (a).

$$[\text{weak mode}] / [\text{strong mode}] = e^{-(\Delta G^\circ_{\text{weak}} - \Delta G^\circ_{\text{strong}})/RT} \quad (\text{a})$$

When two binding modes are exclusive, the measured equilibrium constant (K_{meas}) is

$$K_{\text{meas}} = [\text{complex}] / [\text{protein}][\text{sugar}] \quad (\text{b})$$

$$[\text{complex}] = [\text{strong mode}] + [\text{weak mode}] \quad (\text{c})$$

$$K_{\text{meas}} = ([\text{strong mode}] + [\text{weak mode}]) / [\text{protein}][\text{sugar}] \quad (\text{d})$$

Equation (d) can be rearranged to:

$$K_{\text{meas}} = [\text{strong mode}] (1 + ([\text{weak mode}] / [\text{strong mode}])) / [\text{protein}][\text{sugar}] \quad (\text{e})$$

If we then substitute equation (a) into equation (e) we obtain equation (f).

$$K_{\text{meas}} = [\text{strong mode}] (1 + e^{-(\Delta G^\circ_{\text{weak}} - \Delta G^\circ_{\text{strong}})/RT}) / [\text{protein}][\text{sugar}] \quad (\text{f})$$

Taking natural logarithms of both sides and multiplying by $-RT$ gives:

$$-RT \ln K_{\text{meas}} = -RT \ln([\text{strong mode}] (1 + e^{-(\Delta G^\circ_{\text{weak}} - \Delta G^\circ_{\text{strong}})/RT}) / [\text{protein}][\text{sugar}]) \quad (\text{g})$$

Utilising the property of logarithms that $\ln xy = \ln x + \ln y$:

$$-RT \ln K_{\text{meas}} = -RT \ln([\text{strong mode}] / [\text{protein}][\text{sugar}]) - RT \ln(1 + e^{-(\Delta G^\circ_{\text{weak}} - \Delta G^\circ_{\text{strong}})/RT}) \quad (\text{h})$$

$$\frac{[\text{strong mode}]}{[\text{protein}][\text{sugar}]} = K_{\text{strong}} \text{ (equilibrium constant for strong binding mode)} \quad (i)$$

As $\Delta G^\circ = -RT \ln K$, equation (i) reduces to

$$\Delta G^\circ_{\text{meas}} = \Delta G^\circ_{\text{strong}} - RT \ln(1 + e^{-\Delta \Delta G^\circ / RT}) \quad (j)$$

In the simplest case where the sites are equivalent $\Delta \Delta G^\circ = 0$, $\Delta G^\circ_{\text{meas}} = \Delta G^\circ_{\text{strong}} - RT \ln 2$, the measured ΔG° contains a statistical enhancement of $-0.41 \text{ kcal mol}^{-1}$ relative to the intrinsic binding free energy of the site. At the other extreme where $\Delta \Delta G^\circ$ is very large $\Delta G^\circ_{\text{meas}} = \Delta G^\circ_{\text{strong}}$, effectively collapsing to the simple single site case. The electron density maps we observe are of insufficient quality to accurately experimentally calculate the ratio between the binding modes. However in subunit C, the residual electron density for a second sugar ring appears by inspection to be approximately as strong in both locations. Accepting the error inherent in this imprecise approach may suggest an upper limit of 2:1 for the ratio of occupancy between the two modes. Using this value as a limit, equation (a) yields a value of $\Delta \Delta G^\circ$ (maximum difference in free energy of binding between the two modes at 25°C) of $0.46 \text{ kcal mol}^{-1}$. This value of $\Delta \Delta G^\circ$ gives an enhancement to the intrinsic binding energy of the strong site (equation (j)) of $0.24 \text{ kcal mol}^{-1}$. As $\Delta G^\circ_{\text{meas}}$ is $-7.0 \text{ kcal mol}^{-1}$, we estimate that $\Delta G^\circ_{\text{strong}}$ lies between $-6.59 \text{ kcal mol}^{-1}$ (both sites equal) and $-6.76 \text{ kcal mol}^{-1}$ (maximal suggested free energy difference between the sites). and $\Delta G^\circ_{\text{weak}}$ lies between $-6.59 \text{ kcal mol}^{-1}$ (equal sites) and $-6.3 \text{ kcal mol}^{-1}$. This analysis shows that both modes bind more tightly than methyl α -D-mannopyranoside by at least $1.0 \text{ kcal mol}^{-1}$, although this study cannot identify which is the strong site and which is the weak site.

3.6. Conclusions

This Chapter is a striking illustration of how competing forces are balanced in protein carbohydrate interactions. In one binding mode the secondary sugar makes extensive contacts that are particularly hydrophobic in nature. However in order to maximise these interactions a substantial energy penalty (at least 1 kcal mol⁻¹) is paid in conformational strain and destabilisation of the primary site of attachment. In the other mode, the second sugar avoids the distortion penalty but seems not to harness the full binding potential in terms of hydrogen bonds and van der Waals contacts with the protein. The analysis suggests that both binding modes differ by less than 0.46 kcal mol⁻¹. The extra interactions contribute between 0.54 kcal mol⁻¹ and 1.46 kcal mol⁻¹ to the binding energy (relative to the undistorted mode), to compensate for the distortion. This study has not only graphically highlighted the competing forces involved in carbohydrate recognition but has been able to quantify the forces. The work discussed here provides an exciting opportunity for testing and improving theoretical models of carbohydrate recognition. These results powerfully underline the synergy and necessity in combining crystallography, chemical synthesis, modelling and calorimetry to probe protein carbohydrate recognition.

3.7. Future Work.

As well as improving the resolution of this structure, extending the work from this chapter is linked with continuing work from Chapter 2, obtaining the crystal structure of the α 1-3 and α 1-6 mannobioses and in addition their methylated counterparts.

Chapter 4

2.7Å structure of concanavalin A complexed with a pentasaccharide.

4.1. Summary

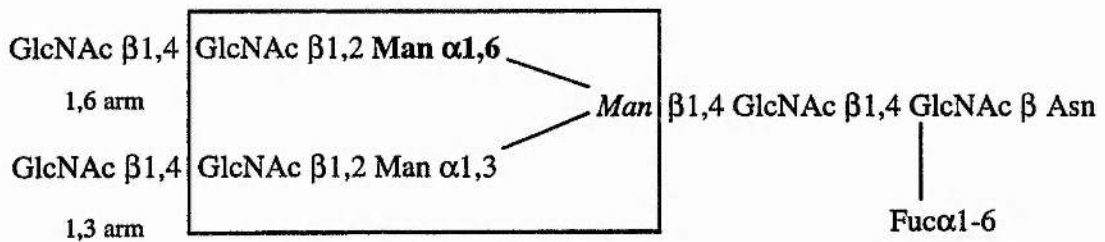
The structure of con A bound to the pentasaccharide GlcNAc β 1-2 Man α 1-6 (GlcNAc β 1-2 Man α 1-3) Man has been determined and refined to 2.7Å. Crystals of the complex form in space group C2 with cell dimensions $a=176.5\text{Å}$, $b=122.8\text{Å}$, $c=124.6\text{Å}$, $\beta=134.2^\circ$. Two tetramers are present in the asymmetric unit. Each of the eight subunits binds one pentasaccharide molecule, one manganese and one calcium ion. In six of eight subunits there is clear density for all five sugar residues and a well ordered binding site. The pentasaccharide adopts the same conformation in all eight subunits. The binding site is a continuous extended cleft on the surface of the protein. Van der Waals interactions and hydrogen bonds anchor the carbohydrate to the protein. Both GlcNAc residues contact the protein. The GlcNAc on the 1-6 arm of the pentasaccharide makes particularly extensive contacts, including two hydrogen bonds. The binding site of the 1-3 arm GlcNAc is much less extensive. Oligosaccharide recognition by con A occurs through specific protein carbohydrate interactions and does not require recruitment of adventitious water molecules. The GlcNAc β 1-2 Man glycosidic linkage ψ torsion angle on the 1-6 arm is rotated by over 50° from that observed in solution. This rotation is coupled to disruption of interactions at the monosaccharide site and may be responsible for reducing the free energy liberated by additional interactions at the 1-6 arm GlcNAc site.

4.2. Introduction.

Modes of oligosaccharide binding by proteins were described in Chapter 1. Among the complex structures available the interaction with high order oligosaccharides are often highly water mediated, with rearrangement of the water molecules on binding playing an

important role in the binding affinity, as seen in complexes of LOL I (Bourne *et al.*, 1992), GS4 (Delbaere *et al.*, 1993), Peanut Lectin (Banerjee *et al.*, 1996) and EcorL (Elgavish and Shaanan, 1998) or occurring through a single saccharide residue until the fucose moiety of the *N*-linked glycan (Figure 4.1.) is present as seen in LOL II and pea Lectin complexes (Rini *et al.*, 1993).

Figure 4.1. The core oligosaccharide of the *N*-linked glycan found on mammalian cell surfaces. Shown boxed is the pentasaccharide used in this study.



Con A does not require the fucose moiety of the *N*-linked carbohydrate or recruitment of water to the binding site for high affinity binding and possesses an extended binding site for interaction. This was first seen experimentally in the structure of con A complexed with the trimannoside to 2.3Å resolution (Naismith and Field, 1996) and has been described in Chapter 1. The trimannoside adopts a favourable conformation and each sugar ring interacts with the protein via hydrogen bonds, polar contacts and van der Waals interactions. This structure provided a rationale for the high affinity displayed by con A for the trimannoside core of the *N*-linked glycan (Table 4.1).

Table 4.1. Thermodynamic data for con A binding to oligosaccharide at the core of the N-linked glycan.

	$-\Delta H^\circ$	$-T\Delta S^\circ$	$-\Delta G^\circ$	K_a
	kcal mol ⁻¹	kcal mol ⁻¹	± 0.1	/M ⁻¹
Me- α -Man	8.2	2.9	5.3	0.82 (± 0.02) x 10 ⁴
Trimannoside	14.1	6.6	7.5	3.37 (± 0.14) x 10 ⁵
Pentasaccharide	10.6	2.2	8.4	1.40 (± 0.10) x 10 ⁶

The pentasaccharide shown boxed in Figure 4.1 binds to con A with a K_a of $1.41 \times 10^6 \text{ M}^{-1}$ (Table 4.1) and is the tightest carbohydrate con A complex (Mandal *et al.*, 1994b). It binds with a three fold higher affinity than the trimannoside although it possesses a ΔH° of only $-10.6 \text{ kcal mol}^{-1}$. The slightly increased enthalpy contribution to binding the trimannoside and pentasaccharide over the monosaccharide indicates extended binding site interactions, seen in the con A - trimannoside complex. It is somewhat surprising then, to observe a less favourable enthalpy for the pentasaccharide binding than the trimannoside binding. Thus the calorimetry data suggests that the pentasaccharide and trimannoside possess different modes of binding, or that they bind in the same way with similar interactions but there is some factor reducing the overall enthalpy change in binding the pentasaccharide. The latter is more probable as further data, obtained by NMRD experiments (Bhattacharyya *et al.*, 1985) have suggested similar mechanisms of binding the two oligosaccharides.

Thus in order to delineate the contribution towards enthalpic and entropic terms, the structure of con A bound to the pentasaccharide is described in this chapter and the

details of the structure which may contribute to the overall free energy of binding are discussed.

4.3. Experimental.

Crystallisation and Data Collection.

Co-crystals of the con A - pentasaccharide complex were obtained after an initial screen consisting of 10, 15, 20 and 25% Polyethylene Glycol (PEG) 6K, pH 4.0, 5.0, 6.0 and 7.0. Optimisation of pH and PEG concentration then followed (Moothoo and Naismith, 1998b). Crystals were obtained from a 10 μ l solution containing 0.6mM con A (Sigma, Poole, UK), 18mM and GlcNAc β 1-2 Man α 1-6 (GlcNAc β 1-2 Man α 1-3) Man (Dextra Laboratories, Reading, UK) 1mM MnCl₂, 1mM CaCl₂, 20mM TRIS pH 7.0 and 0.1M NaCl were mixed with 10 μ l of a reservoir which containing 15% PEG 6K pH 6.0 in a sitting drop tray (Charles Supper Co.) and left to equilibrate. Crystals were obtained as hollow rods varying in size from 0.2 x 0.1 x 0.1mm³ to 0.8 x 0.4 x 0.2mm³.

All data were collected at room temperature from crystals mounted in glass capillaries using the Enraf-Nonius/MacScience DIP2000 dual image plate. Seven crystals were examined ranging in size from 0.2 x 0.2 x 0.4mm³ to 0.8 x 0.4 x 0.2mm³. The average resolution limit for the crystals examined was 3.5Å. The largest crystal diffracted to a resolution of 2.7Å. From this crystal all data were recorded as 200 non-overlapping 20 minute 1.0° oscillations with a crystal to detector distance of 150mm. The data were processed and merged with DENZO and SCALEPACK (Otwinowski, 1993). Autoindexing with DENZO suggested five different lattice types: face centred orthorhombic, body centred orthorhombic, primitive cubic, primitive rhombohedral and c centred monoclinic. This result appeared consistent between crystals and independent of

the orientation of the crystal with respect to the X-ray beam. A stable unit cell matrix could not be obtained in cubic or rhombohedral lattices. The data were integrated and merged in three lattices. Body centred orthorhombic scaled with an R_{merge} of 47.7%, face centred orthorhombic scaled with an R_{merge} of 33.4% and c centred monoclinic with an R_{merge} of 9.4%. The space group was assigned as C2 with unit cell parameters $a=176.5\text{\AA}$, $b=122.8\text{\AA}$, $c=124.6\text{\AA}$, $\beta=134.2^\circ$, with 0.6% of observations rejected. The asymmetric unit contains two tetramers; Matthew's number $2.5\text{ Da}\text{\AA}^{-3}$, approximately 46% solvent. The quality of the data is shown in Table 4.2

Table 4.2. Data quality for the 2.7Å complex of the pentasaccharide - con A complex.

Resolution	No. Reflections	% Complete	R-Merge/ %	Redundancy
25.00-5.8	5208	97.7	4.6	4.0
5.80-4.61	5183	98.6	5.3	4.2
4.61-4.03	5121	98.1	6.3	4.3
4.03-3.66	5114	97.7	8.4	4.3
3.66-3.40	5059	97.2	10.2	4.3
3.40-3.20	5045	97.0	13.3	4.3
3.20-3.04	5066	96.8	16.6	4.3
3.04-2.91	5996	96.2	21.9	4.3
2.91-2.80	5004	96.0	27.4	4.3
2.80-2.70	4967	95.8	33.7	4.1
25.00-2.70	50763	97.1	9.4	4.2

Structure solution.

The structure was determined by the molecular replacement method as implemented in the CCP4 (CCP4, 1994) program AMORE (Navaza, 1994) using all data from 12Å to 3.4Å. The trimannoside con A complex tetramer (Naismith and Field, 1996) was used as the search model (with all metal ions, sugar molecules and water molecules removed and all atoms given full occupancy). A cross rotation function gave only two solutions, $\alpha = 295.68$, $\beta = 65.66$, $\gamma = 132.49$ (Solution 1) and $\alpha = 20.52$, $\beta = 40.16$ and $\gamma = 74.88$ (solution 2). These were 19 and 17 standard deviations above noise respectively. The translation search found one tetramer for each rotation solution with $x = 0.2350$, $y = 0.0000$, $z = 0.0139$ for solution 1 giving a correlation coefficient of 0.42 and an R-factor 45%. Solution 2 with $x = 0.2750$, $y = 0.0000$, $z = 0.0139$ gave a correlation coefficient of 0.33 and an R-factor of 49% (the y co-ordinate is arbitrary for one tetramer in C2). A cross translation search was used to position the two tetramers relative to each other ($\alpha = 284.35$, $\beta = 58.01$, $\gamma = 317.88$, $x = 0.7327$, $y = 0.9996$, $z = 0.5122$ with a correlation coefficient of 0.7 and an R-factor 34%).

Refinement.

The solutions were imported into X-PLOR (Brunger, 1992) and refinement carried out using R_{free} as a guide. R_{free} was calculated on 10% of data which were excluded from all refinement calculations in order to provide an unbiased assessment of the refinement. The raw molecular replacement solution had an initial R factor of 44% and an R_{free} of 44.6%. Rigid body refinement of the two tetramers lowered the R factor to 37.0% and the R_{free} to 36.4%. Rigid body refinement of all eight monomers dropped the R factor to 30.8% and the R_{free} to 30.9%. $F_o - F_c$ and $2 F_o - F_c$ electron density maps were calculated from this model. Strong density was observed for all metal ions and 32 of the 40 possible sugars, with weak density for 6 but ambiguous density for 2. Refinement proceeded smoothly by

alternating cycles of automated X-PLOR (Brunger, 1992) refinement and manual intervention using 'O'. The metal ions were included in the model with zero electrostatic charge and the sugar molecules included when the F_o-F_c map showed density stronger than 3σ for all atoms in the sugar residue. The protein was refined with the Engh & Huber stereochemical parameter dictionary (Engh and Huber, 1991). Non-crystallographic symmetry (NCS) related atoms were restrained to their average positions through positional and B-factor refinement (Table 4.3)

Table 4.3. NCS restraints used in positional and temperature factor refinement.

	weight ^a	σ_{nCS} ^b
Group 1. β -sheet backbone.	100	1.8
Group 2. β -sheet side chains.	60	2.8
Group 3. Non β -sheet backbone.	55	2.5
Group 4. Non β -sheet side chains.	45	3.5
Group 5. Sugars.	40	3.0

^a Energy constant for positional restraints/ kcal mol⁻¹ Å⁻².

^b Target deviation for b factor restraints/ Å².

Table 4.4 summarises the progress of the refinement. Despite the limited resolution restrained temperature factor refinement was carried out because of the presence of 8 fold non-crystallographic symmetry and a clear (>4%) improvement in R_{free} was observed. All individual temperature factors were reset to the overall average temperature factor after each manual intervention. Apart from the 10% of measured data excluded to monitor refinement no cut-offs were applied to the data. A bulk solvent correction was included in the X-PLOR refinement with parameters for solvent density $0.34 \text{ e}^{-\text{Å}^{-3}}$, solvent radius

0.25Å and B factor 50Å². Electron density maps were calculated using SIGMAA (Read, 1986) modified coefficients with all data to 2.7Å. Water molecules were added to the model in batches if they satisfied four criteria: they corresponded to a peak > 3.5σ in the F_o-F_c map; they made hydrogen bonds with reasonable stereochemistry; they reappeared in at least 1σ electron density in subsequently calculated 2F_o-F_c maps and that a drop in R_{free} was observed.

Table 4.4. Progress of refinement of the con A - pentasaccharide complex.

Refinement	R factor/ R _{free}
X-Plor: Rigid Body refinement on each tetramer (30 cycles)	37.0/ 36.9
Rigid Body refinement on each monomer (40 cycles)	30.8/ 30.8
O: Trimannoside, 16 metal ions and 32 waters added. C terminus and >80 residues modelled into density.	
X-Plor: Positional refinement x 30 cycles.	26.6/26.1
NCS Restraints applied, positional x 60 cycles.	
Temperature factor refinement x 50 cycles.	18.2/ 23.3
O: Added 10 out 12 possible terminal GlcNAcs, deleted 13 waters and 54 added. Residues 36, 132, 116, and 151 modelled into electron density.	
X-Plor: Positional refinement x 70 cycles.	18.0/ 22.1
Temperature factor refinement x 20 cycles.	
O: 1 water added, residue 135 modelled into electron density.	
X-Plor: Positional refinement x 10 cycles.	17.9/ 21.9
Temperature factor refinement x 10 cycles.	

4.4. Analysis of the final model.

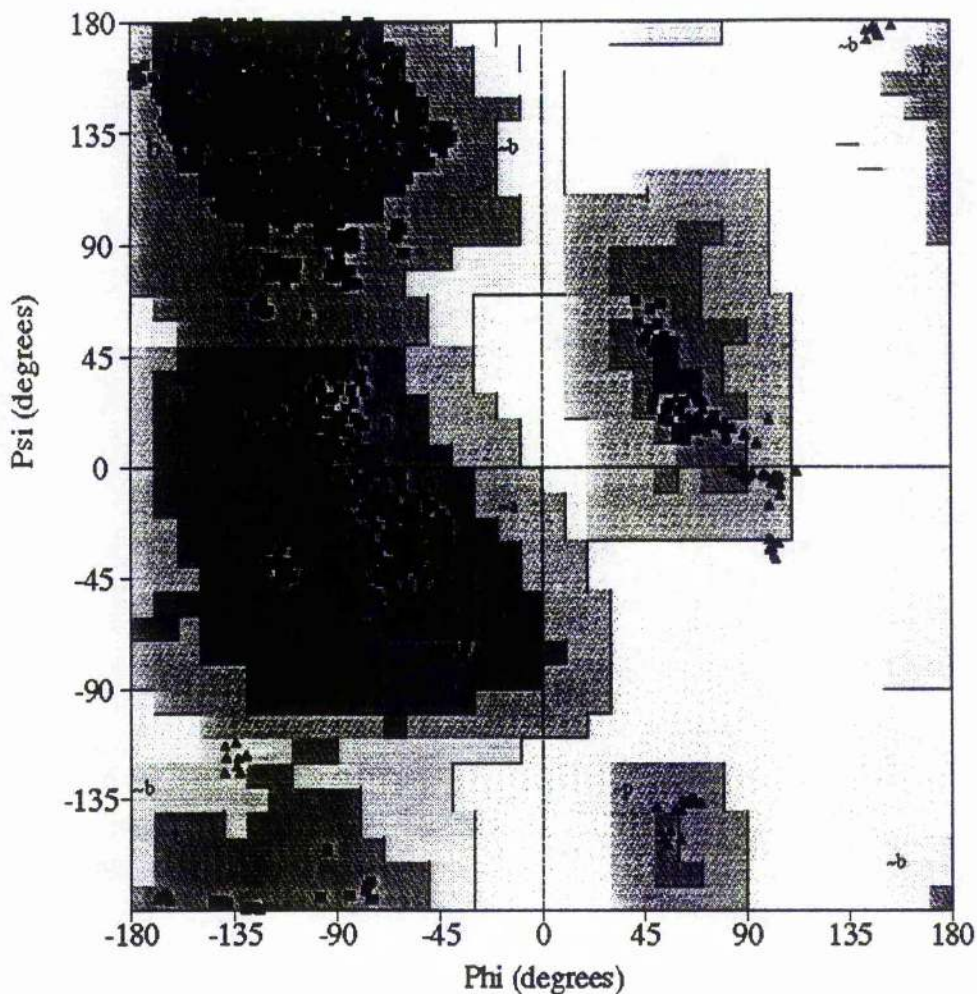
The final model contains 14472 protein atoms (eight subunits of 237 amino acid residues), 496 sugar atoms ($\overline{46}$ residues), 16 metal ions and 74 water molecules. Poorly ordered regions of the structure occur in loop regions at residues 118 to 122, 161, and 204. These regions have been found to be consistently poorly ordered in crystal structures of con A. Table 4.5 shows statistics on the final model. 370 atoms were stereo-chemically modelled and are set to zero occupancy in the final model as they could not be located in electron density. The peptide bond between Ala-207 and Asp-208 is in the unusual *cis* conformation in agreement with the previous con A structures. The Ramachandran plot for the final model is given in Figure 4.2.

Table 4.5. Final model statistics of the 2.7Å con A - pentasaccharide complex.

Resolution range (Å)	25-2.7
R _{free} (%)	21.9
R factor (%)	17.9
Bond r.m.s. deviation (Å) ^a	0.01
Angle r.m.s. deviation (°) ^a	1.746
Non-crystallographic symmetry r.m.s. deviation (Cα atoms) (Å)	0.12
B-factor bonded atoms r.m.s. deviation (Å ²) ^c	3.26
Ramachandran core/additional (%) ^b	85.3/14.7
Protein mean B-factor (Å ²) ^c	29.1
Sugar mean B-factor (Å ²) ^c	35.9
Solvent mean B (Å ²) ^c	22.5

^a r.m.s. deviation from Engh and Huber ideal values (Engh and Huber, 1991).^b Core and additionally allowed regions as defined by PROCHECK (Laskowski *et al.*, 1993). No residues are in the generously allowed or disallowed regions. ^c Calculated with MOLEMAN (G.J. Kleywegt, unpublished program). All stereochemically modelled atoms were removed prior to B-factor analysis, all bonded atoms including those in the sugars are included in the calculation of r.m.s. B-factor deviation for bonded atoms.

Figure 4.2. Ramachandran plot for the final model.^a



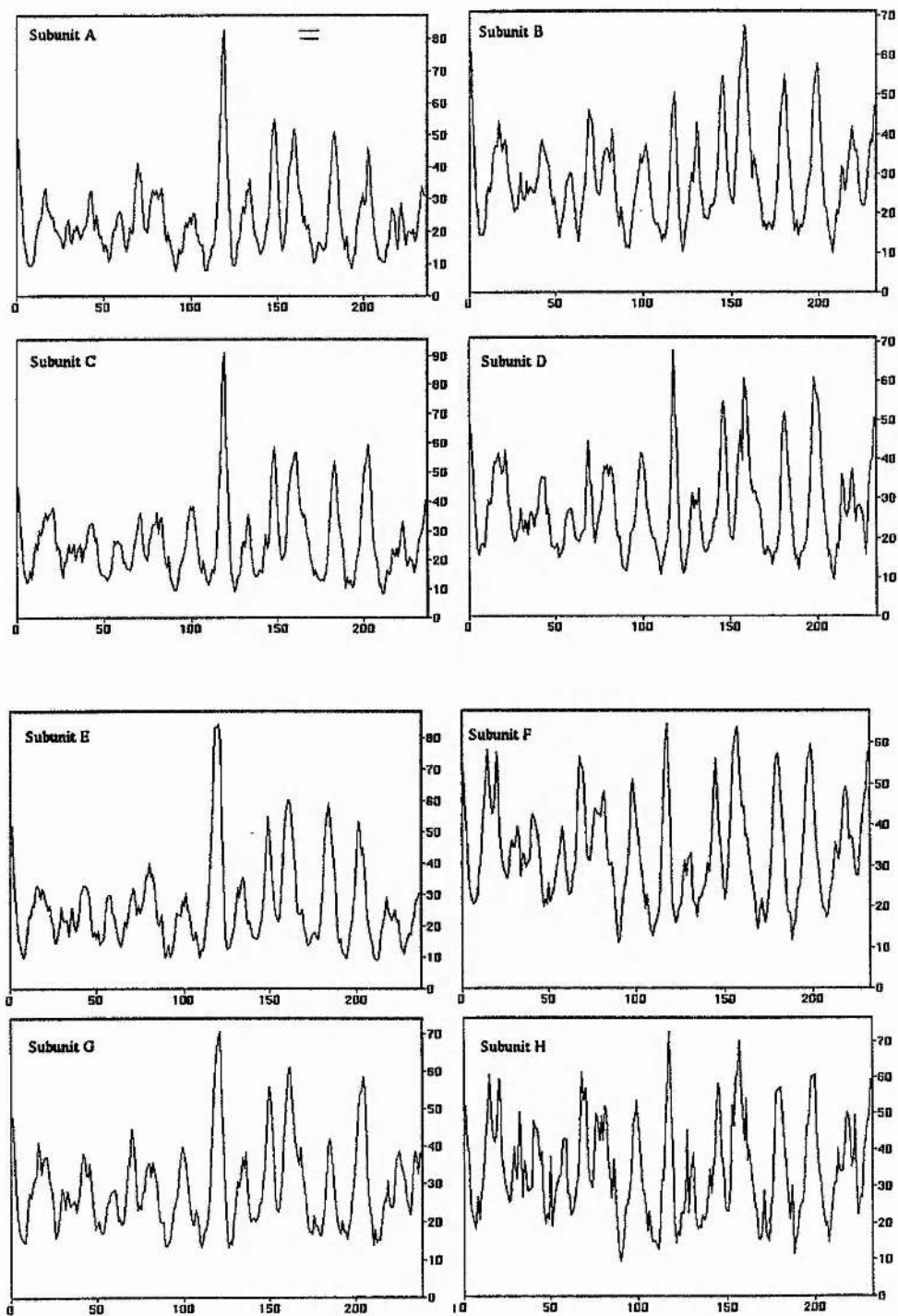
^a See section 2.4 for an explanation of the Ramachandran plot.

Temperature factors. The average temperature factors for each subunit of the final model are given in Table 4.6, and the variation of the temperature factor along the length of the subunits is shown in Figure 4.3. The average temperature factors are considerably higher for subunits F and H. The backbone plots for these two subunits show that the difference is mainly seen in the first eighty residues of these chains.

Table 4.6. Average temperature factors for each subunit.

Subunit	Main Chain	Side chain	All
A	23.5	24.6	24.0
B	28.7	30.1	29.4
C	26.2	27.0	26.6
D	27.5	27.9	27.7
E	25.8	26.8	26.2
F	35.1	35.7	35.4
G	28.8	29.8	29.3
H	33.3	35.5	34.3

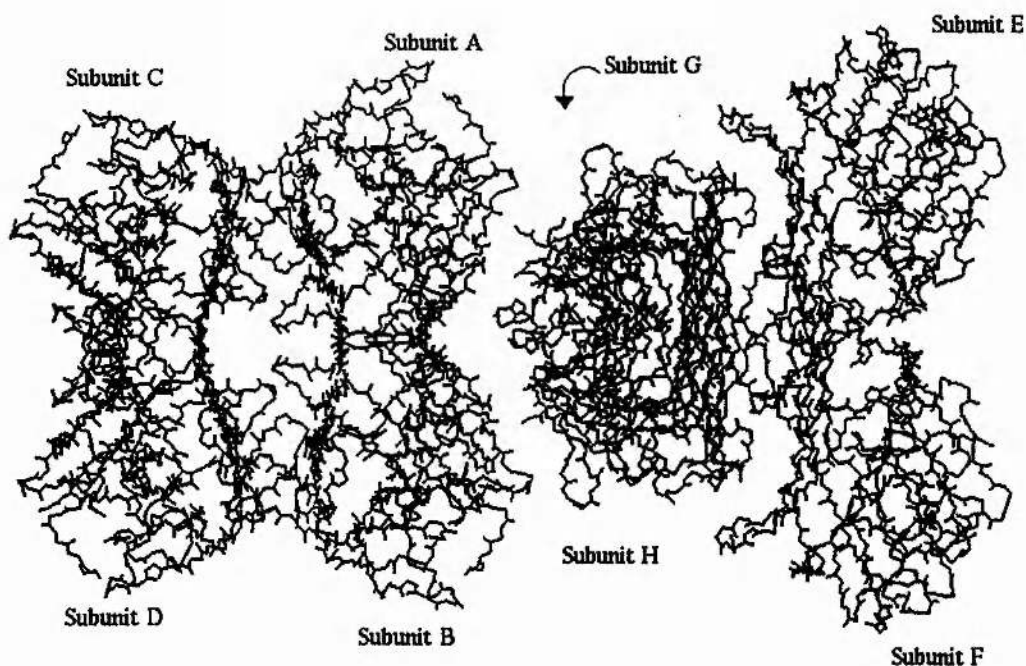
Figure 4.3. Temperature factor plot along the backbone of each subunit. The x-axis represents the average temperature factor for the backbone of each residue, the y-axis represents each residue.



4.5. Results and discussion.

Figure 4.4 shows the packing of the tetramers in the asymmetric unit of the con A - pentasaccharide complex.

Figure 4.4. Packing of the two tetramers found in the asymmetric unit.



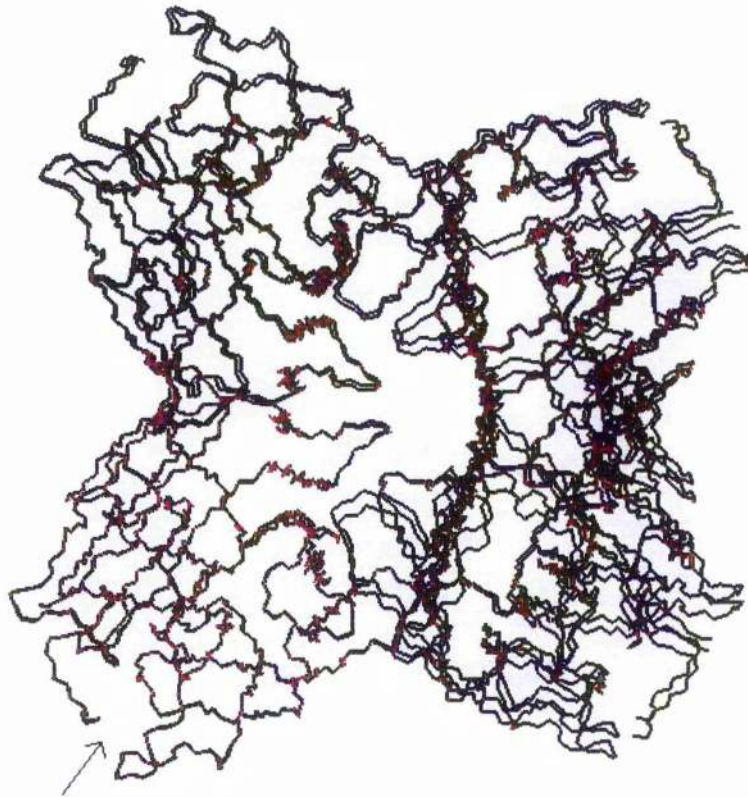
Dimer formation. Formation of the AB dimer buries 1136\AA^2 of subunit A and 1162\AA^2 of subunit B, and comprises 16 hydrogen bond interactions, 4 polar contacts and 138 van der Waals interactions. Formation of the CD dimer buries 1122\AA^2 of subunit C and 1131\AA^2 of subunit D and comprises 15 hydrogen bond interactions, 4 polar contacts and 132 van der Waals interactions. Formation of the EF dimer buries 1115\AA^2 of subunit E, 1148\AA^2 of subunit F and comprises 14 hydrogen bond interactions, 4 polar contacts and 153 van der Waals interactions. Formation of the GH dimer buries 1134\AA^2 of subunit G, 1128\AA^2 of subunit H and comprises 16 hydrogen bond interactions and 140 van der Waals interactions. The dimer formation is very similar to that seen for the MeoMan -

con A complex (Naismith *et al.*, 1994). Dimer formation is comparable between the two tetramers.

Tetramer formation. Tetramer ABCD formation buries a total of 8752\AA^2 comprising 2189\AA^2 from subunit A, 2169\AA^2 from subunit B, 2206\AA^2 from subunit C and 2188\AA^2 from subunit D. Tetramer EFGH formation buries a total of 8489\AA^2 comprising 2122\AA^2 from subunit E, 2096\AA^2 from subunit F, 2138\AA^2 from subunit G and 2133\AA^2 from subunit H. This is considerably less than seen for the native tetramer. Interactions between subunits A and C, and B and D stabilise the ABCD tetramer. There are 5 hydrogen bonds and 46 van der Waals interactions between subunits A and C and 3 hydrogen bonds and 33 van der Waals interactions between subunits B and D. Interactions between subunits E and G, and F and H stabilise the EFGH tetramer. There are 2 hydrogen bonds, 4 polar contacts and 53 van der Waals interactions between subunits E and G and 3 hydrogen bonds, 2 polar contacts and 41 van der Waals interactions between subunits F and H. This too is considerably less than observed for the native tetramer and is in agreement with observations reported in the con A α -D-mannopyranoside complex that the con A tetramer is more tightly packed in the saccharide free structure. The differences in the buried surface area upon tetramer formation, between the two tetramers, is within the limits of variability seen in all of the con A tetramers in this thesis. However, there is a subtle difference in the organisation of the two tetramers. This is illustrated in Figure 4.5. The two tetramers were superimposed via the backbone of one subunit from each tetramer (A and E). The major difference is seen in the dimer - dimer contact. Subtle changes in the organisation of the tetramer have been described in the Mec α Man - con A structure (Naismith *et al.*, 1994). The complex structures available suggest that the native structure lies at one end of a continuum of

tightness of packing. It seems that crystallisation selects a particular spatial arrangement of monomers from a family normally occurring in solution.

Figure 4.5. Differences between the ABCD tetramer and the EFGH tetramer. The tetramers were superimposed through one subunit of each tetramer.



Superimposed subunit

Metal sites. The co-ordination of the metals is the same in each of the eight subunits, and is unchanged from previous con A structures. Table 4.7 gives average metal to ligand distances and temperature factor ratios for each of the ligands and shows that this is a well ordered part of the protein. As seen in the other con A structures, Asp-19, Asn-14 and Tyr-12 have higher temperature factors. The average Ca^{2+} to Mn^{2+} distance is 4.0\AA .

Table 4.7. Metal to ligand distances and ligand temperature factors averaged over the 6 well ordered subunits.

Ligand		Average distance/ Å	Average temperature factor ratio ^a
Mn ²⁺	Glu-8 OE2	2.32	1.13
	Asp-10 OD1	2.06	0.56
	Asp-19 OD1	2.32	1.27
	His-24 NE2	2.39	0.91
	OW	2.07	0.77
	OW	2.18	1.05
	Average	2.22	0.94
Ca ²⁺	Asp-10 OD1	2.50	0.63
	Asp-10 OD2	2.61	0.56
	Tyr-12 O	2.37	1.05
	Asn-14 OD1	2.42	1.30
	Asp-19 OD2	2.48	1.09
	OW	2.33	0.39
	OW	2.52	0.79
Average	2.46	0.83	

^a Average temperature factor/ average temperature factor for the β -sheet residues (defined in Chapter 2).

Sugar binding.

Clear difference electron density is seen in the binding site for all five sugars in six of the eight subunits (Figure 4.6). In the two poorly ordered subunits, electron density for the 1,3 arm terminal GlcNAc was weak and the residue could not be unambiguously

modelled. Figure 4.7 is a schematic representation of the hydrogen bonds made between the oligosaccharide and the protein. All five sugar residues interact with the protein and no water molecules are recruited to the binding site for binding. A summary of the interactions made between the pentasaccharide and the protein is given in Tables 4.8 and 4.9.

Figure 4.6. $F_o - F_c$ Electron density contoured at 3σ and calculated after carrying out simulated annealing at 2000K, Powell minimisation and temperature factor refinement after removal of the sugar from the model.

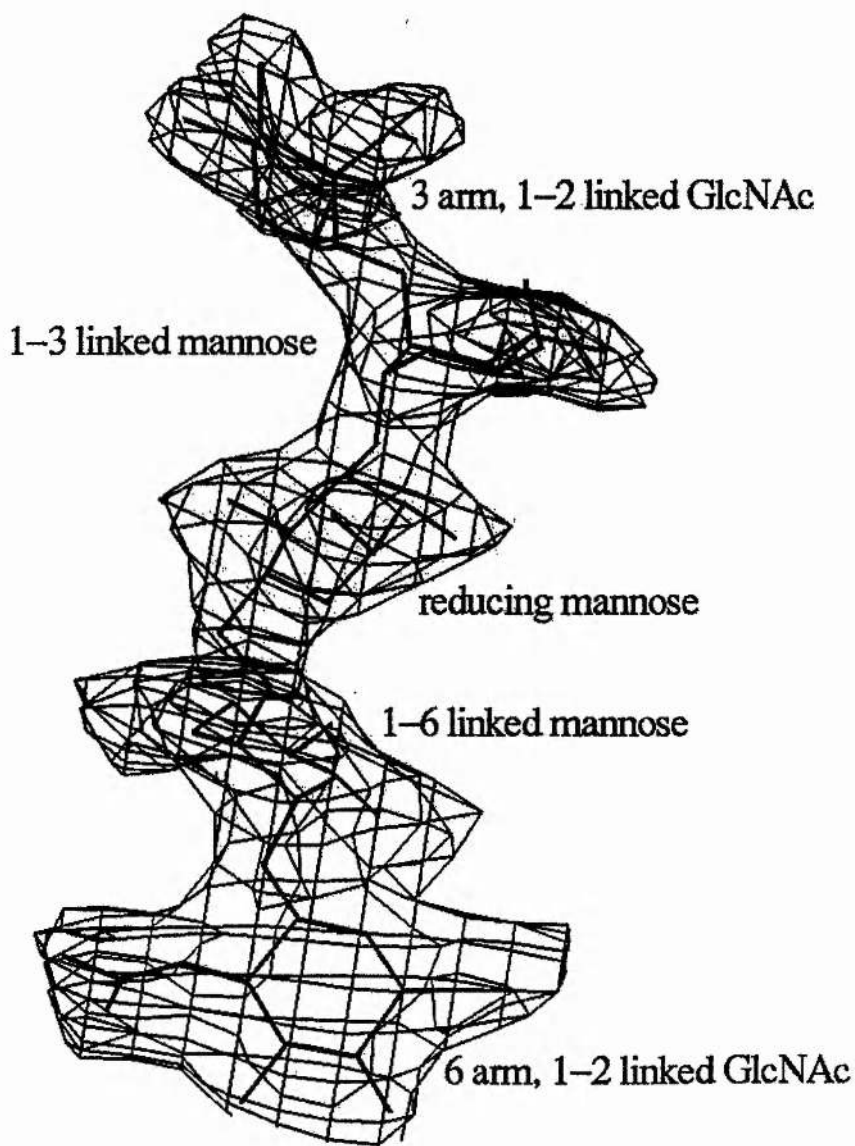


Table 4.8. Hydrogen bonding and polar contact (<3.5 Å) distances between sugar and protein in the six fully ordered sites

Protein	Sugar	Average distance/ Å
GlcNAc, 1-6 arm		
Thr-226 OG1	O3	2.6
Gly-224 O ^a	O4	2.8
Thr226 O ^a	O6	3.4
Ser-168 OG	O7	2.6
1-6 Man		
Arg-228 N	O3	3.0
Asn-14 ND2	O4	2.9
Arg-228 N ^a	O4	3.3
Asp-208 OD1	O4	2.8
Leu-99 N	O5	3.0
Leu-99N ^a	O6	2.9
Tyr-100 N	O6	3.0
Asp-208 OD2	O6	3.0
Reducing Man		
OW	O2	2.7
Asp-16 OD2 ^a	O2	3.3
Tyr-12 OH	O4	2.8

Table 4.8. Continued.

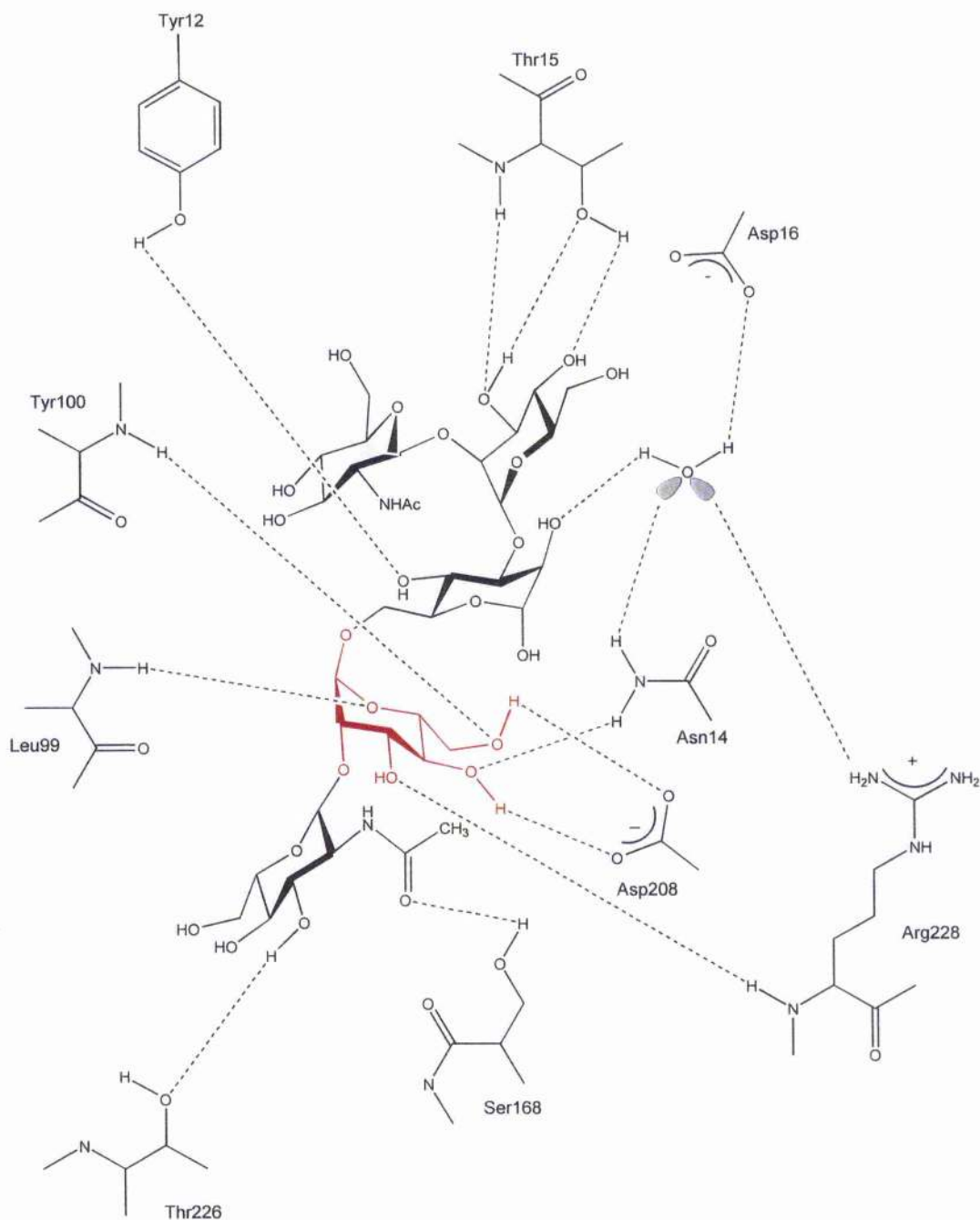
Protein	Sugar	Average distance (typically $\pm 0.1 \text{ \AA}$)
	1-3 Man	
Pro13 O *	$\overline{\text{O3}}$	2.9
Thr-15 N	O3	2.9
Thr-15 OG1	O3	3.0
Thr-15 OG1	O4	2.6
Asp-16 N *	O4	3.0

^a Distance is $< 3.5 \text{ \AA}$ however, the D-H-A geometry deviates from the linearity expected for a hydrogen bond.

Table 4.9. Van der Waals contacts ($< 4.0 \text{ \AA}$) between con A and the pentasaccharide

Sugar	Residue
1-6 arm GlcNAc	Gly-98 (6), Leu-99 (1), Ser-168 (5), Gly-224 (2), Thr-226 (7), Arg-228 (2), Leu-229 (2)
1-6 Man	Tyr-12 (1), Asn-14 (2), Gly-98 (3), Leu-99 (9), Tyr-100 (5), Ala-207 (2), Asp-208 (9), Gly-227 (4), Arg-228 (9)
Reducing Man	Tyr-12 (3), Leu-99 (2), Tyr-100 (1)
1-3 Man	Tyr-12 (2), Pro-13 (3), Asp-14 (4), Thr-15 (10), Asp-16 (4)
1-3 arm GlcNAc	Tyr-12 (2), Pro-13 (2), His-205 (3)

Figure 4.7. Schematic representation of the hydrogen bonds (dotted lines) formed between the pentasaccharide and the protein. The sugar bound at the monosaccharide site is shown in red.



The monosaccharide binding site, first characterised by Helliwell and co-workers (Derewenda *et al.*, 1989) is occupied by the 1,6 arm non-reducing mannose residue as

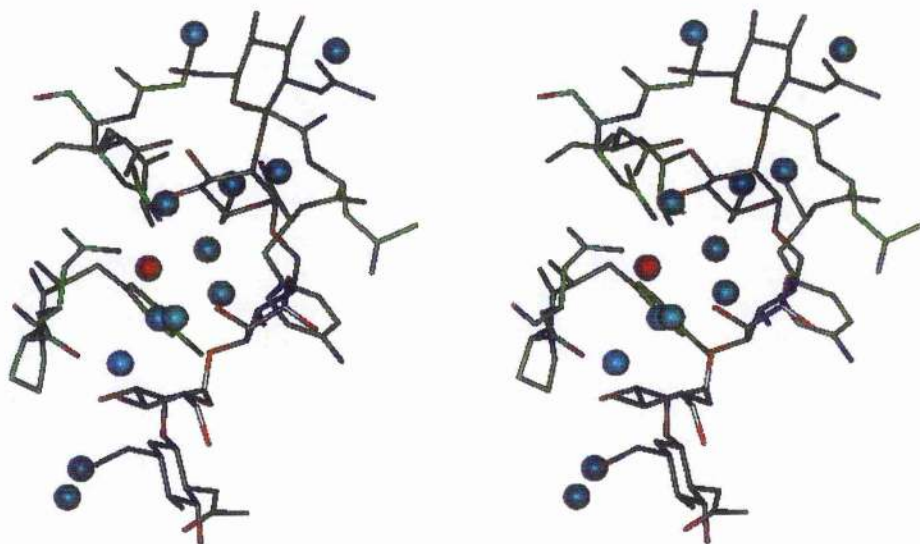
seen in the con A - trimannoside complex (Naismith and Field, 1996). The reducing mannose and the 1,3 arm non-reducing mannose residue are recognised by the protein in an analogous manner to that seen in the structure of the trimannoside complexed with con A (Figure 4.7) (Naismith and Field, 1996). The 1-6 arm GlcNAc site is formed by Gly-98, Ser-168 and the loop Thr-226 to Leu-229. One face of the sugar residue is effectively parallel to the protein surface. The loop from Thr-226 to Leu-229 interacts extensively with the sugar residue and forms one side of the binding site. Gly-98 and Ser-168 sit either side of the N-acetyl group. The 1-3 arm GlcNAc site is much less extensive, the side chain of His 205 caps O4 of the sugar. The ring of the sugar sits on a shallow cleft defined by Pro 13 and the side chain of Tyr 12. There are no hydrogen bonds between this sugar residue and the protein and only seven van der Waals contacts. The average temperature factors for the sugar atoms in subunits B (37.9\AA^2), F (41.0\AA^2), G (40.6\AA^2) and H (48.2\AA^2) are higher than subunits A, C, D and E (22.8, 36.3, 35.9 and 27.7\AA^2 respectively) and correlates with those for the protein atoms. The average temperature factors of the binding loops relative to the core of the protein is shown in Table 4.9.

Table 4.10. Average temperature factor of the binding loops in the six well ordered subunits divided by the average temperature factor for the core (β -sheet) residues.

	Pentasaccharide	0.94 \AA Native
Tyr-12 to Asp-16	1.49	1.39
Gly-98 to Tyr-100	1.61	1.22
Ser-168	1.55	1.36
Ala-207 to Asp-208	1.89	1.28
Gly-224 to Arg-228	1.31	1.21

Water displacement. Displacement of waters from the monosaccharide site has been described (Derewenda *et al.*, 1989; Naismith *et al.*, 1994). In addition to the four waters displaced here, an additional eight water molecules are displaced by the remaining four residues of the pentasaccharide (Figure 4.8).

Figure 4.8. Stereo view of the displacement of waters by the pentasaccharide. The native binding site (green) and the bound pentasaccharide (grey) were superimposed through the C α atoms of the protein. The native waters are shown as blue spheres. The conserved structural water is shown as a red sphere.



Oligosaccharide conformation. As seen in the trimannoside con A complex, the glycosidic conformations of the central Man α 1-6 (Man α 1-3) Man of the pentasaccharide in all eight subunits are very similar to those reported for conformation of the trimannoside molecule by NMR studies (Table 4.11) (Carver *et al.*, 1985).

Table 4.11. Torsion angles for the various oligosaccharides.

	$\varphi^a (^\circ)$	$\psi^a (^\circ)$	$\omega^a (^\circ)$
Methyl α -D-mannopyranoside ^{bc}	63 \pm 5	ND	ND
trimannoside ^c	—		
Manα 1-6 Man	66 \pm 3	-171 \pm 3	73 \pm 5
Man α 1-3 Man ^b	67 \pm 2	-112 \pm 5	ND
pentasaccharide ^c			
1-6 arm GlcNAc β 1-2 Man ^b	-75 \pm 3	-129 \pm 3	ND
1-3 arm GlcNAc β 1-2 Man ^b	-85 \pm 6	-78 \pm 5	ND
Manα 1-6 Man	71 \pm 3	179 \pm 2	72 \pm 3
Man α 1-3 Man ^b	65 \pm 2	-102 \pm 2	ND
Solution values ^d			
GlcNAc β 1-2 Man ^b	-92 \pm 16	-83 \pm 14	ND
Man α 1-6 Man ^b	70 \pm 20	-170 \pm 20	60 \pm 20
Man α 1-3 Man ^b	80 \pm 15	-116 \pm 25	ND

^a See Figure 1.2. Standard deviations are given. The sugar at the monosaccharide site is shown in bold. ^b Angles not defined (ND) for these sugars. ^c Averaged over all ordered fully ordered sugars. ^d (Homans, 1995).

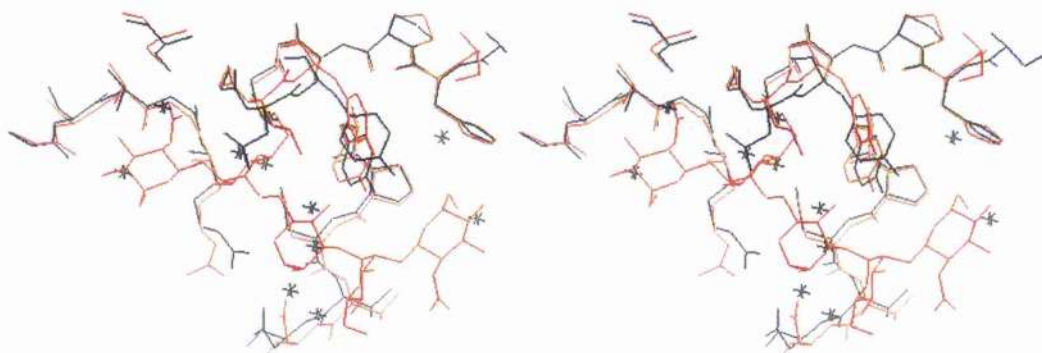
Electron density clearly shows that the anomeric hydroxyl (on the reducing mannose) is found in the alpha configuration (Figure 4.6) this appears to be a result of the anomeric oxygen's participation in crystal contacts. The N-linked glycan does not have a reducing sugar at this position and the 'reducing' mannose is β 1-4 linked to a GlcNAc sugar

(Figure 4.1). There is no structural hindrance to con A binding such a beta linkage in this structure.

In a second study of the trimannoside con A complex (Loris *et al.*, 1996) a second minority solution binding conformation of trimannoside was found in one of four subunits. There is no evidence for any alternate conformation of the pentasaccharide in this structure in any of the eight subunits. The conformation of the 1-3 arm GlcNAc β 1-2 Man glycosidic linkage is very close to that observed in solution (Homans, 1995). However, the ψ dihedral angle (Table 4.11) of the 1-6 arm GlcNAc β 1-2 Man glycosidic linkage is dramatically different (approximately 50°) from the solution minimum (Homans, 1995). This distortion is required to avoid penetration of the GlcNAc sugar residue into protein structure. A distortion in an inter-sugar glycosidic linkage was also seen in Chapter 3 of this thesis in the methyl α 1-2 mannobioside - con A complex. In this case, the distortion was necessary to maximise interactions between the carbohydrate and the protein.

Comparisons with the native structure. The con A - pentasaccharide structure superimposes onto the 0.94Å native structure (Deacon *et al.*, 1997) with an r.m.s. deviation of 0.51Å for all C α atoms. The value is 0.27Å for C α atoms of the β -sheet residues only. This is consistent with other saccharide bound con A structures. Changes in the positions of Tyr-12, Thr-15, Asp-16, Leu-99, Tyr-100, Ser-204 to Pro-206, Gly-224 and Arg-228 are seen when comparing the structures of native and the pentasaccharide complex of con A (Figure 4.8).

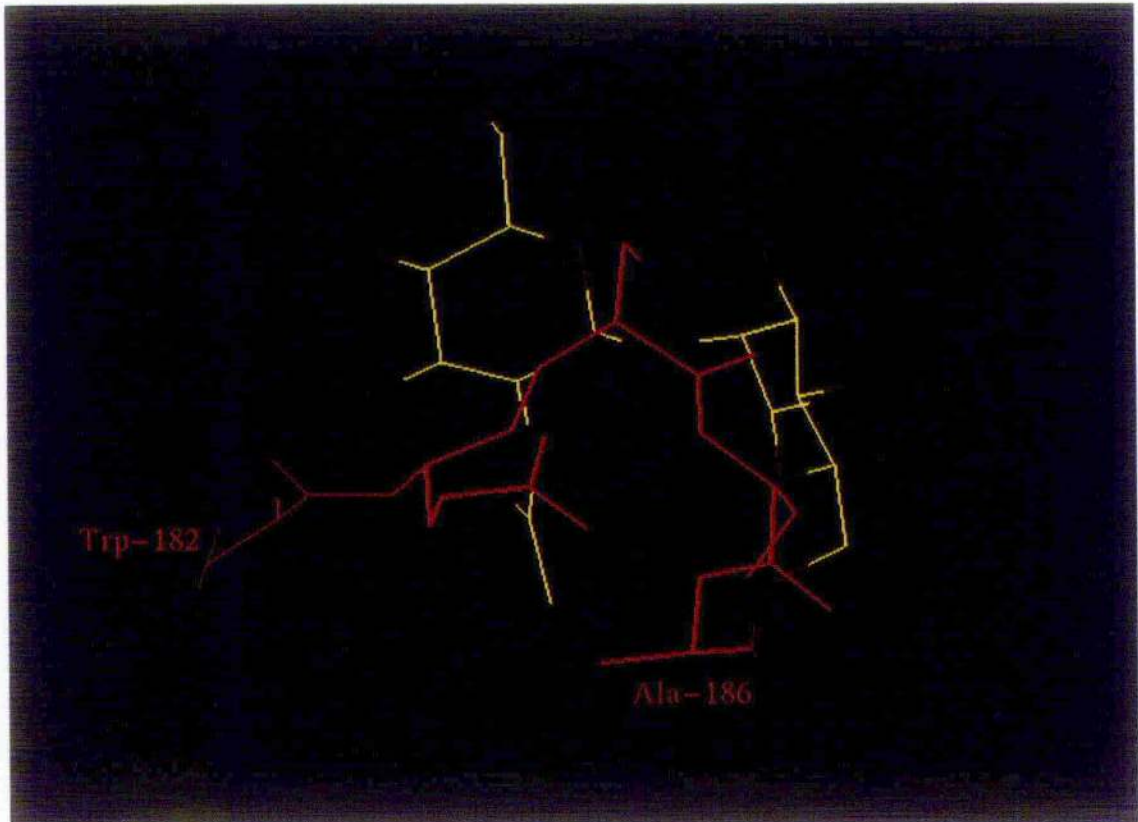
Figure 4.9. Stereo view of the pentasaccharide binding site (blue) superimposed onto the native structure (red). Waters are shown as crosses.



The changes at Tyr-12, Leu-99, Tyr-100 and Arg-228 which represent the monosaccharide site are broadly similar to those seen in the con A - methyl α -D-mannopyranoside complex and have already been discussed in detail by Helliwell and co-workers (Derewenda *et al.*, 1989; Naismith *et al.*, 1994). In addition there are further changes which accommodate the remaining residues of the pentasaccharide. The main chain at Thr-15 moves approximately 0.3Å to accommodate the 1,3 arm non-reducing mannose. The situation for Asp-16 is complicated by crystal contacts in the native and pentasaccharide structures and its position is varied in the subunits of the pentasaccharide structure. The change centred on His-205 is quite pronounced (0.3Å on average) and appears to be a result of accommodating the 1,3 arm terminal GlcNAc residue. The main chain at Gly-224 moves substantially (0.5Å) as result of the carbonyl oxygen forming a polar contact to O4 of the 1-6 arm terminal GlcNAc residue. The two 1-3 arm binding sites are filled by a tight crystal contact in the native structure (Figure 4.10), involving the

loop from Trp-182 to Ser-184. Whether or not this has relevance to the proposed peptide binding site of con A (Scott *et al.*, 1992) is unknown.

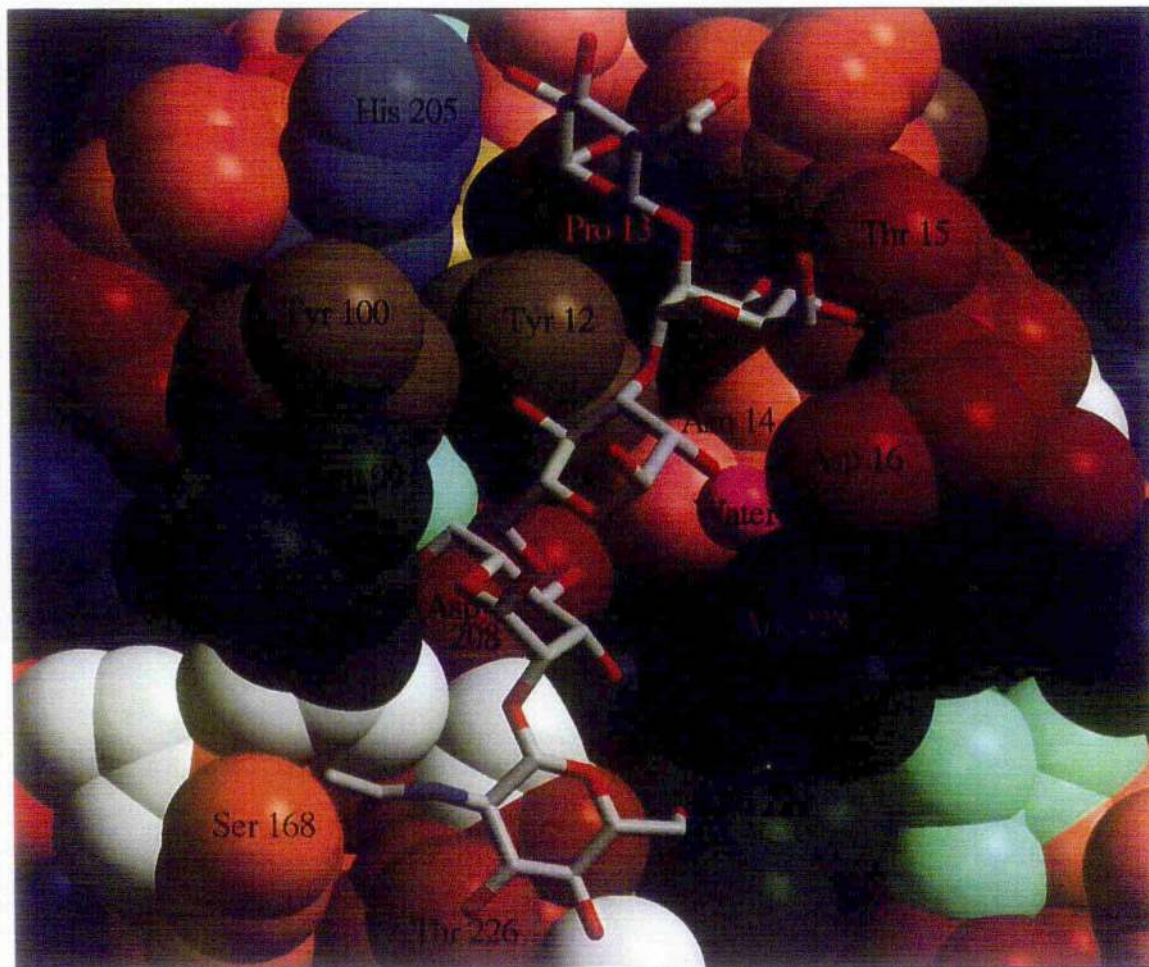
Figure 4.10. Crystal packing in the native con A structure. A crystallographically related loop (red) in the native structure sits in the position of the two 1,3 arm residues of the pentasaccharide.



The pentasaccharide sits in an extended continuous groove on the surface of the protein (Figure 4.11) and buries over 1000\AA^2 of surface area. In terms of buried surface area it is one of the largest continuous protein carbohydrate interfaces characterised. For discussion the continuous carbohydrate binding site is split into five components, in reality these overlap and share several protein residues. The five sites are the 1-3 arm terminal GlcNAc site, the 1-3 Man site, the reducing Man site, the monosaccharide site and the 1-6 arm terminal GlcNAc site. These sites correspond to the cognate sugar

residues shown in Figure 4.1. This is in contrast to the oligosaccharide structures of LOLI and LOLII that have effectively two discrete binding sites on the protein surface (Bourne *et al.*, 1992; Bourne *et al.*, 1994).

Figure 4.11. The Pentasaccharide - con A complex.



The carbohydrate binding site.

The monosaccharide site was first identified by Helliwell and co-workers (Derewenda *et al.*, 1989), the reducing Man and 1-3 Man sites were defined by the determination of the trimannoside complex (Naismith and Field, 1996). The residues forming these sites in the pentasaccharide complex are unchanged from the earlier descriptions.

There are a number of subtle changes in the interactions at the monosaccharide site compared to previous descriptions. A comparison of the monosaccharide site in the pentasaccharide complex and the mannose complex is summarised in Table 4.12.

Table 4.12. Differences among the con A complexes in the contacts at the carbohydrate binding sites.

	H-bonds	Polar contacts	van der Waals contacts
Pentasaccharide complex			
1-6 arm GlcNAc	2	2	25
1-6 Man	6	2	44
Reducing Man	1	1	6
1-3 Man	3	2	23
1-3 arm GlcNAc	0	0	7
Man αOMe complex			
Man α OMe	6	2	48

Superimposing the 237 α protein atoms of the pentasaccharide and methyl α -D mannopyranoside complexes gives an r.m.s. deviation of 0.33Å for the 13 sugar atoms. This distortion in the sugar position relative to the protein is centred around the O2 of the mannose residue which is β 1-2 linked to a GlcNAc residue in the pentasaccharide. Without this movement in the O2 position the GlcNAc residue would sterically clash (approaches < 2.0Å) with residues Thr-222, Ser-168 and Gly-98. The presence of the β 1-2 linkage alters the main chain trace of loop from Gly-98 to Tyr-100, centred at Leu-99 the average shift is 0.2Å. This region is one of the four loops that form the

monosaccharide site. Whilst there is a reduction in the number of van der Waals contacts at the monosaccharide site, the number of hydrogen bonds remains the same.

In the 1,6 arm terminal GlcNAc site, Gly-98 is a key residue, mutation to any other amino acid would abolish binding; as the side chain would intrude into the binding site clashing with the 1-2 linked sugar. Ser-168 hydrogen bonds to the sugar, presumably loss of this H-bond would diminish the interaction and any increase in the size of the residue would inhibit binding of the sugar conformation observed in this crystal structure.

Like the 1,6 arm GlcNAc β 1-2 Man in this structure, LOLI and II also bind sugars 1-2 linked to the monosaccharide site and both have a glycine in an analogous position to Gly-98. However, in the octasaccharide structures (Bourne *et al.*, 1994) which contain a GlcNAc β 1-2 Man disaccharide, the ψ torsion angle of GlcNAc sugar is rotated almost 180° with respect to that seen in our structure, although this is still a low energy conformation. This is because in LOL I and II Ser-168 is replaced by an Asn that would clash with the N-acetyl group. The interactions at this site are much less extensive in LOLI and II than in con A, having less than half the number of van der Waals interactions. The lectin from *Dioclea grandiflora* is homologous to con A and binds the trimannoside with similar affinity (Chervenak and Toone, 1995). It has a Gly at position 98 but does not recognise the pentasaccharide (Gupta & Brewer, 1996). This is due to changes at Ser-168 (to Asn) which creates a steric clash with the N-acetyl group and at Thr-226 (to Gly) which removes a hydrogen bond and van der Waals contacts.

Contribution of the extended binding site to affinity

The pentasaccharide and trimannoside con A complexes suggest that there are in principle two additional mannose binding sites (reducing and 1-3), however calorimetry

(Mandal *et al.*, 1994b) and structural analysis (Derewenda *et al.*, 1989) of the Me α Man - con A complex show that only one sugar is bound. Tabulation of the van der Waals contacts, buried surface area and hydrogen bonds shows that these two Man sites have considerably fewer interactions than the monosaccharide site (Table 4.12). From these Tables it seems that the reducing mannose site contributes little to the protein carbohydrate interaction but that the 1-3 Man will contribute to an increase in affinity. The 1-3 arm GlcNAc residue makes very few contacts with the protein, buries little surface area and its binding site is the least well defined. This sugar residue does induce a clear change in the main chain position centred on His-205 when compared to other con A structures. Given the insubstantial nature of its interaction and distortion in protein main chain it seems that the 1-3 arm GlcNAc residue does not make a large additional contribution to binding. The 1-6 arm GlcNAc site would appear able to make a substantial contribution to binding as it makes an array of van der Waals contacts with the protein, buries a substantial surface area and makes specific hydrogen bonds with the protein.

The disaccharide GlcNAc β 1-2 Man that we predict will bind analogously to the 1-6 arm of the pentasaccharide only liberates 5.2 kcalmol⁻¹ (Mandal *et al.*, 1994b). This is identical to Man which liberates approximately 5.2 kcalmol⁻¹ (Toone, personal communication). It appears that β 1-2 addition of GlcNAc to the monosaccharide mannose is energetically neutral, despite the large increase in the number of protein carbohydrate contacts. Jencks has pointed out that the free energy of binding a single compound AB to a protein is not the algebraic sum of free energy of binding of the isolated components A and B (Jencks, 1981). The difference between the summation of the free energy of binding A and B and the composite molecule AB, is the interaction

energy. The interaction energy has several components, some of which contribute favourably to binding, some of which contribute unfavourably. The most well known of these is the chelate effect, that is the entropy penalty paid by binding AB (loss of translational and rotational degrees of freedom) is less than binding A and B (Page & Jencks, 1971). The chelate effect always contributes favourably to the binding energy. The chelating disaccharide (GlcNAc β 1-2 Man) fails to increase the affinity for the protein over the simple monosaccharide. We believe we have identified structural changes in the pentasaccharide and to the monosaccharide site that contribute to the interaction energy, decreasing the affinity of binding of the pentasaccharide.

Rotation of the 1-6 arm GlcNAc β 1-2 Man glycosidic linkage

A large rotation (approximately 50 $^{\circ}$) in the glycosidic linkage ψ torsion angle from that seen in free solution is observed in the crystal structure. The dynamics observed for this linkage in free solution suggest it is a relatively rigid conformation (Homans, 1995). The energy surface for two glycosidic angles in the simple disaccharide (ϕ and ψ) GlcNAc β 1-2 Man has been calculated (Imberty *et al.*, 1991). Whilst the energetics of the simple disaccharide will be different from the pentasaccharide, it provides a reasonable approximation to the 1-6 arm GlcNAc β 1-2 Man linkage. Bourne *et al.*, saw an alternative conformation for the GlcNAc β 1-2 Man linkage in their LOL I octasaccharide structure (Bourne *et al.*, 1992). However, their observed glycosidic linkages were within $\pm 30^{\circ}$ of a second energetic minimum for this linkage. Plotting the glycosidic torsion angles we have observed on the energy surface, suggest that the conformation seen in the crystal structure is approximately 3 kcalmol $^{-1}$ above the global minimum and is remote from any other minimum.

4.6. Conclusions

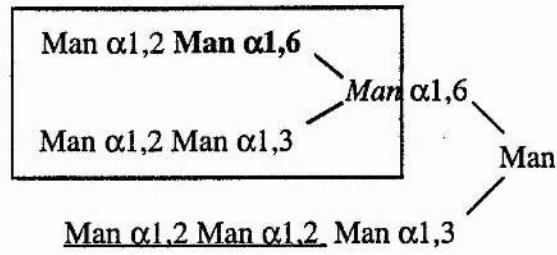
It would follow from this, and previous con A structures that the methyl α -D-mannopyranoside - con A complex represents the energetic minimum of the monosaccharide site and that the combination of the distortions discussed in this chapter decrease the free energy liberated by the monosaccharide site lowering the overall affinity of the oligosaccharide for con A. These are structural manifestations of the interaction energy. It follows that the free energy liberated by the disaccharide GlcNAc β 1-2 Man (5.2 kcalmol^{-1}) is considerably less than algebraic sum of the free energy liberated separately by GlcNAc and Man at their cognate sites. The ΔG° value for Man is 5.2 kcalmol^{-1} and although the value GlcNAc is known (5.2 kcalmol^{-1}) it probably reflects binding at the monosaccharide site. Consequently there is no reliable estimate for the energy liberated by GlcNAc at the 1-6 arm GlcNAc site. However the extent of the contacts between the protein and sugar suggests this is not zero.

4.7. Future Work.

An obvious step forward from this complex, apart from improving the resolution of the structure is to study the binding of the high mannose oligosaccharides with con A, for example the one shown in Figure 4.12. Interestingly *Dioclea grandiflora* and con A both bind the high mannose type structures shown boxed in Figure 4.12 with similar affinity (Mandal, Kishore *et al.*, 1994; Gupta and Brewer, 1996), K_a approximately $9 \times 10^5 \text{ M}^{-1}$. These high mannose sugars precipitate both con A and *Dioclea grandiflora* from solution (Gupta and Brewer, 1996) suggesting they are able to bind two con A molecules. The mode of binding of these sugars is not yet clear but Figure 4.12 shows how high mannose sugars may bind to con A. The α 1-2 linked terminal mannose residues would sit in the 1-

6 and 1-3 GlcNAc sites identified by this study and unlike GlcNAc would not clash with the Asn-168 residue in *Dioclea grandiflora*.

Figure 4.12. Proposed binding of high mannose N-linked glycan to con A. The underlined sugars would possibly bind and cross-link to another con A molecule.



Chapter 5.

Crystal structures of concanavalin A complexed with fructose.

5.1. Summary

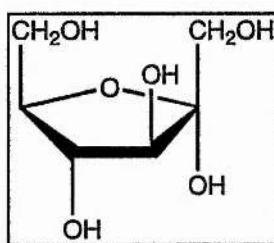
Two crystal forms of concanavalin A complexed with fructose have been obtained from protein solutions containing different concentrations of D-fructofuranose. The crystal form belongs to space group $P2_12_12_1$, with cell dimensions $a = 121.7\text{\AA}$, $b = 119.9\text{\AA}$, $c = 67.3\text{\AA}$ and the crystals diffract to 2.6\AA . A tetramer is present in the asymmetric unit, with evidence for occupancy of the carbohydrate binding sites apparent from the difference electron density in the binding sites. One fructose molecule could be modelled into the structure. The second crystal form belongs to space group $C222_1$, with cell dimensions $a = 103.3\text{\AA}$, $b = 117.9\text{\AA}$, $c = 254.3\text{\AA}$ and diffracts to 2.42\AA . Two dimers are present in the asymmetric unit and partial occupancy of the binding sites is evident from electron density, although a saccharide has not been modelled into the structure.

5.2. Introduction

The thermodynamics of con A binding to pyranose sugars has been well documented (Chervenak and Toone, 1994; Toone, 1994; Chervenak and Toone, 1995; Chervenak and Toone, 1996; Brewer *et al.*, 1997; Gupta *et al.*, 1997), and the structural basis of recognition is being investigated (Derewenda *et al.*, 1989; Loris *et al.*, 1996; Harrop *et al.*, 1996; Naismith and Field, 1996; Moothoo and Naismith, 1998; Bradbrook *et al.*, 1998; Chapters 2, 3 and 4 of this thesis). The binding of con A to furanose sugars has not been as well documented. It is known that con A forms precipitates in the presence of poly-D-arabinofuranosans and poly-D-fructofuranosans (Goldstein *et al.*, 1974). The thermodynamics of the interaction between furanoside sugars and the lectin have not been reported and the only structural investigation with con A has been a crystallisation note of a 2.7\AA complex of con A with arabinose (Kalb *et al.*, 1995). The coordinates for

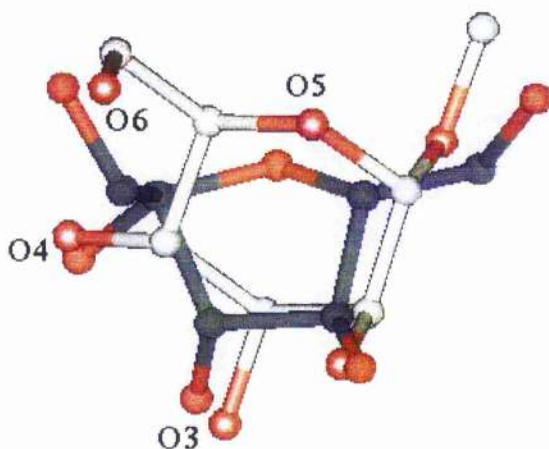
this structure are not available and the structure has not been reported at this time (August 1998). The only lectin - furanose interaction seen in a crystal structure thus far is in the structure of lentil lectin in complex with sucrose (Glc α 1-2Fru) which was determined to 1.9Å (Casset *et al.*, 1995) (described in Chapter 1). In this structure the glucose moiety makes the more extensive interaction with the protein, and the fructose is mainly recognised via water mediated hydrogen bonds.

Figure 5.1. α -D-fructofuranose.



In this Chapter, crystals of two different morphologies grown in the presence of con A and fructose (Figure 5.1) are described. The lack of documentation on the interaction between con A and furanose sugars may be an indication of the difficulty in working with these complexes. However, it is not clear why this should be so. Superimposing α -D-fructofuranose and methyl α -D-mannopyranoside (Figure 5.2) shows that the minimum requirement for binding as described by Goldsteins' Rules (Goldstein *et al.*, 1974) (3-, 4- and 6- hydroxyls) correspond quite closely in such a way that one could expect to see a similar network of hydrogen bonds made between the carbohydrate and the protein as seen for methyl α -D-mannopyranoside.

Figure 5.2. Superposition of α -D-fructofuranose and methyl α -D-mannopyranoside. The atom labels refer to methyl α -D-mannopyranoside.



5.3. Experimental.

Co-crystals were obtained after an initial screen consisting of 10, 15, 20 and 25% Polyethylene Glycol (PEG) 6K, pH 4.0, 5.0, 6.0 and 7.0. Optimisation of pH, PEG and fructose concentration (up to 0.1M) then followed and two crystal forms were obtained from the conditions described below (Moothoo and Naismith, 1998b).

P2₁2₁2₁ crystal form: Crystals were obtained from a protein solution containing 30mM fructose, 0.6mM con A, 1mM MnCl₂, 1mM CaCl₂, 0.1M NaCl and 20mM TRIS pH 7.0. 10 μ l of this solution were mixed with the reservoir in sitting drop wells which contained 20% PEG 6K, 0.1M HEPES pH 6.5 and left to equilibrate. X-rays were generated using an Enraf-Nonius FR591 rotating anode generator and focused using the MacScience mirror system. Data were collected at 20.5°C from a single crystal (0.5 x 0.3 x 0.2mm³) as 90 non-overlapping 25 minute 1.0° oscillations at room temperature to a resolution of

2.6Å. Auto indexing suggested a primitive orthorhombic lattice. The presence of systematic absences in the data at (2n+1, 0, 0) ruled out space group P2₁2₂, and favoured analysis in space groups P2₁2₁2₁ or P2₁2₁2. The cell dimensions are a = 121.7Å, b = 119.9Å, c = 67.3Å. One tetramer is present in the asymmetric unit, Matthew's number 2.41 DaÅ⁻³ (Matthews, 1968). The quality of the data is shown in Table 5.1. The data has a mosaic spread of 0.3°.

Table 5.1. Data quality for the 2.6Å fructose - con A complex.

Resolution	No. Reflections	% Complete	R-Merge/ %	Redundancy
20.00-5.57	2405	73.1	6.2	3.4
5.57-4.43	2627	83.3	6.3	4.0
4.43-3.88	2692	86.5	8.2	4.1
3.88-3.52	2719	88.2	10.2	4.0
3.52-3.27	2761	89.2	13.5	4.0
3.27-3.08	2764	90.4	18.7	3.9
3.08-2.93	2765	90.7	24.7	3.7
3.93-2.80	2711	88.9	31.1	3.6
2.80-2.69	2586	84.8	35.6	3.8
2.69-2.60	1916	62.8	29.1	2.3
20.00-2.60	25946	83.7	11.0	3.7

The structure was determined by the molecular replacement method as implemented in the CCP4 (CCP4, 1994) program AMORE (Navaza, 1994) using all data from 12Å to 3.8Å. The trimannoside con A complex tetramer (Naismith and Field, 1996) was used as the search model (with all metal ions, sugar molecules and water molecules removed and

all atoms set to an occupancy of 1.0). A cross rotation function gave 4 solutions ~13 standard deviations above noise. Solution R1 gave a translation solution ~15 standard deviations above noise in space group $P2_12_12_1$. At this point the space group was assigned as $P2_12_12_1$. The solutions in space group $P2_12_12_1$ are summarised in Table 5.2.

Table 5.2. Rotation and translation solutions for the $P2_12_12_1$ crystal form.

Sol ⁿ	Rotation			Translation			cc ^a	R ^b
	α	β	γ	x	y	z		
R1 ^c	156.60	11.71	92.40	-	-	-	0.39	-
R2	12.55	11.70	93.10	-	-	-	0.38	-
R3	27.07	1.97	221.64	-	-	-	0.37	-
R4	46.27	1.65	239.52	-	-	-	0.37	-
T ^d	156.29	11.71	92.40	0.0100	0.1038	0.1925	0.71	33.8
R ^e	159.28	12.30	89.06	0.0108	0.1028	0.1942	0.79	28.8

a correlation coefficient. ^b R factor. ^c R = Rotation function solution. ^d T = Translation function solution from R solution 1. ^e Refined solution which was imported into CNS.

The solution was imported into CNS (Brunger and Adams, 1998) and refinement carried out using R_{free} as a guide. R_{free} was calculated on 10% of data which were excluded from all refinement calculations. The raw molecular replacement solution had an initial R factor in CNS of 31.6% and a free R factor of 30.8%. Rigid body refinement of the tetramer followed by individual subunits lowered the R factor to 30.1% and the free R factor to 29.8%. F_o-F_c and $2F_o-F_c$ electron density maps were calculated from this model. A number of changes in the protein structure were made manually at this stage using 'O' (Jones *et al.*, 1991) (summarised in Table 5.3). CNS refinement included a bulk solvent correction with parameters for solvent density $0.34 \text{ e}\text{\AA}^{-3}$, solvent radius 0.55\AA and B factor 45.5\AA^2 . The anisotropic B factor correction applied to the data had parameters B11

= 1.73, B22 = 8.03, B33 = -9.75. The non-diagonal terms were 0. The metal ions were included in the model with zero electrostatic charge. Further refinement proceeded smoothly by alternating cycles of automated CNS (restrained positional and temperature factor) refinement (Brunger and Adams, 1998) and manual intervention using 'O' as described in Table 5.3. The protein was refined with the Engh & Huber stereochemical parameter dictionary. Non-crystallographic symmetry (NCS) restraints were applied throughout for both positional and temperature factor refinement (Table 5.4). All individual temperature factors were reset to the overall average after each manual intervention. Apart from the 10% of measured data excluded to monitor the refinement, no cut-offs were applied to the data. Electron density maps were calculated using SIGMAA (Read, 1986) modified coefficients with all data to 2.6Å. Water molecules were added to the model in batches if they satisfied four criteria: they corresponded to a peak $> 3.5\sigma$ in the F_o-F_c map; they made hydrogen bonds with reasonable stereochemistry; they reappeared in at least 1σ in subsequently calculated $2F_o-F_c$ maps and that a drop in the free R factor was observed. α -D-Fructose and β -D-fructose were obtained from the Cambridge Structural database (codes COWDOH01 and KEMXOP) and the structures minimised using X-PLOR.

Table 5.3. Progress of refinement of the $p2_12_12_1$ con A - fructose complex.

Refinement	R factor/ R _{free}
CNS: Rigid body refinement of tetramer (20 cycles) followed by each monomer (20 cycles).	30.1/ 29.8
65 cycles of positional refinement followed by 30 cycles of temperature factor refinement.	24.0/ 22.3
O: Four Mn ²⁺ ions, four Ca ²⁺ ions and 111 waters added. ~ 50 residues modelled into electron density.	
CNS: 80 Cycles of positional refinement, 60 cycles of temperature factor refinement.	23.5/ 21.2
O: 9 Waters deleted 35 waters added, residues 116, 158, 162, 185, 200 and 204 modelled into electron density.	
CNS: 80 Cycles of positional refinement, 50 cycles of temperature factor refinement.	21.6/ 23.1
O: Included 5 membered ring into subunit C.	
CNS: 50 Cycles of positional refinement, 55 cycles of temperature factor refinement.	19.0/ 22.7
a) O: Included α -D-fructose	
CNS: 80 Cycles of positional refinement, 95 cycles of temperature factor refinement.	18.8/ 22.6
b) O: Included β fructose	
CNS: 80 Cycles of positional refinement, 95 cycles of temperature factor refinement.	18.9/ 22.6

Table 5.4. NCS restraints used in positional and temperature factor refinement.

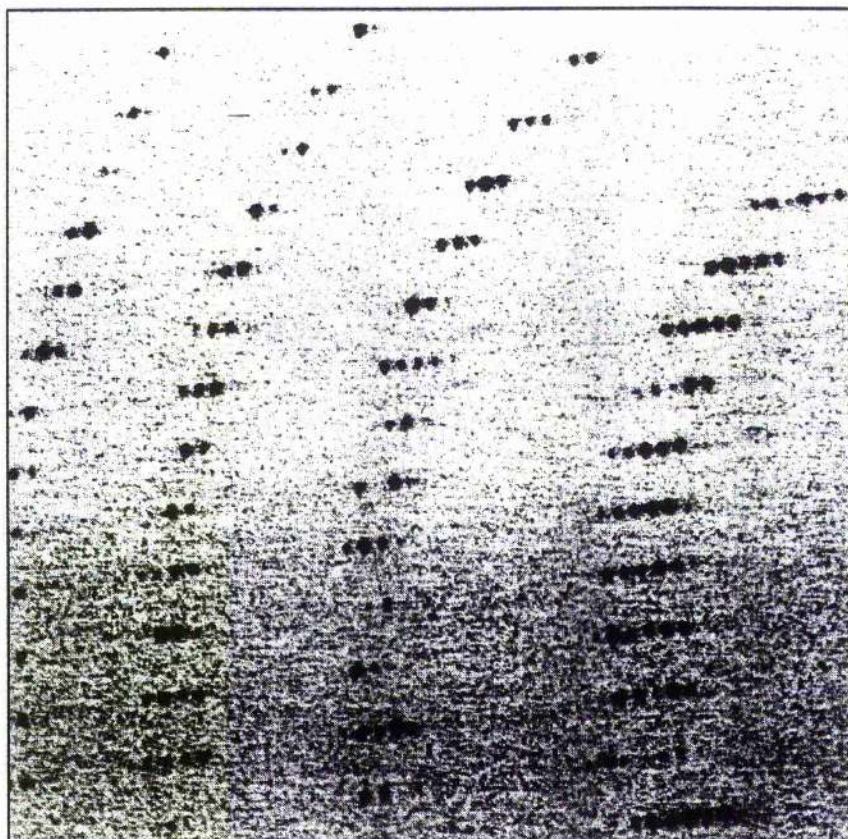
	weight ^a	σ_{nCS} ^b
Group 1. β -sheet backbone.	150	3.0
Group 2. β -sheet side chains.	80	4.0
Group 3. Non β -sheet backbone.	80	4.0
Group 4. Non β -sheet side chains.	60	4.0

^a Energy constant for positional restraints/ kcal mol⁻¹ Å⁻².

^b Target deviation for b factor restraints/ Å².

C222₁ crystal form: Crystals were obtained from a protein solution containing 40mM fructose, 0.6mM con A, 1mM MnCl₂, 1mM CaCl₂, 0.1M NaCl, 20mM TRIS pH 7.0. 10 μ l of this solution were mixed with 10 μ l of a reservoir containing 10% PEG 6K, 0.1M HEPES pH 7.0 in hanging drop trays and left to equilibrate. Large crystals (0.8mm³) were obtained after 2 weeks which diffracted to 2.42Å. X-rays were generated as before and data were collected at room temperature as 135 non-overlapping 20 minute 0.8° oscillations. Figure 5.3 shows one 0.8° oscillation image of this data. Due to the long c cell edge, the spots are very close together.

Figure 5.3. Section of one 0.8° oscillation collected during data collection of the $C222_1$ crystal form of the con A - fructose complex.



Auto-indexing suggested a C orthorhombic lattice. The data were merged in space groups $C222$ and $C222_1$. The cell dimensions are $a = 103.3\text{\AA}$, $b = 117.9\text{\AA}$, $c = 254.3\text{\AA}$. The asymmetric unit contains two dimers, approximately 67% solvent. The quality of the data is shown in Table 5.5. The data has a mosaic spread of 0.3° .

Table 5.5. Quality of data of the C222₁ con A - fructose complex.

Resolution	No. Reflections	% Complete	R-Merge/ %	Redundancy
25.00-5.20	5801	93.6	4.1	4.4
5.20-4.13	5966	99.4	4.4	5.2
4.13-3.61	5966	99.7	6.0	5.5
3.61-3.28	5905	99.7	8.7	5.4
3.28-3.05	5878	99.2	11.5	5.0
3.05-2.87	5842	99.2	14.5	5.2
2.87-2.73	5863	99.2	18.9	5.3
2.73-2.61	5753	97.8	25.8	4.6
2.61-2.51	5692	96.9	25.4	3.5
2.51-2.42	5138	87.0	27.5	3.0
25.00-2.42	57804	97.2	8.1	4.7

Molecular replacement calculations using AMORE from the CCP4 suite were carried out in space groups C222 and C222₁, initially using a native tetramer. This gave one solution in C222₁, however one dimer in the tetramer was crystallographically related to the other. This solution did not pack. Molecular replacement was repeated using the native dimer. In space group C222 only one solution was found at $x = 165.67$, $y = 33.17$, $z = 197.40$ with a correlation coefficient of 0.231 and an R factor of 52.7%. In space group C222₁, two dimers were positioned in the unit cell. At this point the space group was assigned as C222₁. The solutions are summarised in Table 5.6.

Table 5.6. Rotation and translation solutions for the C222₁ crystal form.

Sol ^a	Rotation			Translation			cc ^a	R ^b
	α	β	γ	x	y	z		
R1 ^c	178.50	0.00	0.00	-	-	-	-	-
R2	358.50	0.00	0.00	-	-	-	-	-
T1 ^d	178.50	0.00	0.00	0.4865	0.0244	0.0564	0.34	47.7
T2	358.50	0.00	0.00	0.0139	0.4741	0.0577	0.31	48.6

^a correlation coefficient. ^b R factor. ^c R = Rotation function solution. ^d T = Translation function solution.

The raw solutions were imported into CNS (Brunger and Adams, 1998) and refinement carried out using the free R_{free} as a guide. R_{free} was calculated on 3% of data (1733 reflections) which were excluded from all refinement calculations. The raw molecular replacement solution had an initial R factor in CNS of 30.9% and an R_{free} of 30.2%. Rigid body refinement of the two dimers followed by each subunit lowered the R factor to 24.8% and the free R factor to 26.0%. F_o-F_c and $2 F_o-F_c$ electron density maps were calculated from this model. A number of changes in the protein structure were made manually at this stage using 'O' (Jones *et al.*, 1991) (summarised in Table 5.7). The metal ions were included in the model with zero electrostatic charge. Further refinement proceeded smoothly by alternating cycles of automated CNS (restrained positional and temperature factor) (Brunger and Adams, 1998) refinement and manual intervention using 'O' as described below. The protein was refined with the Engh & Huber stereochemical parameter dictionary (Engh and Huber, 1991). Non-crystallographic symmetry (NCS) restraints were applied throughout for both positional and temperature factor refinement (Table 5.8). CNS refinement included and bulk solvent correction with

parameters for solvent density $0.36 \text{ e}\text{\AA}^{-3}$, solvent radius 0.01\AA and B factor 43.5\AA^2 . The anisotropic B factor correction applied to the data had parameters $B_{11} = -0.13$, $B_{22} = 3.50$, $B_{33} = -3.37$. The non-diagonal terms were 0. All individual temperature factors were reset to the overall average after each manual intervention. Apart from the 3% of measured data excluded to monitor refinement no cut-offs were applied to the data. Electron density maps were calculated using SIGMAA (Read, 1986) modified coefficients with all data to 2.42\AA . Water molecules were added to the model in batches if they satisfied four criteria: they corresponded to a peak $> 3.5\sigma$ in the F_o-F_c map; they made hydrogen bonds with reasonable stereochemistry; they reappeared in at least 1σ in subsequently calculated $2F_o-F_c$ maps and that a drop in R_{free} was observed.

Table 5.7. Progress of refinement of the C222₁ crystal form of the con A - fructose complex.

Refinement	R factor/ R _{free}
CNS: Rigid body refinement of monomers (30 cycles).	24.8/ 26.0
O: Four Mn ²⁺ ions, four Ca ²⁺ ions and 8 waters added.	
CNS: 100 Cycles of positional refinement, 50 cycles of temperature factor refinement.	21.0/ 23.5
O: > 70 Residues moved into electron density, 72 waters added.	
CNS: 60 Cycles of positional refinement, 30 cycles of temperature factor refinement.	20.8/ 23.2
O: 6 Waters deleted, 129 waters added. Residues 135, 158 185 modelled into difference density.	
CNS: 40 Cycles of positional refinement, 40 cycles of temperature factor refinement.	19.6/ 21.1
O: 2 Waters deleted. Regions 203 to 204 modelled into electron density.	
CNS: 30 Cycles of positional refinement, 30 cycles of temperature factor refinement.	19.6/ 21.1
O: 3 Waters deleted.	
CNS: 20 Cycles of positional refinement, 20 cycles of temperature factor refinement.	19.6/ 21.0

Table 5.8. NCS restraints used in positional and temperature factor refinement.

	weight ^a	σ_{nCS} ^b
Group 1. β -sheet backbone.	100	1.8
Group 2. β -sheet side chains.	60	2.8
Group 3. Non β -sheet backbone.	55	2.5
Group 4. Non β -sheet side chains.	45	3.5

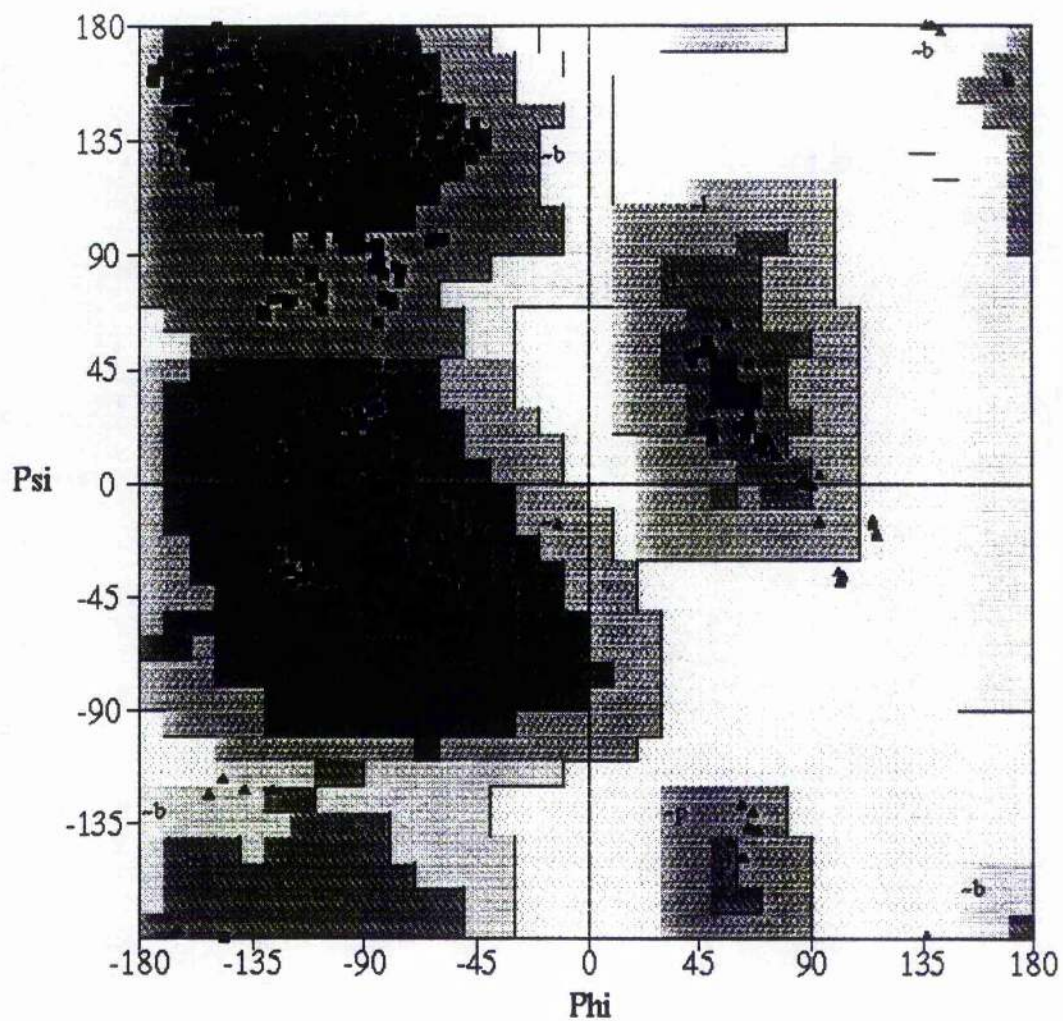
^a Energy constant for positional restraints/ kcal mol⁻¹ Å⁻².

^b Target deviation for b factor restraints/ Å².

5.4. Analysis of the final models.

P2₁2₁2₁ crystal form: The final model contains 4 subunits of 237 residues (7236 protein atoms), 4 Ca²⁺ ions, 4 Mn²⁺ ions, 137 water molecules and 13 sugar atoms. Statistics on the final model are shown in Table 5.9. Poorly ordered regions occur at residues 118 to 122 and 204. These are consistent with poorly ordered regions found in previous con A complex structures (Naismith *et al.*, 1994; Naismith and Field, 1996; Moothoo and Naismith, 1998). 220 atoms in the final model were stereochemically modelled and set to zero occupancy as they could not be located in the electron density. The Ramachandran plot for the final model is given in Figure 5.4. No residues lie outside of the allowed regions.

Figure 5.4. Ramachandran plot for the final model of the $p2_12_12_1$ fructose - con A complex.^a



^a See section 2.4 for explanation of the Ramachandran plot.

Table 5.9. Quality of the final model of the $p2_12_12_1$ con A - fructose complex.

Resolution range (Å)	∞-2.6
R _{free} (%) alpha/ beta	22.6/ 22.6
R factor (%) alpha/ beta	18.8/ 18.9
Bond r.m.s. deviation (Å) ^a —	0.009
Angle r.m.s. deviation (°) ^a	1.628
Non-crystallographic symmetry r.m.s. deviation(Å) (cα atoms)	0.08
B-factor bonded atoms r.m.s. deviation (Å ²) ^b (main chain)	1.557
B-factor bonded atoms r.m.s. deviation (Å ²) ^b (side chain)	2.183
Ramachandran core/additional/ (%) ^c	84.6/ 15.4
Protein mean B-factor (Å ²) ^d (all)	32.43
Protein mean B-factor (Å ²) ^d (main chain)	32.01
Protein mean B-factor (Å ²) ^d (side chain)	32.90
Sugar mean B-factor (Å ²) ^d alpha/ beta	87.68/ 91.04
Solvent mean B (Å ²) ^d	36.48

^a r.m.s. deviation from Engh and Huber ideal values (Engh and Huber, 1991). ^b B-factor deviation for bonded atoms. ^c core and additionally allowed regions as defined by PROCHECK (Laskowski *et al.*, 1993).

^d Calculated with MOLEMAN (G.J. Kleywegt, unpublished program). All stereochemically modelled atoms were removed prior to B-factor analysis, all bonded atoms including those in the sugars are included in the calculation of r.m.s.

Figure 5.5 shows the temperature factors along the backbone of the protein for each subunit, with the highest peaks corresponding with the poorly ordered regions of the structure. The poorly ordered regions of the structure have consistently been found to be disordered in con A crystal structures. Table 5.10 gives the average temperature factors for each subunit and those of the core (β -sheet residues) for each subunit. The temperature factors for the B subunit are higher than for the other three subunits. The A subunit also has slightly higher temperature factors than the C and D subunits.

Figure 5.5. Temperature factor plots along the backbone of each subunit in the $P2_12_12_1$ con A - fructose complex.

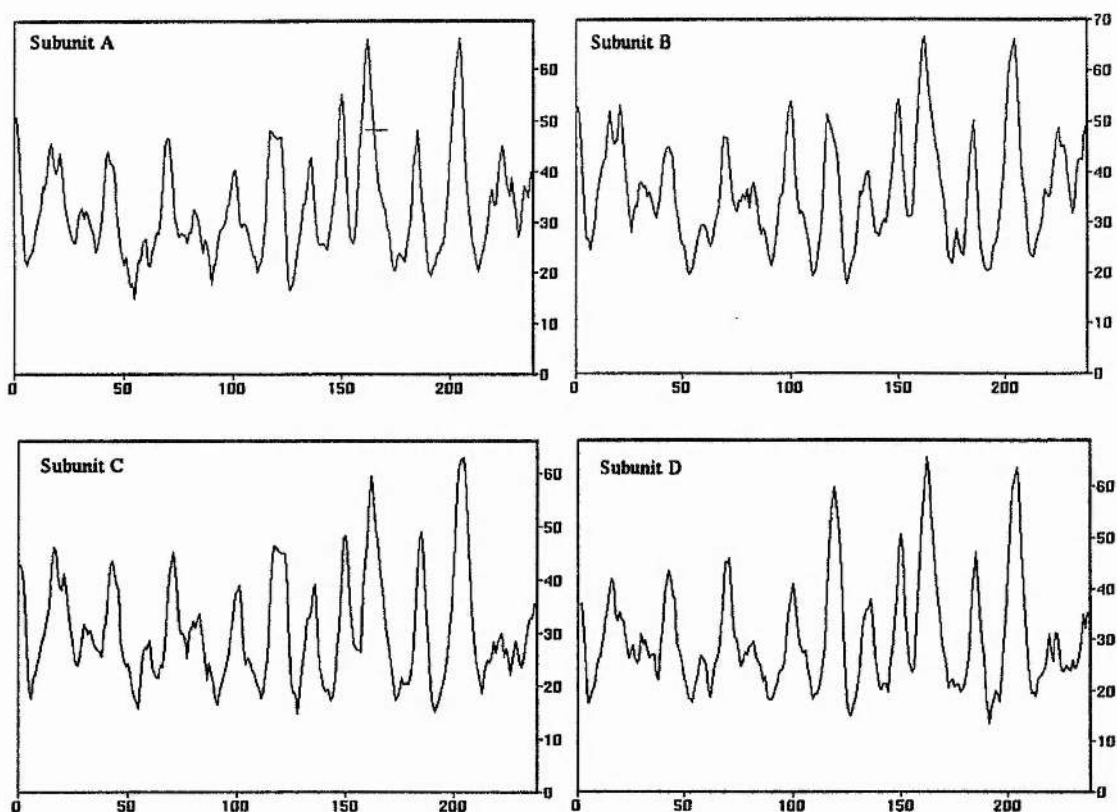


Table 5.10. Temperature factors for each subunit of the $P2_12_12_1$ fructose - con A complex.

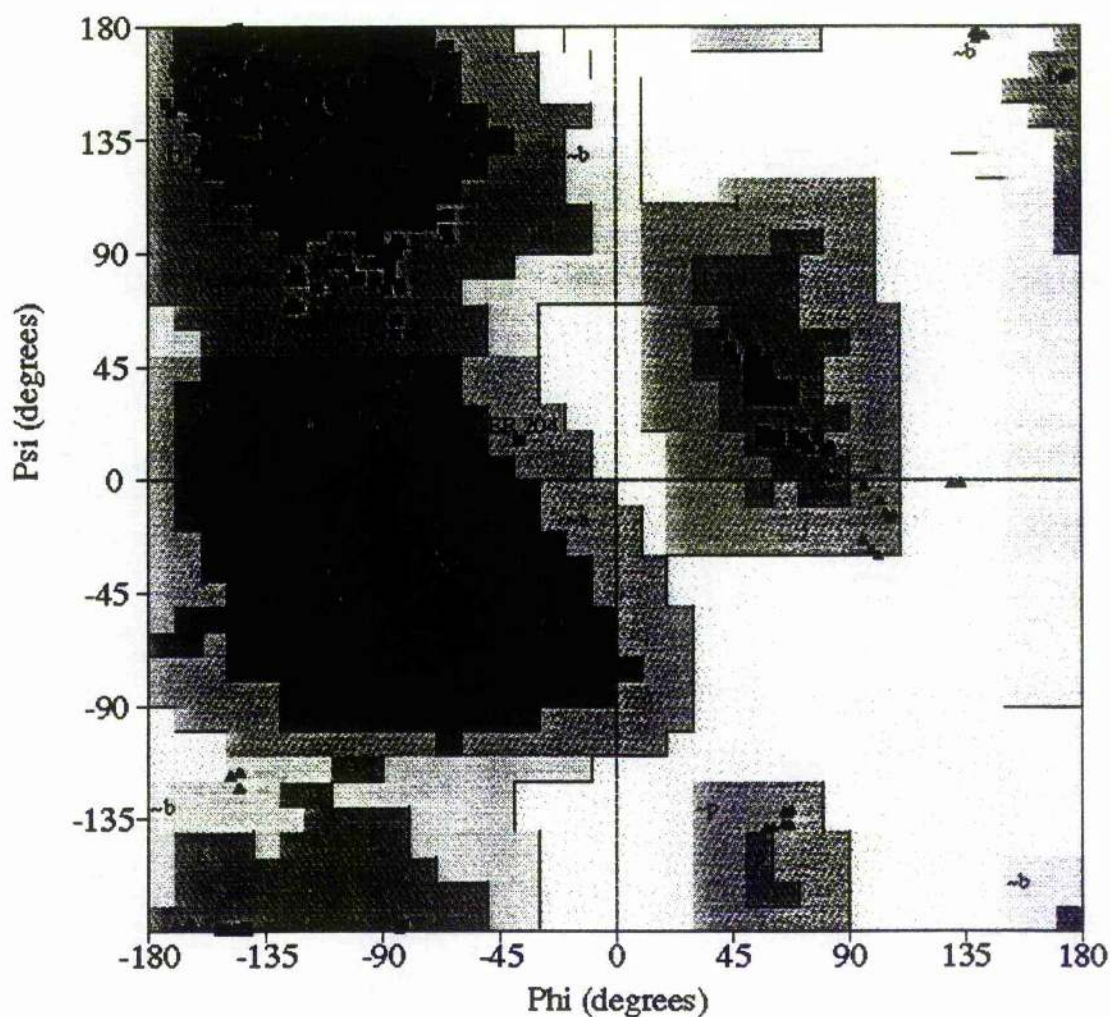
Subunit	Main chain (Å)/ core ^a	Side chain (Å)/ core ^a	All atoms (Å)/ core ^a
A	32.6/ 27.06	33.3/ 27.3	32.9/ 27.2
B	35.6/ 30.0	36.3/ 30.7	35.9/ 30.3
C	30.0/ 25.1	31.3/ 25.8	30.6/ 25.5
D	29.9/ 24.8	30.7/ 25.1	30.8/ 24.9

^a b factor calculated for β -sheet residues only.

$C222_1$ crystal form: The final structure contains 4 subunits (7236 protein atoms) of which 197 atoms were set to zero occupancy in the final model as they could not be located in electron density. Also included are four Mn^{2+} ions, four Ca^{2+} ions and 288 water

molecules. The Ramachandran plot for the final model is shown in Figure 5.6. No residues are found outside the allowed regions. The one residue which lies in the generously allowed regions is in a poorly ordered part of the protein. Statistics on the final model are shown in Table 5.11.

Figure 5.6. Ramachandran plot for the final model of the C222₁ con A - fructose complex.^a



^a See section 2.4 for explanation of the Ramachandran plot.

Table 5.11. Statistics on the final model of the C222₁ fructose - con A complex.

Resolution range (Å)	∞-2.42
R _{free} (%)	21.0
R factor (%)	19.6
Bond r.m.s. deviation (Å) ^a —	0.01
Angle r.m.s. deviation (°) ^a	1.7
Non-crystallographic symmetry r.m.s. deviation (cα atoms) (Å)	0.09
B-factor bonded atoms r.m.s. deviation (Å ²) ^b	1.1
Ramachandran core/ additional/ generously allowed (%) ^c	85.2/ 14.7/ 0.1
Protein mean B-factor (Å ²) ^b (all)	26.6
Protein mean B-factor (Å ²) ^b (main chain)	26.3
Protein mean B-factor (Å ²) ^b (side chain)	26.9
Solvent mean B (Å ²)	37.6

^a r.m.s. deviation from Engh and Huber ideal values (Engh and Huber, 1991). ^b Calculated with MOLEMAN (G.J. Kleywegt, unpublished program). All stereochemically modelled atoms were removed prior to B-factor analysis, all bonded atoms including those in the sugars are included in the calculation of r.m.s. B-factor deviation for bonded atoms. ^c core and additionally allowed regions as defined by PROCHECK (Laskowski *et al.*, 1993).

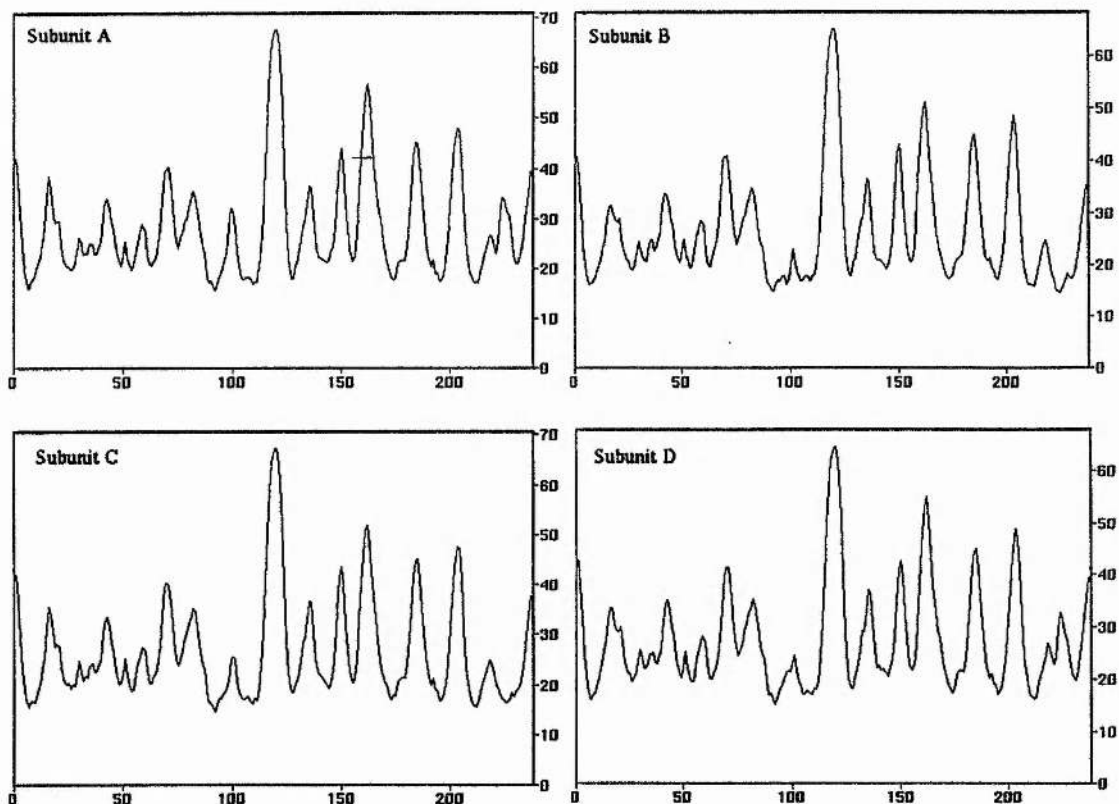
Table 5.12 shows the temperature factors for each subunit, for all residues and for the core (β-sheet) residues. Subunit A has slightly higher temperature factors than the other 3 subunits. Figure 5.7 shows the temperature factor plots along the backbone of each subunit. The highest peaks correspond with poorly ordered regions of the protein.

Table 5.12. Temperature factors for each subunit of the C222₁ fructose - con A complex.

Subunit	Main chain (Å)/ core ^a	Side chain (Å)/ core ^a	All atoms (Å)/ core ^a
A	27.3/ 23.1	27.8/ 23.8	27.5/ 23.4
B	25.6/ 22.3	26.5/ 23.3	26.0/ 22.8
C	25.7/ 22.3	26.2/ 23.1	25.9/ 22.7
D	26.7/ 22.8	27.2/ 23.6	26.9/ 23.2

^a b factor calculated for β-sheet residues only.

Figure 5.7. Temperature factor plots along the protein backbone for each subunit of the C222₁ con A - fructose complex.



5.5. Results and Discussion

P2₁2₁2₁ crystal form:

Tetramer Formation. The AB dimer interaction comprises 16 hydrogen bonds and 165 van der Waals interactions, and the dimer CD interaction comprises 17 hydrogen bonds, one polar contact and 164 van der Waals interactions. 2332Å² of surface accessible area is buried upon dimer AB formation and 2301Å² upon dimer CD formation. The tetramer is stabilised by 3 hydrogen bonds, 1 polar contact and 53 der Waals interactions between subunits A and C and 4 hydrogen bonds, 2 polar contacts and 57 van der Waals interactions between subunits C and D. 8901Å² of surface accessible area is buried upon formation of the tetramer comprising 2223Å² from subunit A, 2214Å² from subunit B, 2223Å² from subunit C and 2241Å² from subunit D.

As seen in other con A - carbohydrate structures, the tetramer is less tightly packed than the native tetramer.

Metal sites.

The co-ordination of the metals is unchanged from previous con A structures. Average metal to ligand distances and temperature factors are given in Table 5.13, showing that the metal binding sites in this structure are higher relative to the core residues compared with the other con A structures.

Table 5.13. Metal to ligand distances and temperature factors in the $P2_12_12_1$ con A fructose complex.

Metal	Ligand	Distance /Å	Temperature factor ratio ^a
Ca ²⁺	Asp10 OD1	2.5	1.1
	Asp10 OD2	2.3	1.0
	Tyr12 O	2.5	1.2
	Asn14OD1	2.5	1.5
	Asp19 OD2	2.2	1.4
	WA O	2.3	1.2
	WB O	2.2	1.0
Average		2.4	1.0
Mn ²⁺	Glu8 OE2	2.3	1.0
	Asp10 OD2	2.1	1.0
	Asp19 OD1	2.3	1.4
	His24 NE2	2.3	1.0
	WC O	2.0	0.8
	WD O	2.1	0.6
Average		2.2	1.0

^a Temperature factor/ temperature factor for β -sheet residues only.

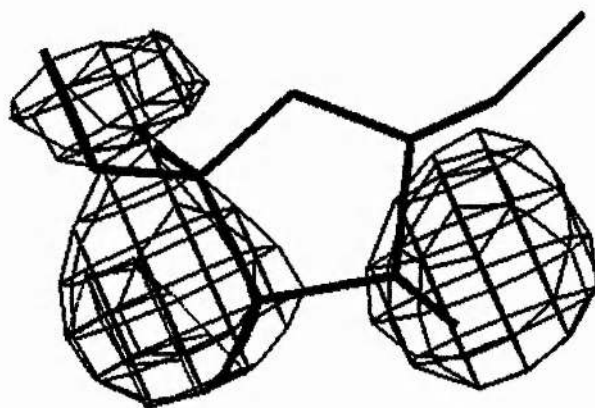
Sugar binding.

Difference electron density is very patchy in three of the four subunits allowing unambiguous modelling of a fructose molecule in only one of the binding sites. Changes in the binding sites are similar to changes seen in other con A - carbohydrate complexes, in particular Leu-99 and Arg-228 adopt the conformations seen in the other bound

structures and are markedly different to their positions in the native structure, strongly suggesting at least partial occupancy of the binding sites.

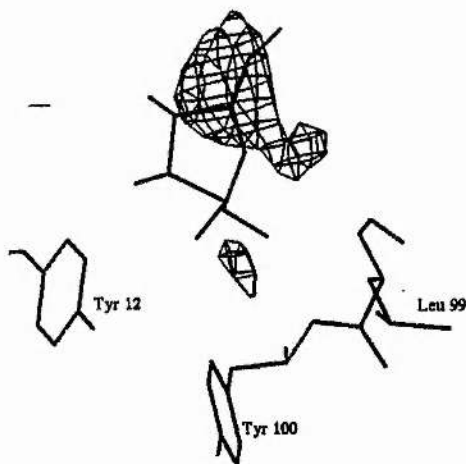
In the A subunit, crystal contacts are in very close proximity to the binding site. A symmetry related His-205 partially occludes the monosaccharide binding site. However there is enough room to accommodate fructose in a beta conformation and there is some difference density in the binding site, however it is impossible to say whether this is water molecules or a saccharide partially occupying the binding site. Figure 5.8 shows the difference electron density seen in the binding site in this subunit.

Figure 5.8. F_o-F_c electron density contoured at 2σ seen in the binding site of subunit A.



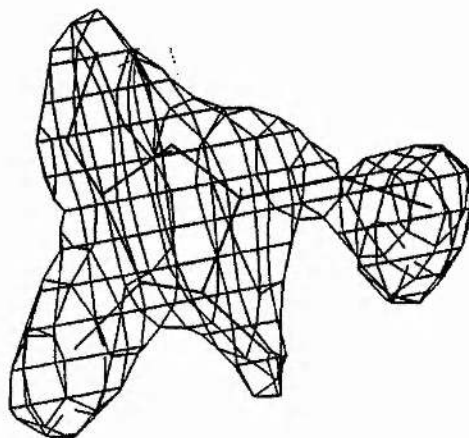
In the B subunit, again there is some density where the fructose would be located. In this subunit however, the density does not resemble that seen for water molecules in the binding site. In Figure 5.9, fructose has been modelled in to the position it occupies in subunit C. The binding site here is not restricted at all by crystal contacts.

Figure 5.9. F_o-F_c electron density contoured at 2σ seen in the binding site of subunit B. α -D-fructofuranose has been superimposed in a possible position.



In the C subunit, there is sufficient difference density to attempt to model the monosaccharide into the binding site. Figure 5.10 shows the difference density in this binding calculated before any sugar had been included in the model and with α -D-mannose superimposed in its bound position in the con A - methyl α -D-mannopyranoside complex.

Figure 5.10. F_o-F_c electron density contoured at 2σ seen in the binding site of subunit C prior to inclusion of fructose.



The structure was refined at first with a 5 membered ring, with no substituents. The difference density was not clear enough to distinguish between α or β conformation. After a cycle of CNS refinement (restrained positional and temperature factor) with both configurations the only difference in the model statistics was a 0.1 increase in the R factor with the beta conformation. In agreement with the poor density at this site, the resulting temperature factors of the saccharides are extremely high. When the α conformation is modelled into the structure, difference density appears for a possible β conformation which would make sensible hydrogen bonds with the protein (Figure 5.11), although this could be a water molecule. When the β conformation is modelled, no such density appears for the alternate conformation. Figure 5.12 shows the $2F_o-F_c$ density calculated after alpha fructose had been included in the refinement. Tables 5.14 and 5.15 show hydrogen bonds and van der Waals interactions between fructose and con A made by each of the configurations. The tentative suggestion is that the β - conformation is bound.

Figure 5.11. F_o-F_c electron density contoured at 2σ seen in the binding site of subunit C, calculated after α -D-fructose is included in the model.

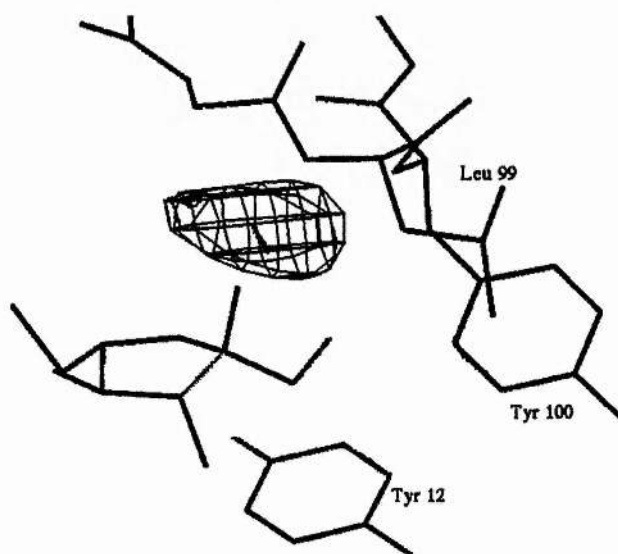


Figure 5.12. $2F_o-F_c$ density of the subunit C binding site. The map is contoured at 1σ and was calculated after α -D-fructose had been included in the refinement.

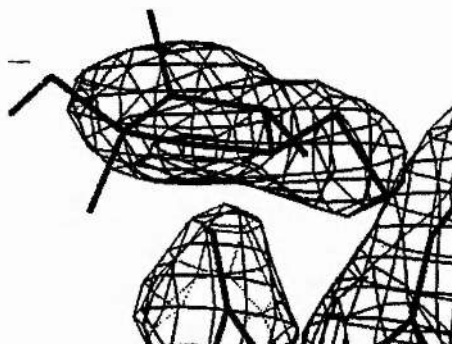


Table 5.14. Hydrogen bonds and polar contacts ($< 3.5\text{\AA}$) between fructose and con A.

Sugar	Protein	Distance/ \AA (Alpha)	Distance/ \AA (Beta)
O1	Leu-99 N	3.4	2.4
O1	Leu-98 N	-	3.2 ^a
O1	Asp-208 OD2	-	3.5 ^a
O2	Leu-99 N	3.3	-
O2	Asp-208 OD2	3.3	-
O2	Asp-208 OD1	3.5 ^a	-
O3	Asn-14 ND2	2.8	2.7 ^a
O4	Asn-14 ND2	3.1 ^a	3.1 ^a
O4	Asp-208 OD1	2.3	2.3
O4	Asp-208 OD2	3.3	3.3
O4	Arg-228 N	2.5	2.5
O6	Arg-228 N	3.1 ^a	3.0 ^a

^aDistance is $< 3.5\text{\AA}$ but geometry deviates ($> 90^\circ$) from the linearity expected for a hydrogen bond.

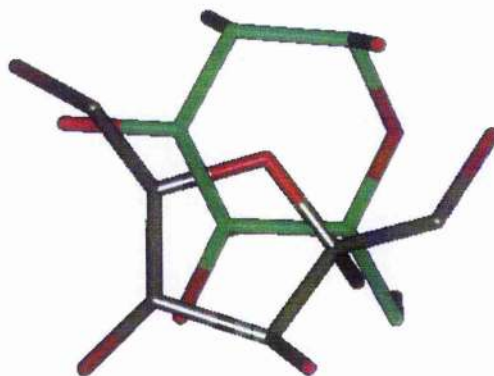
Table 5.15. Van der Waals interaction (< 4.0Å) between fructose and con A.

Sugar	Contacts	Residue
Alpha	54	Tyr-12, Asn-14, Gly-98, Leu-99, Asp-208, Gly-277, Arg-228
Beta	58	Tyr-12, Asn-14, Gly-98, Leu-99, Asp-208, Gly-277, Arg-228

Buried surface area analysis shows that in the α -D-fructose - con A model, 262Å² of polar surface accessible area and 116Å² of polar surface area is buried upon complex formation. For β -D-fructose, the values are 256Å² and 110Å² respectively for α -D-fructose and β -D-fructose. These values do not allow prediction of one conformation over the other.

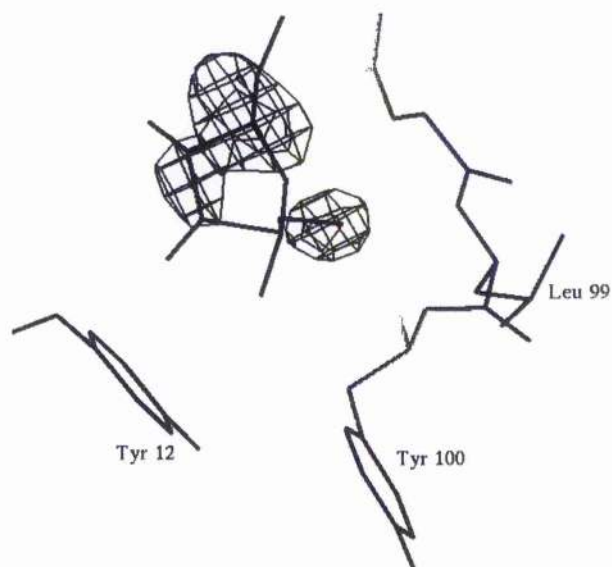
Superimposing subunit C onto the Me α Man - con A complex reveals that the fructose, as modelled here, is not positioned in the expected orientation (Figure 5.1). Figure 5.13 illustrates the position of the bound fructose as modelled into subunit C compared with bound Me α Man.

Figure 5.13. Comparison of the positions of bound fructose (grey) and bound Me α Man (green). The two subunits were superimposed through the C α atoms of the protein backbone.



In subunit D, there is also partial density for fructose but not good enough to model a saccharide in. Figure 5.14 shows the density in this binding site with β -D-fructose superimposed in its position in subunit C.

Figure 5.14. F_o-F_c electron density seen in the binding site of subunit D of the $P2_12_12_1$ fructose - con A complex.



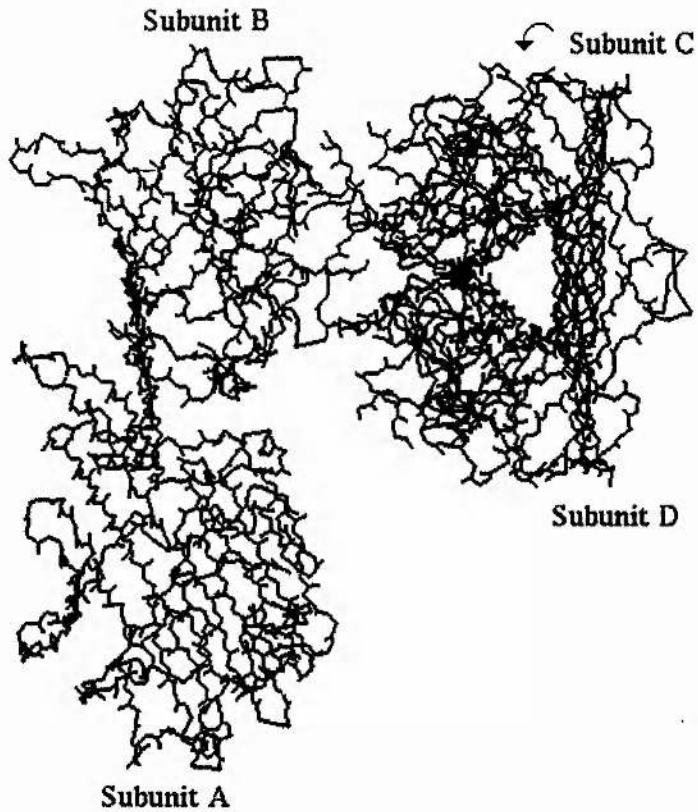
C2221 crystal form:

Tetramer organisation.

The organisation of the two dimers in the asymmetric unit is shown in Figure 5.15. Formation of the AB dimer buries 2280\AA^2 of surface accessible area and involves 16 hydrogen bonds, 1 polar contacts and 139 van der Waals interactions. Formation of the CD dimer buries 2256\AA^2 of surface accessible area and involves 17 hydrogen bonds, 1 polar contacts and 135 van der Waals interactions. The conventional tetramer is formed by applying the transformation (x, -y, -z) (a two fold rotation axis) on one of the dimers. 9229\AA^2 Of surface accessible area is buried upon formation of the tetramer, comprising 2344\AA^2 from subunit A, 2292\AA^2 from subunit B, 2300\AA^2 from subunit C and 2293\AA^2 from subunit D. The tetramer is stabilised by 10 hydrogen bonds, 2 polar contacts and 140 van der Waals interactions between subunits A and C and 12 hydrogen bonds, and 83 van der Waals interactions between subunits B and D.

This multimer organisation is comparable with the packing seen in the other carbohydrate bound con A structures, ie the tetramer is much less tightly packed than the native tetramer.

Figure 5.15. Organisation of the two dimers present in the asymmetric unit.



Metal sites.

As seen in the other carbohydrate bound con A structures (Derewenda *et al.*, 1989; Naismith *et al.*, 1994; Naismith and Field, 1996; Moothoo and Naismith, 1998) (see also Chapters 2, 3 and 4), there is no change in metal coordination in this complex. The average metal to ligand distances and ligand temperature factors are given in Table 5.16.

Table 5.16. Metal to ligand distances and temperature factors in the C222₁ con A - fructose complex.

Metal	Ligand	Distance /	Temperature factor/
		Å	Å ²
Ca ²⁺	Asp10 OD1	2.4	18.0
	Asp10 OD2	2.5	17.3
	Tyr12 O	2.3	23.1
	Asn14OD1	2.4	24.9
	Asp19 OD2	2.3	25.1
	WA O	2.4	18.4
	WB O	2.5	20.4
Average		2.4	21.0
Mn ²⁺	Glu8 OE2	2.3	20.7
	Asp10 OD2	2.2	16.8
	Asp19 OD1	2.3	25.17
	His24 NE2	2.3	20.01
	WC O	2.2	16.9
	WD O	2.1	16.7
Average		2.2	19.4

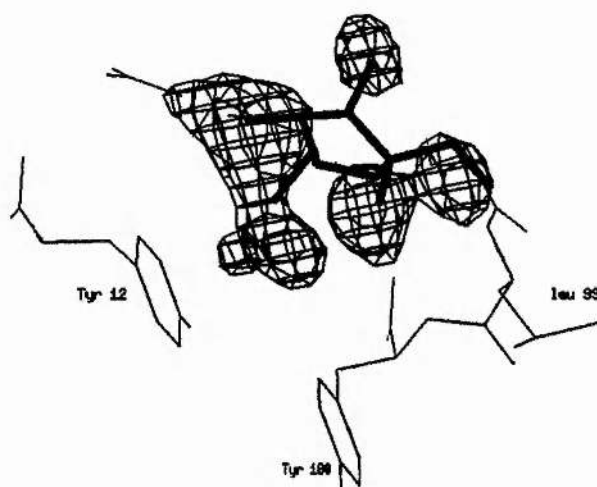
Sugar binding.

There is noisy difference electron density in the binding sites of the protein. In some cases it is difficult to say whether this is water or a carbohydrate of partial occupancy. In all subunits, changes in the protein structure at the binding site, with reference to the native structure, are similar to those seen in other con A carbohydrate structures. This is

particularly clear for in Leu-99 and Arg-228, strongly suggesting occupancy of the binding site.

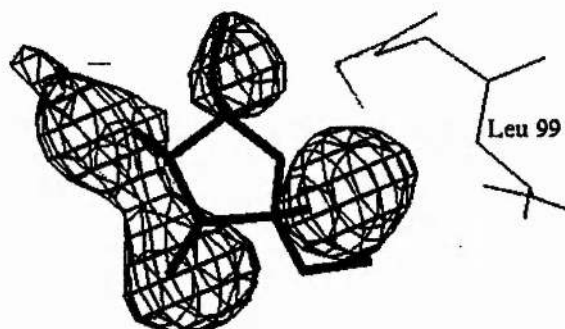
Subunit A: In this subunit there is density present in the binding site which can correspond with positions of the fructose oxygen atoms (Figure 5.16). This density however, also corresponds with the positions of four water molecules seen in the 0.94Å native structure (Deacon *et al.*, 1997) and it is difficult to tell simply from the density whether this is a binding site partially occupied by fructose or occupied by water molecules.

Figure 5.16. Difference electron density contoured at 2σ observed in the subunit A binding site with α-D-fructose superimposed.



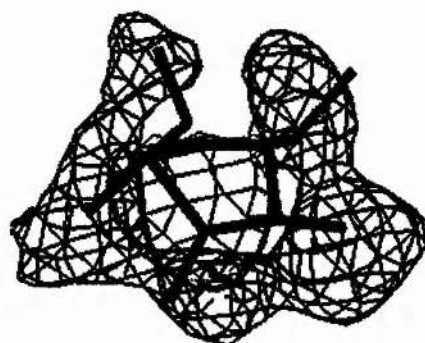
Subunit B: In this subunit, again the density corresponds with the positions of water molecules (Figure 5.17), however there is additional density which is not accounted for by water molecules seen in the native structure. The binding site is surrounded by crystal contacts, in particular a symmetry related His-205 which sits in the position of the reducing sugar in the α1-2 mannobiose complex (Chapter 2).

Figure 5.17. Difference electron density contoured at 2σ observed in the subunit B binding site with α -D-fructose superimposed.



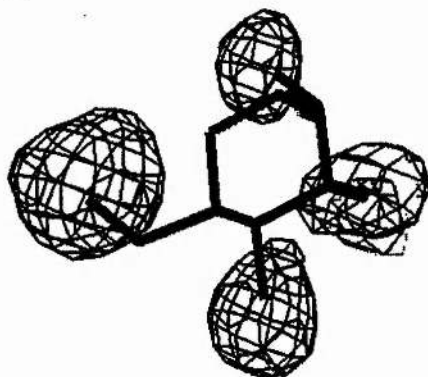
Subunit C: The binding site in subunit C shows the strongest difference density (Figure 5.18). The binding site has no crystal contacts in close proximity and the density seen cannot all be accounted for by water molecules, although it is still impossible to unambiguously model a saccharide into the structure.

Figure 5.18. Difference density contoured at 2σ seen in the binding site of subunit C with β -D-fructose superimposed.



Subunit D: This subunit contains the most ambiguous difference density (Figure 5.19). The density seen corresponds exactly with waters seen in this region in the native structure and there is no evidence from the density for any ring structure.

Figure 5.19. Difference electron density contoured at 2σ seen in the binding site of subunit D with methyl α -D-mannopyranoside superimposed in its position seen in the methyl α -D-mannopyranoside - con A structure.



5.6. Conclusions.

P2₁2₁2₁ complex: In this crystal form of the fructose con A complex, fructose can be clearly seen to be occupying the binding site of one subunit. It is not obvious whether the conformation of the sugar is α or β . A slightly lower R factor results when refined with α , however difference density appears in the position for the β configuration when refined with α . In addition to this the α configuration makes a stronger hydrogen bond

tetramer, to varying degrees. In addition to this, the presence of water molecules cannot account for all of the difference electron density seen in the binding sites. In particular subunit C. Lastly, the changes seen in the binding site loops with respect to the native structure are similar with changes seen in other carbohydrate bound con A structures.

Two different crystal forms of con A grown in the presence of fructose have been obtained with different ratios of carbohydrate to protein, and both give a similar result, ie partial occupancy of the monosaccharide binding site. The packing of the tetramers indicate that these are carbohydrate bound structures. However, inspection of difference electron density reveals that the carbohydrate is present in low occupancy. It is impossible to quantify this as the density is too weak to even model a carbohydrate in to the binding site apart from in one subunit of the $P2_12_12_1$ structure.

In comparing the structure of fructose with mannose it is apparently clear that the atoms known to be essential for carbohydrate binding to con A (Goldstein *et al.*, 1974) are present in equivalent positions in fructose to mannose. Why equivalent binding of fructose is not observed is not clear from these structures.

5.7. Future work.

There is an obvious need for further work investigating the interaction between con A and furanose sugars in order to rationalise the poor occupancy of the con A binding sites by D-fructose in these structures by both calorimetric methods and higher resolution structures.

Appendix 1.
Rerefinement of the
con A - methyl α -D-mannopyranoside structure.

The 2.0Å methyl α -D-mannopyranoside - con A structure (Naismith *et al.*, 1994) was refined using CNS (Brunger and Adams, 1998) in order to be able to make a useful comparison with the con A complexes described in this thesis which were refined using CNS. The coordinates from the PDB (accession number 5CNA) were imported into CNS and the model refined according to the protocol followed in Chapters 2 to 5 of this thesis. Statistics on the initial and final model are shown in Table A1.

Table A1. Statistics on the con A - methyl α -D-mannopyranoside structure before and after CNS refinement.

	Initial	Rerefined
Rfree (%)	19.1	23.5
R factor (%)	19.0	21.1
Bond r.m.s. deviation (Å) ^a	0.02	0.007
Angle r.m.s. deviation (°) ^a	2.9	1.7
Dihedrals r.m.s. deviation (°)	26.7	26.0
Impropers r.m.s. deviation (°)	1.3	0.87
NCS deviation	0.37	0.08
B-factor bonded atoms r.m.s. deviation (Å ²) ^b	4.0	1.4
Ramachandran core/additional/ generously allowed/ disallowed (%) ^c	85.0/ 14.5/ 0.2/ 0.2	86.6/ 13.3/ 0.1/ 0.0
Protein mean B-factor (Å ²) ^b (all)	32.4	31.9
Protein mean B-factor (Å ²) ^b (main chain)	30.4	31.3
Protein mean B-factor (Å ²) ^b (side chain)	34.6	32.7
Sugar mean B-factor (Å ²)	31.7	31.8
Solvent mean B (Å ²)	49.6	49.3

^a r.m.s. deviation from Engh and Huber ideal values (Engh and Huber, 1991). ^b Calculated with MOLEMAN (G.J. Kleywegt, unpublished program). All stereochemically modelled atoms were removed prior to B-factor analysis, all bonded atoms including those in the sugars are included in the calculation of r.m.s. B-factor deviation for bonded atoms. ^c core and additionally allowed regions as defined by PROCHECK (Laskowski *et al.*, 1993).

The initial R factors increased immediately upon positional and temperature factor refinement. Application of NCS restraints, two rounds of CNS refinement and one round of manual refinement in which a number of small changes were made in the protein structure lead to a stable R factor which was higher than at the start of refinement. However the model is of improved quality. This means that although the original structure was a better fit of the model to the experimental data, the protein contained more regions distorted from ideality in order to achieve this. The new model was used to make comparisons in Chapter 2. This model makes a more useful comparison as it has had the bulk solvent correction and anisotropic B factor correction applied to it in the same way as the refined structures in this thesis.

Bibliography.

Allen, K. N., Lavie, A., Glasfeld, A., Tanada, T. N., Gerrity, D. P., Carlson, S. C., Farber, G. K., Petsko, G. A. and Ringe, D. (1994) Role of the Divalent Metal-Ion in Sugar Binding, Ring-Opening, and Isomerization By D-Xylose Isomerase - Replacement of a Catalytic Metal By an Amino-Acid, *Biochemistry*, **33**, 1488-1494.

Bains, G., Lee, R. T., Lee, Y. C. and Freire, E. (1992) Microcalorimetric Study of Wheat-Germ-Agglutinin Binding to N- Acetylglucosamine and Its Oligomers, *Biochemistry*, **31**, 12624-12628.

Baker, E. N. and Hubbard, R. E. (1984) Hydrogen-Bonding in Globular-Proteins, *Progr. Biophys. Mol. Biol.*, **44**, 97-179.

Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. and Vijayan, M. (1996) Conformation, Protein-Carbohydrate Interactions and a Novel Subunit Association in the Refined Structure of Peanut Lectin-Lactose Complex, *J. Mol. Biol.*, **259**, 281-296.

Bhattacharyya, L., Brewer, C. F., Brown, R. D. and Koenig, S. H. (1985) Proton and Deuteron Nuclear Magnetic-Relaxation Dispersion Studies of Ca-2&-Mn-2&-Lentil Lectin and Ca-2&-Mn-2&-Pea Lectin - Evidence For a Site of Solvent Exchange in Common With Concanavalin-a, *Biochemistry*, **24**, 4985-4990

Bhattacharyya, L. and Brewer, C. F. (1988) Lectin-Carbohydrate Interactions - Studies of the Nature of Hydrogen- Bonding Between D-Galactose and Certain D-Galactose-Specific Lectins, and Between D-Mannose and Concanavalin-A, *Eur. J. Biochem.*, **176**, 207-212.

Bouckaert, J., Poortmans, F., Wyns, L. and Loris, R. (1996) Sequential Structural-Changes Upon Zinc and Calcium-Binding to Metal- Free Concanavalin-a, *J. Biol. Chem.*, **271**, 16144-16150.

Bourne, Y., Ayoub, A., Rouge, P. and Cambillau C. (1994) Interaction of a Legume Lectin With 2 Components of the Bacterial- Cell Wall - a Crystallographic Study, *J. Biol. Chem.*, **269**, 9429-9435.

Bourne, Y., Bolgiano, B., Liao, D. L., Strecker, G., Cantau, P., Herzberg, O., Feizi, T. and Cambillau, C. (1994) Cross-Linking of Mammalian Lectin (Galectin-1) By Complex Biantennary Saccharides, *Nat. Struct. Biol.*, **1**, 863-870.

Bourne, Y., Mazurier, J., Legrand, D., Rouge, P., Montreuil, J., Spik, G. and Cambillau, C. (1994) Structures of a Legume Lectin Complexed With the Human Lactotransferrin N2 Fragment, and With an Isolated Biantennary Glycopeptide - Role of the Fucose Moiety, *Structure*, **2**, 209-219.

Bourne, Y., Rouge, P., and Cambillau, C. (1990) X-Ray Structure of a (Alpha-Man(1-3)Beta-Man(1-4)GlcNAc)-Lectin Complex At 2.1-Å Resolution - the Role of Water in Sugar-Lectin Interaction, *J. Biol. Chem.*, **265**, 18161-18165.

Bourne, Y., Rouge, P. and Cambillau, C. (1992) X-Ray Structure of a Biantennary Octasaccharide-Lectin Complex Refined At 2.3Å Resolution, *J. Biol. Chem.*, **267**, 197-203.

Bourne, Y., Rouge, P. and Cambillau, C. (1990) 3-Dimensional Structures of Complexes of *Lathyrus ochrus* Isolectin I With Glucose and Mannose - Fine Specificity of the Monosaccharide- Binding Site, *Proteins*, **8**, 365-376.

Bradbrook, G. M., Gleichman, T., Harrop, S. J., Habash, J., Raftery, J., Kalb, J., Yariv, J., Hillier, I. H. and Helliwell, J. R. (1998) X-ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes - Relating structure to thermodynamics of binding, *J. Chem. Soc. Faraday Trans.*, **94**, 1603-1611.

Brewer, C. F. (1993) Studies of the Extended Binding-Site Specificity of Concanavalin-a, *Abs. Am. Chem. Soc.*, **205**, 2-CARB.

Brewer, C. F. and Brown III, R. D. (1979) Mechanism of Binding of Mono- and Oligosaccharides to Concanavalin A: A Solvent Proton Magnetic Relaxation Dispersion Study, *Biochemistry*, **18**, 2555-2562.

Brewer, C. F., Dam, T. K., Oscarson, S., Rozwarski, D. and Sacchettini, J. (1997) Thermodynamics of binding deoxy carbohydrate analogs of the core trimannoside to concanavalin A and Dioclea grandiflora lectin, *Glycobiology*, **7**, 108-108.

Brunger, A. T. (1992). \bar{X} -PLOR (Version 3.1) Manual: A system for X-ray crystallography and NMR. New Haven, U.S.A., Yale University Press.

Brunger, A. T. and Adams, P. D. (1998). CNS: Crystallography and NMR system, Development Version 0.1.

Burmeister, W. P., Ruigrok, R. W. H. and Cusack, S. (1992) The 2.2-Å Resolution Crystal-Structure of Influenza-B Neuraminidase and Its Complex With Sialic-Acid, *EMBO J.*, **11**, 49-56.

Carver, J. P., MacKenzie, A. E. and Hardman, K. D. (1985) Molecular Model for the Complex Between Concanavalin A and a Biantennary-Complex Class Glycopeptide, *Biopolymers*, **24**, 49-63.

Casset, F., Hamelryck, T., Loris, R., Brisson, J. R., Tellier, C., Daothi, M. H., Wyns, L., Poortmans, F., Perez, S. and Imberty, A. (1995) Nmr, Molecular Modeling, and Crystallographic Studies of Lentil Lectin-Sucrose Interaction, *J. Biol. Chem.*, **270**, 25619-25628.

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

Cheetham, J. C., Artymiuk, P. J. and Phillips, D. C. (1992) Refinement of an Enzyme Complex With Inhibitor Bound At Partial Occupancy - Hen Egg-White Lysozyme and Tri-N-Acetylchitotriose At 1.75-Angstrom Resolution, *J. Mol. Biol.*, **224**, 613-628.

Chervenak, M. C., Oas, T. G. and Toone, E. J. (1992) Structure and Energetics of Concanavalin A Oligosaccharide Binding, *Abs. Am. Chem. Soc.*, **204**, Pt1, 102-CARB.

Chervenak, M. C. and Toone, E. J. (1994) A Direct Measure of the Contribution of Solvent Reorganization to the Enthalpy of Ligand-Binding, *J. Am. Chem. Soc.*, **116**, 10533-10539.

Chervenak, M. C. and Toone, E. J. (1995) Calorimetric Analysis of the Binding of Lectins With Overlapping Carbohydrate-Binding Ligand Specificities, *Biochemistry*, **34**, 5685-5695.

Chervenak, M. C. and Toone, E. J. (1996) Analysis of the binding specificities of oligomannoside-binding proteins using methylated monosaccharides, *Bioorg. Med. Chem.*, **4**, 1963-1977.

Colman, P. M., Varghese, J. N. and Laver, W. G. (1983) Structure of the Catalytic and Antigenic Sites in Influenza-Virus Neuraminidase, *Nature*, **303**, 5912, 41-44.

Crennell, S., Garman, E., Laver, G., Vimr, E. and Taylor, G. (1994) Crystal-Structure of Vibrio-Cholerae Neuraminidase Reveals Dual Lectin-Like Domains in Addition to the Catalytic Domain, *Structure*, **2**, 535-544.

Cyglar, R., Rose, D. R. and Bundle, D. R. (1991) Recognition of a Cell-Surface Oligosaccharide of Pathogenic *Salmonella* by an Antibody Fab Fragment, *Science*, **253**, 442-445.

Deacon, A., Gleichmann, T. Kalb, J., Price, H., Rafferty, J., Bradbrook, G., Yariv, J. and Heeliwell, J. R. (1997) The structure of concanavalin A and its bound solvent structure determined with small molecule accuracy at 0.94Å resolution, *J. Chem. Soc., Faraday Trans.*, **53**, 4305-4312.

Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S. and Dauter, Z. (1993) Structures of the Lectin-IV of Griffonia-Simplicifolia and Its Complex With the Lewis-B Human Blood-Group Determinant At 2.0- Angstrom Resolution, *J. Mol. Biol.*, **230**, 950-965.

Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb, A. J., Dodson, E. J., Papiz, M. Z., Wan, T., and Campbell, J. (1989) The Structure of the Saccharide-Binding Site of Concanavalin-A, *EMBO J.*, **8**, 2189-2193.

Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F., Sacchettini, J. C. (1995) X-Ray Crystal-Structure of the Soybean Agglutinin Cross-Linked With a Biantennary Analog of the Blood-Group-I Carbohydrate Antigen, *Biochemistry*, **34**, 4933-4942.

DiGabriele, A. D., Lax, I., Chen, D. I., Svahn, C. M., Jaye, M., Schlessinger, J., Hendrickson, W. A. (1998) Structure of a heparin-linked biologically active dimer of fibroblast growth factor, *Nature*, **393**, 6687, 812-817.

Dimick, S. M., Powell, S. C., McMahon, S. A., Moothoo, D. N., Naismith, J. H. and Toone, E. J. (1998) On the Nature and Origin of the Cluster Glycoside Effect, *J. Am. Chem. Soc.*, Accepted.

Drickamer, K. (1992) Engineering Galactose-Binding Activity Into a C-Type Mannose-Binding Protein, *Nature*, **360**, 6400, 183-186.

Dunitz, J. D. (1995) Win Some, Lose Some - Enthalpy-Entropy Compensation in Weak Intermolecular Interactions, *Chemistry & Biology*, **2**, 709-712.

Elgavish, S. and Shaanan, B. (1998) Structures of the Erythrin coralloidendron Lectin and of its Complexes with Mono- and Disaccharides, *J. Mol. Biol.*, **277**, 917-932.

Engh, R. A. and Huber, R. (1991) Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement, *Acta Cryst. A*, **47**, 392-400.

Finer-Moore, J. S., Kossiakoff, A. A., Hurley, J. H., Earnest, T. and Stroud, R. M. (1992) Solvent Structure in Crystals of Trypsin Determined By X-Ray and Neutron-Diffraction, *Proteins: Struct., Funct., Genet.*, **12**, 203-222.

Goldstein, I. J., Reichert, C. M. and Misaki, A. (1974) Interaction of Concanavalin A with Model Substrates, *Ann. NY Acad. Sci.*, **234**, 283-295.

Goldstein, I. J., Hollerman, C. E. and Smith, E. E. (1965) Protein - Carbohydrate Interaction II. Inhibition studies on the Interaction of concanavalin A with Polysaccharides, *Biochemistry*, **4**, 876-883.

Goldstein, I. J. and Poretz, R. D. (1986). Isolation and Chemical Properties of Lectins. The lectins: Properties, Functions and Applications. In *Biology and Medicine*. I. E. Leiner, N. Sharon and I. J. Goldstein. eds. Orlando, Academic Press Inc., pp. 35-244.

Grunwald, E. and Steel, C. (1995) Solvent Reorganization and Thermodynamic Enthalpy-Entropy Compensation, *J. Am. Chem. Soc.*, **117**, 5687-5692.

Gupta, D. and Brewer, C. F. (1996) Thermodynamic Binding-Studies of Concanavalin-A, *Abs. Am. Chem. Soc.*, **211**, Pt1, 76-CARB.

Gupta, D., Cho, M. J., Cummings, R. D. and Brewer, C. F. (1996) Thermodynamics of Carbohydrate-Binding to Galectin-1 From Chinese- Hamster Ovary Cells and 2 Mutants - a Comparison With 4 Galactose- Specific Plant-Lectins, *Biochemistry*, **35**, 15236-15243.

Gupta, D., Dam, T. K., Oscarson, S. and Brewer, C. F. (1997) Thermodynamics of Lectin Carbohydrate Interactions. Binding of the Core Trimannoside of Asparagine Linked Carbohydrates and Deoxy analogs to Concanavalin A, *J. Biol. Chem.*, **272**, 6388-6392.

Gupta, D., Rao, N., Puri, K. D., Matta, K. L. and Surolia, A. (1992) Thermodynamic and Kinetic-Studies On the Mechanism of Binding of Methylumbelliferyl Glycosides to Jacalin, *J. Biol. Chem.*, **267**, 8909-8918.

Habash, J., Raferty, J., Weiserberger, S., Cassetta, A., Lehmann, M. S., Hoghoj, P., Wilkinson, C., Campbell, J. W. and Helliwell, J. R. (1997) Neutron Laue Diffraction

Study of Concanavalin A: The Proton of Asp28, *J. Chem. Soc., Faraday Trans.*, **93**, 4313-4317.

Harrop, S. J., Helliwell, J. R., Wan, T. C. M., Kalb, A. J., Tong, L. and Yariv, J. (1996) Structure Solution of a Cubic-Crystal of Concanavalin A Complexed With Methyl Alpha-D-Glucopyranoside, *Acta Cryst. D*, **52**, 143-155.

Hester, G. and Wright, C. S. (1996) The Mannose-Specific Bulb Lectin From *Galanthus-Nivalis* (Snowdrop) Binds Monomannoside and Dimannoside At Distinct Sites - Structure- Analysis of Refined Complexes At 2.3-Angstrom and 3.0-Angstrom Resolution, *J. Mol. Biol.*, **262**, 516-531.

Homans, S. W. (1995) Conformational Studies on Oligosaccharides. In *Biomolecular NMR spectroscopy*. J. N. S. Evans, ed. Oxford, Oxford University Press, pp 230-257.

Imberty, A., Delage, M. M., Bourne, Y., Cambillau, C. and Perez, S. (1991) Data bank of three-dimensional structures of disaccharides: Part II, N-acetyllactosamine type N-glycans, Comparison with the crystal structure of a biantennary octasaccharide, *Glycoconj. J.*, **8**, 456-483.

Imberty, A., Gerber, S., Tran, V. and Perez, S. (1990) Data Bank of Three-Dimensional Structures of Disaccharides, A Tool to Build 3-D Structures of Oligosaccharides. Part I. Oligomannose Type N-Glycans, *Glycoconj. J.*, **7**, 27-54.

Johnson, L. N., Acharya, K. R., Jordan, M. D. and McLaughlin, P. J. (1990) Refined Crystal-Structure of the Phosphorylase Heptulose-2-Phosphate Oligosaccharide AMP Complex, *J. Mol. Biol.*, **211**, 645-661.

Jones, T. A., Zou, J.-Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models, *Acta Cryst. A*, **47**, 110-119.

Kaku, H. and Goldstein, I. J. (1992) Interaction of Linear Manno-Oligosaccharides With 3 Mannose-Specific Bulb Lectins - Comparison With Mannose Glucose-Binding Lectins, *Carbohydr. Res.*, **229**, 337-346.

Kalb, A. J., Frolow, F., Yariv, J. and Eisenstein, M. (1995) Properties of a Crystal of the Complex of Methyl α -D- Arabinofuranoside With Concanavalin-a, *Acta Cryst. D*, **51**, 1077-1079.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures, *J. Appl. Cryst.*, **26**, 548-558.

Lavie, A., K. Allen, N., Petsko, G. A. and Ringe, D. (1994) X-Ray Crystallographic Structures of D-Xylose Isomerase-Substrate Complexes Position the Substrate and Provide Evidence For Metal Movement During Catalysis, *Biochemistry*, **33**, 5469-5480.

Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H. and Rini, J. M. (1993) X-Ray Crystal-Structure of the Human Dimeric S-Lac Lectin, L-14-Ii, in Complex With Lactose At 2.9-Angstrom Resolution, *J. Biol. Chem.*, **268**, 27034-27038.

Loris, R., Casset, F., Bouckaert, J., Pletinckx, J., Daothi, M. H., Poortmans, F., Imberty, A., Perez, S. and Wyns, L. (1994) The Monosaccharide Binding-Site of Lentil Lectin - an X-Ray and Molecular Modeling Study, *Glycoconj. J.*, **11**, 507-517.

Loris, R., Maes, D., Poortmans, F., Wynsm L. and Bouckaert, J. (1996) A Structure of the Complex Between Concanavalin A and Methyl-3,6-di-O-(α -D-mannopyranosyl)- α -D-mannopyranoside Reveals Two Binding Modes, *J. Biol. Chem.*, **271**, 30614-30618.

Mandal, D. K., Bhattacharyya, L., Koenig, S. H., Brown, R. D., Oscarson, S. and Brewer, C. F (1994) Studies of the Binding-Specificity of Concanavalin-A - Nature of the Extended Binding-Site For Asparagine-Linked Carbohydrates, *Biochemistry*, **33**, 1157-1162.

Mandal, D. K., Kishore, N. and Brewer, C. F. (1994) Thermodynamics of Lectin-Carbohydrate Interactions - Titration Microcalorimetry Measurements of the Binding of N-Linked Carbohydrates and Ovalbumin to Concanavalin A, *Biochemistry*, **33**, 1149-1156.

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

Merritt, E. A., Sarfaty, S., Vandenakker, F., Lhoir, C., Martial, J. A. and Wol, W. G. J. (1994) Crystal-Structure of Cholera-B Pentamer Bound to Receptor G(M1) Pentasaccharide, *Protein Science*, **3**, 166-175.

Miller, D. M., Olson, J. S., Pflugrath, J. W. and Quioco, F. A. (1983) Rates of Ligand-Binding to Periplasmic Proteins Involved in Bacterial Transport and Chemotaxis, *J. Biol. Chem.*, **258**, 3665-3672.

Moothoo, D. N. and Naismith, J. H. (1998a) Concanavalin A distorts the β -GlcNAc-(1-2)-Man linkage of β -GlcNAc-(1-2)- α -Man-(1-3)-[β -GlcNAc-(1-2)- α -Man-(1-6)]-Man upon binding, *Glycobiology*, **8**, 173-181.

Moothoo, D. N. and Naismith, J. H. (1998b) A general method for co-crystallisation of concanavalin A with carbohydrates, *Acta. Cryst. D*, In press.

Moothoo, D. N., McMahon, S. A., Dimick, S. M., Toone, E. J. and Naismith, J. H. (1998) Crystallisation of succinylated concanavalin A bound to a synthetic bivalent ligand and preliminary structural analysis, *Acta. Cryst. D*, **54**, In press.

Mortell, K. H., Weatherman, R. V. and Kiessling, L. L. (1996) Recognition Specificity of Neoglycopolymers Prepared By Ring-Opening Metathesis Polymerization, *J. Am. Chem. Soc.*, **118**, 2297-2298.

Muller, N (1991) Model-Calculations of Changes of Thermodynamic Variables For the Transfer of Nonpolar Solutes From Water to Water-D₂, *J. Sol. Chem.*, **20**, 669-680.

Nagahora, H., Harata, K., Muraki, M. and Jigami, Y. (1995) Site-Directed Mutagenesis and Sugar-Binding Properties of the Wheat- Germ-Agglutinin Mutants Tyr73phe and Phe116tyr, *Eur. J. Biochem.*, **233**, 27-34.

Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb, A. J. and Yariv, J. (1994) Refined Structure of Concanavalin-A Complexed With Methyl Alpha-D- Mannopyranoside At 2.0 Angstrom Resolution and Comparison With the Saccharide-Free Structure, *Acta Cryst. D*, **50**, 847-858.

Naismith, J. H. and Field, R. A. (1996) Structural Basis of Trimannoside Recognition By Concanavalin-A, *J. Biol. Chem.*, **271**, 972-976.

Naismith, J. H., Habash, J., Harrop, S., Helliwell, J. R., Hunter, W. N., Wan, T. C. M., Weisgerber, S., Kalb, A. J. and Yariv, J. (1993) Refined Structure of Cadmium-Substituted Concanavalin-A At 2.0 Angstrom Resolution, *Acta Cryst. D*, **49**, 561-571.

Navaza, J. (1994) AMoRe: An Automated Package for Molecular Replacement, *Acta Cryst. A*, **50**, 157-163.

Otwinowski, Z. (1993). Oscillation Data Reduction Program. In *Proceedings of the CCP4 study weekend: Data collection and processing.*, Warrington, U.K., Daresbury Laboratory.

Page, M. L. and Jencks, W. P. (1971) Entropic Contributions to Rate Accelerations in Enzymatic and Intramolecular Reactions and the Chelate Effect, *Proc. Natl. Acad. Sci. USA*, **68**, 1678-1683

Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. and Suguna, K. (1998) Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 angstrom resolution, *J. Mol. Biol.*, **276**, 787-796.

Quesenberry, M. S., Lee, R. T. and Lee, Y. C. (1997) Difference in the binding mode of two mannose-binding proteins: Demonstration of a selective minicluster effect, *Biochemistry*, **36**, 2724-2732.

Quioco, F. A. (1988) Molecular-Features and Basic Understanding of Protein Carbohydrate Interactions - the Arabinose-Binding Protein Sugar Complex, *Cur. Topics Microbiol. and Immunol.*, **139**, 135-148.

Quioco, F. A. (1993) Probing the Atomic Interactions Between Proteins and Carbohydrates, *Biochem. Soc. Trans.*, **21**, 442-448.

Ramakrishnan, C. and Ramachandran, G. N. (1965) Stereochemical criteria for polypeptide and protein chain conformation, *Biophys. J.*, **5**, 909-933.

Ravishankar, R., Ravindran, M., Suguna, K., Surolia, A. and Vijayan, M. (1997) Crystal structure of the peanut lectin - T-antigen complex. Carbohydrate specificity generated by water bridges, *Curr. Sci.*, **72**, 855-861.

Read, R. J. (1986) Improved Fourier Coefficients for Maps Using Phases From Partial Structures With Errors, *Acta Cryst. A*, **42**, 140-149.

Reddy, V. S. and Rao, V. S. R. (1992) Modes of Binding of Alpha(1-2) Linked Manno-Oligosaccharides to Concanavalin-A, *Int. J. Biol. Macromol.*, **14**, 185-192.

Rini, J. M. (1995) Lectin Structure, *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 551-577.

Rini, J. M., Hardman, K. D., Einspahr, H., Suddath, F. L. and Carver, J. P. (1993) X-Ray Crystal-Structure of a Pea Lectin-Trimannoside Complex At 2.6 Angstrom Resolution, *J. Biol. Chem.*, **268**, 10126-10132.

Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. and Vijayan, M. (1996) A Novel Mode of Carbohydrate-Recognition in Jacalin, a Moraceae Plant Lectin With a Beta-Prism Fold, *Nat. Struct. Biol.*, **3**, 596-603.

Sauter, N. K., Hanson, J. E., Glick, G. D., Brown, J. H., Crowther, R. L., Park, S. J., Skehel, J. J. and Wiley, D. C. (1992) Binding of Influenza-Virus Hemagglutinin to Analogs of Its Cell- Surface Receptor, Sialic-Acid - Analysis By Proton Nuclear-Magnetic- Resonance Spectroscopy and X-Ray Crystallography, *Biochemistry*, **31**, 9609-9621.

Schwarz, F. P., Misquith, S. and Surolia, A. (1996) Effect of Substituent On the Thermodynamics of D-Glucopyranoside Binding to Concanavalin-A, Pea (*Pisum-Sativum*) Lectin and Lentil (*Lens-Culinaris*) Lectin, *Biochem. J.*, **316**, 123-129.

Scott, J. K., Longanathan, D., Easley, R.B., Gong, X. and Goldstein, I.J. (1992) A Family of Concanavalin A Binding Peptides from a hexapeptide Epitope Library, *Proc. Natl. Acad. Sci. USA*, **89**, 5398-5402.

Shaanan, B., Lis, H. and Sharon, N. (1991) Structure of a Legume Lectin With an Ordered N-Linked Carbohydrate in Complex With Lactose, *Science*, **254**, 862-866.

Sharma, V. and Surolia, A.(1997) Analysis of carbohydrate recognition by legume lectins: Size of the combining site loops and their primary specificity, *J. Mol. Biol.*, **267**, 433-445.

Sharon, N. and Lis, H. (1989) In *Lectins*. London, Chapman and Hall.

Sharon, N. and Lis, H. (1989) Lectins as cell recognition molecules, *Science*, **246**, 227-233.

Sharon, N. and Lis, H. (1990) Legume Lectins - A Large Family of Homologous Proteins, *FASEB J.*, **4**, 3198-3208.

Shibuya, N., Goldstein, I. J., Vandamme, E. J. M. and Peumans, W. J. (1988) Binding-Properties of a Mannose-Specific Lectin From the Snowdrop (*Galanthus-Nivalis*) Bulb, *J. Biol. Chem.*, **263**, 728-734.

Sigurskjold, B. W. and Bundle, D. R. (1992) Thermodynamics of Oligosaccharide Binding to a Monoclonal-Antibody Specific For a Salmonella O-Antigen Point to Hydrophobic Interactions in the Binding-Site, *J. Biol. Chem.*, **267**, 8371-8376.

Sixma, T. K., Pronk, S. E., Kalk, K. H., Vanzanten, B. A. M., Berghuis, A. M.

Hol, W. G. J. (1992) Lactose Binding to Heat-Labile Enterotoxin Revealed By X-Ray Crystallography, *Nature*, **355**, 561-564.

Stehle, T., Yan, W., Benjamin, T. L. and Harrison, S. C. (1994) Structure of Murine Polyomavirus Complexed With an Oligosaccharide Receptor Fragment, *Nature*, **369**, 160-163.

Stein, P. E., Boodhoo, A., Armstrong, G. D., Heerze, L. D., Cockle, S. A., Klein, M. H. and Read, R. J. (1994) Structure of a Pertussis Toxin Sugar Complex As a Model For Receptor- Binding, *Nat. Struct. Biol.*, **1**, 591-596.

Strynadka, N. C. J. and James, M. N. G. (1991) Lysozyme Revisited - Crystallographic Evidence For Distortion of an N-Acetylmuramic Acid Residue Bound in Site-D, *J. Mol. Biol.*, **220**, 401-424.

Sumner, J. B. (1919) The Globulins of the Jack Bean *Canavalia ensiformis*, *J. Biol. Chem.*, **37**, 137-142.

Surolia, A., Sharon, N. and Schwarz, F. P. (1996) Thermodynamics of Monosaccharide and Disaccharide Binding to Erythrina Corallodendron Lectin, *J. Biol. Chem.*, **271**, 17697-17703.

Toone, E. J. (1994) Structure and Energetics of Protein Carbohydrate Complexes, *Curr. Opin. Struct. Biol.*, **4**, 719-728.

Vermersch, P. S., Tesmer, J. J. G. and Quioco, F. A. (1992) Protein Ligand Energetics Assessed Using Deoxy and Fluorodeoxy Sugars in Equilibrium Binding and High-Resolution Crystallographic Studies, *J. Mol. Biol.*, **4**, 923-929.

Vyas, M. N., Vyas, N. K. and Quioco, F. A. (1994) Crystallographic Analysis of the Epimeric and Anomeric Specificity of the Periplasmic Transport Chemosensory Protein-Receptor For D-Glucose and D-Galactose, *Biochemistry*, **33**, 4762-4768.

Watowich, S. J., Skehel, J. J. and Wiley, D. C. (1994) Crystal-Structures of Influenza-Virus Hemagglutinin in Complex With High-Affinity Receptor Analogs, *Structure*, **2**, 719-731.

Weis, W., Brown, J. H., Cusack, S., Paulson, J. C., Skehel, J. J. and Wiley, D. C. (1988) Structure of the Influenza-Virus Hemagglutinin Complexed With Its Receptor, Sialic-Acid, *Nature*, **333**, 426-431.

Weis, W. I., K. Drickamer and Hendrickson, W. A. (1992) Structure of a C-Type Mannose-Binding Protein Complexed With an Oligosaccharide, *Nature*, **360**, 127-134.

Withers, S. G., Street, I. P. and Rettig, S. J. (1986) The Preferred Conformation of 2-Fluoro-2-Deoxy Beta-D-Mannopyranosyl Fluoride - an X-Ray Crystallographic and Two-Dimensional Proton Nuclear Magnetic-Resonance Study, *Can. J. Chem.*, **64**, 232-236.

Wright, C. S. (1984) Structural Comparison of the 2 Distinct Sugar Binding-Sites in Wheat- Germ Agglutinin Isolectin-II, *J. Mol. Biol.*, **178**, 91-104.

Wright, C. S. (1990) 2.2Å Resolution Structure Analysis of Two Refined N-acetylneuraminyllactose Wheat Germ Agglutinin Isolectin Complexes, *J. Mol. Biol.*, **215**, 635-651.

Wright, C. S. (1992) Crystal-Structure of a Wheat-Germ-Agglutinin Glycophorin-Sialoglycopeptide Receptor Complex - Structural Basis For Cooperative Lectin-Cell Binding, *J. Biol. Chem.*, **267**, 14345-14352.

Wright, C. S. and Hester, G. (1996) The 2.0 Angstrom Structure of a Cross-Linked Complex Between Snowdrop Lectin and a Branched Mannopentaose - Evidence For 2 Unique Binding Modes, *Structure*, **4**, 1339-1352.

Wright, C. S. and Kellogg, G. E. (1996) Differences in Hydrophobic Properties of Ligand-Binding At 4 Independent Sites in Wheat-Germ Agglutinin-Oligosaccharide Crystal Complexes, *Prot. Sci.*, **5**, 1466-1476.

Zdanov, A., Li, Y., Bundle, D. R., Deng, S. -J., MacKenzie, C. R., Narang, S. A., Young, N. M. and Cygler, M. (1994) Structure of a Single-Chain Antibody Variable Domain (Fv) Fragment Complexed with a Carbohydrate Antigen at 1.7Å Resolution, *Proc. Natl. Acad. Sci. USA*, **91**, 6423-6427.

Zhu, X., Hsu, B. T. and Rees, D. C. (1993) Structural studies of the anti-ulcer drug sucrose octasulfate to acidic fibroblast growth factor, *Structure*, **1**, 27-34.

ISSN 0959-6653
CODEN GLYCEJ

GLYCOBIOLOGY

Volume 8 Number 2

February 1998



Official Journal of the Society for Glycobiology
Oxford University Press

Concanavalin A distorts the β -GlcNAc-(1 \rightarrow 2)-Man linkage of β -GlcNAc-(1 \rightarrow 2)- α -Man-(1 \rightarrow 3)-[β -GlcNAc-(1 \rightarrow 2)- α -Man-(1 \rightarrow 6)]-Man upon binding

Davina N. Moothoo and James H. Naismith¹

Centre for Biomolecular Sciences, Purdie Building, The University, St. Andrews, Scotland KY16 9ST, United Kingdom

Received on June 26, 1997; revised on August 19, 1997; accepted on August 31, 1997

¹To whom correspondence should be addressed

Carbohydrate recognition by proteins is a key event in many biological processes. Concanavalin A is known to specifically recognize the pentasaccharide core (β -GlcNAc-(1 \rightarrow 2)- α -Man-(1 \rightarrow 3)-[β -GlcNAc-(1 \rightarrow 2)- α -Man-(1 \rightarrow 6)]-Man) of N-linked oligosaccharides with a K_a of 1.41×10^6 M⁻¹. We have determined the structure of concanavalin A bound to β -GlcNAc-(1 \rightarrow 2)- α -Man-(1 \rightarrow 3)-[β -GlcNAc-(1 \rightarrow 2)- α -Man-(1 \rightarrow 6)]-Man to 2.7 Å. In six of eight subunits there is clear density for all five sugar residues and a well ordered binding site. The pentasaccharide adopts the same conformation in all eight subunits. The binding site is a continuous extended cleft on the surface of the protein. Van der Waals interactions and hydrogen bonds anchor the carbohydrate to the protein. Both GlcNAc residues contact the protein. The GlcNAc on the 1 \rightarrow 6 arm of the pentasaccharide makes particularly extensive contacts and including two hydrogen bonds. The binding site of the 1 \rightarrow 3 arm GlcNAc is much less extensive. Oligosaccharide recognition by Con A occurs through specific protein carbohydrate interactions and does not require recruitment of adventitious water molecules. The β -GlcNAc-(1 \rightarrow 2)-Man glycosidic linkage PSI torsion angle on the 1 \rightarrow 6 arm is rotated by over 50° from that observed in solution. This rotation is coupled to disruption of interactions at the monosaccharide site. We suggest destabilization of the monosaccharide site and the conformational strain reduces the free energy liberated by additional interactions at the 1 \rightarrow 6 arm GlcNAc site.

Key words: carbohydrate conformation/Con A saccharide complex/crystal structure/molecular recognition/thermodynamics

Introduction

Protein carbohydrate interactions underpin an immense and diverse range of biological processes including immune response, cell differentiation, inflammation, and infection (Ni and Tizard, 1996; Lowe and Ward, 1997). Given the key role of protein carbohydrate recognition in so many processes, it is unsurprising that there is considerable interest in potential therapeutic strategies to control, utilize, or block protein carbohydrate binding. The very ubiquity of carbohydrates that makes them so important as therapeutic targets, presents an enormous problem of selectivity as there is considerable potential for unwanted interactions. Understanding the molecular basis of high affinity oligosaccharide

protein interactions is required for the rational design of biologically active oligosaccharide analogues.

Con A binds to mannose and glucose with weak affinity (K_a (association constant) 0.82×10^4 M⁻¹) (Mandal *et al.*, 1994b; Williams *et al.*, 1992). The structure of the protein was reported in the 1970s (Hardman and Ainsworth, 1972; Reeke *et al.*, 1975) and has now been extended to 1.0 Å (Helliwell and Helliwell, 1996). The 2.9 Å α -Man-OMe Con A complex was reported in 1989 (Derewenda *et al.*, 1989) and provided a clear structural basis for the mannose/glucose selectivity (Goldstein and Poretz, 1986). This structure was extended to 2.0 Å resolution (Naismith *et al.*, 1994), and a high resolution study of the glucose complex has been completed (Harrop *et al.*, 1996). The sugar is firmly anchored to the protein by direct hydrogen bonds and van der Waals contacts burying some 75% of its accessible surface. The site of interaction between α -Man-OMe and Con A is known as the monosaccharide binding site. Such detail was not available at the oligosaccharide level until the trimannoside (α -Man-(1 \rightarrow 3)-[α -Man-(1 \rightarrow 6)]-Man) complex (K_a 3.37×10^5 M⁻¹; Mandal *et al.*, 1994b) of Con A was reported (Naismith and Field, 1996). This structure showed that trisaccharide recognition does not require recruitment of adventitious water molecules to mediate protein sugar hydrogen bonds. It also showed that all three sugar residues were involved in direct contacts (van der Waals and hydrogen bonding) to the protein. The 1 \rightarrow 6 terminal mannose binds in the monosaccharide site in a very similar manner as that observed in α -Man-OMe Con A complex (Derewenda *et al.*, 1989; Naismith *et al.*, 1994) and probably serves as the anchor to bind the other sugar residues. A second report of the trimannoside Con A complex has appeared (Loris *et al.*, 1996) and confirms the first except that in one of the four subunits the trimannoside apparently adopts a different conformation. No functional explanation of this occurrence could be given, and it is unclear whether this is an artefact of crystallization.

Con A belongs to the homologous plant lectin protein family. The function of these lectins *in vivo* remains unclear. They have, however, proved to be a valuable source of fundamental information on protein carbohydrate recognition. The structures of lectin oligo- and/or monosaccharide complexes have recently been reviewed (Rini, 1995). Plant lectins are not true monosaccharide binding proteins. They are designed to recognize high order carbohydrates, but will bind smaller fragments of the oligosaccharides, such as mannose or galactose. The structures of oligosaccharide complexes of *Lathyrus ochrus* isolectins I and II (LOLI and LOLII; Bourne *et al.*, 1992, 1994a,b), *Griffonia simplicifolia* (Delbaere *et al.*, 1993), and *Erythrina corolladendron* (Shaanan *et al.*, 1991) provided experimental evidence that recruitment of water to mediate hydrogen bonds may play an important role in binding oligosaccharides to plant lectins. Mediation of hydrogen bonds by water is a common feature of

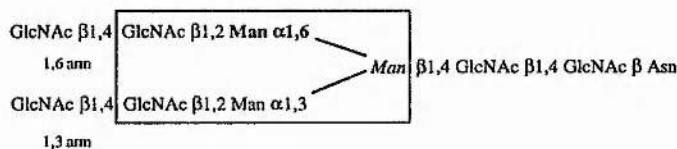


Fig. 1. The N-glycan found on the surface of many mammalian cells. The pentasaccharide is boxed, the sugar bound by the monosaccharide site is shown in boldface type. The reducing sugar of the pentasaccharide is shown in italics.

transport proteins (Mowbray and Cole, 1992; Sharff *et al.*, 1993; Quiocho and Ledvina, 1996), which completely envelop the sugar and form extremely tight complexes (K_a values in the 10^9 M^{-1} range). However, oligosaccharide complexes of other proteins (Sixma *et al.*, 1992; Weis *et al.*, 1992; Wright, 1992; Merritt *et al.*, 1994a,b; Stein *et al.*, 1994; Hester *et al.*, 1995; Hester and Wright, 1996; Vandenakker *et al.*, 1996; Wright and Hester, 1996; Wright and Kellogg, 1996) do not appear to require recruitment of additional water molecules for specific protein saccharide interactions. Interestingly structures of a FAB fragment on its own and complexed to carbohydrates from *Salmonella typhimurium* (Rose *et al.*, 1990; Cygler *et al.*, 1991, 1992; Zdanov *et al.*, 1994) contain a single structurally conserved water molecule which mediates protein oligosaccharide interactions in an analogous manner to the structural water seen in Con A (Naismith and Field, 1996). A wealth of microtitration calorimetry data are available for protein carbohydrate complexes, principally on Con A (Chervenak and Toone, 1995; Mandal *et al.*, 1994a; Mandal and Brewer, 1993; Toone, 1994; Weatherman *et al.*, 1996; Williams *et al.*, 1992) although other lectins and proteins have been studied (Bains *et al.*, 1992; Bundle *et al.*, 1994; Schwarz *et al.*, 1991, 1993; Surolija *et al.*, 1996). These data are complimentary to that obtained by crystallography and quantify the binding affinity. The precise contributions of hydrogen bonding, van der Waals contacts, sugar conformational strain, water rearrangement and surface area occlusion to binding affinity are the subject of intensive investigation as it holds the key to progress in developing modeling methodologies. Combining structural studies with calorimetry data is a powerful way to start delineating the contributions of each of these processes (Gupta *et al.*, 1997).

The pentasaccharide shown in Figure 1 is found in the core of N-linked glycans of the complex type. It binds to Con A with an affinity K_a 1.41×10^6 M^{-1} and is the tightest carbohydrate Con A complex (Mandal *et al.*, 1994b). We now report the crystal structure of concanavalin A complexed to the pentasaccharide at 2.7 Å resolution. All five sugar residues were modeled into electron density, four sugars make direct protein carbohydrate hydrogen bonds and all five make van der Waals contacts with the protein. The structural basis of the specificity of Con A for the pentasaccharide is clear. Compared to other Con A carbohydrate complexes there appear to be subtle changes in the structure of the protein and in the interactions between the sugar residues and the protein. The differences are particularly evident at the monosaccharide site and may partly explain the thermodynamic characteristics of oligosaccharide binding. The conformation of the pentasaccharide is rotated with respect to the observed solution conformation and the predicted energetic minima.

Results

Overall structure of the pentasaccharide protein complex

Two tetramers of Con A are found in the asymmetric unit of the unit cell, giving eight independent 237 residue molecules. Each monomer is a sandwich of two β sheets, the overall fold being identical to the native structure of the protein first reported in the 1970s (Hardman and Ainsworth, 1972; Reeke *et al.*, 1975). The underlying structure changes little on carbohydrate binding. Specific differences in protein structure are dealt with later. The protein main chain is well ordered except for the loop from Asn 118 to Glu 122, this region is consistently disordered or weakly ordered in all structural studies of the protein. The saccharide binding site is remote from the subunit interface.

Differences to native structure

Changes in the positions of Tyr 12, Thr 15, Asp 16, Leu 99, Tyr 100, Ser 204 to Pro 206, Gly 224, and Arg 228 are seen when comparing the structures of native and the pentasaccharide complex of Con A. The changes at Tyr 12, Leu 99, Tyr 100, and Arg 228 are broadly similar to those seen for the α -Man-OME complex and have already been discussed in detail by Helliwell and coworkers (Derewenda *et al.*, 1989; Naismith *et al.*, 1994). The main chain at Thr 15 moves approximately 0.3 Å to accommodate the 1 \rightarrow 3 Man. The situation for Asp 16 is complicated by crystal contacts in the native and pentasaccharide structures and its position is varied in the subunits of the pentasaccharide structure. The change centered on His 205 is quite pronounced (0.3 Å on average) and appears to be a result of accommodating the 1 \rightarrow 3 arm GlcNAc residue. The main chain at Gly 224 moves substantially (0.5 Å) as result of the carbonyl oxygen forming a polar contact to O4 of the 1 \rightarrow 6 arm GlcNAc residue. Displacement of water molecules (and crystal contacts) at the monosaccharide site has already been reported (Derewenda *et al.*, 1989; Naismith *et al.*, 1994); we see a similar pattern for the other four sites. Water molecules or hydrogen bonding crystal contacts are often close to the positions of the hydrogen bonding groups of the sugar. The two 1 \rightarrow 3 binding sites are filled by a tight crystal contact in the native structure, involving the loop from Trp 182 to Ser 184. Whether or not this has relevance to the proposed peptide binding site of Con A (Scott *et al.*, 1992) is unknown. In summary, the binding site is reformed in the native structure with water molecules (or crystal contacts) filling the positions of the various sugar residues in the native structure.

Conformation of the bound pentasaccharide

The Con A pentasaccharide complex displays well defined electron density for all five sugar residues in six of the eight monomers. In the other two monomers (both in the same tetramer) which have higher average B-factors; the GlcNAcs on the 1 \rightarrow 3 arm are not well defined. As with our trimannoside Con A complex, the glycosidic conformations of the central α -Man-(1 \rightarrow 3)-[α -Man-(1 \rightarrow 6)]-Man of the pentasaccharide (Table I) in all eight subunits are very similar to those reported for conformation of the trimannoside molecule by NMR studies (Homans, 1995). Electron density clearly shows that the anomeric hydroxyl (on the reducing mannose) is found in the alpha configuration (Figure 2) this appears to be a result of the anomeric oxygen's participation in crystal contacts. The N-linked glycan does not have a reducing sugar at this position and the "reducing"

mannose is β 1 \rightarrow 4 linked to a GlcNAc sugar (Figure 1). We can see no structural hindrance to Con A binding such a beta linkage. In a second study of the trimannoside Con A complex (Loris *et al.*, 1996), a second minority solution binding conformation of trimannoside was found in one of four subunits, in direct contrast to our result, where the same majority solution conformation of sugar was found in all four subunits (Naismith *et al.*, 1994). There is no evidence for any alternate conformation of the pentasaccharide in this structure in any of the eight subunits. The conformation of the 1 \rightarrow 3 arm β -GlcNAc-(1 \rightarrow 2)-Man glycosidic linkage is very close to that observed in solution (Homans, 1995). However, the ψ dihedral angle (Table I) of the 1 \rightarrow 6 arm β -GlcNAc-(1 \rightarrow 2)-Man glycosidic linkage is dramatically different ($\sim 50^\circ$) from the solution minimum (Homans, 1995). This distortion is required to avoid penetration of the GlcNAc sugar residue into protein structure; the consequences of this are discussed later.

Discussion

The interactions between Con A and the pentasaccharide, the structural basis of specificity

The pentasaccharide sits in an extended continuous groove on the surface of the protein (Figure 3) and buries over 1000 \AA^2 of surface area. In terms of buried surface area it is one of the largest continuous protein carbohydrate interfaces characterized. For discussion we have split the continuous carbohydrate binding site into five components; in reality these overlap and share several protein residues. The five sites are the 1 \rightarrow 3 arm GlcNAc site, the 1 \rightarrow 3 Man site, the reducing Man site, the monosaccharide site, and the 1 \rightarrow 6 arm GlcNAc site. These sites correspond to the cognate sugar residues shown in Figure 1. This is in contrast to the oligosaccharide structures of LOLI and LOLII that have effectively two discrete binding sites on the protein surface (Bourne *et al.*, 1992, 1994a,b).

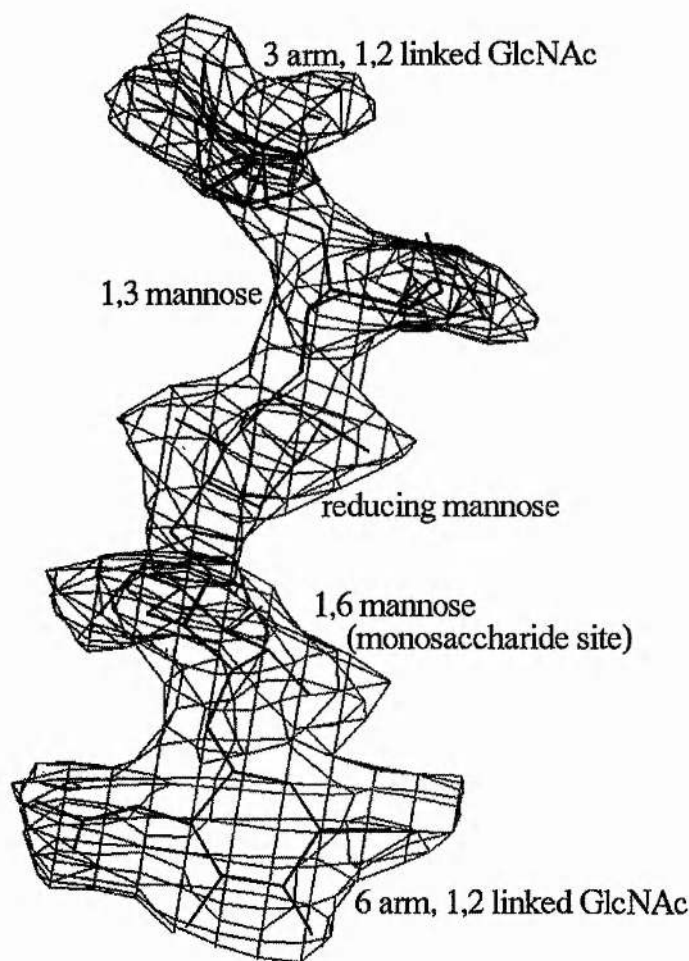


Fig. 2. The difference electron density map contoured at 3σ was calculated by removing the sugar from the model, then refining with simulated annealing (2000K) followed by Powell minimization and B-factor protocols. All five sugars are in good electron density (figure produced using O (Jones *et al.*, 1991)).

Table I. Torsion angles of the glycosidic linkages

	φ^a ($^\circ$)	ψ^a ($^\circ$)	ω^a ($^\circ$)
α -Man-OMe ^{b,c}	63 \pm 5	ND	ND
Trimannoside ^b			
α -Man-(1 \rightarrow 6)-Man	66 \pm 3	-171 \pm 3	73 \pm 5
α -Man-(1 \rightarrow 3)-Man ^b	67 \pm 2	-112 \pm 5	ND
Pentasaccharide ^c			
1-6 arm β -GlcNAc-(1 \rightarrow 2)-Man ^b	-75 \pm 3	-129 \pm 3	ND
1-3 arm β -GlcNAc-(1 \rightarrow 2)-Man ^b	-85 \pm 6	-78 \pm 5	ND
α -Man-(1 \rightarrow 6)-Man	71 \pm 3	179 \pm 2	72 \pm 3
α -Man-(1 \rightarrow 3)-Man ^b	65 \pm 2	-102 \pm 2	ND
Solution values ^d			
β -GlcNAc-(1 \rightarrow 2)-Man ^b	-92 \pm 16	-83 \pm 14	ND
α -Man-(1 \rightarrow 6) Man ^b	70 \pm 20	-170 \pm 20	60 \pm 20
α -Man-(1 \rightarrow 3) Man ^b	80 \pm 15	-116 \pm 25	ND

^aThe torsion angles φ , ψ and ω are: φ is O5 - C1 - OX - CX, ψ is C1 - OX - CX - C(X-1), and ω is OX - CX - C(X-1) - C(X-2). Standard deviations are given. The sugar at the monosaccharide site is shown in boldface type.

^bOne or more angles are not defined (ND) for these sugars.

^cAveraged over all ordered fully ordered sugars.

^dHomans, 1995.

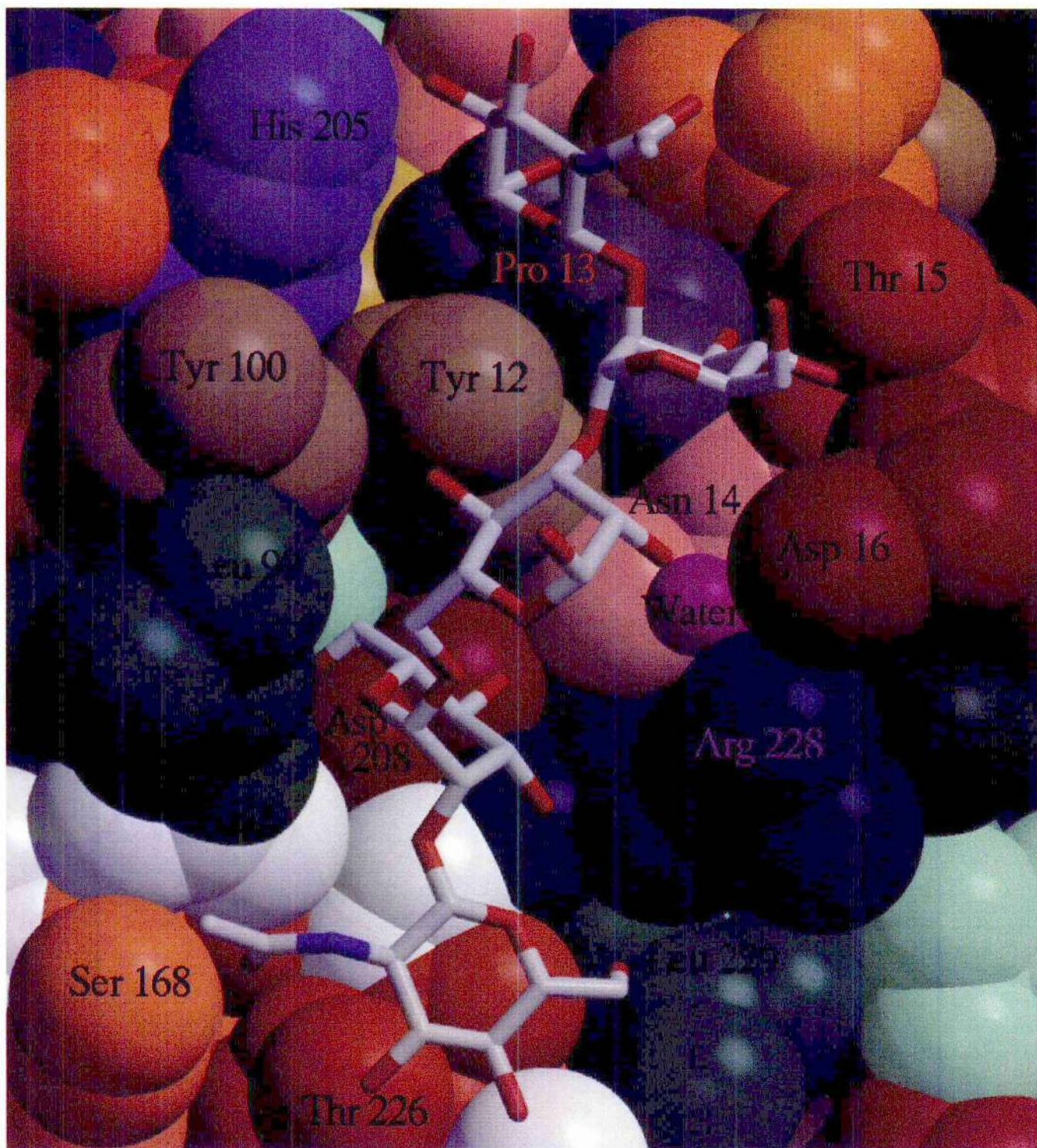


Fig. 3. (a) The pentasaccharide Con A complex. The binding site is an extended groove on the surface of the protein. (Figure generated by RASTER3D; Merritt and Murphy, 1994).

The monosaccharide site was first identified by Helliwell and coworkers (Derewenda *et al.*, 1989); the reducing Man and 1→3 Man sites were defined by the determination of the trimannoside complex (Naismith and Field, 1996). The residues forming these sites in the pentasaccharide complex are unchanged from these earlier descriptions. The interactions at these sites are listed in

Tables II and III. There are a number of subtle changes in the interactions at the monosaccharide site compared to previous descriptions that we deal with more fully later.

The 1→6 arm GlcNAc site is formed by Gly 98, Ser 168 and the loop Thr 226 to Leu 229. One face of the sugar residue is effectively parallel to the protein surface. The loop from Thr 226

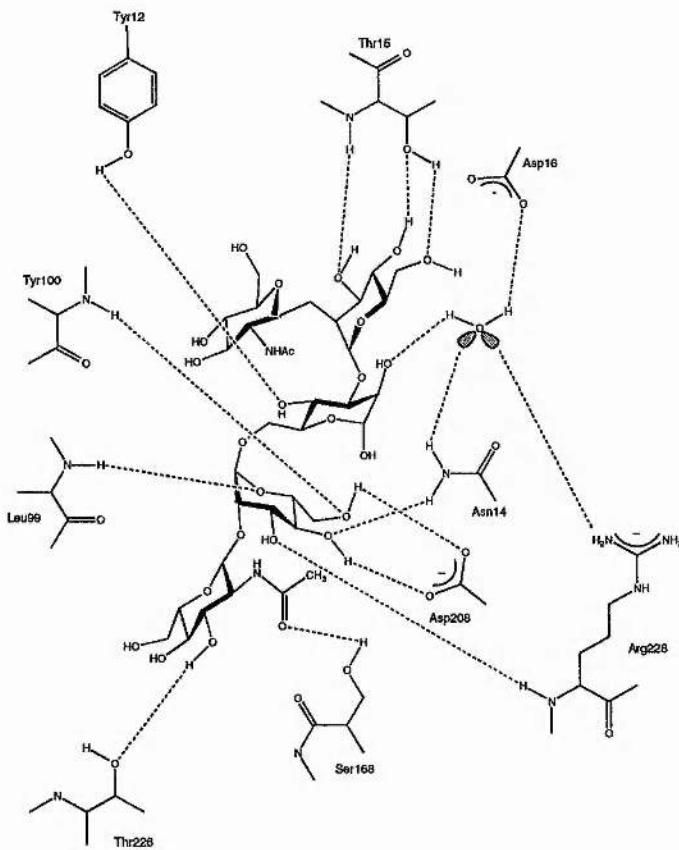


Fig. 3. (b) A schematic representation of the hydrogen bonds between the protein and the pentasaccharide.

to Leu 229 interacts extensively with the sugar residue and forms one side of the binding site. Gly 98 and Ser 168 sit either side of the N-acetyl group, their contacts are listed in Tables II and III. Gly 98 is a key residue, mutation to any other amino acid would abolish binding; as the side chain would intrude into the binding site clashing with the 1→2 linked sugar. Ser 168 hydrogen bonds to the sugar; presumably loss of this H-bond would diminish the interaction and any increase in the size of the residue would inhibit binding of the sugar conformation we observe. The 1→6 arm GlcNAc sugar makes a fairly extensive interaction with the protein (Tables II, III). The 1→3 arm GlcNAc site is much less well defined, the side chain of His 205 caps O4 of the sugar (Figure 3). The ring of the sugar sits on a shallow cleft defined by Pro 13 and the side chain of Tyr 12. There are no hydrogen bonds between this sugar residue and the protein and only 7 van der Waals contacts.

LOLI and II also bind sugars 1→2 linked to the monosaccharide site and both have a Gly in an analogous position to Gly 98. However, in the octasaccharide structures (Bourne *et al.*, 1992, 1994b) which contain a β-GlcNAc-(1→2)-Man disaccharide, the PHI torsion angle of GlcNAc sugar is rotated almost 180° with respect to that in our structure. This is because in LOLI and II Ser 168 is replaced by an Asn that would clash with the N-acetyl group. The interactions at this site are much less extensive in LOLI and II than in Con A, having less than half the number of van der Waals interactions. The lectin from *Dioclea grandiflora*

is homologous to Con A and binds the trimannoside with similar affinity (Chervenak and Toone, 1995). It has a Gly at position 98 but does not recognize the pentasaccharide (Gupta and Brewer, 1996). This is due to changes at Ser 168 (to Asn), which creates a steric clash with the N-acetyl group, and at Thr 226 (to Gly), which removes a hydrogen bond and van der Waals contacts. Interestingly *Dioclea grandiflora* and Con A both bind the high mannose type structures shown in Figure 4 with similar affinity (Mandal *et al.*, 1994b; Gupta and Brewer, 1996) approximately $K_a 9 \times 10^5 \text{ M}^{-1}$. These high mannose sugars precipitate both Con A and *Dioclea grandiflora* from solution (Gupta and Brewer, 1996), suggesting they are able to bind two Con A molecules. The mode of binding of these sugars is not yet clear but we suggest in Figure 4 how high mannose sugars may bind to Con A. The α 1→2 linked terminal mannose residues would sit in the 1→3 and 1→6 GlcNAc sites identified by this study and unlike GlcNAc would not clash with the Asn 168 residue in *Dioclea grandiflora*.

Table II. Hydrogen bonding and polar contact (<3.5 Å) distances between sugar and protein in the six fully ordered sites

Protein	Sugar	Average distance (typically ±0.1Å)
GlcNAc, 1→6 arm		
Thr-226 OG1	O3	2.6
Gly-224 O ^a	O4	2.8
Thr-226 O ^a	O6	3.4
Ser-168 OG	O7	2.6
1→6 Man		
Arg-228 N	O3	3.0
Asn-14 ND2	O4	2.9
Arg-228 N ^a	O4	3.3
Asp-208 OD1	O4	2.8
Leu-99 N	O5	3.0
Leu-99N ^a	O6	2.9
Tyr-100 N	O6	3.0
Asp-208 OD2	O6	3.0
Reducing Man		
OW ^b	O2	2.7
Asp-16 OD2 ^{a,c}	O2	3.3
Tyr-12 OH	O4	2.8
1→3 Man		
Pro-13 O ^a	O3	2.9
Thr-15 N	O3	2.9
Thr-15 OG1	O3	3.0
Thr-15 OG1	O4	2.6
Asp-16 Na	O4	3.0

^aThese contacts are within hydrogen bonding distance; however, the donor-H-acceptor geometry differs substantially from the linearity expected for a hydrogen bond. The sugar at the monosaccharide site is shown in boldface type.

^bOW is the structurally conserved water molecule. This is hydrogen bonded to Asn-14, Asp-16, and Arg-228.

^cThis contact is absent in some units due to crystal contacts.

Table III. Van der Waals contacts (<4.0Å) between Con A and the pentasaccharide^a

Sugar	Residue
1→6 arm GlcNAc	Gly-98 (6), Leu-99 (1), Ser-168 (5), Gly-224 (2), Thr-226 (7), Arg-228 (2), Leu-229 (2)
1→6 Man	Tyr-12 (1), Asn-14 (2), Gly-98 (3), Leu-99 (9), Tyr-100 (5), Ala-207 (2), Asp-208 (9), Gly-227 (4), Arg-228 (9)
Reducing Man	Tyr-12 (3), Leu-99 (2), Tyr-100 (1)
1→3 Man	Tyr-12 (2), Pro-13 (3), Asp-14 (4), Thr-15 (10), Asp-16 (4)
1→3 arm GlcNAc	Tyr-12 (2), Pro-13 (2), His-205 (3)

^aAveraged over the six well ordered binding sites. The sugar at the monosaccharide site is shown in boldface type.

Contribution of the other sites to affinity

The pentasaccharide and trimannoside Con A complexes suggest that there are in principle two additional mannose binding sites (reducing and 1→3); however, calorimetry (Mandal *et al.*, 1994b) and structural analysis (Derewenda *et al.*, 1989) of the α -Man-OMe Con A complex show that only one sugar is bound. Tabulation of the van der Waals contacts, buried surface area and hydrogen bonds shows that these two Man sites have considerably fewer interactions than the monosaccharide site (Table II, III). From these tables we suggest that the reducing mannose site contributes little to the protein carbohydrate interaction but that the 1→3 Man will contribute to an increase in affinity. The 1→3 arm GlcNAc residue makes very few contacts with the protein, buries little surface area and its binding site is the least well defined. This sugar residue does induce a clear change in the main chain position centered on His 205 when compared to other Con A structures. Given the insubstantial nature of its interaction and distortion in protein main chain we would argue that the 1→3 arm GlcNAc residue does not make a large additional contribution to binding. The 1→6 arm GlcNAc site would appear able to make a substantial contribution to binding as it makes an array of van der Waals contacts with the protein, buries a substantial surface area, and makes specific hydrogen bonds with the protein.

The 1→6 arm GlcNAc site

The disaccharide β -GlcNAc-(1→2)-Man that we predict will bind analogously to the 1→6 arm of the pentasaccharide only liberates 5.2 kcalmol⁻¹ (Mandal *et al.*, 1994b). This is identical to Man which liberates approximately 5.2 kcalmol⁻¹ (E.J.Toone, personal communication). It appears that β 1→2 addition of GlcNAc to the monosaccharide mannose is energetically neutral,

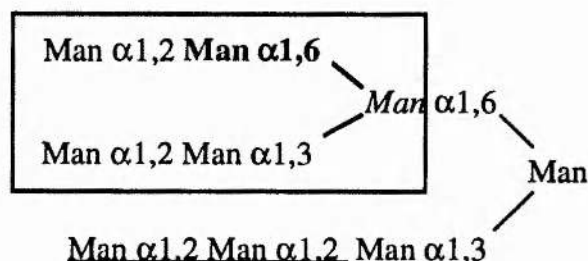


Fig. 4. The high mannose N-linked glycan. We predict that the five boxed sugars mimic the pentasaccharide of this study; the sugar we predict that is bound by the monosaccharide site is shown in boldface type. The sugars which we propose bind and cross-link to another Con A molecule are underlined.

despite the large increase in the number of protein carbohydrate contacts. Jencks has pointed out that the free energy of binding a single compound AB to a protein is not the algebraic sum of free energy of binding of the isolated components A and B (Jencks, 1981). The difference between the summation of the free energy of binding A and B and the composite molecule AB, is the interaction energy. The interaction energy has several components, some of which contribute favorably to binding, some of which contribute unfavorably. The most well known of these is the chelate effect, that is, the entropy penalty paid by binding AB (loss of translational and rotational degrees of freedom) is less than binding A and B (Page and Jencks, 1971). The chelate effect always contributes favorably to the binding energy. The chelating disaccharide (β -GlcNAc-(1→2)-Man) fails to increase the affinity for the protein over the simple monosaccharide. We believe we have identified structural changes in the pentasaccharide and to the monosaccharide site that contribute to the interaction energy, decreasing the affinity of binding of the pentasaccharide.

Rotation of the 1→6 arm β -GlcNAc-(1→2)-Man glycosidic linkage

We have identified a large rotation ($\sim 50^\circ$) in the glycosidic linkage PSI torsion angle from that seen in free solution. The dynamics observed for this linkage in free solution suggest it is a relatively rigid conformation (Homans, 1995). The energy surface for two glycosidic angles in the simple disaccharide (PHI and PSI) β -GlcNAc-(1→2)-Man has been calculated (Imberty *et al.*, 1991). While the energetics of the simple disaccharide will be different from the pentasaccharide, it provides a reasonable approximation to the 1→6 arm β -GlcNAc-(1→2)-Man linkage. Bourne *et al.*, saw an alternative conformation for the β -GlcNAc-(1→2)-Man linkage in their LOL I octasaccharide structure (Bourne *et al.*, 1992). However, their observed glycosidic linkages were within $\pm 30^\circ$ of a second energetic minimum for this linkage. Plotting the glycosidic torsion angles we have observed on the energy surface suggests that the conformation we see is ~ 3 kcalmol⁻¹ above the global minimum and is remote from any other minimum. The crystal structure of the Se155-4 antibody Fab fragment in complex with α -Gal-(1→2)-[α -Abe-(1→3)]-Man, the α -Gal-(1→2)-Man ϕ torsion angle is rotated by about 40° from the theoretical global energy minimum, about 3 kcalmol⁻¹ above the minimum (Cygler *et al.*, 1992). However, Cygler *et al.* (1992) suggested that this conformation is favored by a hydrogen bond within the sugar molecule which was not predicted by theoretical calculations. We find no such "compensating" internal hydrogen bond to explain the rotation we observe. We have not found any other examples of the sugar conformation in the bound state being so energetically distant from the conformation seen or predicted for the uncomplexed sugar.

Table IV. Differences among the Con A complexes at the contacts at the carbohydrate binding sites

	H-bonds	Polar contacts ^a	Van der Waals contacts
Pentasaccharide complex^b			
1→6 arm GlcNAc	2	2	25
1→6 Man	6	2	44
Reducing Man	1	1	6
1→3 Man	3	2	23
1→3 arm GlcNAc	0	0	7
Trisaccharide complex			
1→ Man	6	2	46
Reducing Man	1	1	8
1→3 Man	3	2	23
α-Man-OMe complex^c			
α-Man-OMe	6	2	48

^aPolar contacts are those within hydrogen bonding distance ($< 3.5 \text{ \AA}$) which have geometry inconsistent with a hydrogen bond. The sugar at the monosaccharide site is shown in boldface type.

^bAveraged over the six well ordered sites in the pentasaccharide complex.

^cThe C and D subunits in this complex are distorted by protein sugar crystal contacts are not included in this comparison (Naismith *et al.*, 1994).

Distortion of the monosaccharide site

A comparison of the monosaccharide site in the pentasaccharide complex and the mannose complex is summarized in Table IV. Superimposing the 237 α protein atoms of the pentasaccharide and α -Man-OMe complexes give an r.m.s. deviation for the 13 sugar atoms of 0.33 \AA . This distortion in the sugar position relative to the protein is centered around the O2 of the mannose residue which is β 1→2 linked to a GlcNAc residue in the pentasaccharide. Without this movement in the O2 position the GlcNAc residue would sterically clash (approaches $< 2.0 \text{ \AA}$) with residues Thr 222, Ser 168, and Gly 98. The presence of the β 1→2 linkage alters the main chain trace of the loop from Gly 98 to Tyr 100, centered at Leu 99 the average shift is 0.2 \AA . This region is one of the four loops that form the monosaccharide site. We note that while there is a reduction in the number of van der Waals contacts at the monosaccharide site, the number of hydrogen bonds remains the same (Table IV).

We hypothesize that the α -Man-OMe complex represents the energetic minimum of the monosaccharide site. We believe the combination of the distortions (discussed above) decrease the free energy liberated by the monosaccharide site lowering the overall affinity of the oligosaccharide for Con A. In short, we suggest these are structural manifestations of the interaction energy. We predict that the free energy liberated by the disaccharide β -GlcNAc-(1→2)-Man ($5.2 \text{ kcal mol}^{-1}$) is considerably less than algebraic sum of the free energy liberated separately by GlcNAc and Man at their cognate sites. The ΔG^0 value for Man is $5.2 \text{ kcal mol}^{-1}$, and although the value GlcNAc is known ($5.2 \text{ kcal mol}^{-1}$) it probably reflects binding at the monosaccharide site. Consequently, we have no reliable estimate for the energy liberated by GlcNAc at the 1→6 arm GlcNAc site. However, the extent of the contacts between the protein and sugar suggests this

is not zero. Quantitation of the components of the interaction energy that are likely to include solvent reorganization is beyond our reach at present. However, the combination of calorimetry and structural work continues to generate a significant resource for theoretical approaches to oligosaccharide recognition

Materials and methods

Con A was obtained from Sigma (Poole, United Kingdom) and β -GlcNAc-(1→2)- α -Man-(1→3)-[β -GlcNAc-(1→2)- α -Man-(1→6)]-Man was obtained from Dextra Laboratories (Reading, United Kingdom). Crystals of the protein carbohydrate complex were obtained from a sitting drop using a solution containing a 30-fold excess of sugar over Con A (protein concentration 6 mg ml^{-1}), 1 mM MnCl_2 , and 1 mM CaCl_2 . This was equilibrated against a reservoir buffered at pH 6.0 containing 15% polyethylene glycol 6K as the precipitant. Data were recorded as 200 nonoverlapping $20 \text{ min } 1^\circ$ oscillations. The data were processed and merged with DENZO and SCALEPACK (Otwinowski, 1993). The space group was assigned as C2 with unit cell parameters $a = 176.5 \text{ \AA}$, $b = 122.8 \text{ \AA}$, $c = 124.6 \text{ \AA}$, $\beta = 134.2^\circ$. The asymmetric unit contains two tetramers; Matthew's number 2.5 Da \AA^{-3} , ~46% solvent. Data collection and reduction statistics are shown in Table V. The structure was determined by the molecular replacement method as implemented in the CCP4 (CCP4, 1994) program AMORE (Navaza, 1994) using all data from 12 \AA to 3.4 \AA . The trimannoside Con A complex of Con A (PDB code 1CVN) (Naismith and Field, 1996) was used as the search model (with all metal ions, sugar molecules, and water molecules removed). Strong difference electron density was observed for all metal ions and 32 of the 40 possible sugars, with weak density for 6 but ambiguous density for 2. A number of changes in protein structure were made manually at this stage using "O" (Jones *et al.*, 1991). The metal ions were included in the model with zero electrostatic charge and the sugar molecules included when the $F_o - F_c$ map showed density stronger than 3σ for all atoms in the sugar residue. Further refinement proceeded smoothly by alternating cycles of automated X-PLOR (Brunger, 1992) refinement and manual intervention using "O". The protein was refined with the Engh and Huber stereochemical parameter dictionary. Noncrystallographic restraints were applied throughout for both positional and B-factor refinement. Despite the limited resolution we did carry out restrained B-factor refinement because we had 8-fold non-crystallographic symmetry and saw a clear ($>4\%$) improvement in R_{free} . All individual B-factors were reset to the overall average temperature factor after each manual intervention. Apart from the 10% of measured data excluded to monitor refinement no cut-offs were applied to the data. A bulk solvent correction was included in the X-PLOR refinement with parameters for solvent density 0.34 e- \AA^{-3} , solvent radius 0.25 \AA and B factor 50 \AA^2 . Electron density maps were calculated using SIGMAA (Read, 1986) modified coefficients with all data to 2.7 \AA and included the X-PLOR bulk solvent correction. Water molecules were added to the model in batches if they satisfied four criteria: they corresponded to a peak $> 3.5\sigma$ in the $F_o - F_c$ map; they made hydrogen bonds with reasonable stereochemistry; they reappeared in at least 1σ in subsequently calculated $2F_o - F_c$ maps and that a drop in the free R factor was observed. Statistics on the final model are shown in Table V. The coordinates and structure factors have been deposited with Protein Data Bank (Bernstein *et al.*, 1977) (entry codes 1TEI and 1TEISF).

Table V. Crystallographic data collection statistics and refinement statistics

Unique reflections	53516
Completeness of data (%) (25.0 - 2.7Å/2.8-2.7Å)	97.1/95.8
R _{merge} (I) (%) (25.0 - 2.7Å/2.8-2.7Å) ^a	9.4/33.7
Average data redundancy (25.0 - 2.7Å/2.8-2.7Å)	4.2/4.1
% of data > 1σ (25.0 - 2.7Å/2.8-2.7Å)	93/83
Refinement	
Resolution range (Å)	25-2.7
R _{free} (%) ^b	21.9
R factor (%)	17.9
Bond r.m.s. deviation (Å) ^c	0.01
Angle r.m.s. deviation (°) ^c	1.746
Non-crystallographic symmetry r.m.s. deviation (cα atoms) (Å)	0.12
B-factor bonded atoms r.m.s. deviation (Å ²) ^e	3.26
Ramachandran core/additional (%) ^d	85.3/14.7
Protein mean B-factor (Å ²)	29.1
Sugar mean B-factor (Å ²)	35.9
Solvent mean B (Å ²)	22.5
Number of protein atoms	14472
Number of sugar atoms	496
Number of solvent atoms	74
Number of metal ions	16

^aR_{merge}(I) = $\sum_{hkl} \sum_i |I_i - I(hkl)| / \sum_{hkl} \sum_i I_i(hkl)$.

^bR_{free} is calculated on 10% of data excluded during refinement.

^cr.m.s. deviation from Engh and Huber ideal values (Engh and Huber, 1991).

^dCore and additionally allowed regions as defined by PROCHECK (Laskowski *et al.*, 1993). No residues are in the generously allowed or disallowed regions. All B-factor calculations exclude the 370 protein atoms (2.6%) and 21 sugar atoms (4.2%) (1→3 arm GlcNAc in F and H subunits) which were stereochemically modeled.

^eCalculated with MOLEMAN (G.J.Kleywegt, unpublished program). All stereochemically modeled atoms were removed prior to B-factor analysis, all bonded atoms including those in the sugars are included in the calculation of r.m.s. B-factor deviation for bonded atoms.

Acknowledgments

We thank the Wellcome Trust (043586/Z/95/Z/MP/RF/PK) for equipment support. We are grateful to Rob Field, John Helliwell, Steve Homans, Charlie Weller, and Trevor Rutherford for discussions, encouragement, and advice. The research is supported by the B.B.S.R.C. (B08307).

Abbreviations

Con A, concanavalin A; GlcNAc, glucosamine; LOLI, *Lathyrus ochrus* isolectin I; LOLII, *Lathyrus ochrus* isolectin II; Man, mannose; r.m.s., root mean square.

References

- Bains,G., Lee,R.T., Lee,Y.C. and Freire,E. (1992) Microcalorimetric study of wheat-germ-agglutinin binding to N-acetylglucosamine and its oligomers. *Biochemistry*, **31**, 12624-12628.
- Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Myer,E.F.Jr., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: a computer based archive file system for macromolecular structures. *J. Biol. Chem.*, **112**, 535-542.

- Bourne,Y., Ayoub,A., Rouge,P. and Cambillau,C. (1994a) Interaction of a legume lectin with 2 components of the bacterial cell wall—a crystallographic study. *J. Biol. Chem.*, **269**, 9429-9435.
- Bourne,Y., Mazurier,J., Legrand,D., Rouge,P., Montreuil,J., Spik,G. and Cambillau,C. (1994b) Structures of a legume lectin complexed with the human lactotransferrin N2 fragment, and with an isolated biantennary glycopeptide—role of the fucose moiety. *Structure*, **2**, 209-219.
- Bourne,Y., Rouge,P. and Cambillau,C. (1992) X-Ray structure of a biantennary octasaccharide-lectin complex refined at 2.3 Å resolution. *J. Biol. Chem.*, **267**, 197-203.
- Brunger,A.T. (1992) *X-PLOR (Version 3.1) Manual: A System for X-Ray Crystallography and NMR*. Yale University Press, New Haven.
- Bundle,D.R., Eichler,E., Gidney,M.A.J., Meldal,M., Ragauskas,A., Sigurskjold,B.W., Sinnott,B., Watson,D.C., Yaguchi,M. and Young,N.M. (1994) Molecular recognition of a salmonella trisaccharide epitope by monoclonal-antibody Se155-4. *Biochemistry*, **33**, 5172-5182.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760-763.
- Chervenak,M.C. and Toone,E.J. (1995) Calorimetric analysis of the binding of lectins with overlapping carbohydrate-binding ligand specificities. *Biochemistry*, **34**, 5685-5695.
- Cyglar,M., Wu,S., Zdanov,A., Bundle,D.R. and Rose,D.R. (1992) Carbohydrates, shapes and biological recognition. *Biochem. Soc. Trans.*, **21**, 437-441.
- Cyglar,R., Rose,D.R. and Bundle,D.R. (1991) Recognition of a cell-surface oligosaccharide of pathogenic *Salmonella* by an antibody Fab fragment. *Science*, **253**, 442-445.
- Delbaere,L.T.J., Vandonselaar,M., Prasad,L., Quail,J.W., Wilson,K.S. and Dauter,Z. (1993) Structures of the lectin-IV of *Griffonia-Simplicifolia* and its complex with the Lewis-B human flood-group determinant at 2.0- angstrom resolution. *J. Mol. Biol.*, **230**, 950-965.
- Derewenda,Z., Yariv,J., Helliwell,J.R., Kalb (Gilboa),A.J., Dodson,E.J., Papiz,M.Z., Wan,T. and Campbell,J. (1989) The structure of the saccharide binding site of concanavalin A. *EMBO J.*, **8**, 2189-2193.
- Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr. A*, **47**, 392-400.
- Goldstein,I.J. and Poretz,R.D. (1986) Isolation and chemical properties of lectins. In Leiner,I.E., Sharon,N. and Goldstein,I.J. (eds.), *The Lectins*. Academic Press, New York, pp. 35-244.
- Gupta,D. and Brewer,C.F. (1996) Thermodynamic binding-studies of concanavalin-a. *Abstr. Pap. Am. Chem. Soc.*, **211**, 76-CARB.
- Gupta,D., Dam,T.K., Oscarson,S. and Brewer,C.F. (1997) Thermodynamics of lectin carbohydrate interactions. Binding of the core trimannoside of asparagine linked carbohydrates and deoxy analogs to concanavalin A. *J. Biol. Chem.*, **272**, 6388-6392.
- Hardman,K.D. and Ainsworth,C.F. (1972) Structure of concanavalin Å at 2.4Å resolution. *Biochemistry*, **11**, 4442-4447.
- Harrop,S.J., Helliwell,J.R., Wan,T.C.M., Kalb,A.J., Tong,L. and Yariv,J. (1996) Structure solution of a cubic-crystal of concanavalin a complexed with methyl alpha-D-glucopyranoside. *Acta Crystallogr. D*, **52**, 143-155.
- Helliwell,J.R. and Helliwell,M. (1996) X-Ray crystallography in structural chemistry and molecular-biology. *Chem. Commun.*, **14**, 1595-1602.
- Hester,G., Kaku,H., Goldstein,I.J. and Wright,C.S. (1995) Structure of mannose-specific snowdrop (*Galanthus-Nivalis*) lectin is representative of a new plant lectin family. *Nature Struct. Biol.*, **2**, 472-479.
- Hester,G. and Wright,C.S. (1996) The mannose-specific bulb lectin from *Galanthus-Nivalis* (snowdrop) binds monomannoside and dimannoside at distinct sites—structure-analysis of refined complexes at 2.3-angstrom and 3.0-angstrom resolution. *J. Mol. Biol.*, **262**, 516-531.
- Homans,S. W. (1995) Conformational studies on oligosaccharides. In Evans,J.N.S. (ed.), *Biomolecular NMR Spectroscopy*. Oxford University Press, Oxford, pp. 230-257.
- Imberty,A., Delage,M.M., Bourne,Y., Cambillau,C. and Perez,S. (1991) Data bank of three-dimensional structures of disaccharides. II. N-Acetylglucosamine type N-glycans, comparison with the crystal structure of a biantennary octasaccharide. *Glycoconjugate J.*, **8**, 456-483.
- Jencks,W.P. (1981) On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. USA*, **78**, 4046-4050.
- Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110-119.
- Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 548-558.
- Loris,R., Maes,D., Poortmans,F.L.W. and Bouckaert,J. (1996) A structure of the complex between concanavalin A and methyl-3,6-di-O-(alpha-d-mannopyranosyl)-alpha-d-mannopyranoside reveals two binding modes. *J. Biol. Chem.*, **271**, 30614-30618.

- Lowe, J.B. and Ward, P.A. (1997) Therapeutic inhibition of carbohydrate-protein interactions *in vivo*. *J. Clin. Invest.*, **99**, 822–826.
- Mandal, D.K., Bhattacharyya, L., Koenig, S.H., Brown, R.D., Oscarson, S. and Brewer, C.F. (1994a) Studies of the binding-specificity of concanavalin-a—nature of the extended binding-site for asparagine-linked carbohydrates. *Biochemistry*, **33**, 1157–1162.
- Mandal, D.K. and Brewer, C.F. (1993) Differences in the binding affinities of dimeric concanavalin-a (including acetyl and succinyl derivatives) and tetrameric concanavalin-a with large oligomannose-type glycopeptides. *Biochemistry*, **32**, 5116–5120.
- Mandal, D.K., Kishore, N. and Brewer, C.F. (1994b) Thermodynamics of lectin-carbohydrate interactions—titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. *Biochemistry*, **33**, 1149–1156.
- Merritt, E.A. and Murphy, M.E.P. (1994) Raster 3D version 2.0: a program for photorealistic graphics. *Acta Crystallogr. Sect D*, **50**, 869–873.
- Merritt, E.A., Sarfaty, S., Vandenaeker, F., Lhoir, C., Martial, J.A. and Wol, W.G.J. (1994a) Crystal-structure of cholera-B pentamer bound to receptor G(M1) pentasaccharide. *Protein Sci.*, **3**, 166–175.
- Merritt, E.A., Sixma, T.K., Kalk, K.H., Vanzanten, B.A.M. and Hol, W.G.J. (1994b) Galactose-binding site in *Escherichia coli* heat-labile enterotoxin (Lt) and cholera-toxin (Ct). *Mol. Microbiol.*, **13**, 745–753.
- Mowbray, S.L. and Cole, L.B. (1992) 1.7 Å X-Ray structure of the periplasmic ribose receptor from *Escherichia coli*. *J. Mol. Biol.*, **225**, 155–175.
- Naismith, J.H., Emmerich, C., Habash, J., Harrop, S.J., Helliwell, J.R., Hunter, W.N., Raftery, J., Kalb, A.J. and Yariv, J. (1994) Refined structure of concanavalin-a complexed with methyl alpha-D-mannopyranoside at 2.0 angstrom resolution and comparison with the saccharide-free structure. *Acta Crystallogr. D*, **50**, 847–858.
- Naismith, J.H. and Field, R.A. (1996) Structural basis of trimannoside recognition by concanavalin-a. *J. Biol. Chem.*, **271**, 972–976.
- Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
- Ni, Y. and Tizard, I. (1996) Lectin-carbohydrate interactions in the immune system. *Vet. Immunol. Immunopathol.*, **55**, 205–223.
- Otwinowski, Z. (1993) Oscillation data reduction program. In Sawyer, L., Isaacs, N. and Bailey, S. (eds.), *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*. Daresbury Laboratory, Warrington, U.K., pp. 56–62.
- Page, M.L. and Jencks, W.P. (1971) entropic contributions to rate accelerations in enzymatic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. USA*, **68**, 1678–1683.
- Quioco, F.A. and Ledvina, P.S. (1996) Atomic structure of specificity of bacterial periplasmic receptors for active transport and chemotaxis; variation of common themes. *Mol. Microbiol.*, **20**, 17–25.
- Read, R.J. (1986) Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A*, **42**, 140–149.
- Reeke, G.N.J., Becker, J.W. and Edelman, G.M. (1975) The covalent and three dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding and quaternary structure. *J. Biol. Chem.*, **250**, 1525–1547.
- Rini, J.M. (1995) Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 551–577.
- Rose, D.R., Cygler, M., To, R.J., Przybylska, M., Sinnott, B. and Bundle, D.R. (1990) Preliminary crystal-structure analysis of an FAB specific for a *Salmonella* O-polysaccharide antigen. *J. Mol. Biol.*, **215**, 489–492.
- Schwarz, F.P., Puri, K. and Surolia, A. (1991) Thermodynamics of the binding of galactopyranoside derivatives to the basic lectin from winged bean (*Psophocarpus-Tetragonolobus*). *J. Biol. Chem.*, **266**, 24344–24350.
- Schwarz, F.P., Puri, K.D., Bhat, R.G. and Surolia, A. (1993) Thermodynamics of monosaccharide binding to concanavalin-a, pea (*Pisum sativum*) lectin, and lentil (*Lens culinaris*) lectin. *J. Biol. Chem.*, **268**, 7668–7677.
- Scott, J.K., Longanathan, D., Easley, R.B., Gong, X. and Goldstein, I.J. (1992) A family of concanavalin A binding peptides from a hexapeptide epitope library. *Proc. Natl. Acad. Sci. USA*, **89**, 5398–5402.
- Shaanan, B., Lis, H. and Sharon, N. (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*, **254**, 862–866.
- Sharff, A.J., Rodseth, L.E. and Quioco, F.A. (1993) Refined 1.8-angstrom structure reveals the mode of binding of beta-cyclodextrin to the maltodextrin binding-protein. *Biochemistry*, **32**, 10553–10559.
- Sixma, T.K., Pronk, S.E., Kalk, K.H., Vanzanten, B.A.M., Berghuis, A.M. and Hol, W.G.J. (1992) Lactose binding to heat-labile enterotoxin revealed by x-ray crystallography. *Nature*, **355**, 561–564.
- Stein, P.E., Boodhoo, A., Armstrong, G.D., Heerze, L.D., Cockle, S.A., Klein, M.H. and Read, R.J. (1994) Structure of a pertussis toxin sugar complex as a model for receptor-binding. *Nature Struct. Biol.*, **1**, 591–596.
- Surolia, A., Sharon, N. and Schwarz, F.P. (1996) Thermodynamics of monosaccharide and disaccharide binding to erythrina coralodendron lectin. *J. Biol. Chem.*, **271**, 17697–17703.
- Toone, E.J. (1994) Structure and energetics of protein carbohydrate complexes. *Curr. Opin. Struct. Biol.*, **4**, 719–728.
- Vandenaeker, F., Steensma, E. and Hol, W.G.J. (1996) Tumor-marked disaccharide d-Gal-beta-1,3-GalNAc complexed to heat-labile enterotoxin from *Escherichia coli*. *Protein Sci.*, **5**, 1184–1188.
- Weatherman, R.V., Mortell, K.I., Chervenak, M., Kiessling, L.L. and Toone, E.J. (1996) Specificity of C-glycoside complexation by mannose glucose specific lectins. *Biochemistry*, **35**, 3619–3624.
- Weis, W.I., Drickamer, K. and Hendrickson, W.A. (1992) Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature*, **360**, 127–134.
- Williams, B.A., Chervenak, M.C. and Toone, E.J. (1992) Energetics of lectin-carbohydrate binding—a microcalorimetric investigation of concanavalin a-oligomannoside complexation. *J. Biol. Chem.*, **267**, 22907–22911.
- Wright, C.S. (1992) Crystal-structure of a wheat-germ-agglutinin glycoprotein-sialoglycopeptide receptor complex—structural basis for cooperative lectin-cell binding. *J. Biol. Chem.*, **267**, 14345–14352.
- Wright, C.S. and Hester, G. (1996) The 2.0 angstrom structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose—evidence for 2 unique binding modes. *Structure*, **4**, 1339–1352.
- Wright, C.S. and Kellogg, G.E. (1996) Differences in hydrophobic properties of ligand-binding at 4 independent sites in wheat-germ agglutinin-oligosaccharide crystal complexes. *Protein Sci.*, **5**, 1466–1476.
- Zdanov, A., Li, Y., Bundle, D.R., Deng, S.-J., MacKenzie, C.R., Narang, S.A., Young, N.M. and Cygler, M. (1994) Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7 Å resolution. *Proc. Natl. Acad. Sci. USA*, **91**, 6423–6427.

Crystallization of succinylated concanavalin A bound to a synthetic bivalent ligand and preliminary structural analysis

DAVINA N. MOOTHOO,^a STEPHEN A. MCMAHON,^a SARAH M. DIMICK,^b ERIC J. TOONE^b AND JAMES H. NAISMITH^{a*} at ^aCentre for Biomolecular Science, Purdie Building, The University, St Andrews KY16 9ST, Scotland, and ^bDepartments of Biochemistry and Chemistry, Duke University, North Carolina, NC 27708–0346, USA. E-mail: naismith@st-and.ac.uk

(Received 2 February 1998; accepted 9 March 1998)

Abstract

Crystals have been obtained of succinylated concanavalin A complexed to a novel bidentate synthetic ligand. The crystals are the first example of a lectin with a synthetic multivalent ligand and the first report of crystallization of succinylated concanavalin A. The crystals were obtained by sitting-drop vapour diffusion equilibrating with a solution of 20% polyethylene glycol, pH 5, 293.5 K. Crystals are orthorhombic, belonging to space group $C222_1$ with unit-cell dimensions of $a = 99.1$, $b = 127.4$, $c = 118.9$ Å. The asymmetric unit contains a dimer, with over 65% of the volume occupied by water. The ligand cross links concanavalin A monomers. Succinylated concanavalin A is known to be a dimer in solution, yet it is found as the typical concanavalin A tetramer in the crystal. The contacts holding together the tetramer appear extensive and suggest that a fine balance between dimer and tetramers exists. Data to 2.65 Å have been collected and the structure determined by the molecular replacement method.

1. Introduction

Lectins comprise a varied family of sugar-binding proteins; they are found in all types of organisms. Plant lectins in particular have been of intense interest, because they exhibit exquisite specificity for oligosaccharides and unlike their mammalian counterparts are much more tractable to characterization by a broad range of biophysical techniques. Thus, these proteins, such as the lectin from *Canavalia ensiformis*, concanavalin A (con A), have served as models for more complex species (Rini, 1995). In contrast to the exquisite selectivity at the oligosaccharide level, the proteins are relatively promiscuous at the monosaccharide level, falling into two broad classes; mannose specific and galactose specific. The affinity of the lectin for monosaccharides is correspondingly lower, of the order $K_a = 1 \times 10^3 M^{-1}$ compared with $K_a = 1 \times 10^6 M^{-1}$ for oligosaccharides (Chervenak & Toone, 1995; Mandal *et al.*, 1994; Toone, 1994). The advent of calorimetric data on lectin (especially con A) carbohydrate complexes and the realisation that *ab initio* modelling methods were failing to accurately model protein–carbohydrate interactions has reinvigorated the structural study of lectin–carbohydrate complexes. A number of oligosaccharide–protein complexes have been determined (Rini, 1995), most recently the structures of con A with its cognate trisaccharide and pentasaccharide (Naismith & Field, 1996; Moothoo & Naismith, 1998).

Protein–saccharide interactions are appealing therapeutic targets in many diseases, particularly those involving infection and inappropriate immune response (Dwek, 1996). Therapeutics designed for this task would principally be aimed at

disrupting or interfering with molecular recognition processes rather than being cytotoxic. Oligosaccharides are poor therapeutics, however, as they are too polar for satisfactory uptake. Monosaccharides, however, bind too weakly to be of value. One possible solution to this conundrum has been the use of polyvalent ligands. This is based on the observation that in nature many carbohydrate proteins appear to function as oligomers. Chemical synthesis has produced families of such multivalent ligands, which on the basis of agglutination assays appear to have dramatic results (Kanai *et al.*, 1997; Sigal *et al.*, 1996). Several structures are now known of lectins complexed to naturally occurring carbohydrates that cross link the protein in the crystalline phase (Bourne *et al.*, 1994; Dessen *et al.*, 1995; Wright, 1992; Wright & Hester, 1996). Important questions remain as to the thermodynamic basis of these polyvalent interactions (Roy, 1996; Toone, 1994). As part of a program to combine calorimetry, crystallography and chemical synthesis we report the crystallization and structure determination of the first synthetic multivalent ligand–lectin complex.

2. Crystallization and X-ray data collection

Succinylated con A was purchased from Sigma (Poole, UK). The synthesis of the bidentate ligand [1,3-di-(*N*-propyloxy- α -D-mannopyranosyl)-carbonyl 5-methylazido-benzene, Fig. 1] will be described elsewhere. Crystallization trials were performed by the means of the sitting-drop vapour-diffusion method (Ducruix & Giegé, 1992). A solution of succinylated con A (1.2 mM)/ligand (18 mM) was prepared in 20 mM Tris pH 7, 100 mM NaCl, 1 mM CaCl₂ and 1 mM MnCl₂ and equilibrated against 20% polyethylene glycol ($M_r = 6000$), 100 mM citric acid, pH 5 in sitting-drop trays (Charles Supper, USA) at 293.5 K. Small crystals of dimensions 0.1 × 0.2 × 0.2 mm were initially obtained. Optimization of the ligand concentration (5 mM) yielded block-shaped crystals of up to 1.0 × 0.8 × 0.6 mm. Crystal growth was complete in 14 d. 12 crystals were examined, mounted in a thin-walled glass capil-

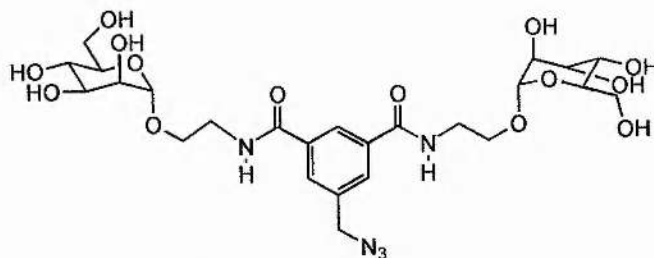


Fig. 1. The bivalent ligand [1,3-di-(*N*-propyloxy- α -D-mannopyranosyl)-carbonyl 5-methylazido-benzene] which cross links con A.

lary, by exposure to X-rays. The resolution limit varied from 2.65 to 3.5 Å. All diffraction data were collected at 293.5 K from a single crystal (0.6 × 0.5 × 0.3 mm) which diffracted to the highest resolution. Larger crystals did not give higher resolution data. All data were recorded on the Nonius/MacScience DIP2000 dual image plate. X-rays were generated using a Enraf-Nonius FR951 rotating-anode generator and focused using the MacScience mirror system, through a 0.5 mm collimator. Data were collected as 92 non-overlapping 25 min λ oscillations with a crystal-to-detector distance of 140 mm. The resolution limit of the data was 2.65 Å and no significant crystal decay was observed during data collection. The programs *DENZO* and *SCALEPACK* (Otwinowski, 1993) were used to process the data. The crystal was indexed in a centred orthorhombic space group with unit-cell dimensions $a = 99.1$, $b = 127.4$, $c = 118.9$ Å. Analysis of diffraction data identified systematic absences consistent with space group $C222_1$. A dimer of molecular mass 49 kDa, gives rise to a V_m (Matthews, 1968) of $3.8 \text{ \AA}^3 \text{ Da}^{-1}$ and indicates a solvent content of over 65%. The data are 96% complete from 26 to 2.65 Å, with an R_{merge} 7.2% with a redundancy of 2.2 and a total of 84% of possible data are greater than 1σ . For the high-resolution shell (2.75–2.65 Å), the corresponding values are 98% complete, R_{merge} 20.2%, redundancy 2.2 and 83% greater than 1σ .

3. Structure solution

The structure was solved by molecular replacement, using *AMoRe* (Navaza, 1994) as implemented in the *CCP4* package (Collaborative Computational Project, Number 4, 1994). The conventional con A dimer (monomers *A* and *B*) from the trimannoside complex of con A (1CVN) stripped of metals, waters and sugars was used as the search model. The rotation function found a single solution with a correlation coefficient of 0.28 and the translation function produced a final solution with a correlation coefficient of 0.84. A translation search in $C222_1$ gave no solution. In order to visualize the interaction of the ligand with each protein monomer, the asymmetric unit was redefined such that one monomer of the con A dimer (monomer *A*) was linked *via* the ligand to another monomer

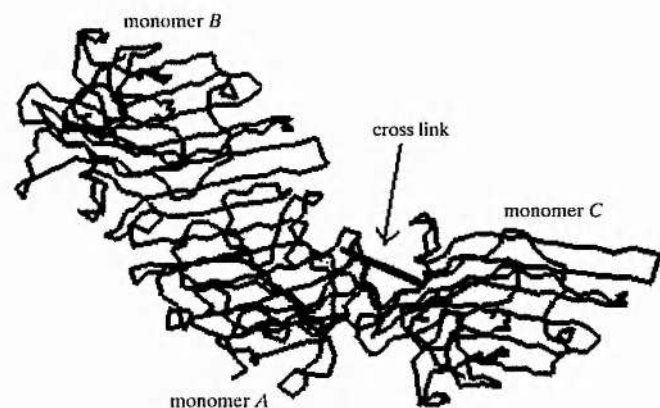


Fig. 2. The conventional dimer of con A is formed by monomer *A* and monomer *B*, the ligand cross links monomer *A* to monomer *C*. Monomer *C* is related by the crystallographic transformation $(\frac{1}{2} - x, \frac{1}{2} - y, z + \frac{1}{2})$ to monomer *B*. The cross-linking ligand is shown as a straight line between the two sugar-binding sites.

(monomer *C*). Monomer *C* was generated by applying the symmetry transformation $(\frac{1}{2} - x, \frac{1}{2} - y, z + \frac{1}{2})$ to monomer *B* (Fig. 2). Other crystallographic operators generate the 'typical' con A tetramer. The initial free R factor was 38% to 2.65 Å. Rigid-body refinement of the monomers gave a free R factor of 30%. Clear density is visible for the metal ions and ligand. The ligand is currently being built into this electron density and a stereochemical dictionary constructed. Once the ligand has been built into density, refinement will recommence. A fully refined structure will be reported elsewhere.

That succinylated con A is found as a tetramer in this crystal is puzzling; the electron density shows evidence of succinylation on two lysine residues. Dimerization of the dimers occludes 4300 \AA^2 of surface area which compares to 5200 \AA^2 for native con A (Deacon *et al.*, 1997; Reeke *et al.*, 1975) and 4800 \AA^2 for the mannose con A complex (Naismith *et al.*, 1994). There are 120 protein-protein contacts less than 4.0 Å between the two dimers in this structure, compared with 260 for the native and 160 for the mannose complex. Although this tetramer is clearly less tightly packed than the other two, by a standard set of standard crystallographers' criteria (buried surface area and contacts) one would identify the protein to be a functional tetramer. However, solution studies show unambiguously that succinylated con A is a dimer (Gunther *et al.*, 1973). It appears that the energy barrier between dimers and tetramers for succinylated con A is small enough that crystal-packing forces can drive formation of the tetramer. We feel this result poses an interesting test for those who wish to devise automatic methods of identifying protein oligomerization state from static crystal structures.

JHN acknowledges BBSRC support (B08307) and we thank John Helliwell for encouragement.

References

- Bourne, Y., Bolgiano, B., Liao, D. L., Strecker, G., Cantau, P., Herzberg, O., Feizi, T. & Cambillau, C. (1994). *Nature Struct. Biol.* **1**, 863–870.
- Chervenak, M. C. & Toone, E. J. (1995). *Biochemistry*, **34**, 5685–5695.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Deacon, A., Gleichmann, T., Kalb (Gilboa), A. J., Price, H., Rafferty, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* **53**, 4305–4312.
- Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F. & Sacchettini, J. C. (1995). *Biochemistry*, **34**, 4933–4942.
- Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins*. Oxford: IRL Press.
- Dwek, R. A. (1996). *Chem. Rev.* **96**, 683–720.
- Gunther, G. R., Wang, J. L., Yahara, I., Cunningham, B. A. & Edelman, G. (1973). *Proc. Natl Acad. Sci. USA*, **70**, 1012–1016.
- Kanai, M., Mortell, K. H. & Kiessling, L. L. (1997). *J. Am. Chem. Soc.* **119**, 9931–9932.
- Mandal, D. K., Kishore, N. & Brewer, C. F. (1994). *Biochemistry*, **33**, 1149–1156.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Moothoo, D. N. & Naismith, J. H. (1998). *Glycobiology*, **8**, 173–181.
- Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb, A. J. & Yariv, J. (1994). *Acta Cryst.* **D50**, 847–858.
- Naismith, J. H. & Field, R. A. (1996). *J. Biol. Chem.* **271**, 972–976.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Otwinowski, Z. (1993). In *Proceedings of the CCP4 study weekend: Data collection and processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

- Reeke, G. N. J., Becker, J. W. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1525–1547.
- Rini, J. M. (1995). *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577.
- Roy, R. (1996). *Curr. Opin. Struct. Biol.* **6**, 692–702.
- Sigal, G. B., Mammen, M., Dahmann, G. & Whitesides, G. M. (1996). *J. Am. Chem. Soc.* **118**, 3789–3800.
- Toone, E. J. (1994). *Curr. Opin. Struct. Biol.* **4**, 719–728.
- Wright, C. S. (1992). *J. Biol. Chem.* **267**, 14345–14352.
- Wright, C. S. & Hester, G. (1996). *Structure*, **4**, 1339–1352.

Acta Cryst. (1998). D54,

A general method for co-crystallization of concanavalin A with carbohydrates

DAVINA N. MOOTHOO AND JAMES H. NAISMITH* at Centre for Biomolecular Science, Purdie Building, The University, St. Andrews KY16 9ST, Scotland. E-mail: naismith@st-and.ac.uk

(Received 11 May 1998; accepted 25 June 1998)

Abstract

A small grid of conditions has been developed for co-crystallization of the plant lectin concanavalin A (conA) and polysaccharides. Crystals have been obtained of complexes of conA with α 1-2 mannoside, 1-methyl α 1-2 mannoside, fructose, a trisaccharide and a pentasaccharide. The crystals diffract to resolutions of 1.75–2.7 Å using a copper rotating-anode source. The crystals are grown in the presence of polyethylene glycol 6K [10–20% (w/v)] at around pH 6.0. Optimization for each particular carbohydrate requires small adjustments in the conditions; however, all complexes give some crystalline precipitate in this limited grid. The α 1-2 mannoside complex crystals diffract to 1.75 Å with space group *I*222 and cell dimensions $a = 91.7$, $b = 86.8$, $c = 66.6$ Å. One monomer is present in the asymmetric unit. The 1-methyl α 1-2 mannoside complex crystallizes in space group *P*2₁2₁2₁, cell dimensions $a = 119.7$, $b = 119.7$, $c = 68.9$ Å and diffract to 2.75 Å. One tetramer is present in the asymmetric unit. Two crystal forms of the conA–fructose complex have been obtained. The first has space group *P*2₁2₁2₁, cell dimensions $a = 121.7$, $b = 119.9$, $c = 67.3$ Å with a tetramer in asymmetric unit and diffracts to 2.6 Å. The second crystallizes in space group *C*222₁, cell dimensions $a = 103.3$, $b = 117.9$, $c = 254.3$ Å with two dimers in the asymmetric unit and diffracts to 2.42 Å. Structures and crystallization of the trisaccharide–conA and pentasaccharide–conA complexes have already been reported. In all complexes, the protein is found as tetramer, although varying combinations of non-crystallographic and crystallographic symmetry are involved in generating the tetramer. The precise packing of the tetramer varies from crystal to crystal and it is likely that this variability facilitates crystallization.

1. Introduction

Protein–carbohydrate interactions are ubiquitous in biology and are important in inflammation, immune response and infection (Dwek, 1996; Ni & Tizard, 1996; Lowe & Ward, 1997). As such, these interactions are clear targets for therapeutic intervention. Current modelling approaches are unsatisfactory for protein–carbohydrate complexes. Understanding the basis of protein specificity for oligosaccharides not only has application in designing appropriate force fields for modelling, but is also of considerable intellectual interest. This knowledge can only be obtained by probing the contributions of surface area, hydrogen bonding, van der Waals interactions, sugar conformation and solvent reorganization to binding energy. This requires thermodynamic and structural characterization of a wide range of protein–carbohydrate complexes. Chemical synthesis of novel sugars serves to extend the range of carbohydrates beyond naturally occurring sugars and to test specific hypotheses. If crystallography is to provide

valuable information, it must be able to do so quickly and reliably. We have chosen to work with the plant lectin, concanavalin A (conA). The protein has several advantages for a study of protein–carbohydrate interactions: it is readily available and possesses an increasingly well characterized specificity for oligosaccharides.

Although conA was crystallized in a form suited to X-ray diffraction in the 1970s (Hardman & Ainsworth, 1972) well diffracting crystals of conA in complex with a sugar proved elusive and were not reported until the conA–glucose complex in 1987 (Yariv *et al.*, 1987). This was followed in 1989 by the structure of conA complexed with α -D-mannopyranoside (Derewenda *et al.*, 1989). These two complexes were obtained under very different crystallization conditions and the delay between native and complex crystallization reflects the difficulty in obtaining co-crystals. Soaking of the protein crystals in carbohydrate-containing solution results in crystal disruption. The mannose–conA complex unambiguously identified the binding site and the interactions involved in monosaccharide recognition by the protein. This work has been developed by higher resolution studies of the mannose and glucose complexes (Naismith *et al.*, 1994; Harrop *et al.*, 1996). The lectin itself has now been studied at atomic resolution (Deacon *et al.*, 1997).

Whereas the monosaccharides (glucose and mannose) are bound weakly ($K_a = 0.82 \times 10^4 M^{-1}$) by conA, oligosaccharides are bound more strongly (K_a up to $1.41 \times 10^6 M^{-1}$ for the pentasaccharide; Mandal *et al.*, 1994; Toone, 1994; Chervenak & Toone, 1995) and are specific to a particular lectin. Structures of the cognate trisaccharide and pentasaccharide complexes of conA have been reported (Loris *et al.*, 1996; Naismith & Field, 1996; Moothoo & Naismith, 1998). These studies highlighted the extended nature of the binding site and the role of a single structurally conserved water. The pentasaccharide structure showed that conA is capable of significant distortion of carbohydrate conformation. Relating these observations to the thermodynamic data obtained by calorimetry is the crux of developing a rational understanding of protein–carbohydrate interactions.

We now report a general method for obtaining carbohydrate complexes of conA and describe the preliminary X-ray characterization of these complexes. This method will allow rapid X-ray analysis and thus allow us to relate structural results to thermodynamic data.

2. Crystallization and data collection

Concanavalin A and fructose were obtained from Sigma (Poole, UK), α 1-2 mannoside was obtained from Dextra Laboratories (Reading, UK) and 1-methyl α 1-2 mannoside was synthesized at the University of St. Andrews. The general

protocol for crystallization and structure determination of the complexes follows.

Solutions were made up containing 0.6 mM conA, 18 mM oligosaccharide, 1 mM MnCl₂, 1 mM CaCl₂, 20 mM Tris pH 7.0 and 50 mM NaCl. An equal volume of this solution was mixed with the reservoir solution and the resulting mixture was equilibrated against the reservoir solution at 293.5 K.

The reservoir grid consists of 10, 15, 20 and 25% polyethylene glycol (PEG) 6 K, pH 4.0, 5.0, 6.0, 7.0 and 8.0. This grid is repeated in the presence of 1.0 M LiCl. Optimization of PEG 6K concentration, pH and sugar concentration then follows. Data were collected on a Nonius DIP2000 from crystals sealed in glass capillaries with a small amount of the mother liquor. Data were processed and merged using DENZO and SCALEPACK (Otwinowski, 1993). Structures were determined using the molecular-replacement procedure AMoRe (Navaza, 1994) as implemented in the CCP4 suite (Collaborative Computational Project, Number 4, 1994).

Details are given below for crystallization and structure determination for each complex currently being studied. In the absence of carbohydrate, native conA crystallizes under these conditions. The trimannoside and pentasaccharide complexes, which have already been published (Naismith & Field, 1996; Moothoo & Naismith, 1998), are not discussed here but were obtained in an analogous manner.

2.1. α 1-2 Mannobiose

Crystals were obtained with a reservoir containing 10% PEG 6K and 0.1 M citric acid pH 5.0. Crystals which grew in one week were prone to cracking, whilst slower growing crystals, which took up to two weeks to grow, were less prone to cracking. Data were collected from a crystal of dimensions 1.2 × 1.0 × 0.8 mm which diffracted to a resolution of 1.75 Å. Data were recorded as 141 non-overlapping 25 min 0.75° oscillations. A second low-resolution data-collection pass enabled low-resolution data which was obscured by the backstop to be collected. The data were processed and merged in an *I*-centred orthorhombic lattice with unit-cell dimensions $a = 91.7$, $b = 86.8$, $c = 66.6$ Å and gave an R_{merge} of 6.1%, an average I/σ of 11.3, a completeness of 95% and a multiplicity of 3.9. The highest resolution shell (1.81–1.75 Å) has an R_{merge} of 23.9%, a completeness of 80% and a multiplicity of 2.8. The asymmetric unit contains one monomer with a Matthews' number of 2.59 Da Å⁻³ (Matthews, 1968). Molecular-replacement calculations confirmed the space group as *I*222. In the first difference Fourier maps, electron density was visible for the carbohydrate. The structure is being refined to 1.75 Å; this, and a comparison to the atomic resolution native structure (Deacon *et al.*, 1997), will be reported elsewhere.

2.2. 1-Methyl α 1-2 mannobioside

Crystals were obtained with a reservoir containing 13.5% PEG 6K, 1.0 M LiCl and 0.1 M Tris pH 7.0. The crystals were elongated rods and varied considerably in size. Some crystals grew as hollow rods. The crystals took 2–4 weeks to grow and a crystal with dimensions 0.5 × 0.2 × 0.2 mm was used for X-ray analysis. Data to 2.75 Å were recorded as 190 non-overlapping 32 min 0.8° oscillations. Autoindexing with DENZO suggested two different lattice types: primitive tetragonal and primitive orthorhombic. Data were integrated in both lattices. Frames 1–30 showed a high *R* factor and low I/σ compared with the remaining data. Inspection of the images showed that the cell

contracted ~1 Å along the *c* dimension during the first 30° of data collection, after which the cell ceased to contract. The diffraction intensity increased as the cell decreased. Frames 1–30 were excluded from all further analysis. Merging the data in Laue groups *4/m*, *4/mmm* and *2/mmm* gave R_{merge} values of 47.4, 50.8 and 10.1%, respectively. The space group was assigned as *P*2₁2₁2₁ (on the basis of systematic absences) with unit-cell parameters $a = 119.7$, $b = 119.7$, $c = 68.9$ Å to give an R_{merge} of 10.1%, an average I/σ of 8.9, a multiplicity of 3.29 and a completeness of 99.7%. The highest resolution shell (2.85–2.75 Å) has an R_{merge} of 29.0%, 99.5% completeness and a multiplicity of 4.9. The asymmetric unit contains one tetramer and approximately 49% solvent. Molecular-replacement calculations confirm the space group and density for the sugar is visible in preliminary maps. Refinement of the complex is proceeding and will be reported elsewhere.

2.3. Fructose

Crystals were obtained from a protein solution containing 30 mM fructose equilibrated against a reservoir containing 20% PEG 6K and 0.1 M HEPES pH 6.5. Data were collected from one crystal of dimensions 0.5 × 0.3 × 0.2 mm as 90 non-overlapping 25 minute 1.0° oscillations to a resolution of 2.6 Å. Autoindexing suggested a primitive orthorhombic lattice. On the basis of systematic absences the space group was assigned as *P*2₁2₁2₁, with cell dimensions $a = 121.7$, $b = 119.9$ and $c = 67.3$ Å. Merging the data gave an R_{merge} of 11%, a completeness of 83.7%, an average I/σ of 7.1 and a multiplicity of 3.7. The highest resolution shell (2.69–2.60 Å) has an R_{merge} of 29.1%, a completeness of 62.8% and a multiplicity of 2.3. The asymmetric unit contains one tetramer with a Matthews' number of 2.41 Da Å⁻³ (Matthews, 1968). Molecular-replacement calculations confirm the space group. Electron density is visible for the sugar in one subunit. Refinement is proceeding and details will be reported elsewhere.

Crystals of a second conA–fructose complex were obtained from a protein solution containing 40 mM fructose equilibrated against 10% PEG 6K and 0.1 M HEPES pH 7.0 in a hanging drop. Large crystals (0.8 mm³) were obtained after 2 weeks and diffracted to 2.42 Å. Data were collected as 135 non-overlapping 20 min 0.8° oscillations. Fig. 2 shows one 0.8° oscillation image of this data. Autoindexing with this data suggested a *C* orthorhombic lattice with unit-cell parameters $a = 103.3$, $b = 117.9$, $c = 254.3$ Å. The data merged with an R_{merge} of 8.1%, a completeness of 97.2%, an average I/σ of 8.3 and a multiplicity of 4.7. The highest resolution shell (2.51–2.42 Å) has an R_{merge} of 27.5%, a completeness of 87% and a multiplicity of 2.9. The asymmetric unit contains two dimers, approximately 67% solvent. Molecular-replacement calculations using a dimer as the search model gave two solutions in space group *C*222₁. Electron density is visible for the saccharide in the binding sites. Refinement of this structure is ongoing and will be reported elsewhere.

3. Discussion

A general protocol involving optimization of a narrow range of parameters for the crystallization of conA–saccharide complexes has been used to obtain several structures of these complexes. The protein is consistently found as a tetramer. The sugar always binds at a region of crystal contacts and leads to a variation in the crystal packing. We hypothesize that it is this

flexibility which leads to the ease of crystallization of conA with oligosaccharides of varying sizes. The combination of current thermodynamic studies with the information gained from the rapid structure determination of these complexes provide a powerful insight into the factors which govern protein-carbohydrate interactions.

Fig. 1. (a) α 1-2 Mannobiose, R = H, Me. (b) Fructose. (c) The core of the N-linked glycan found on mammalian cell surfaces. The pentasaccharide is shown in bold.

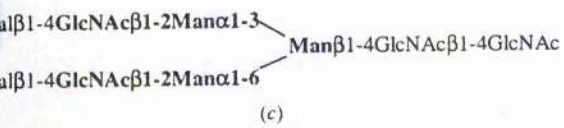
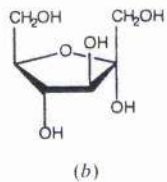
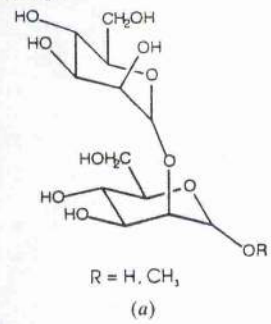
Fig. 2. Section of a 0.8° oscillation diffraction pattern from a C22₁ crystal of the conA-fructose complex. Due to the long c cell edge (254.3 Å) the spots are close together.

We thank Rob Field for synthesizing 1-methyl α 1-2 mannobiose. The research is supported by the BBSRC (B08307).

References

- Collaborative Computing Project Number 4 (1994). *Acta Cryst.* D50, 760-763.
- Chervenak, M. C. & Toone, E. J. (1995). *Biochemistry*, 34, 5685-5695.
- Deacon, A., Gleichmann, T., Kalb (Gilboa), A. J., Price, H., Rafferty, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* 53, 4305-4312.
- Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb, A. J., Dodson, E. J., Papiz, M. Z., Wan, T. & Campbell, J. (1989). *EMBO J.* 8(8), 2189-2193.
- Dwek, R. A. (1996). *Chem. Rev.* 96, 683-720.
- Hardman, K. D. & Ainsworth, Initials? (1972). *Biochemistry*, 11, 4442-4447.
- Harrop, S. J., Helliwell, J. R., Wan, T. C. M., Kalb, A. J., Tong, L. & Yariv, J. (1996). *Acta Cryst.* D52(1), 143-155.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A47, 110-119.
- Loris, R., Maes, D., Poortmans, F. L. W. & Bouckaert, J. (1996). *J. Biol. Chem.* 271(48), 30614-30618.
- Lowe, J. B. & Ward, P. A. (1997). *J. Clin. Invest.* 99(5), 822-826.
- Mandal, D. K., Kishore, N. & Brewer, C. F. (1994). *Biochemistry*, 33(5), 1149-1156.
- Matthews, B. W. (1968). *J. Mol. Biol.* 33, 491-497.
- Moothoo, D. N. & Naismith, J. H. (1998). *Glycobiology*, 8(2), 173-181.
- Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb, A. J. & Yariv, J. (1994). *Acta Cryst.* D50(6), 847-858.
- Naismith, J. H. & Field, R. A. (1996). *J. Biol. Chem.* 271(2), 972-976.
- Navaza, J. (1994). *Acta Cryst.* A50, 157-163.
- Ni, Y. & Tizard, I. (1996). *Vet. Immunol. Immunopath.* 55(1-3), 205-223.
- Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Isaacs and S. Bailey, pp. 56-62. Warrington: Daresbury Laboratory.
- Toone, E. J. (1994). *Curr. Opin. Struct. Biol.* 4(5), 719-728.
- Yariv, J., Kalb, A. J., Papiz, M. Z., Helliwell, J. R., Andrews, S. J. & Habash, J. (1987). *J. Mol. Biol.* 195(3), 759-760.

296.Fig. 1



296.Fig. 2

