Neutron and X-ray protein crystallography studies of lectin saccharide interactions

A thesis submitted to the University of Manchester
for the Degree of Doctor of Philosophy
in the Faculty of Science and Engineering

2003

M. P. Blakeley

Department of Chemistry

Contents

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification at this or any other university or institute of learning.

Copyright

The Author

The author graduated from the University of Manchester in July 1999 with a Bachelor of Science (Hons.) in Chemistry and since October 1999 has been engaged in research under the supervision of Professor J. R. Helliwell in the Department of Chemistry, Manchester. The second and third years of this project were spent at the Institut Laue Langevin, Grenoble, under the supervision of Dr. D. A. A. Myles.

Acknowledgements

I would like to express my sincere thanks to my supervisors for their encouragement and guidance throughout this project. In particular, thanks to John Helliwell for giving me the opportunity to work in Grenoble and to Dean Myles for endless discussions, free coffees and keeping it to himself when I made stupid mistakes.

I'd like to thank EPSRC and ILL for providing the necessary funding and invaluable facilities and expertise to conduct this research. Thanks also to the members of staff at the Department of Chemistry, the ILL and the EMBL who gave me valuable help and advice.

Many thanks to all my friends, in particular Jim, Jo, Cookie, Leila, Ryder and Bob who all let me stay at their respective abodes while I was writing up. Thanks to Karl for helping me with some of the figures and Flora and Ingrid who made me feel at home in France. Cheers to Dave and FB for always helping me when I need it. Plus can't forget the Kro Bar which has provided many moments of relaxation.

Special thanks go to Anuja for putting up with me while I was stressed and to my Mum for always being there.

Abstract

Chapter 1 reviews carbohydrate-binding proteins and the different types of interactions involved in binding. The lectins, a class of carbohydrate-binding proteins, are described in more detail. *Chapter 2* is a review of neutron protein crystallography and the technical advances in the field. The chapter also describes the identification of hydrogen or deuterium atoms by high-resolution neutron crystallography, yielding information that is complementary and supplementary to that gained by X-ray protein crystallography. *Chapter 3* describes my key findings from the three-dimensional structure X-ray analysis of the complex between peanut agglutinin (PNA) and T-antigen at room temperature to 2.25 Å resolution. The results of the lower resolution study[143] have been confirmed, in that the enhanced affinity for T-antigen over lactose is generated by extra water mediated interactions. Furthermore, from analysis of the water structure in general, it has been shown how important the bound waters are for the effective functioning of this protein. *Chapters 4* and *5* describe how I was able to successfully cryo-cool a large concanavalin A crystal and collect both a neutron data set at 12K to 2.5 Å and an X-ray data set at 100K to 1.65 Å resolution. The resulting 12K neutron structure has then been compared to the 293K neutron structure[51]. Many more waters as D_2O are identified at 12K than at 293K and those that are conserved at both temperatures have been identified. The definition of the nuclear density of the waters bound in the saccharide binding site has been improved at 12K relative to 293K. In contrast, increased disorder at the metal sites was observed at 12K compared to 293K in line with current results from EPR experiments in which the local symmetry at the Mn^{2+} site deviates from the crystal symmetry at low temperatures (4K, 120K and 135K)[258]. *Chapter 6* describes my three-dimensional structure X-ray analysis of the complex between methyl-α-D-glucopyranoside and D_2O cubic concanavalin A to 2.1 Å resolution. The D_2O cubic structure was then compared against the previous H_2O cubic X-ray structure[8] and the two structures were found to be similar. Metal to ligand distances and sugar-lectin hydrogen bond interactions were compared and found to be the same within the precision of the structures. Thus, the D_2O soaking did not perturb the structure (at this resolution at least). However, the B-factors of the D_2O structure are higher than for the H_2O structure and this warrants further study. *Chapter 7* describes possible future work based on the work in this thesis, such as yet larger unit cell problems, data collection on fully deuterated protein crystals and perhaps most exciting, neutron studies of cryo-cooled large protein-crystals.

Glossary of Abbreviations

Abbreviations

Ala	Alanine
Arg	Arginine
Asp	Aspartic acid
Asn	Asparagine
Cys	Cysteine
Glu	Glutamic acid
Gln	Glutamine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
Gal	Galactose
Glc	Glucose
Man	Mannose
Lac	Lactose
Fuc	Fucose
L-Fuc	L-Fucose
Abe	Abequose
Gal*N*Ac	*N*-aceytlgalactosamine
Glc*N*Ac	*N*-acetylglucosamine
Lac*N*Ac	*N*-acetyllactosamine
ILL	Institut Laue Langevin

ESRF	European Synchrotron Radiation Facility
HEW	hen egg-white
r.m.s.	root-mean-square
g	Gramme(s)
LADI	Laue diffractometer
σ	sigma
Hz	Hertz
UHR	Ultra-high resolution
SR	Synchrotron radiation
ESR	Electron Spin Resonance
M	Molar
EPR	Electron Paramagnetic Resonance
Me	Methyl group
ENDOR	Electron Nuclear Double Resonance
MHz	Megahertz
ml	Millilitre(s)
mmol	Millimole(s)
mol	Mole(s)
PDB	protein data bank
Cd	cadmium
CW	continuous wave
(*v/v*)	volume in volume
n	Integer
con A	concanavalin A
PNA	peanut agglutinin
SBA	soybean agglutinin
GSIV	*Griffonia simplicifolia* lectin IV
NMR	Nuclear magnetic resonance
WBAI	winged bean agglutinin I
WBAII	winged bean agglutinin II
EcorL	*Erythrina corallodendron* lectin
VVL-B4	*Vicia villosa* isolectin B4
pH	$-\log_{10}[H_3O]^+$

ECL	*Erythrina cristagalli* lectin
RPbAI	*Robinia pseudoacacia* bark agglutinin I
DBL	*Dolichos biflorus* seed lectin
PSL	pea lectin
LOLI	*Lathyrus ochrus* isolectin I
LOLII	*Lathyrus ochrus* isolectin II
FRIL	*Dolichos lablab* lectin
LCL	lentil lectin
Con Br	*Canavalia brasiliensis* lectin
DGL	*Dioclea grandiflora* lectin
favin	*Vicia faba* lectin
PHA-L	*Phaseolus vulgaris* agglutinin
MAL	*Maackia amurensis* leukoagglutinin
DB58	*Dolichos biflorus* stem and leaf lectin
UEA-I	*Ulex europaeus* lectin I
UEA-II	*Ulex europaeus* lectin II

Chapter 1

Carbohydrate-binding proteins

and legume lectins

1.1   Carbohydrates

There are 4 major classes of biomolecules; carbohydrates, lipids, nucleic acids and proteins. Of these, the carbohydrates make up most of the organic matter on earth because of their multiple roles in all forms of life[1].

In the early nineteenth century, individual sugars were often named after their source, *e.g.* grape sugar (Traubenzucker) for glucose, cane sugar (Rohrzucker) for saccharose (the name sucrose was coined much later). The term 'carbohydrate' was applied originally to monosaccharides, in recognition of the fact that their empirical composition can be expressed as $C_n(H_2O)_n$. However, the term is now used generically in a wider sense. The generic term 'carbohydrate' includes monosaccharides, oligosaccharides and polysaccharides as well as substances derived from monosaccharides by reduction of the carbonyl group (alditols), by oxidation of one or more terminal groups to carboxylic acids, or by replacement of one or more hydroxyl group(s) by a hydrogen atom, an amino group, a thiol group or similar heteroatomic groups. It also includes derivatives of these compounds. The term 'sugar' is frequently applied to monosaccharides and lower oligosaccharides[2].

Emil Fischer[3] began his fundamental studies on carbohydrates in 1880. Within ten years, he could assign the relative configurations of most known sugars and had also synthesized many sugars. This led to the necessity to name the various compounds. Fischer and others laid the foundations of a terminology still in use, based on the terms triose, tetrose, pentose, and hexose. He also endorsed Armstrong's proposal to classify sugars into aldoses and ketoses.

The concept of stereochemistry, developed since 1874 by van't Hoff and Le Bel, had a great impact on carbohydrate chemistry because it could easily explain isomerism. Rosanoff in 1906 selected the enantiomeric glyceraldehydes as the point of reference; any sugar derivable by chain lengthening from what is now known as D-glyceraldehyde belongs to the D-series, a convention still in use.

D-Glyceraldehyde   L-Glyceraldehyde
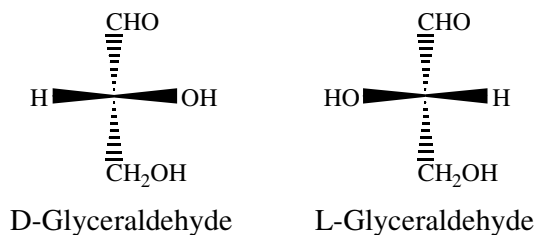
Figure 1.   The two enantiomers of glyceraldehyde.

Towards the end of the nineteenth century it was realized that the free sugars existed as cyclic hemiacetals or hemiketals. Cyclic hemiacetals or hemiketals of sugars with a five-membered (tetrahydrofuran) ring are called furanoses, those with a six-membered (tetrahydropyran) ring pyranoses. The following is a schematic representation of pyranose ring closure in D-glucose.
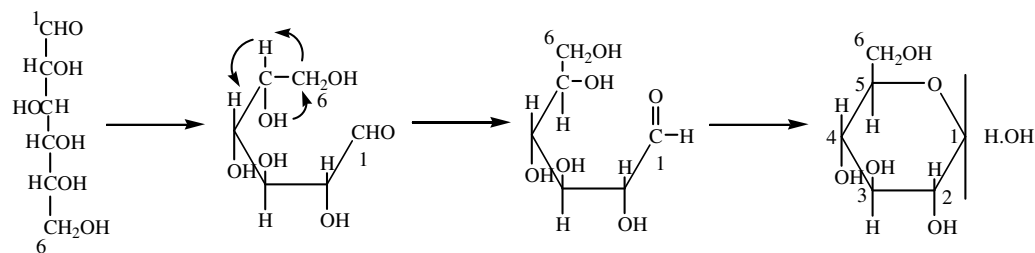
Figure 2.   Haworth representation of D-glucopyranose.

The new centre of chirality generated by hemiacetal ring closure is called the anomeric centre. The two stereoisomers are referred to as anomers, designated α or β according to the configurational relationship between the anomeric centre and a specified anomeric reference atom. For example below are shown the two anomers for D-glucopyranose.

α-D-Glucopyranose β-D-Glucopyranose
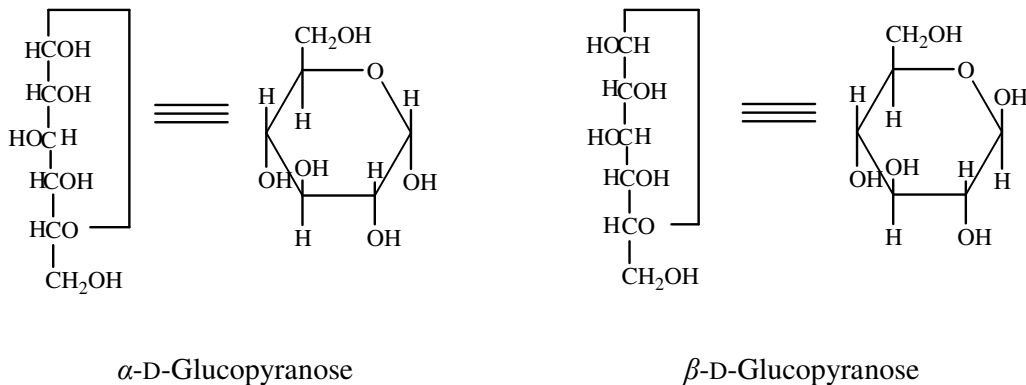
Figure 3.   The two anomers of D-glucopyranose.

The Haworth representation implies a planar ring. However, monosaccharides assume conformations that are not planar. For example, β-D-glucopyranose assumes a chair conformation:
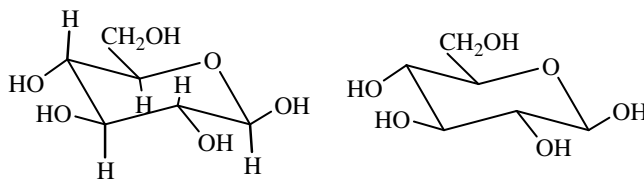
Figure 4. β-D-Glucopyranose in a chair conformation.

Glycosides are mixed acetals formally arising by elimination of water between the hemiacetal or hemiketal hydroxyl group of a sugar and a hydroxyl group of a second compound. The bond between the two components is called a glycosidic bond. Below is an example of a glycoside.
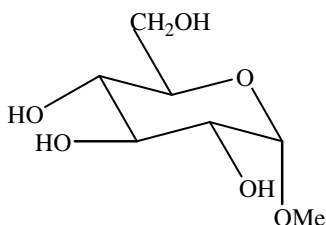
Figure 5.   Methyl-*α*-D-glucopyranoside.

Oligosaccharides are compounds in which monosaccharide units are joined by glycosidic linkages. According to the number of units, they are called disaccharides, trisaccharides, tetrasaccharides, pentasaccharides etc.
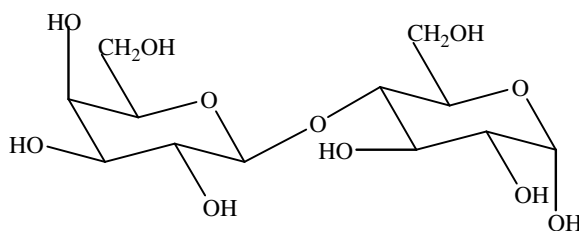
Figure 6.   A   disaccharide,   *β*-D-Galactopyranosyl-(14)-*α*-D-glucopyranose [*β*-D-Gal*p*-(14)-*α*-D-Glc*p*]   or   4-*O*-*β*-D-galactopyranosyl-*α*-D-glucopyranose (trivial name *α*-lactose).
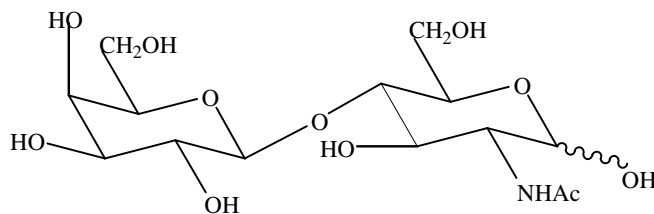
Figure 7.   *β*-D-Galactopyranosyl-(14)-*N*-acetyl-D-glucosamine   (trivial   name *N*-acetyllactosamine; Lac*N*Ac).
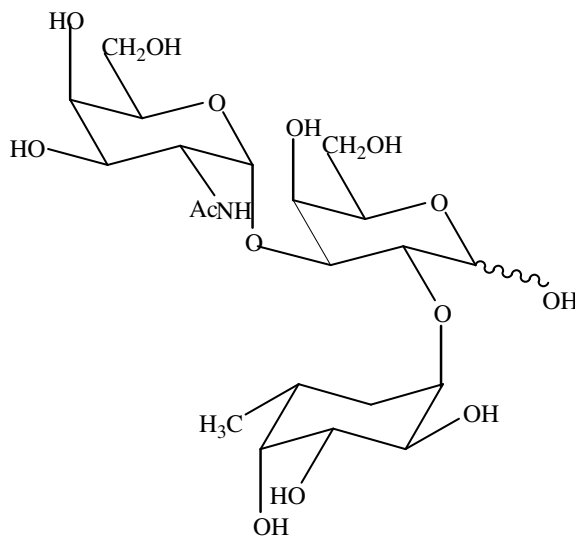
Figure 8.   2-Acetamido-2-deoxy-*α*-D-galactopyranosyl-(1→3)-[*α*-L-
                  fucopyranosyl-(1→2)]-D-galactopyranose   {*α*-D-Gal*p*NAc-(1→3)-[*α*-L-
                  Fuc*p*-(1→2)]-D-Gal*p*} (blood group A trisaccharide).

The borderline with polysaccharides cannot be drawn strictly; however, the term
'oligosaccharide' is commonly used to refer to a defined structure as opposed to a
polymer of unspecified length or a homologous mixture. Polysaccharide (glycan) is
the name given to a macromolecule consisting of a large number of monosaccharide
(glycose) residues joined to each other by glycosidic linkages. Oligosaccharide and
polysaccharide structures occur not only in free form but often as parts of
glycopeptides or glycoproteins or of glycolipids. The ability of carbohydrates to
form complex, diverse structures arises from their ability to bind together *via* any of
their many hydroxyl groups. The resultant linkages are all stereospecific and
branching can occur.

1.2 Carbohydrate-binding proteins

In living cells, carbohydrates are used as a source of energy, as building blocks and as recognition elements. Consequently, carbohydrate-binding proteins are required for cellular processing of carbohydrates to achieve functions such as transportation, degradation, biosynthesis, storage, signalling, antigen-binding and cell-cell adhesion[4,5]. The binding affinities and specificities of these proteins for carbohydrates play a key role in many intracellular and species-specific intercellular processes such as cell fertilisation, development and maturation[6]. In addition, the attachment of bacteria, viruses and toxins to animal cells is mediated through protein-carbohydrate interactions[7].

Nature has selected carbohydrates as recognition elements probably because of their unique ability to generate a vast repertoire of stereoisomers. There is therefore, considerable interest in seeking an understanding at the atomic level of the reasons for the differences observed in specificity and affinity in the interaction between carbohydrates and proteins[6]. Moreover, the unravelling of the relationship between the structure of proteins, carbohydrates and their complexes and the thermodynamics of formation of these complexes, will highlight features important in rational drug design[8].

Due to the rapid advances in many areas of protein crystallography and glycobiology, the number of well-refined structures of complexes of a variety of proteins with carbohydrates has increased considerably, providing a wealth of information[6]. Carbohydrate-binding proteins include lectins, anti-carbohydrate antibodies (*e.g.* monoclonal Fab), bacterial periplasmic binding proteins (*e.g.* L-arabinose binding protein, D-galactose binding protein) and some enzymes (*e.g.* glycogen phosphorylase, lysozyme, xylose-glucose isomerase[9]. It is possible to broadly divide the carbohydrate-binding proteins into two groups[10]. Proteins belonging to the first, such as the bacterial periplasmic binding proteins and glycogen phosphorylase, have binding sites that are buried and essentially inaccessible to the bulk solvent. On the other hand, the second group, represented by lectins, Fabs,

lysozyme and xylose-glucose isomerase have carbohydrate binding sites that are located in shallow grooves close to the protein surface. Furthermore, naturally, the first group have high affinity carbohydrate binding sites whereas the second group tend to have sites with lower affinity[11-14].

1.2.1 Protein-carbohydrate hydrogen bonds

Despite these differences between the two groups, general trends in protein-saccharide interactions are observed for all carbohydrate-binding proteins. Hydrogen bonds confer specificity and affinity upon protein-carbohydrate interactions. Three classes of hydrogen bond are observed: cooperative hydrogen bonds, bidentate hydrogen bonds and hydrogen bond networks[10,15-17].

Cooperative hydrogen bonds result from the simultaneous participation of a sugar hydroxyl group as a donor and acceptor of hydrogen bonds. This is possible as the hydroxyl oxygen is sp^3 hybridised, resulting in a tetrahedral arrangement of two lone pairs and a proton. With few exceptions the cooperative hydrogen bonds follow the simple scheme:

$$(NH)_n \rightarrow OH_c \rightarrow O$$
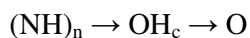
where OH_c represents a carbohydrate hydroxyl group, n = 1 or 2, and $(NH)_n$ and O represent donor and acceptor groups of protein residues (main-chain or side-chain), respectively. According to this scheme, there is a preference of a sugar hydroxyl to accept a hydrogen bond from amide groups and to donate to oxygen atoms. Main-chain amide groups, the asparagine side-chain amide group and sometimes glutamine serve as hydrogen bond donors. In addition, charged side-chains such as arginine act as hydrogen bond donors, and in some cases hydroxyl containing residues such as tyrosine and 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. Infrequently, protein hydroxyl groups have been observed to serve as hydrogen bond donors, as observed in the binding of an inhibitor to the catalytic site of glycogen phosphorylase[12].

Bidentate hydrogen bonds are those in which two adjacent hydroxyls of a hexapyranoside each interact with different atoms of the same polar side-chain residue. Amide groups of consecutive peptides can also play a similar role[18]. A bidentate hydrogen bond is more stable than two monodentate interactions. The chelating interaction is more stable as two hydrogen bonds are formed while only one side-chain is immobilised. The entropic penalty is therefore much less; this is the chelate effect and is well known in biology[19]. The bidentate hydrogen bonds observed thus far are those formed when adjacent pairs of hydroxyls are both equatorial, and when one is equatorial and the other is axial[16]. The sugar ring oxygen atom O5, together with the O4 hydroxyl or the O6 hydroxyl in hexapyranosides can also participate in bidentate hydrogen bonds[16,20,21]. Freedom of torsional rotations of C5-C6 and C6-O6 bonds allows the O6 hydroxyl of hexapyranosides to be paired with either O5 or the O4 hydroxyl (axial or equatorial).
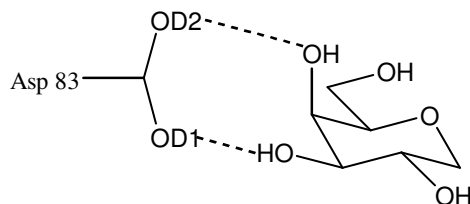
Figure 9.   Example of bidentate hydrogen bond between an Aspartic acid side chain and two sugar hydroxyl groups.

In all bidentate cases, the distance between the two sugar oxygen atoms is ~2.8 Å, ideal for bidentate bonding with the planar side-chains of aspartic acid, asparagine, glutamic acid and arginine. In fact, the majority of hydrogen bonds observed in protein-carbohydrate interactions are formed by these very amino acids. Networks of hydrogen bonds in protein-carbohydrate interactions result from further involvement of binding-site polar groups in hydrogen bonding interactions with other polar groups[10]. Aspartic acid residues in particular play a crucial role in

protein-carbohydrate complexes.   In the D-galactose binding protein[16] (or chemoreceptor protein) of *Escherichia coli*, the binding site is designed for tight binding and sequestering of either the *α*- or *β*-anomer of the D-stereoisomer of the C4-epimers galactose and glucose. The binding specificity and affinity are conferred primarily by polar side-chains that form intricate networks of cooperative and bidentate hydrogen bonds with the sugar substrate. Furthermore, each of the pairs of anomeric hydroxyls and epimeric hydroxyls is recognized by a distinct aspartic acid residue.   A similar situation is observed in the L-arabinose binding protein complex[15], whereby an aspartic acid residue is precisely aligned to accept a hydrogen bond from either the *α*- or *β*-anomeric hydroxyl.
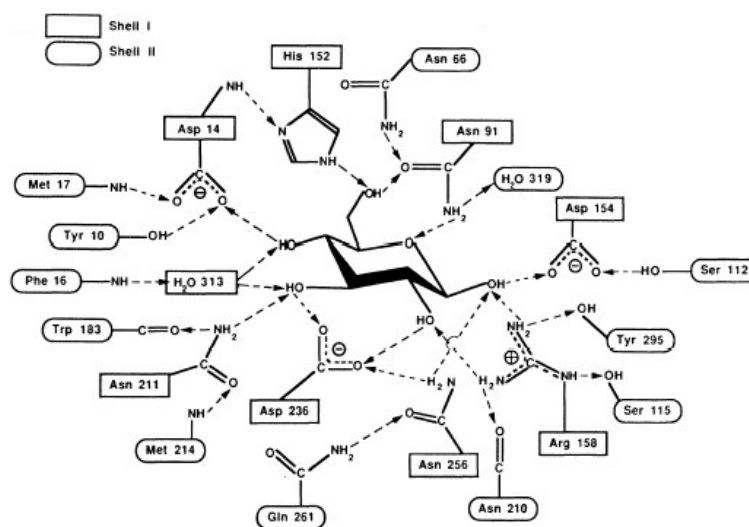
Figure 10.   Schematic representation of the hydrogen bond network in D-galactose binding protein[16].   Residues in shell I are directly hydrogen bonded to the sugar and other residues within the same shell. Residues in shell II hydrogen bond to those in shell I. The directions of the hydrogen bonds (arrows are from donors to acceptors) were dictated solely by the nature of the structure.
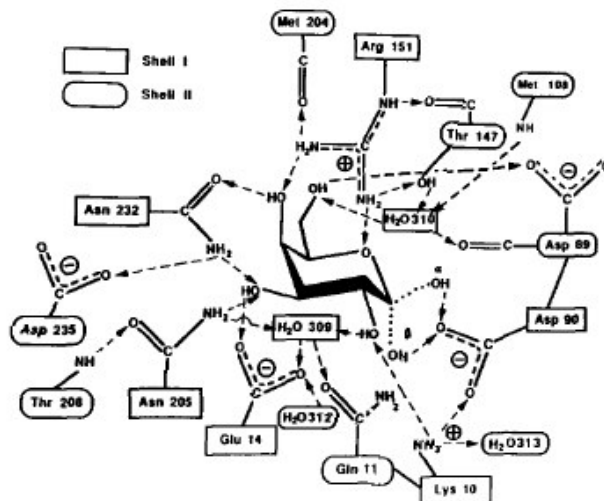
Figure 11.   Schematic diagram of the hydrogen bonds between L-arabinose binding protein and *α*- or *β*-D-galactose[15]. The *α* and *β* anomeric hydroxyls are distinguished by the dotted bonds.   Shell I and II description as in **Figure 10**.

In the refined crystal structure of a glycogen phosphorylase-oligosaccharide complex[12] all the polar groups of the sugar are hydrogen bonded to certain polar groups on the enzyme, with the exception of the ring oxygen. In the refined structure of lysozyme in complex with a trisaccharide[22], hydrogen bonds are shown to help bind all three sugar rings to the protein, utilizing residues such as aspartic acid, asparagine, glutamic acid and also main-chain atoms of alanine and valine.   In particular it has been shown[22] that the distortion of sugar-ring D is aided by a strong hydrogen bond between the NH group of Val 109 and the O6 sugar atom of ring D. In   slight   contrast,   in   the   2.05 Å   resolution   crystal   structure   of   a dodecasaccharide-Fab complex[13] the carbohydrate recognition site is found to be defined primarily by aromatic amino acids rather than the carboxylic acid and amide side-chains that are the main features of other carbohydrate-binding proteins. Nevertheless there are still significant hydrogen bonding interactions between the sugar and protein, although in this case there are less polar residues present in the binding site.
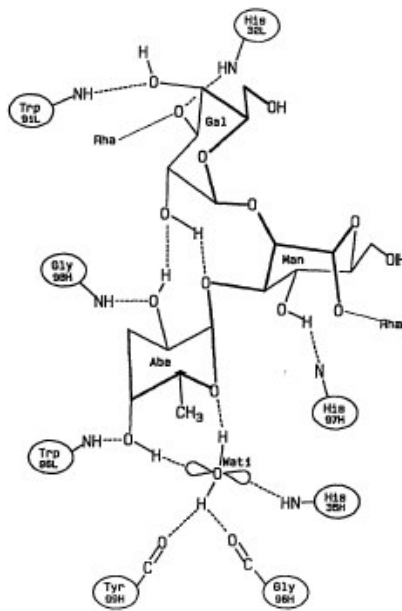
Figure 12.   Schematic representation of hydrogen bonds involving *Salmonella* serogroup type B oligosaccharide and the Fab antibody[13]. The *α*-D-Gal[*α*-D-Abe]*α*-D-Man epitope that is visible in the crystal is shown as well as a water molecule that helps bind the oligosaccharide.

1.2.2 Protein-carbohydrate van der Waals interactions

A second major feature of protein-saccharide interactions in general is that the hydrophobic patch of the sugar-ligand interacts through van der Waals interactions with aromatic residues of the protein such as tryptophan and phenylalanine. This is observed in both the carbohydrate-binding proteins with buried binding-sites, and those with binding-sites exposed to the surface; and is found to be a key feature of protein-carbohydrate interactions[10,13,17,23]. For example, in the glycogen phosphorylase complex with an oligosaccharide[12], stacking of non-polar groups against the oligosaccharide is observed, while in the Fab-dodecasaccharide complex[13] stacking of a tryptophan residue against C3, C4, C5 and C6 of abequose and C1 and C2 of galactose is observed. Another tryptophan residue in the Fab

antibody protein stacks against C6 of abequose and C3 of mannose in a manner similar to the stacking interactions in L-arabinose binding protein[17]. In the lysozyme-trisaccharide complex[22] hydrophobic interactions are observed between tryptophan residues 62 and 108 and sugar rings B and C respectively. Such stacking interactions contribute to the stability and specificity of the complexes.
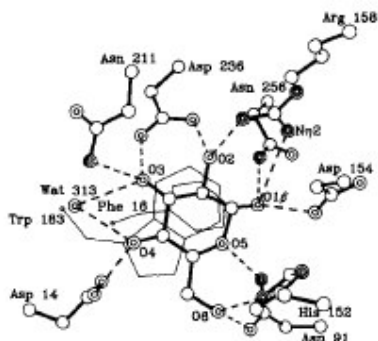
Figure 13.   Illustrates the stacking interaction and the hydrogen bonds between D-galactose binding protein and *β*-D-glucose in the complex[16]. A phenylalanine residue and a tryptophan residue are partially stacked against the A-face and B-face of the glucose respectively.

1.2.3 Protein-carbohydrate water mediated interactions

The hydrogen bond interactions between the hydroxyl groups of the sugar and polar residues of the protein binding site, coupled with the van der Waals interactions between aromatic amino acids and the hydrophobic face of the sugar ring, appear to be the major characteristics of protein-saccharide interactions. However, water-mediated interactions are also evident in most of the protein-saccharide complexes and can be crucial in terms of how tightly a sugar-ligand is bound. In L-arabinose binding protein[15], two isolated water molecules form hydrogen bonds between the sugar and protein. In glycogen phosphorylase[12] contacts between the product and co-factor are mediated by a water molecule and in the lysozyme-trisaccharide complex referred to above, three water molecules link the

sugar ring B to the binding surface of the protein. In the Fab-dodecasaccharide complex[13] the role of a single buried water molecule is pivotal; in addition to coordinating four hydrogen bonds (one bifurcated) it enhances Fab-antigen surface complementarity (see **Figure 12**).
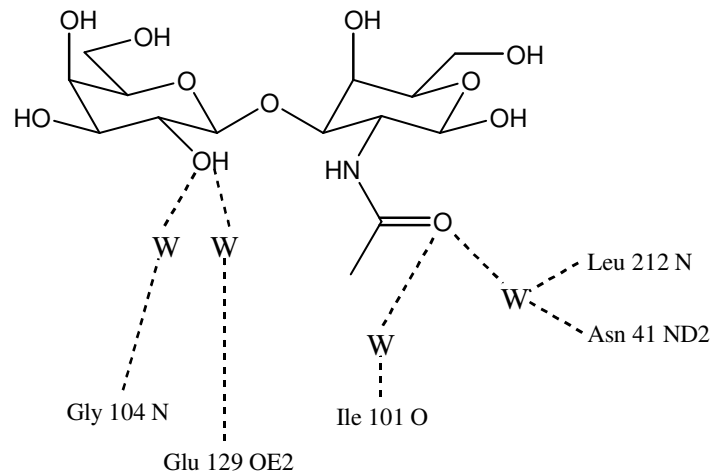
Figure 14.   Example of water mediated hydrogen bond interactions between a disaccharide (T-antigen) and a carbohydrate-binding protein (peanut agglutinin).

Well refined structures of a variety of protein-carbohydrate complexes have contributed immensely to our understanding of the atomic interactions between proteins and carbohydrates.   They clearly show the importance of polar and non-polar residues as well as water molecules for carbohydrate recognition and affinity[6].

1.3   Lectins

Although lectin research had its origins at the turn of the 20th century, these cell-agglutinating and sugar-specific proteins gained wide attention during the late 1960s. This was prompted largely by studies which demonstrated that lectins are extremely useful reagents for the detection and isolation of glycoproteins and for the partial characterization of their carbohydrate moieties, as well as for following changes that occur in cell-surface sugars in such processes as development, differentiation and neoplastic transformation[24].

The term lectin is from the Latin *legere*, which means to pick out or choose. This was first defined[25] to describe a class of proteins of plant origin, which agglutinate cells and exhibit antibody-like sugar-binding specificity. However, the discovery of many lectins in both animal as well as plant tissue prompted several further attempts to reach a common agreement as to how to define a lectin[26-28]. Goldstein *et al.,* (1980)[26] defined lectins as carbohydrate-binding proteins of non-immune origin that agglutinate cells or precipitate polysaccharides or glycoconjugates. This definition implies that lectins are multivalent, that is, they possess at least two sugar binding-sites, which enable them to agglutinate animal and plant cells and/or to precipitate polysaccharides, glycoproteins etc. Accordingly this definition would exclude a wide variety of other saccharide-binding proteins, such as sugar-specific enzymes like glycosidases and glycosyltransferases. The emphasis on non-immune origin is included in the definition in order to distinguish lectins from anti-carbohydrate antibodies, which may act as cell agglutinins[29].

There are four main families of lectin which have been analysed by X-ray crystallography; animal, bacterial, plant and viral. During the last few years, lectin research has grown enormously. While in the past the emphasis was almost exclusively on lectins from plants and their application in biology and medicine, recently there has been as much interest in lectins from animal and bacteria, and the biological significance of their interactions with carbohydrates[29,30]. Lectins from different families all combine with carbohydrates by hydrogen bonds and

hydrophobic interactions, but the amino acids involved may differ even if the specificity of the lectins is the same. It appears therefore that, nature finds diverse solutions for the design of binding sites for structurally similar ligands, such as mono- or oligosaccharides[31].

1.3.1 Animal lectins

Animal lectins have been shown to be implicated in the life of cells, mediating cell-recognition and cell-adhesion[32]. Moreover, they are implicated in direct first-line defence against pathogens, immune regulation and prevention of autoimmunity[33]. Some proteins we now regard as animal lectins were in fact discovered before plant lectins, though many were not recognised as carbohydrate-binding proteins for many years after first being reported. For example, as early as 1853, Charcot and Robin[34] observed strange crystal-like structures in pathological tissues. Similar structures were observed in the sputum of asthmatics by Leyden[35] in 1872. It was later discovered that the structures observed represent a virtually pure form of a single protein (CLCP)[36] which has been shown[37] to recognise and bind carbohydrate, and has been designated as galectin-10.

The first animal (mammalian) lectin activity observed was probably found in snakes. In 1902, a detailed study[38] was published on the agglutination and lysis of erythrocytes and leukocytes by a variety of snake venoms. They attributed the agglutinating activity to "intermediary bodies", or antibody-like factors. As recently as 1988, it was concluded that most animal lectins belonged to one of two structural families: the C-type lectins *i.e.* requiring Ca^{2+} for activity (but excluding Mn^{2+} seen regularly in plant lectins), and the S-type lectins[39] (sulfydryl-dependent or *β*-galactosidase binding). Both families had a conserved domain of approximately 120 amino-acid residues, although the C-type and S-type domains were totally unrelated to each other. The few exceptions at that time were a heterogeneous group referred to as 'N-type' (not C or S). Nowadays there is considerably more structural information available and so the perception has changed[40,41]. It is clear that there are

many Ca^{2+}-dependent lectins that do not possess the defining C-type conserved primary sequence, and S-type lectins (now widely known as galectins) that are not sulfydryl-dependent. Thus the genuine C-type lectins are more diverse than first thought, and can be classified into subgroups depending on gene structure and the nature of additional non-lectin domains.

Above all it is clear that there are many different animal lectin families. If we define a lectin family as a group of two or more carbohydrate-binding proteins with significant sequence identity (other than species homologues), it is possible to list at least 10 families in addition to the galectins and C-type lectins (**Table 1**). Classification is not always straightforward. For example, the ficolins are clearly distinct from tachylectins 5A and 5B and from *Limas flavus* agglutinin, but all share a homologous fibrinogen-like domain that is responsible for carbohydrate binding.

C-type
S-type (galectins)
I-type (siglecs and others)
P-type (phosphomannosyl receptors)
Pentraxins
(Trout) egg lectins
Calreticulin and calnexin
ERGIC-53 and VIP-36
Discoidins
Eel agglutinins
Annexin lectins
Ficolins
Tachylectins   5A   &   5B   ⎫
Limas flavus agglutinin   ⎬ fibrinogen-type

Table 1.   Structural families of animal lectins.

Many animal lectins are not unique molecules, but rather are members of a group of homologues that differ between species. For example, the same term 'galectin-1' is used to describe molecules that differ considerably in primary sequence in those that

have been studied in detail in humans, mice, and rats. These various galectins-1 are nevertheless species homologues and do possess a great deal in common with respect to properties. Paradoxically however, very minor differences in amino acid sequence can be responsible for surprisingly large changes in characteristics[33]. One of the most striking examples of this is between tachylectin-1 and tachylectin-P from the Japanese horseshoe crab. Despite sharing 218/221 (99%) amino acid residues, tachylectin-1 agglutinates bacteria but not erythrocytes, while tachylectin-P is a haemagglutinin without affinity for bacteria. The bacterial agglutinating activity of tachylectin-1 cannot be inhibited by simple sugars, but the erythrocyte agglutination mediated by tachylectin-P is readily inhibited by several *N*-acetylated sugars[42].

Another feature of some animal lectins is their similarity in shape to certain plant lectins; the tertiary fold of some galectins (see **Figures 15/16**) and pentraxins, resemble that of legume lectin structures. These animal lectins have nothing in common with each other or with legume lectins at the amino acid level, and so this appears to be an extraordinary example of convergent evolution[33] or serendipitous use by nature of a molecule involved for a function in another domain.
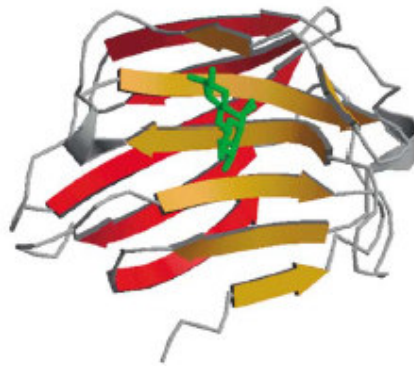
Figure 15.   The human galectin-7 subunit in complex with galactose, with its legume lectin-like fold[41]. Bound galactose is shown in green ball-and-stick representation.

1.3.2 Plant lectins

The plant lectins are sub-divided into those from cereal, bulb, plant toxin or legume sources. Despite their abundance in the seed or bulb of the plant, plant lectins are enigmatic in that their true physiological functions are not clearly defined[43,44] in contrast to mammalian lectins[32]. It has been suggested however, that such large amounts of the lectin are needed by the plant in order to form an impervious cross-linked matrix as a defence mechanism against fungal attack. Such plant defence could be a valuable tool in agriculture and is one of the approaches that has been considered for genetic modification of crops to confer pest resistance[45].

Plant lectins can be purified easily and their specificities make them useful as components of affinity columns for the separation of glycoconjugates or as markers of blood groups and tumour cell lines[46]. Their ability to bind to animal cells and discriminate between different carbohydrates has led to an enormous body of research over the past 70 years.
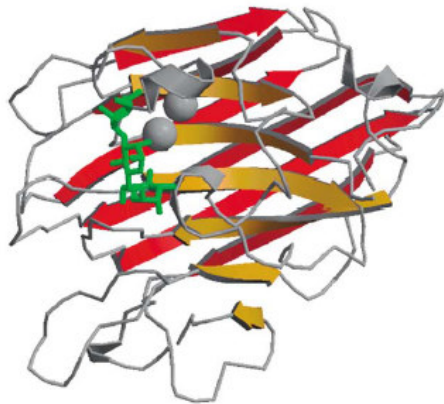
Figure 16.   One of the plant lectins (concanavalin A) in complex with the trisaccharide [Man(α1–3)]Man(α1–6)Man). Metal ions are shown as grey spheres, bound carbohydrate is shown in green ball-and-stick representation[41].

1.4   Legume lectins

Of the plant lectins, those from leguminous sources have been studied the most extensively. Concanavalin A was the first legume lectin to be characterised early in the 20th century. It was isolated from the Jack bean (*Canavalia ensiformis*) as 'bisphenoid' crystals by J. B. Sumner in 1919[47]. To date, concanavalin A is the most well studied legume lectin. It was the first of the lectins to have its three-dimensional structure solved by X-ray crystallography[48], one of the first to have a carbohydrate-bound structure solved[49], and the first lectin to be solved to atomic resolution[50]. It is also the only lectin yet to be studied by high resolution neutron diffraction techniques[51,52].

Figure 17.   Concanavalin A is isolated from jack beans; approx 10% by weight of protein is concanavalin A.

More than 100 lectins from various leguminous sources have been isolated, mostly from seeds, and characterised to varying degrees[31,53]. Because they are readily obtainable in purified form and exhibit an amazing variety of sugar specificities, these lectins are eminently suitable for tackling the questions of how proteins bind carbohydrate. Fundamental insights obtained from the study of these lectins, has been widely applied to lectins from other families, such as the pharmacologically important C-type animal lectins.

1.4.1 Legume lectin structure

Legume lectins resemble each other in their physicochemical properties. The legume lectin subunits, all of which are nearly superimposable, are in the shape of a dome. The carbohydrate-binding sites of the legume lectins, similarly to those of other lectins[54] are in the form of shallow depressions on the surface of the protein. The subunits consist largely of three anti-parallel β-sheets and possess a topology similar to that of the 'jelly roll fold' commonly found in viral coat proteins[55,56]. This has become known as the 'legume lectin fold'[57]. The fold is devoid of α-helices with the residues not involved in β-sheets being part of loops and turns. The three anti-parallel β-sheets consist of a 6-stranded flat sheet at the back of the sub-unit (sheet 1), a 7-stranded curved sheet towards the front of the sub-unit (sheet 2) and a small 5-stranded sheet (sheet 3) at the top of the subunit. The 5-stranded β-sheet plays a major role in connecting the 6-stranded β-sheet and the 7-stranded β-sheet. Furthermore, the loops that connect the strands at the two ends of the 7-stranded β-sheet curve towards and interact with each other to produce a second hydrophobic core, in addition to the one between the two larger sheets[58].The small differences between legume lectin structures are confined primarily to the loops which make up approximately 50% of the structures.
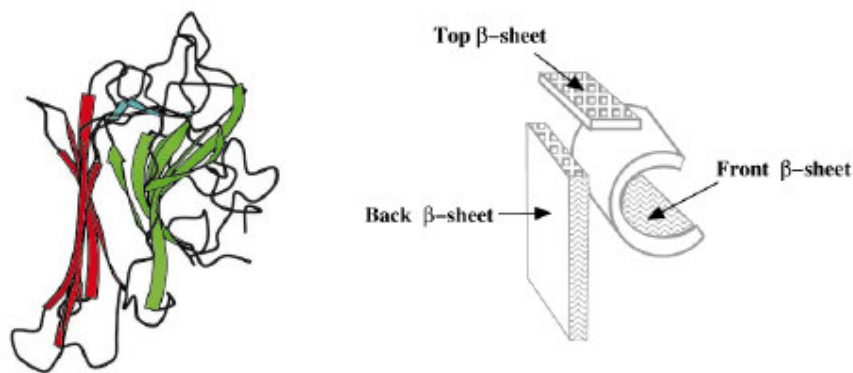
Figure 18.   A ribbon model of the legume lectin tertiary structure and a schematic representation of the legume lectin subunit in terms of sheets.

The legume lectins consist of two or four sub-units of 25-30 kDa, each sub-unit with a single carbohydrate-binding site. As well as displaying a remarkable repertoire of carbohydrate specificities, the legume lectins also display considerable variation in quaternary association, making them of interest also for the analysis of protein-protein interaction (see **Section 3.2.3**). Furthermore, the quaternary structure relates to activity, highly ordered homogeneous cross-linked lattices are formed when lectins are mixed with multivalent carbohydrates, resulting in a higher form of specificity than can be achieved on the level of the monomer[59].

The subunits are most commonly composed of a single polypeptide chain, generally around 250 residues in length. The subunits of some (*e.g.* pea lectin, lentil lectin and *Vicia faba* lectin), are composed of two different polypeptide chains; an *α*-chain of ~6 kDa and a *β*-chain of ~18 kDa. The majority, but not all, are glycosylated and carry up to three asparagine-linked oligosaccharides per subunit[60].

The carbohydrate-binding site is accessible not just to monosaccharides but to oligosaccharides as well. The sites appear to be preformed as few conformational changes occur upon carbohydrate binding. The interaction with carbohydrate requires tightly bound Ca^{2+} and a transition metal ion, usually Mn^{2+}. The Ca^{2+} and the Mn^{2+} are found ~4 Å apart and are in close proximity (9-13 Å) to the carbohydrate-binding site. The transition metal binding site (*S*1) has a slightly distorted octahedral geometry, the metal to ligand distances in the square plane being almost equal with the other two being significantly longer. The Ca^{2+} binding site (*S*2) has a pentagonal bipyramidal geometry. These metal binding sites were first described for concanavalin A[61] and have been found to be extremely well conserved in all other legume lectin structures. A schematic drawing of the metal binding sites is given in **Figure 19**. The two metal sites are bridged by two aspartic acid residues and both metals have four protein ligands and two conserved water ligands. One of the water ligands from the calcium ion forms a bridge with a carbonyl group of an aspartic acid residue (Asp 208 in concanavalin A) which is preceded by a *cis*-peptide bond. This *cis*-peptide bond is crucial for the binding of saccharide as it creates the correct spatial arrangement of protein residues required by reorienting the side-chain

of a key aspartic acid residue (Asp 208 in concanavalin A) into the binding site region. Quite recently the structure of demetallised concanavalin A has been refined at 2.5 Å resolution and the structures of several intermediate steps in the remetallisation process have been presented[62,63]. The key event in the metal-induced activation of concanavalin A is the *trans* to *cis* isomerization of the Ala 207 to Asp 208 peptide bond.
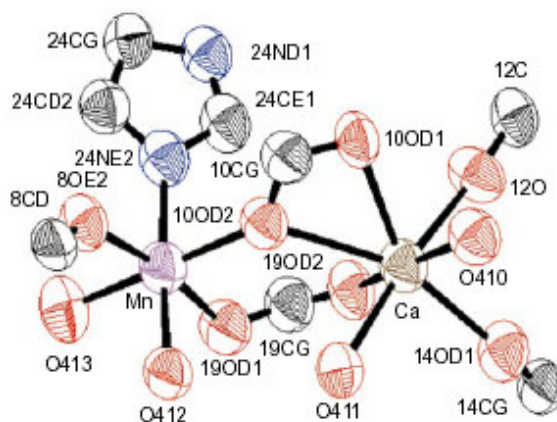
Figure 19.   The metal binding sites *S*1 and *S*2 in the 0.94 Å resolution structure of concanavalin A, depicted in terms of atomic displacement ellipsoids[50].

Many of the primary structures of legume lectins have been established by chemical or molecular genetics techniques. They exhibit reasonable homologies, in general with over 20% of invariant residues and close to 20% similar ones. In certain cases lectins exhibit up to 99% of invariant residues between them, as is the case with concanavalin A and *Canavalia brasiliensis* lectin. These invariant residues within the family include several of those involved in binding of monosaccharides through hydrogen bonds and hydrophobic interactions, and almost all the residues that co-ordinate to the metal ions[24].

1.4.2 Specificity and affinity

Legume lectins can be classed on the basis of the type of carbohydrates they bind to, *viz.* mannose/glucose, galactose/*N*-acetylgalactosamine, L-fucose, *N*-acetylglucosamine/chitobiose or complex sugars (*i.e.* no simple mono- or disaccharide is capable of inhibiting agglutination)[64]. The crystal structures available for the legume lectins at the time of writing, are given below.

Galactose/Gal*N*Ac-binding lectins

Erythrina corallodendron lectin (EcorL)[65-70], peanut agglutinin (PNA)[58,71], soybean agglutinin (SBA)[72], winged-bean agglutinin I (WBAI)[64,73], winged-bean agglutinin II (WBAII)[74-77], *Vicia villosa* isolectin B4 (VVL-B4)[78,79], *Griffonia simplicifolia* isolectins 1-A and 1-B [GSI-A(Gal*N*Ac/GSI-B(Gal)][57,80], *Erythrina cristagalli* lectin (ECL)[81], *Robinia pseudoacacia* bark agglutinin I (RPbAI)[82,83], *Dolichos biflorus* seed lectin (DBL)[84-86]

Mannose/Glucose-binding lectins

concanavalin A (con A)[48,50], pea lectin (PSL)[87,88], *Lathyrus ochrus* isolectin I (LOLI)[18,89-92], *Lathyrus ochrus* isolectin II (LOLII)[93], *Dolichos lablab* lectin (FRIL)[94], lentil lectin (LCL)[95-97], *Dioclea guianensis* seed lectin[98], *Canavalia brasiliensis* lectin (con Br)[99], *Dioclea grandiflora* lectin (DGL)[100-105], *Vicia faba* lectin (favin)[106], *Pterocarpus angolensis* lectin[107]

Complex binding lectins

Griffonia simplicifolia lectin 4(GSIV)[108], *Phaseolus vulgaris* agglutinin (PHA-L)[109], *Maackia amurensis* leukoagglutinin (MAL)[60,110,111], *Dolichos biflorus* stem and leaf lectin (DB58)[84-86]

Chitobiose/*N*-acetylglucosamine binding lectins

Ulex europaeus lectin II (UEA-II)[112,113]

L-Fucose binding lectins

Ulex europaeus lectin I (UE-I)[114]

Even though the lectins can be classified in terms of which type of carbohydrate they bind, subtle but significant differences in carbohydrate specificities exist within each class[64]. Certain lectins exhibit much greater affinities to specific sugars than other members of the same family. For instance, within the mannose/glucose binding legume lectins, concanavalin A, con Br, DGL and *Dioclea guianensis* lectin have been shown to exhibit increased affinity (a 60-fold increase in affinity is observed for concanavalin A) for a trimannoside core [3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside], relative to their monosaccharide binding affinities[115]. This is in contrast to other members of the group such as pea lectin, lentil lectin and LOLI, which exhibit no such enhanced binding to the trimannoside while their monosaccharide binding is very similar to that of concanavalin A[116].

1.4.3 Hydrogen bonding & van der Waals interactions in legume lectins

It has been found that in general, irrespective of their specificity, legume lectins bind monosaccharides through the side-chains of a constellation of three invariant binding site residues; an aspartic acid, an asparagine, and an aromatic residue (phenylalanine, tryptophan, or tyrosine)[54]. The only exceptions to this generalization are *Dolichos biflorus* lectin (DBL)[85] and PHA-L[109], in the binding site of which the invariant aromatic residue is replaced by leucine, and the *Maackia amurensis* lectins[60] MAL and MAH, in which the conserved asparagine is replaced by aspartic acid. Glycine, the fourth amino acid that is invariant in the binding site of almost all legume lectins, participates in monosaccharide binding *via* its main-chain amide; it is replaced by arginine in concanavalin A[49] and by lysine in MAL and MAH[60]. The placement of the aromatic residues appears to be in part responsible for conferring sugar specificity upon legume lectins as they disallow the binding of particular sugar epimers through steric hindrance and/or by creating an unfavourable non-polar environment[10,16,23]. Replacement of the aromatic amino acid by site-directed mutagenesis, with a smaller non-aromatic residue has been shown to produce an

inactive protein for EcorL[67] and *Robinia pseudoacacia* lectin[117], illustrating the essential role of this kind of interaction for ligand binding.
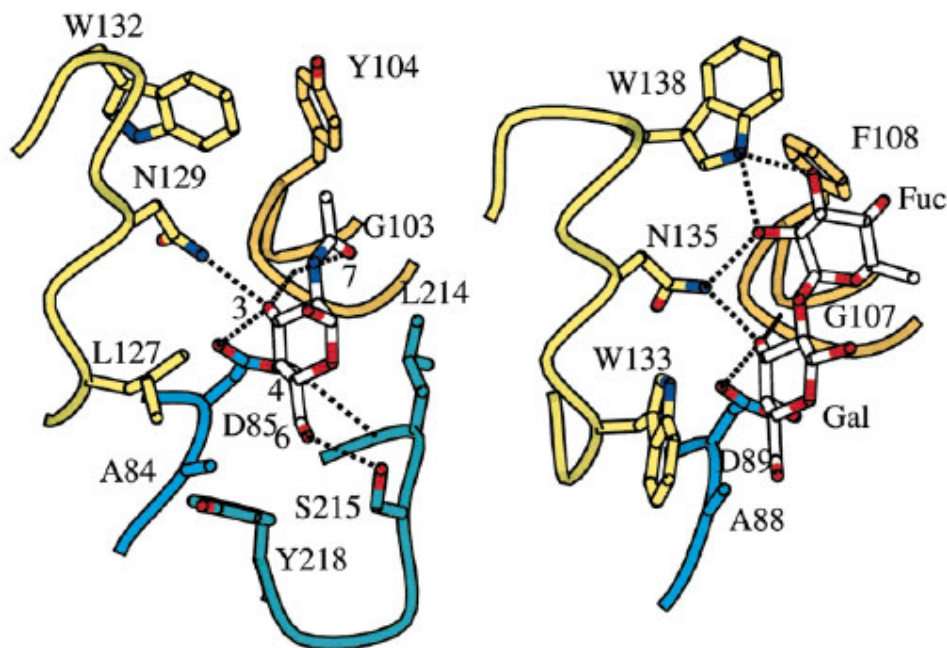
Figure 20.   *Left*, a view on the monosaccharide-binding site of the DBL-blood group A trisaccharide complex[85]. The visible Gal*N*Ac residue and the side-chains of sugar-binding residues are shown as ball-and-stick models. The sugar-binding residues belong to four different stretches, each shown in a different color. Nitrogen atoms are shown in blue; oxygen atoms in red. Hydrogen bonds between sugar and protein residues are shown as broken lines. The sugar hydroxyl groups that are involved in hydrogen bonds with the protein are labelled. *Right*, GSIV in complex with the Lewis b tetrasaccharide. For clarity, only the Fuc(α1-2)Gal moiety is shown.

Furthermore, from investigations of DBL complexes[85] it has been shown that the reason for this lectin's specificity for Gal*N*Ac, as opposed to Gal, is due to the fact that DBL has a leucine residue (Leu 127) at the position normally occupied by an aromatic residue in the majority of legume lectins. The reduction in the contact

surface area as a result of the presence of Leu 127 is compensated by additional hydrophobic interactions between the lectin and the methyl group of the *N*-acetyl moiety of Gal*N*Ac and by an additional hydrogen bond between the *N*-acetyl group and Gly 103. Such a mechanism whereby the lack of stacking of a large aromatic side-chain against the sugar ring is replaced by compensating interactions with the *N*-acetyl group, seems to be a recurring feature in protein-carbohydrate recognition and was recently observed in the complex of the animal lectin tachylectin-2 with Glc*N*Ac[118].

Returning to the key hydrogen bonding amino acid side-chains, it was found that substitution, by site-directed mutagenesis, of the conserved aspartic acid or asparagine in EcorL (Asp 89 and Asn 133 in EcorL) by alanine resulted in the loss of sugar-binding ability[67]. This is in accord with X-ray crystallographic data which shows that Asp 89 forms hydrogen bonds with the 3- and 4-OH groups of the galactose ligand and that Asn 133 is hydrogen bonded with the 4-OH of this sugar, and also with the inability of EcorL to interact with 3- or 4-fluorogalactose. Mutation of these two key amino acids in other legume lectins, such as PHA-L[119], pea lectin[120], and GS-II[117] abolished their binding activity.

1.4.4 How do legume lectins discriminate between galactose and mannose?

The fact that the key amino acids involved in the binding of monosaccharides are highly conserved in all legume lectins and have an identical spatial disposition raises the question of how the legume lectins distinguish between the C4-epimers. By comparison of the crystal structures of monosaccharide complexes of galactose and mannose/glucose specific legume lectins, it can be seen that discrimination is achieved by positioning the ligand in a different orientation in the two groups. For instance, in concanavalin A[8,49], LOLI[18] and lentil lectin[121], mannose and glucose are oriented so that the oxygen atoms of the aspartic acid side-chain form hydrogen bonds to 6-OH and 4-OH of the monosaccharide, whereas the amide of the

asparagine side-chain forms a bond to the 4-OH. In addition, the main-chain amide of glycine (or of arginine in the case of concanavalin A and of the other *Diocleae* lectins) hydrogen bonds with the 3-OH of the monosaccharide. This is in contrast to the legume lectins specific for galactose, such as EcorL[67,69], SBA[72] and PNA[71], where the oxygen atoms of the invariant aspartic acid are hydrogen bonded with the 4-OH and 3-OH, and the amide of the asparagine and the amide of glycine are both bonded to the 3-OH. In this way the same constellation of three highly conserved residues in legume lectins, correctly positioned by the two metal ions either directly or *via* water molecules, provides the framework required for binding diverse monosaccharides[31].
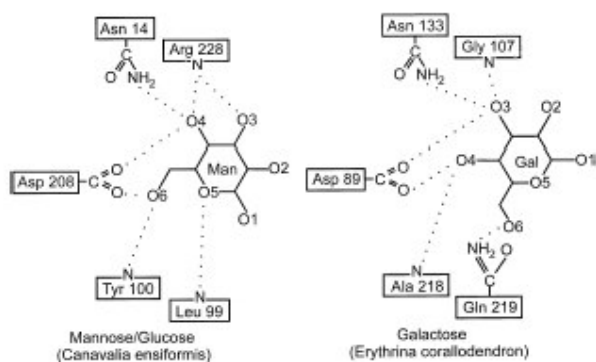
Figure 21.   Monosaccharide recognition. *Left*, schematic diagram of the binding of mannose to concanavalin A. *Right*, similar diagram for the recognition of galactose by the lectin from *Erythrina corallodendron*[59].

It seems fair to say that the invariant amino-acids are not the major determinant of specificity, but this is rather determined by the exact shape of the binding site, and the precise spatial distribution of the surrounding residues. Small changes in the structure of sites may therefore, result in marked changes in protein specificity[24].
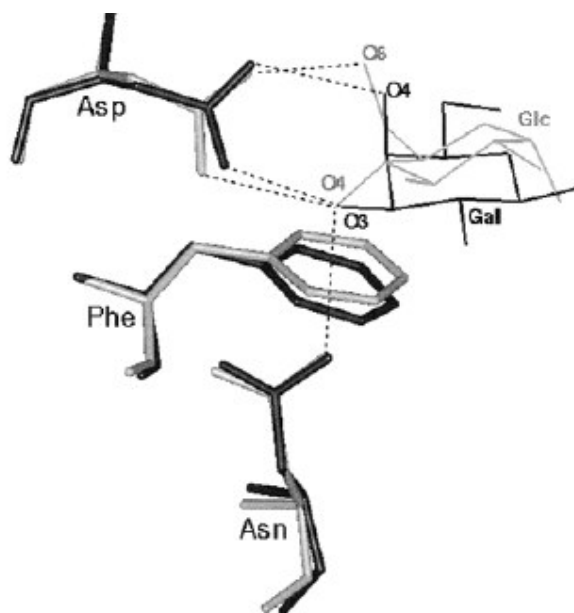
Figure 22.   Superposition of galactose in the combining site of EcorL with glucose in the combining site of LOLI. Note that the galactose is hydrogen bonded to the conserved combining site residues via its 3- and 4-OH groups, whereas the glucose is bonded to the same residue via its 4- and 6-OH groups. The stacking of the ligands on the aromatic phenylalanine residue is also shown[31].

Another way to describe the specificities seen in legume lectins has been suggested[122]. It has been noted on several occasions of the existence of 4 loops that make up the monosaccharide binding site in legume lectins. Through analyses of the size and sequence variability of these loops, (described here as A, B, C, and D) the existence of a common theme that subserves to define their binding specificities was revealed. The most striking observation was the existence of an explicit correlation between the length of loop D and the monosaccharide specificity of the lectins. Thus, loop D of all the Glc/Man-binding legume lectins, despite having significant sequence variation are identical in size (gap of 7 residues), whereas lectins that recognise Gal*N*Ac or only complex glycan chains, (except those belonging to the *Cisalpiniodeae* sub-family) have a corresponding loop length with a gap of 5 residues or less. Most Glc*N*Ac, L-Fuc and Gal-specific lectins exhibit a gap of 6

residues in loop D but varied in the lengths of loop C. Therefore, this implies that the size of the backbone of this loop is a primary determinant of the carbohydrate specificities of legume lectins. In loop D, the sequence variability appears to be inconsequential to a large extent in determining the primary specificity. However, the overall variability in the carbohydrate specificity of legume lectins has, in addition, contributions from the variation in the size of loop C, which further influences the size and openness of the binding site, as well as the sequence differences of loop C, which essentially fine-tune the specificities[122].
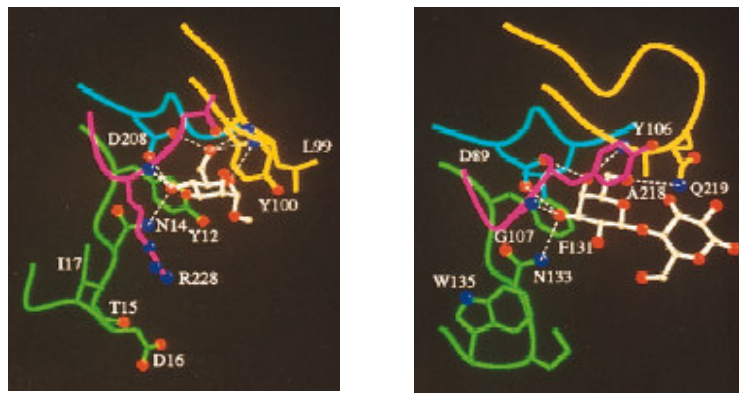
Figure 23.   The binding sites of two legume lectins. Similar orientations and identical colour scheme for the four loops are used for both. *Left*, the Man/Glc-specific lectin concanavalin A (bound to Me*α*Man) and *right*, the Gal*N*Ac-specific EcorL (bound to lactose). Hydrogen bonds between the lectin and sugar are shown as white dotted lines. Side-chains of some of the residues are not shown for clarity[122].

1.5   Summary

Carbohydrates and proteins are ubiquitous in nature. They have many biological roles and have therefore become the focus of much scientific investigation. The interactions between carbohydrates and proteins have been identified as of therapeutic value, and thus the carbohydrate-binding proteins have been studied extensively by means of theoretical, structural and thermodynamic techniques. Detailed and complementary structural information derived from X-ray and neutron diffraction data, combined with calorimetric data allow us to estimate the driving forces behind protein-carbohydrate interactions.

The lectins, a class of carbohydrate-binding proteins, are excellent model systems for investigation into the factors involved in protein-saccharide interactions; they bind saccharides through non-covalent interactions such as direct and water mediated hydrogen bonds and van der Waals interactions, with other factors playing an important role in binding affinity such as carbohydrate conformation, solvent reorganisation and changes in the protein binding site.

Chapter 2

Protein water structure

and neutron crystallography

2.1   Protein water structure

The role of water in biological systems cannot be overemphasised. Proteins almost invariably exist and function in an aqueous environment. Water molecules are seen as mobile molecules, which occupy sites either in the interior of the protein or distributed over the surface. During their residence on these sites, water molecules establish hydrogen bonds with the amino acid residues thus contributing to the stability of the three-dimensional structure and the effective functioning of the protein. Water molecules can make up to four hydrogen bonds and play crucial roles in protein-carbohydrate interactions, the stabilisation of secondary and tertiary protein structures and in the stabilisation of the conformation of oligosaccharides in binding[90,91].

In an X-ray or neutron diffraction experiment, the favoured time-averaged positions occupied in the crystal by water molecules are revealed. The sites range from highly localized sites (high occupancy/low B-factor) in which a water molecule is virtually always present (although possibly rapidly exchanging) to sites which only show a weak preference and cannot be distinguished[123]. This is in contrast to NMR experiments which reveal only water molecules with long (> 300 ps) residence times[124]. For example, in a neutron diffraction study of carbonmonoxymyoglobin, 89 solvent molecules with their hydrogen bonding were revealed[125], whereas NMR studies observed only a few hydration sites[126]. A molecular dynamics simulation of carbonmonoxymyoglobin[127], using the neutron model (described above) as the starting point, revealed that only a few of the waters are tightly bound to protein atoms throughout the whole time frame of simulations. Therefore, most of the bound water sites observed in the neutron maps do not have residence times long enough to be detected by NMR experiments. NMR experiment and neutron diffraction give us very important but different information regarding protein hydration. Neutron diffraction provides an average picture of the interactions between water and protein atoms and measures the average occupancies while NMR experiments are sensitive to the dynamic stability of hydration and measures only cases where water molecules are tightly bound to protein atoms and do not tumble for a relatively long time.
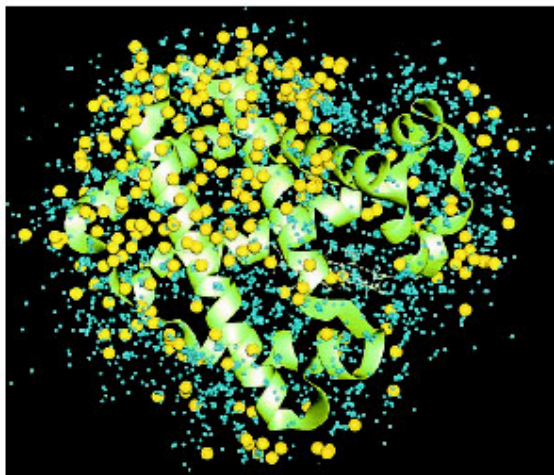
Figure 24.   A view of the myoglobin hydration sites generated in the MD simulation[127] (yellow dots) compared to the collection of the hydration sites contained in the Protein Data Bank (blue dots)[306].

In small crystal structures such as amino-acids, the waters are usually well-ordered, and are few in number. However, as the size of the molecule increases, so do the number of water molecules, therefore, a macromolecule may contain several hundred water molecules. In protein crystals, which can contain anywhere between 25-80% solvent[128], a large proportion of these are found within solvent channels and voids situated between the protein molecules, and thus can be very disordered. Nevertheless, some of the waters are found to be buried within the protein or found hydrogen bonded to amino-acid residues on the surface. In these situations the waters are found to be generally more ordered, as they make interactions with the protein itself as well as solvent-solvent interactions. One or more hydration shells may be present within the protein systems. The first shell generally appears as fairly well-ordered peaks, but as the distance away from the protein increases the shells become progressively more disordered and so in most cases it is rare to observe more than the first two hydration shells[129].

Water molecules in small-molecule crystal hydrates are usually in accurately defined positions. Due to the accuracy of the water positions, useful information regarding the geometries of water hydrogen bonding is gained. The distances and angles of the

standard hydrogen bonded system, X─H----Y, where X and Y are electronegative atoms, have been well established[130,131] in the small-molecule crystal hydrates. General surveys[130-133] of the available small-molecule neutron structures have indicated that:

(1) The X─H----Y angle of the hydrogen bond tends to be close to 180°, with the majority being greater than 150°.

(2) The X─H----Y angle tends to decrease as the X----Y distance increases.

(3) Non-covalent H----Y distances of less than the sum of the van der Waals radii of H and Y are considered as hydrogen bonding.

When X and Y are oxygen atoms, H----O distances of less than 2.4 Å are usually regarded as hydrogen bonds, though becoming very much weaker at distances greater than 2.0 Å. Further analyses of results from hydrated small-molecule neutron structures[134] show that for Y = O, the O----O distances generally range from 2.5 to 3.1 Å (mean = 2.805 Å), while the H----O distances range from 1.5 to 2.3 Å (mean = 1.857 Å). The covalent bond lengths and angles are seen to be fairly widely distributed with means of 0.96 Å and 107°.

A generally accepted definition for hydrogen bonds is that they occur whenever a proton is shared between two electronegative atoms. Hence, only hydrogen atoms bonded to nitrogen or oxygen atoms are usually considered in analyses of protein hydrogen bond networks. However, X-ray and neutron diffraction studies[135,136] have shown that crystals of various organic compounds exhibit close C-H----X contacts (where X is an electronegative atom, in most cases oxygen) which show all the stereochemical hallmarks of hydrogen bonds[137]. Among the hydrogen atoms involved in these contacts, the most common are those attached to α-carbon atoms. This is consistent with the fact that these hydrogen atoms are more acidic than others. In particular, close contacts found between C_α-H groups and main-chain oxygen atoms in adjacent strands of β-sheets are the most ubiquitous.

From calculations of hydrogen bond energies, it has been shown[127] that hydrogen bonding is a significant factor to the stability of the water structure and that at least two hydrogen bonds with a total energy much larger than *kT* are needed to 'bind' a water molecule. The occupancy and residence times of water molecules at hydration sites are clearly correlated with the number of hydrogen bonds that a water molecule shares with the protein atoms.

Although correlations between hydrogen bond distances and angles may be found in small-molecule crystal structures, these geometrical characteristics are generally 'too soft' to be used as stereochemical restraints in refining the solvent structure in protein crystals. From investigations[138] of the water structure around a variety of crystal hydrates, which focused on the short range non-bonded contacts around water positions, it was revealed that there are several structural regularities. These features relate to the non-bonded O----O, H----O, and H----H interactions, and give rise to a set of repulsive restrictions that are seen to be very much stronger stereochemical restraints than those associated with hydrogen bonding. In more general terms, the inclusion of the non-bonding relationships as well as the attractive hydrogen bonding interactions, leads to a significant increase in our understanding of water structure(s). The repulsive restrictions can be used as stereochemical restraints in the interpretation of solvent structures within larger hydrated systems, such as protein crystals[139].

2.1.1 Conserved waters in legume lectins

An analysis of the water structure in the crystals of different legume lectins was carried out by R. Loris *et al.*[140]. They found that a significant proportion of the water oxygen positions appear to be conserved in all independent crystal structures of a given protein. Moreover, some of these water molecules are conserved in crystal structures of homologous proteins as well. Seven water sites are found to be completely conserved in all legume lectin crystal structures, independent of their degree of sequence homology or saccharide specificity. Of these, four water molecules are found as ligands to the metals (two to each of them) and one water molecule is located in the saccharide binding site interacting with the invariant aspartic acid and asparagine residues. Of the remaining two conserved water molecules, one stabilises a *β*-hairpin structure and the other stabilises a *β*-bulge of the backsheet[140] as shown in **Figure 25**.

Conserved waters can significantly affect the specificity and affinity of specific protein-saccharide interactions. It has been shown that a hydrogen bond between carbohydrate and protein, which is mediated by a single water molecule, can be as strong as a direct protein-carbohydrate interaction[141]. Naismith and Field, (1995)[142] provided an example of a legume lectin that utilises a conserved water to enhance specificity, when they elucidated the lectin-sugar crystal structure and details of binding for concanavalin A with the trimannoside core of an *N*-linked glycan. Concanavalin A was found to specifically recognise the trimannoside core of many complex glycans with enhanced affinity over glucose or mannose monosaccharides. This was found to be due, in part, to a conserved water molecule that mediates a hydrogen bond between the protein and trimannoside core. Another example was identified from the crystal structure solution of the EcorL-lactose complex[65]. In this complex it was shown that a water molecule bridges between the 6-OH of the galactose subunit of the lactose and the main-chain oxygen of Leu 86; This water-mediated hydrogen bond aids in stabilization of the complex and increases the affinity of EcorL towards the lactose disaccharide. In the crystal structures of PNA in combination with two disaccharides, it has been shown that there are four waters

that occupy the binding site in both complexes, which play a major role in the observed differences in specificity of PNA for the two disaccharides[58,143] (see **Section 3.3/3.7**). In the LOLI-biantennary octasaccharide complex[91], several water molecules are involved in stabilizing the complex through hydrogen bonds. The saccharide-lectin complex is stabilized by 21 hydrogen bonds, 14 directly between the protein and the saccharide and seven by the way of a single water molecule. In addition, 14 water molecules are involved in indirect interactions linking the octasaccharide to the lectin or to itself. Therefore, water molecules play an important role in both the interactions between saccharides within the octasaccharide itself and between the octasaccharide and the lectin. These examples illustrate the important roles of water molecules in defining carbohydrate specificity and affinity and also the oligosaccharide conformations.

Water molecules are not only utilized to help stabilize the saccharide conformation, tether the sugar-ligand to the protein, and act as ligands to the metal sites, they are also involved in the binding of saccharides through thermodynamic considerations[12,15,144,145]. In the saccharide-free form of many lectins, water molecules have been observed that are positioned at the same positions as the hydroxyl groups of the sugars on binding. These conserved waters are expelled on binding, which can give rise to a favourable contribution to the entropy of the reaction and thereby to the overall free energy of the reaction.
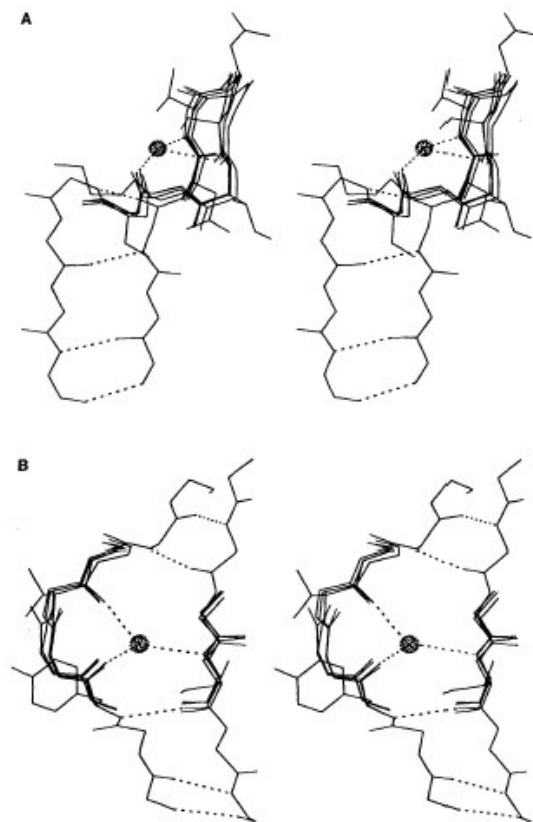
Figure 25.   Stereo view of two hydration sites conserved in all legume lectin crystal structures[140].   Shown are the coordinates of monomer A of orthorhombic lentil lectin, with relevant parts of the backbone atoms of concanavalin A, GSIV and EcorL superimposed.   **A**, conserved water stabilizing a β-hairpin structure.   **B**, a conserved water that stabilizes a β-bulge structure.

2.2 Neutron protein crystallography

2.2.1 Overview of technique

The determination of the structures of biological macromolecules and their bound solvent by X-ray diffraction has been immensely useful in gaining an understanding of biological processes, such as ligand-binding and molecular function, enzymatic catalysis and general protein folding problems[146]. Routine analysis of protein structures is performed extensively at synchrotron radiation sources using high-performance detectors and cryoprotected small crystals. Nevertheless, despite the pre-eminence of the X-ray technique, it has been shown that much information can be gained from neutron diffraction that is complementary and supplementary to that obtained by X-ray diffraction. Neutron crystallography can be used to directly determine the positions of hydrogen atoms of both the protein and its bound solvent at lower resolutions than required with X-rays, and thereby additional information can be gained to that from an X-ray diffraction experiment. The ability to locate hydrogen atom positions in a protein will be crucial for a full comprehension of many reaction pathways and structure-function relationships in biology. In addition, water molecules located in both the interior and on the surface of a protein have been shown to play several diverse and important roles for the efficient functioning of proteins and thus vital information on hydrogen bonding interactions can be gained from neutron data. Using X-rays the ability to locate the hydrogen atoms of a protein and its bound solvent is only possible at extremely high resolution (greater than 1 Å) and where the relevant atoms are ordered sufficiently well. Using complex algorithms, the details of hydrogen atom positions have often been based on calculated position assignments. However, often the positions of the hydrogen atoms calculated on the basis of chemical and energetic considerations are incorrect[51]. The detailed location of hydration and protonation sites can at best only be inferred using these methods.
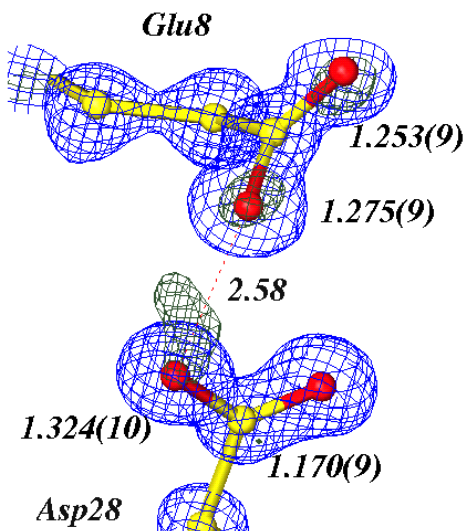
Figure 26.   Direct visualization of hydrogen atoms is possible with X-rays, however, only at very high resolution (> 1 Å) and where the atoms are sufficiently well ordered[50].

For X-rays, the scattering centres are the atomic electrons. Consequently, X-rays are scattered in proportion to the number of electrons thus leading to a strict correlation with the atomic position in the periodic table. Therefore, hydrogen, with only one electron, scatters X-rays very weakly. For neutrons, on the other hand, the scattering centres are the atomic nuclei. Each nucleus has a characteristic interaction with a neutron. There is considerably less variation between the elements and furthermore, the interaction can be different for different isotopes of the same element[147]. The hydrogen atoms have a negative neutron scattering length b_H of –0.374 which is of similar magnitude but opposite in sign to C (b_H = 0.667), N (b_H = 0.94), O (b_H = 0.5805) *etc*. The scattering length is a measure of how efficiently a given atom diffracts neutrons.

Atom type	Scattering length b (10^{-12} cm)
H	-0.374
D	0.667
C	0.665
N	0.936
O	0.580
S	0.285
Mn	-0.373
Ca	0.470

Figure 27.   Neutron scattering lengths for some certain atom types.

The negative scattering factor for hydrogen is caused by a phase shift of a scattered neutron which is 180° different compared with those for most other nuclei. The practical effect of this sign difference is that hydrogen atoms appear (at sufficiently high resolution) as negative peaks in neutron Fourier maps, in contrast to the majority of atoms, which appear as positive peaks[146]. This can help distinguish hydrogen from other atoms. Unfortunately, at medium resolution, negative nuclear density for the hydrogen atoms can simply cancel the positive nuclear density of the atom to which it is attached[52]; techniques such as protein perdeuteration have been developed to overcome this problem (see **Section 2.2.3**).

If we search the protein data bank (date; 11/08/03, http://www.rcsb.org/pdb) we see that out of ~18600 X-ray structures deposited, only around 140 of these have a resolution limit of 1 Å or better, whereas nearly 15000 of these have resolution limit of 2.5 Å or better. Most structures within the PDB then are determined at resolutions around 1.5-2.5 Å, with nearly three-quarters (74.4%) of the structures deposited found in this resolution range (see **Figure 28**). It therefore seems reasonable to suggest that, for many of the proteins investigated, in order to directly observe hydrogen atoms it may be more viable to determine the neutron crystal structure at a

lower resolution than to try and achieve ultra-high resolution data collection with X-rays. To this end suitable neutron protein crystallography facilities are being developed worldwide to serve the need and demand.
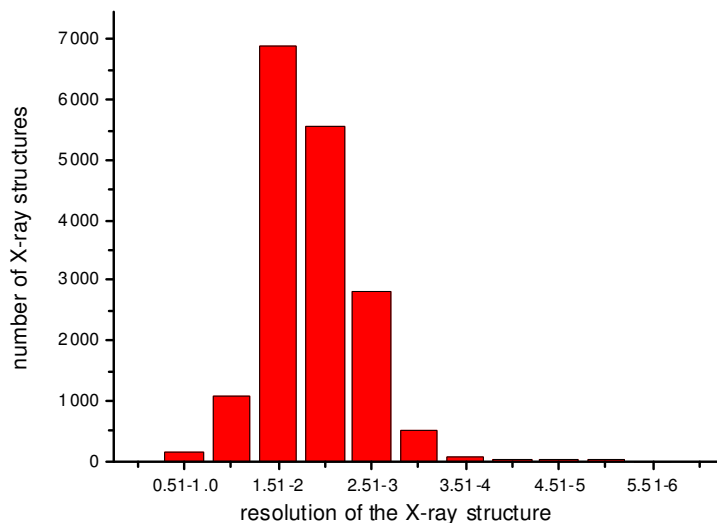
Figure 28. Indicates the number of X-ray structures determined at specific resolutions.

In neutron diffraction studies the neutrons do not cause any significant radiation damage to the crystal. This means that the longer exposure times necessary with neutrons do not harm the crystal and moreover the same crystal can be used again for either further neutron experiments or an X-ray diffraction study. However, there are various limitations associated with neutron protein crystallography that make a successful experiment a lot more demanding than its X-ray counterpart. The low flux of even the most powerful reactor sources, some $10^6 – 10^7$ times less intense than the photon fluxes from X-ray synchrotron sources[146] is a major drawback in performing and increasing the number of successful neutron experiments. Data collection times of several weeks to several months have been necessary in order to achieve reasonable counting statistics.

2.2.2 Technical developments

Neutron protein crystallography nowadays has been considerably improved by developments in perdeuterated protein expression, instrumentation, detection systems, data processing and data analysis. In particular, the development of the neutron Laue technique and instrumentation designed specifically to harness the power of the Laue technique has greatly improved data acquisition times and made many experiments that were considered too expensive (in terms of the amount of beam time needed) become reality.

One specific development is neutron Laue diffraction. Like its X-ray counterpart it uses a stationary crystal during exposure to the incident beam, with the reflections from the crystal extracting the appropriate wavelength from the spectrum emitted from the source[148]. In Laue geometry the sample is illuminated by all available neutrons, thereby maximising the flux at the sample and stimulating large numbers of reflections simultaneously over all incident wavelengths[149]. This, combined with a very large image plate detector, speeds up data collection so that data collection can be reduced to days or even hours[51,52]. Moreover, faster data collection allows the behaviour of protons under H/D exchange conditions, or change in pH, to be monitored more readily through collection of multiple data sets[148]. It is difficult however to employ the whole spectrum of the incident white beam, as this leads to very high background and possible extensive overlap between Bragg reflections[150]. Therefore, quasi-Laue methods[151], where a restricted wavelength band is extracted from the white beam with Ti-Ni multiple spacing multilayers of high reflectivity have been used[152]. Technical developments made at the Institut Laue Langevin, Grenoble, over recent years are several fold, and impressive. The neutron <u>L</u>aue <u>di</u>ffractometer, LADI, incorporates a large area image plate detector of cylindrical geometry which fully encircles the protein sample; the total collection area is 400 x 800 mm^2 so that with a pixel area of 0.2 x 0.2 mm^2, the total number of pixels available to record the neutron Laue diffraction pattern is 2000 x 4000 with this device[153]. This position-sensitive detector approach thus greatly increases the capability to simultaneously record many neutron reflections. A neutron image plate

detector is very similar to an X-ray image plate, in that it consists of photostimulable luminescence materials such as BaFBr and is sensitive to X-rays and gamma rays. However, in order to be sensitive to neutrons the image plate must be doped with a neutron converter such as gadolinium (Gd)[154,155]. Neutron image plates have been successfully produced by mixing the neutron converter, Gd with a photostimulated luminescence (PSL) material on a flexible plastic support[156]. Four Gd_2O_3 doped $BaF(Br.I):Eu^{2+}$ Fuji neutron image plates are bonded onto the outer surface of the cylindrical camera drum of LADI. Diffracted neutrons pass through the thin aluminium drum, entering the phosphor image plate from the back. Reciprocal space is surveyed in step scan mode by rotating the sample about the cylinder axis[149].
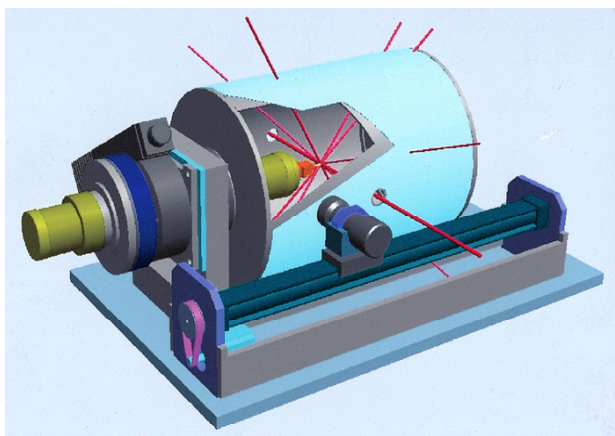
Figure 29.   The Laue Diffractometer at the Institut Laue Langevin.

Traditionally it was feared that with a white beam of synchrotron X-rays or neutrons, the Laue diffraction spots would always be multiple (that is overlapping) thus confounding any quantitative structure analysis by Fourier techniques. These perceived multiplicity problems were addressed in detail by workers using polychromatic SR sources, where very fast data collection opportunities with white X-ray beams became possible[148]. Detailed analysis showed that Laue patterns were, after all, dominated by single, and not multiple reflections, greatly reducing the complexity of the detector requirement[157,158]. However, spatial overlap of spots does occur and becomes more severe as the unit cell length increases[148]. LADI utilises a spectral bandpass of the neutron beam which can be tailored by filters to obtain various $\delta\lambda/\lambda$ *e.g.* 20%. Selection of these filter characteristics (narrow band), are

needed in order to minimise spot-spot overlap when collecting data from large unit cell systems and also to lower the background count, compared to that of broader band white beams[159]. In a neutron diffraction experiment, a large crystal is needed due to the weaker neutron flux relative to its X-ray counterpart, and high beam divergence. However, in neutron Laue diffraction, we can now use smaller crystals, or work on larger unit cells (*i.e.* study larger proteins).

2.2.3 Hydrogen and deuterium

The total scattering cross-section of an atom is the sum of two terms, the coherent and incoherent cross-sections. It is the coherent cross-section that gives rise to interference, while the incoherent cross-section serves only to increase the background[160]. One major problem associated with neutron diffraction of proteins is the large incoherent scattering cross-section of hydrogen. As hydrogen atoms make up approximately 50% of the atoms of a protein structure, a large background signal is then observed. This large background reduces the signal-to-noise ratio and therefore, degrades the resolution limit. Furthermore, at the resolution limits attainable for protein neutron crystallography (around 2 Å), the negative nuclear peaks indicative of hydrogen atoms can be cancelled out by the positive scattering of the atom it is bound to. For example, the positive scattering peak of a water-oxygen atom can cancel out the negative scattering peaks of its associated hydrogen atoms, particularly if the water is disordered. Both of these problems can be improved by the exchange of hydrogen for deuterium. Exchanging hydrogen for deuterium has 3 advantages[51]. Firstly, reduction of the incoherent scattering is possible as deuterium has a much smaller incoherent scattering cross-section than hydrogen. Therefore, the incoherent background signal is lowered which aids the resolution limit of the data. Secondly, since deuterium has a positive scattering length of 0.667, which is of the same sign and of similar magnitude to those of the other atoms of the protein, whereas H has a scattering length of −0.374, contrast-variation studies[161] can be performed. By varying the relative values of protein and solvent scattering densities in a crystal, it is possible to obtain information on the shape and dimensions of

protein molecular envelopes. Neutron diffraction methods are ideally suited to these contrast variation studies because H/D exchange leads to large differential changes in the protein and solvent scattering densities and crucially is structurally non-perturbing[162,163]. Thirdly, deuteration also improves the location of D atoms of both the protein and bound solvent. For the protein, cooperation between the positive density for a D atom and that of the atom to which it is bound *e.g.* the peptide N atom, enhances the density of these atoms and so it is possible to assign both their positions correctly. The same situation occurs with respect to the water molecules. As the scattering lengths of D and O are both positive in sign, the density for water molecules is enhanced and aids in assigning the correct D_2O orientation[160].

Hydrogen incoherent scattering

A-DNA/H2O   A-DNA/D2O
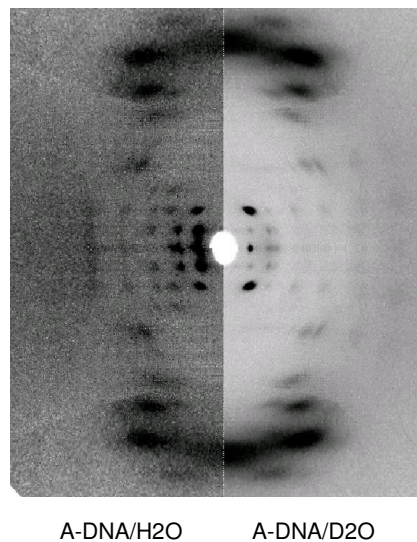
Figure 30.   The larger neutron diffraction pattern background scatter for hydrogens in DNA, as example, vs deuterated DNA (fibres). Figure courtesy of Dr T. Forsyth, Grenoble.

In order to exchange hydrogen atoms within the sample for deuterium, it has been common-place to soak hydrogenated protein crystals in a D_2O buffer or to use a vapour-diffusion method. These methods, however, only exchange those hydrogen atoms which are attached to N or O atoms, and not those bonded to carbon atoms, such as main-chain hydrogen atoms. Moreover, these techniques will only exchange

those hydrogen atoms that are generally exposed and free to exchange *i.e.* the hydrogen atoms buried in the core of the protein may be inaccessible to the D_2O buffer. Also, hydrogen atoms involved in secondary structure stabilization, such as *α*-helices and *β*-sheets appear to be more resistant to exchange. The fact that certain areas of the protein and its bound solvent preferentially exchange with respect to other areas can be used to inspect which regions of the protein are accessible, and thereby provide information on the breathing characteristics of different areas of the protein. However, it is preferential to exchange as many hydrogen atoms as possible in order to reduce the background on the detector. Therefore, in some cases crystals are now grown from D_2O originally, increasing the level of deuterium in the crystal. Growing crystals from D_2O increases the level of exchange but still does not exchange the hydrogen atoms attached to carbon atoms. In order to achieve this, the protein must be expressed from bacteria grown on deuterated media, so that fully perdeuterated proteins can be synthesized[162]. In this case, (almost) all the hydrogen atoms are now replaced by deuterium and therefore, the incoherent hydrogen background scattering is (almost) avoided, the signal-to-noise improved, and thus the resolution limit of the data improved. It must be pointed out that the elimination of the incoherent scattering is very sensitive to the absolute degree to which the protein and the rest of the crystal   (mother liquor, inhibitors, *etc*) is perdeuterated. For instance, a fully perdeuterated system can offer as much as a 40-fold reduction in sample derived background, however, a 5% contamination of hydrogen into the crystal system reduces the improvement to around 12-fold. Thus the degree to which the backgrounds are actually reduced must wait for experimental verification and will vary between protein crystal systems.

The ILL, in collaboration with the EMBL outstation in Grenoble, has established a joint laboratory (DLAB) to support the deuteration of biological molecules for neutron scattering experiments. This initiative aims to provide users with the tools and facilities required for the specific and selective isotopic-labelling of complex bio-molecules such as proteins, nucleic acids, lipids and sugars. The provision of these deuterated molecules should greatly enhance both the quality and quantity of

neutron experiments that can be done at ILL and in many cases will make feasible new more sophisticated experiments than can presently be performed.

An important point to emphasize regarding the exchange of hydrogen for deuterium is that no modification to the structure of the macromolecule has been observed from comparison of deuterated and non-deuterated (protiated) protein crystals[162,163]. Whilst the effects of D_2O on proteins is well documented and indeed is a method used for the determination of enzyme mechanisms[164], there is little structural information on the impact of perdeuteration. Most of the published work on perdeuteration was performed in the 1960's by Crespi and Berns[165,166]. Their experiments, as well as more recent data, indicate that perdeuteration has a destabilizing effect on proteins. Conversely biological properties such as ligand-binding appear not to be significantly affected[167-169]. An investigation was made by S. J. Cooper *et al*., (1998)[162] on the structural comparison of protiated and perdeuterated elongation factor Tu. Data were collected from both a protiated (native) crystal and a (95%) perdeuterated crystal to resolutions of 2.5 Å and 4.0 Å respectively. Unfortunately the fairly low resolution limit for the perdeuterated protein did not allow a very detailed examination of the differences found between the two structures, however, it did allow a comparison between the overall domain arrangement, as well as secondary structure. Elongation factor Tu is quite a large protein of 393 amino acids and has a multiple domain structure. It therefore allows a comparison between the overall domain to domain arrangement, as well as secondary structure. Overall it was shown that there are no major domain shifts (*i.e.* none greater than 1.5 Å) and both structures have the same fold and global arrangement of domains. It was therefore concluded that, even though there are observed physicochemical differences[167] between protiated and perdeuterated proteins, they are not of an overall structural nature and are presumably, kinetic. Another investigation into the effects on structure of perdeuteration was performed by Gamble *et al*., (1994)[163] when they determined the structure of perdeuterated *Staphylococcal* Nuclease (SNase). The native and perdeuterated forms of SNase behaved similarly throughout the purification procedure and crystallized under the same experimental conditions. X-ray diffraction was used to compare the diffraction

characteristics and structure of the native and deuterated protein, firstly direct comparison of the diffraction data, then by analysis of difference Fourier maps made between the two data sets, and lastly by independent refinement and comparison of the two final structures at 1.9 Å resolution. All of these comparisons confirm that there are no significant structural differences between native and perdeuterated SNase.

2.3   Early   neutron   protein   crystallography   studies (1980-1991)

Early neutron studies of proteins illustrate the important information that can be gained from a single crystal neutron diffraction experiment which has vastly improved our understanding of many biological mechanisms and of the dynamics of protein systems. At the same time as being ground breaking studies, they illustrate the demanding requirements of a single crystal neutron diffraction experiment, such as crystal size and data collection time.

2.3.1 Myoglobin

Myoglobin (Mb) is found in vertebrate muscle, its function being to pick up oxygen from the blood and pass it to the mitochondria. It consists of a single polypeptide chain of 153 residues, folded into eight helical regions (A to H), with a haem group wedged between helices E and F. The haem group contains an iron (Fe) in the Fe (II) state which binds the oxygen. In the Met form of myoglobin the iron has been oxidized to the Fe (III) state. Until the haem iron atom is reduced back to the Fe (II) state, myoglobin cannot combine with or transport oxygen. Metmyoglobin (met-Mb) was the first protein structure to be determined by X-ray crystallography by J. C. Kendrew and co-workers[170].

The X-ray structure of oxygen bound myoglobin (oxy-Mb) was determined to 1.6 Å by S. E. V. Phillips, (1980)[171]. The iron atom of the haem group is coordinated to four porphyrin nitrogen atoms and the nitrogen atom of an invariant 'proximal' histidine residue, the sixth coordination position being the site of reversible binding of molecular oxygen. The oxygen ligand lies in a pocket surrounded by 3 groups, two hydrophobic (Phe CD1 and Val E11) and the imidazole ring of a 'distal' histidine residue (E7). This distal histidine is present in myoglobin and haemoglobin in many different organisms. Its function in myoglobin was not evident from the X-ray structure, although it was shown that the residue presents steric hindrance to

linear ligands such as CO and favours "bent" ones, such as molecular oxygen. At physiological pH, the distal histidine may have a hydrogen atom bonded to either the Nε or the Nδ atom and interaction with oxygen may therefore be a hydrogen bond or a simple van der Waals contact.
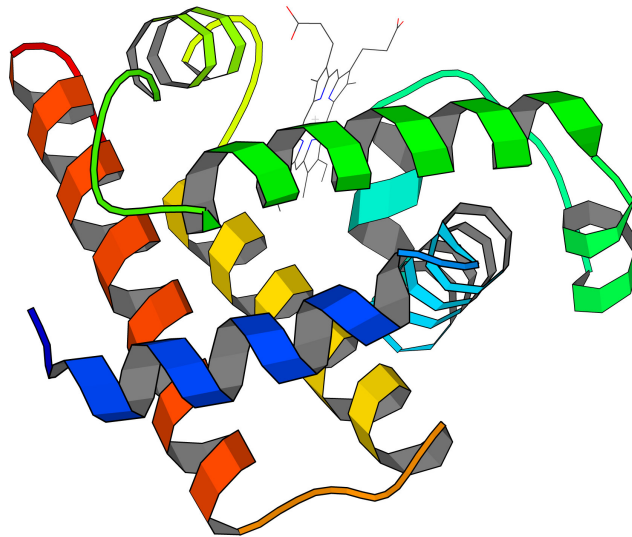
Figure 31.   Ribbon diagram of the helices which make up the subunit of myoglobin, with the haem group shown as bonds.

A neutron diffraction study[172] on oxy-Mb showed that a strong $F_o - F_c$ peak adjacent to the Nε atom indicated the presence of a deuterium atom at this position. A deuterium atom bound to the Nε atom of the imidazole ring therefore could be involved in a hydrogen bond with the oxygen molecule, thereby stabilizing the oxygen-bound conformation. The geometry of the hydrogen bond indicated a medium-strength interaction, such a bond contributing several kcal/mol to the enthalpy of oxygen binding. Two crystals, each ~8mm^3 were used to collect the neutron diffraction data, having been transferred to deuterated mother liquor 3 months before data collection. The data were collected at Brookhaven National Laboratory in 42 days to 2.0 Å resolution. The X-ray structure at 1.6 Å was used as the starting model to calculate phases and amplitudes for the neutron data. The

refinement involved a combined crystallographic and conformational energy method described by Jack and Levitt[173].

An early neutron diffraction study[174] of carbonmonoxymyoglobin (MbCO), in which carbon monoxide is bound at the sixth coordination site instead of oxygen found that the distal histidine Nε atom in this case is deprotonated (not deuterated) in contrast to the situation in oxymyoglobin. Therefore, no hydrogen bond exists between the CO ligand and the imidazole ring of the distal histidine. The study paid particular attention to the water structure of the protein and identified 40 water molecules as D_2O *i.e.* the coordinates for the oxygen and the two deuterium atoms for each water molecule were elucidated. These water molecules were shown to be mainly involved in bridges between protein atoms (intra- and intermolecular bridges), thereby helping stabilize the overall protein quaternary structure. From an analysis of which main-chain amide protons had exchanged it was clear that less exchange occurs in hydrogen bonded peptides than in peptides not hydrogen bonded. Therefore, less exchange was observed for the *α*-helical regions than any other parts of the protein. Significant deviations from linearity were shown to exist in the hydrogen bond geometries in helical regions. The mean N-D⋯O angle was found to be 154°, while the mean C-O⋯D angle was 145°. The MbCO crystals were grown from 70% saturated ammonium sulphate solution at pH 5.7 in a pressure vessel. They were then soaked in D_2O mother liquor for several months prior to data collection. Data were collected from a single crystal of $24mm^3$ on a 4-circle diffractometer at Brookhaven[175]. The initial coordinates were based on the 2 Å multiple isomorphous replacement phased X-ray maps of J. C. Kendrew *et al*., (1960)[170]. The refinement employed a real-space method of Diamond, (1971)[176].

The neutron data for the MbCO crystal were later re-analysed[125,177] using a least-squares reciprocal-space refinement technique[178,179] modified to include restraints on the configuration of water binding[138], coupled with a solvent analysis technique[180] which includes the low-order (low angle) diffraction data. The improved phasing greatly improved the localisation and orientation of water molecules and ions, and also improved the fit to the data (lower R-factor). Two hydration shells were observed, with 87 water molecules (as D_2O) and 5 ions

identified. The 87 waters were all found to be bound to polar or charged groups. The study also gave details about the location of non-exchanging (or slowly exchanging) protons of the main-chain. Eleven main-chain amide hydrogen atoms did not exchange for deuterium. All of these unexchanged groups were found to be positioned on the inside faces of helices, facing into the hydrophobic core of the protein. The protonation states of the histidine residues were re-analysed and confirmed the result from the earlier refinement that the distal histidine residue is deprotonated at the Nε atom in MbCO. Details of hydrogen bonding geometry gave a mean N-D····O angle of 147° for helical regions.
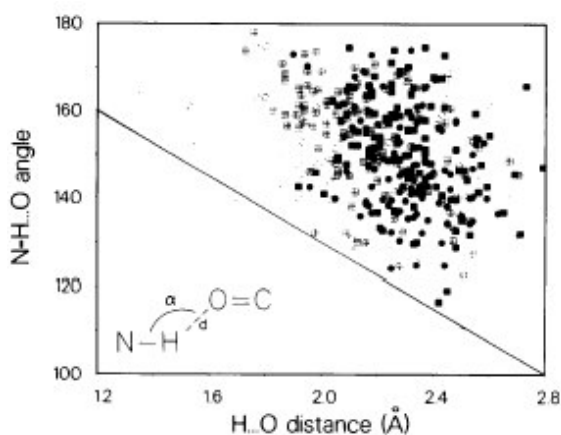
Figure 32. Plot of the dependence of the hydrogen bonded angle at the H atom (N-D---O) on the H---O distance[281].

The structure of acid metmyoglobin (pH 5.6) and its bound water have also been investigated by neutron diffraction techniques[181]. In this study, the distal histidine Nε atom was found to have a deuterium atom bound to it, as in oxymyoglobin. From an analysis of the contributions of the bulk solvent structure factors, it was shown that:

(1) For the low-angle data the contributions from the protein and the solvent are of the same order of magnitude.

(2) That the phase of the bulk solvent is always different from that of the protein.

(3) The inclusion of the solvent in the structure factor calculation helped reduce the R-factor.

An estimation of H/D exchange was performed through occupancy refinement and showed that greater exchange occurred in this form of myoglobin than in the other forms investigated, with no amide hydrogen atoms substantially unexchanged. The authors stated that this is possibly due to the large errors associated with the occupancy refinement and/or the fact that the acid metmyoglobin crystals had been grown from deuterated ammonium sulphate and then stored in the mother liquor for ten years. In the other early neutron studies on myoglobin the crystals had been grown from H_2O and then soaked in D_2O for only a matter of months, not years.

2.3.2 Trypsin

There are 3 invariant residues which characterize the catalytic site of serine proteases, a histidine (His 57), an aspartic acid (Asp 102) and a serine (Ser 195). Serine proteases hydrolyse peptide bonds by a base catalysed nucleophilic attack on the carbonyl carbon of the substrate by the hydroxyl group of Ser 195[182,183]. Coincident with this attack, the hydroxyl proton of the serine is transferred to the imidazole of His 57. The unresolved mechanistic question was whether His 57 is the actual chemical base in the hydrolysis reaction or whether the histidine acts as an intermediary through which Asp 102 functions as the base. Attempts to solve this problem have involved a number of indirect techniques, including NMR, isotope exchange, difference IR and QM calculations. However, these attempts have resulted in different identifications of the group[184,185].

Neutron diffraction lends itself to this problem, due to its ability to directly locate hydrogen atoms experimentally. A form of bovine trypsin inhibited by a monoisopropylphosphoryl group (MIP) was chosen as the most suitable candidate for analysis because it has been found that the bound MIP group closely mimicks the expected structure and electrostatic properties of a real substrate-enzyme intermediate[186,187]. Neutron data to 2.2 Å were collected[187] at the single-crystal station at Brookhaven High Flux Beam Reactor[188]. The crystal of bovine trypsin had a volume of 1.5mm³ and was soaked for 1 month in D_2O solution. The initial

phasing model was calculated by applying the appropriate neutron scattering lengths to the refined trypsin X-ray coordinates[189]. Refinement was performed using the constrained difference Fourier technique[190,191]. A difference map calculated with only the deuterium between the His 57 and Asp 102 omitted, revealed a large peak at the position of a deuterium bound to the imidazole nitrogen of His 57[187,192].
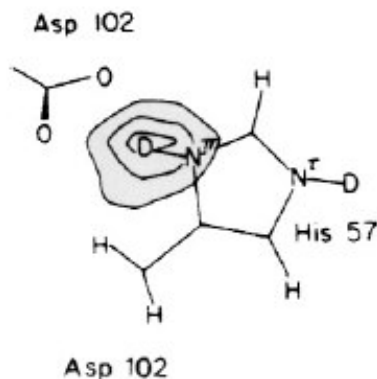
Figure 33.   A difference map[187] showing the N-D---O hydrogen bond between His 57 and the side chain of Asp 102.

This result therefore, supports the conclusion that the mechanistically important proton is coordinated to the imidazole of His 57, and that at physiological pH the protonated species in the tetrahedral intermediate is the imidazole of His 57 rather than the carboxylate of Asp 102. As well as identifying His 57 as the catalytic base in the transition state of the reaction, details of other aspects of protein structure which are observable only by neutron diffraction techniques were discussed. These include the orientation of well-ordered amide side-chains, location and orientation of water molecules and hydrogen exchange properties of bovine trypsin[192].

2.3.3 Crambin

Crambin is a small protein, 46 residues in length, of which 46% are in helices and 17% are in β-sheet. It is hydrophobic, that is to say, organic solvents are required to solubilize it. It has been shown to diffract to extremely high resolution in X-ray

diffraction[193-196]. The protein possesses remarkable temperature stability[197] and has a large number of intramolecular main-chain to main-chain and main-chain to side-chain hydrogen bonds[198]. An analysis of the packing features of crambin show that the hydrophobic surface of the protein is shielded from the solvent by the interaction with a symmetry-related molecule. Such a close intermolecular association results in a very efficient packing, with only 32% of the crystal volume occupied by solvent. Furthermore, in the crystal only 10-15% of the solvent can be classified as bulk solvent. Hence, crambin is an excellent system for the detailed study of the bound water structure.
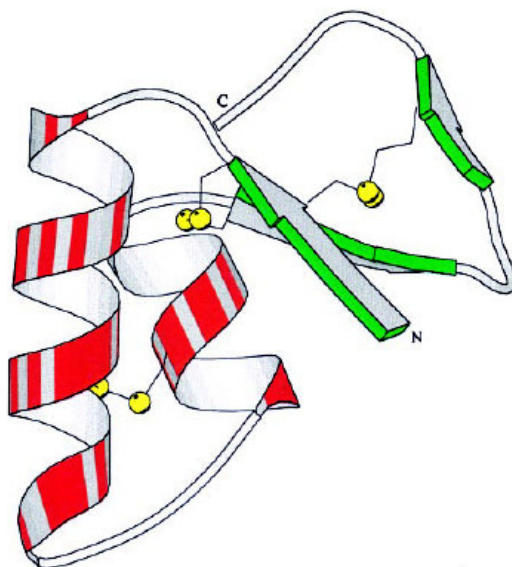
Figure 34.   Ribbon diagram indicating the general fold of Crambin. The disulfide bridges are shown in yellow, β-sheets and extended chain are coloured green and helices in red[282].

A neutron study[199,200] was performed by Teeter and Kossiakoff on a crambin crystal of 1.4mm^3 that had been soaked in deuterated mother liquor. A neutron data set was collected to 1.5 Å, comprising 3871 unique reflections (73% completeness). From analyses of these data, much additional information was gained on the following particular features: methyl rotors, ring structures, exchange of amide protons, an arginine-carboxyl salt bridge and the overall bound water structure. From an analysis of the maps, most of the methyl rotors were assigned by the method of

Kossiakoff and Spencer[192]. Out of 26 methyl rotors, 24 appeared to be in a staggered conformation. This is consistent with rotation being a quantized event consisting of rapid reorientations of ~120° steps to positions of highest stability[200].

The observed H/D exchange pattern is worthy of note as it does not appear to follow the same general trend for other proteins studied by neutron diffraction such as myoglobin, trypsin and ribonuclease. In these proteins, there is a definite exchange pattern for the α-helical regions. The protons that are unexchanged are found to be located predominantly on the inside of the helices, which is less accessible. In crambin on the other hand, nearly all the peptide protons in the α-helical regions are found to be unexchanged wherever they are located, even though 90% of the amide protons are within 3 Å of the solvent. It was suggested that the tight molecular packing and furthermore, the large amount of intramolecular interactions may play some role in the resistance of crambin to hydrogen exchange. However, it is also possibly due to the fact that the crystal used for analysis in this study was only soaked for days rather than months.

A total of 73 of the 90 water molecules identified in the neutron refinement were identified in the X-ray data also, illustrating the good agreement between the two techniques. The positions of the waters provided new information on the hydrogen bonding networks of crambin.

2.3.4 Subtilisin

The structure of a mutant variant of subtilisin, in which methionine 222 was changed to glutamine, has been studied by neutron diffraction[201]. The mutant was used for the study as unfortunately large crystals could not be grown for the native molecule. Subtilisin contains a diverse set of secondary and tertiary structure types, making it possible to analyze in one protein structure the comparitive breathing characteristics of segments of sheets, helices, loops and turns. A protein of 275 amino acids (27 kDa) it is found to have a considerable hydrophobic core. The core consists of an extended seven-stranded parallel β-sheet unit, which is sandwiched between four

lengths of α-helix. The hydrogen exchange pattern of subtilisin identified the β-sheet and α-helix secondary structure elements to be the most resistant to exchange. Fifty-five percent of the peptide amide hydrogen atoms were fully exchanged, 15% unexchanged and 30% partially exchanged. The largest concentration of unexchanged sites was in the seven stranded parallel β-sheet, in which there were 11 fully protected groups. As judged by solvent accessibility calculations, the main-chain atoms of the residues contained in the sheet are, with the exception of strand 1, well shielded from direct contact with solvent. 57 water molecules were also identified from the neutron maps and the position of these agreed well with the low B-factor water positions from the X-ray analysis.

A large crystal (2.7mm^3) was soaked in D_2O mother liquor solution for a month prior to data collection. The initial phasing model was calculated by applying the neutron scattering lengths to the refined X-ray coordinates[202]. The refinement was carried out using a combination of the constrained difference Fourier technique[191] and an energy minimization routine[203]. The protonation states of the six histidine residues were assigned, with the active site histidine (His 64) found to be neutral at the pH of the analysis (pH 6.1).
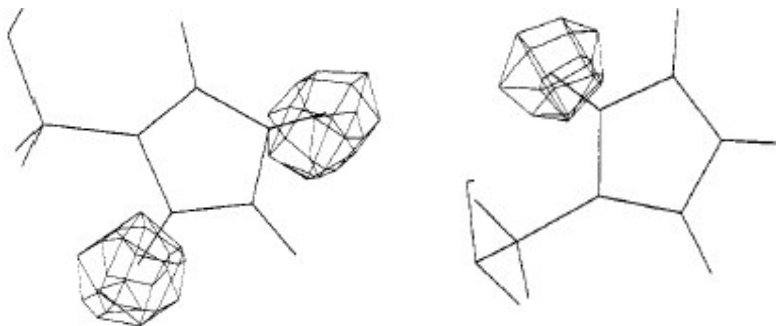
Figure 35.   Difference maps of histidine residues indicate their protonation states. *Left*, Positive difference neutron density indicates that His 17 is protonated and *right*, a single deuterium atom on ND1 indicates that His 64 is neutral[201].

Hydroxyl rotors (Oγ-Dγ) were also investigated. They possess rotational freedom to align optimally within their electrostatic environment. Thus, their character provides data as to how the energetic components are partitioned between non-bonded and torsional effects, and under what circumstances each of these effects dominate the observed geometry. The dihedral conformations of a majority of the hydroxyl rotors were assigned, the preferred orientation being trans (180°) with the other two low-energy conformers (60°, 300°) about equally populated. Deviations from low-energy states were explained in terms of local electrostatic effects. For the serine residues, about 21% of the hydroxyls were found to act exclusively as hydrogen bond acceptors and 37% as hydrogen bond donors, and in 42% the group functions as both.

2.3.5 Ribonuclease-A

Bovine pancreatic ribonuclease-A is one of the most extensively studied proteins. It was the subject of many pioneering studies in protein chemistry and enzymology, and was only the third enzyme whose structure was solved by X-ray diffraction[204,205]. It has continued to be the focus of considerable attention, and a vast quantity of kinetic, mechanistic, thermodynamic and structural information has been collected[206].

A full refinement of ribonuclease-A was carried out by a restrained least-squares procedure, in which, both X-ray and neutron data to 2 Å resolution were included simultaneously[207]. Since both types of data result from the same crystal structure, the number of observations can be increased by combining the two types of information in the same refinement. This can lead to a better overall description of the structure due to improvements of the data-to-parameter ratio. The influence of systematic errors can also be reduced since they are unlikely to be similar in the two types of data.
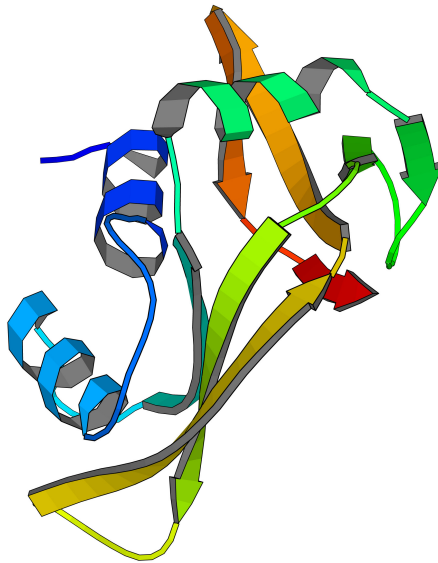
Figure 36.   Ribbon diagram of the subunit of ribonuclease-A. As can be seen both α-helices and β-sheets are present within the subunit.

A large crystal ($30mm^3$) was soaked in deuterated synthetic mother liquor, with several complete exchanges of the solvent taking place over 6 months. Using this crystal, neutron data were collected over a period of 6 months to 2 Å resolution. Three crystals were used to collect the X-ray data to 2 Å resolution. The starting model was the refined X-ray structure of Wlodawer *et al*., (1982)[208]. Major modifications were made to the bound solvent with 128 D_2O molecules identified and one phosphate molecule bound in the active site. The correct orientation of asparagine and glutamine side-chains were assigned, with some having to be rotated by 180° in order to agree with the maps. Several other side-chain alterations were made, including the side-chain of the catalytically active Lys 41. This new side-chain position for Lys 41 indicated that the residue makes a salt-link to the phosphate molecule in the active site. The orientation of the lysine side-chains is aided by the strong scattering observed from an ND_3 group in the neutron maps. The assignment of main-chain hydrogen bonding interactions were almost unchanged from the results of the X-ray refinement of Wlodawer *et al*., (1982). However, many of the interactions involving side-chain groups were changed as a result of the joint refinement. Hydrogen exchange has also been studied[207,209] in ribonuclease-A. The

degree of exchange of amide hydrogen atoms was calculated in a subsequent refinement with the neutron data alone. Of the 120 peptide amide hydrogen atoms, 28 were found to be at least partially protected from exchange. Most of the protected hydrogen atoms were found to be involved in hydrogen bonds with main-chain carbonyls groups. Not surprisingly, amide hydrogen atoms corresponding to non-polar amino acids were the most highly protected, with more than half of the methionine, valine, isoleucine and leucine residues protected. A non-uniform distribution of protected amides was observed. Of the two distinct areas surrounding the cleft near the active site, one appears to be much more flexible than the other. In particular, the section of the *β*-sheet containing residues 75, 106-109, 116 and 118 appears to be much more highly protected from exchange than any other part of the protein. Only the central part of the *β*-sheet found in the other half of the molecule contains protected amides. Each of the helices within the structure were found to have two or three unexchanged amide hydrogen atoms.

2.3.6 Lysozyme

Lysozyme is an enzyme of 129 amino-acid residues, for which, the mechanism of action in cleaving polysaccharides is well understood. The states of protonation of the side-chains of Glu 35 and Asp 52 are central to the mechanism. In the accepted reaction mechanism, the side-chain of Glu 35 is protonated, whereas the side-chain of Asp 52 must be ionized. In order to verify this, and investigate the H/D exchange characteristics of lysozyme, a neutron diffraction study was performed[210].

Neutron diffraction data were measured on a large triclinic lysozyme crystal to 1.4 Å resolution. The crystal had previously been soaked for about 2 months in deuterated buffer at pH 4.2. Data were also collected on a non-deuterated crystal at pH 4.6 to 1.8 Å resolution. Refinement was undertaken using restrained least-squares program, PROLSQ[178]. From analysis of the neutron maps, it was shown that the data are completely consistent with the protonation of Glu 35 and ionization of Asp 52 at pH 4.2 to 4.6. Furthermore, the degree of exchange was given for almost all of the

backbone amide hydrogen atoms in the crystal. The amides in residues 28 to 34, which make up most of the buried helix are all protected, whereas in the helix centred around residues 90 to 100 about half are protected. Even in the small stretches of helix around residues 114 and 122, there is some degree of protection.
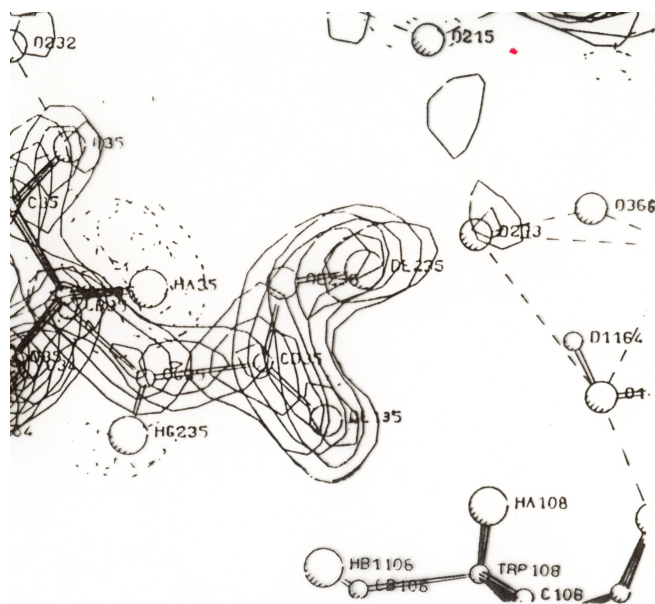
Figure 37.   A neutron difference map shows that the side-chain of Glu 35 is deuterated in triclinic crystals of lysozyme[210].

2.3.7 Insulin

Insulin was one of the first polypeptide hormones to be crystallized[211] and characterized by X-ray diffraction[212]. Although many hydrogen atoms have been observed in difference X-ray Fourier maps for insulin, in order to make their location more precise a neutron diffraction study[213] was performed. A single crystal of porcine 2Zn insulin (1 x 1.5 x 2.3 mm) was slowly deuterated by addition of synthetic mother liquor for over 4 months. Neutron diffraction data were collected to 2.2 Å resolution. A joint neutron/X-ray restrained least-squares refinement was undertaken using the neutron data as well as a 1.5 Å resolution X-ray data set collected previously[214]. Overall the refined model was seen to be essentially the

same as the original X-ray model. However, this was not the case for the refined solvent sites. Analysis of the $2F_o$ - F_c maps revealed that nearly all of the more ordered solvent sites (mainly those forming hydrogen bonds to the protein) in the neutron and X-ray densities were very similar, but many of the remaining sites were located in either in X-ray density alone, or in regions of overlapping X-ray and neutron densities for which the centroids for the two densities were far apart (> 1 Å). Moreover, several of the solvent sites appeared to move into areas of no density. The occupancies of the amide H/D exchange sites were estimated from refinement using the neutron data alone. The pattern of exchange was found to be very different in insulin compared to other proteins studied. The only extended region of significant protection under the experimental conditions was found in the centre of the helical region in each of the B-chains. No significant protection of amide hydrogen atoms was observed in the areas of *β*-sheet.
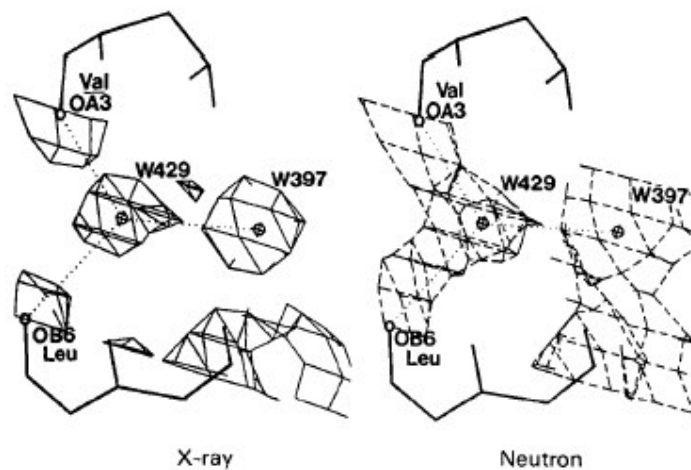
Figure 38.　　X-ray and positive neutron density in the region of two water molecules, numbered 397 and 429. Continuous density of the neutron map is due to unresolved D-atoms and indicates the directions of the hydrogen bonds[213].

2.3.8 Vitamin B$_{12}$ coenzyme

The hydrated crystal structure of vitamin B$_{12}$ coenzyme provides an excellent system in which the distribution of the solvent molecules and their general interactions at the interface can be studied to atomic resolution using diffraction methods. Seventeen to eighteen water molecules are present per coenzyme[215]. High-resolution neutron (0.95 Å) and X-ray diffraction data (0.94 and 1.1 Å) were obtained for vitamin B$_{12}$ coenzyme[139,216,217]. All of the H- and D-atom positions for the coenzyme molecule and approximately 65% of the solvent D atoms were located from the neutron difference Fourier maps. The solvent regions were analysed in two stages: first, main sites were assigned to the well defined regions of solvent density and refined using least-squares; second, continuous sites were assigned representing the more disordered, diffuse and elongated regions of solvent density. Water networks were formulated and extend throughout all the solvent regions of the crystal.
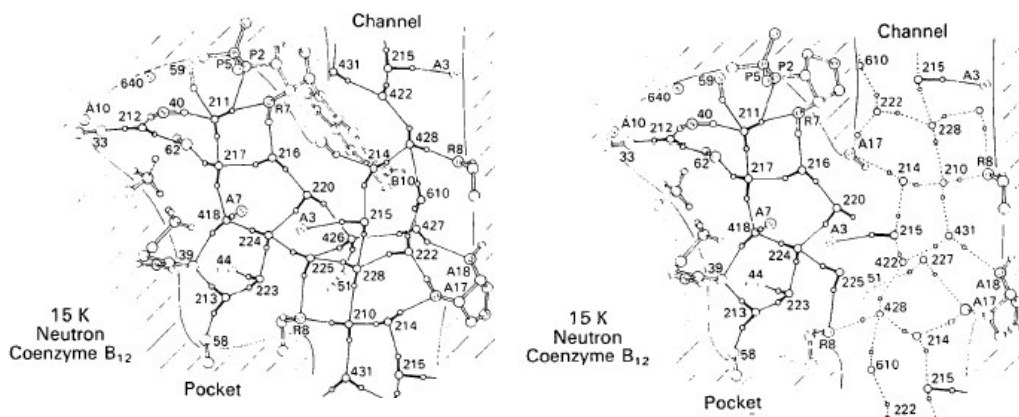
Figure 39.   Water networks in the vitamin B$_{12}$ coenzyme crystal at 15K[283]. *Left*, network 1 is represented by solid lines between solvent molecules. Partial clathrate cage of water molecules around the CB10 methyl of the benzimidazole group. *Right*, network 2 represented by solid lines between waters in the pocket region and dashed lines in the channel region.

Interesting orientational arrangements of the individual water molecules around both polar and non-polar groups of the coenzyme molecule were observed. The individual hydrogen bond geometries varied over wide ranges, similar to those observed in small-molecule crystal structures. Furthermore, large deviations from tetrahedral coordination were found around the majority of water molecules. Of the eleven methyl groups present in vitamin B_{12} coenzyme molecule, six are well ordered and five disordered.

2.3.9 Summary of early neutron studies

Both neutron and X-ray diffraction pose the inherent phase problem. A single experiment measures only intensities and does not provide the phase information needed to calculate density maps. Neutron diffraction, similar to the X-ray case does offer the opportunity of determining phases experimentally through the use of heavy atom derivatives or anomalous dispersion effects using isotopes of gadolinium, cadmium or samarium *etc*. The use of these approaches in neutron diffraction studies is, however, time and neutron consuming[147]. In each of the studies mentioned above, the high-resolution X-ray crystal structure was used as an initial phasing model for refinement of the neutron data. Several different refinement strategies have been employed, including restrained difference Fourier techniques[191], the Jack-Levitt combined crystallographic and conformational energy refinement method[173], and the restrained least-squares conjugate gradient method[178] of Hendrickson and Konnert. In each of these methods modification to the code were implemented to account for the variable occupancy of exchangeable H/D atoms and to incorporate neutron scattering lengths in place of X-ray scattering factors. In addition, Hendrickson and Wlodawer introduced a method for the use of both X-ray and neutron data in a joint-refinement strategy. Although many different strategies of refinement have been applied, when the neutron data is of high quality, whichever strategy has been employed has produced a well-refined neutron structure which has provided supplementary information to the X-ray structure. In early experiments, only monochromatized neutron beams were used to collect Bragg reflections from

protein crystals. Therefore, the data collection times for these early protein neutron diffraction studies were sometimes very long indeed. The crystal sizes were often very large (> 10 mm^3) and many of the protein systems studied relatively small *i.e.* have small unit-cell size. These factors significantly aided the success of these experiments, but restricted the choice of protein systems that could be studied. With the new technical advances described (sections **2.2.2/2.2.3**) in neutron diffraction, data collection times and crystal-size requirements have been reduced and protein systems with more challenging, larger, unit cells can be studied.

2.4 Recent neutron protein crystallography studies (1991-2003)

2.4.1 Lysozyme

The first protein to be tested using the neutron Laue diffractometer (LADI) at ILL, was a large triclinic crystal of hen egg-white (HEW) lysozyme[218]. The unit cell parameters are relatively small ($a = 27.2$ Å, $b = 31.9$ Å, $c = 34.3$ Å) and consequently this gave a strong diffraction pattern to 3 Å in just 2.5 hrs, using a neutron white beam, in the wavelength range 2.5 to 8.5 Å.

The first full crystal structure report[150] of the use of the LADI detector was made by Niimura *et al.* for HEW tetragonal lysozyme crystals ($a = b = 79.1$ Å and $c = 36.6$ Å) to a resolution of 2 Å. Enzymatic activity in lysozyme is maximal at pH 5.0 when the carboxylate group of Glu 35 is protonated; it is then this proton that is transferred to the sugar substrate during the hydrolysis process. During the reaction the other important catalytic site residue, Asp 52 remains in a dissociated state. The crystal in the neutron study was grown at pH 7.0, and it is known[219] that lysozyme is less active at pH 7.0. Therefore, it was of interest to determine the protonation states of these two crucial residues.

The complete data set was collected in 10 days, illustrating the improvements gained by using quasi-Laue methods and a large area neutron image plate detector. A single crystal of dimensions 2 x 2 x 1.5 mm was grown in D_2O solvent, and diffraction measurements made in the wavelength range 3 to 4 Å. Since the wavelength range was narrow the number of Bragg reflections that were multiple was small (50). The HEW lysozyme structure at 1.33 Å resolution[220] was used as the initial model for the refinement against the neutron data. The positions of 960 hydrogen atoms (696 as hydrogen, 264 as deuterium) in the protein and 157 bound water molecules were determined. More specifically, the neutron maps displayed no indication of D-atoms around the carboxylate atoms of the catalytic site residues, Asp 52 and Glu 35.

These results are consistent and further confirm Phillips' proposed model of lysozyme's activity[221].
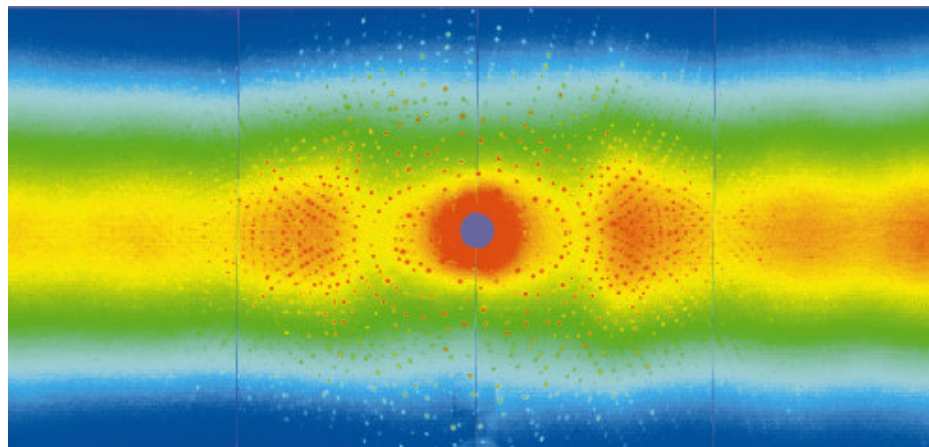
Figure 40.   A neutron quasi-Laue diffraction pattern (wavelength range 3 to 4 Å) from a tetragonal HEW lysozyme crystal recorded over 12 hours[150].

The deuteration states of the hydrogen atoms and bound water structure of triclinic crystals of HEW lysozyme (M_r = 14.6 kDa) have also been studied[222] by quasi-Laue methods. Large crystals were grown over several years and H/D exchange was performed by successive replacements of the mother liquor with D_2O buffer solutions. The data were collected on LADI using a neutron bandpass wavelength range of 2.7 to 3.5 Å. The average measuring time per frame was 10 hours, giving a total data collection time of 2 weeks. The ultra-high resolution X-ray structure[223] to 0.95 Å resolution was used as the starting model and throughout the refinement the non-H atoms remained fixed at their X-ray positions. Largely diverging R- and R_{free} occur if all the atoms are refined, as seen in the above neutron analysis of tetragonal lysozyme[150].

In the 1.7 Å resolution neutron structure determined, 244 water molecules form approximately one layer of water around the protein with far fewer water molecules located further away. Water molecules around apolar patches are found in tangential layers 4-5 Å away from the surface or form C-H----O contacts. Many of the waters are orientationally disordered, with only 115 out of 244 waters in mean single

80

orientations. A total of 214 out of the 268 potentially exchangeable H atoms had been exchanged. By inspection of the carboxyl groups of Glu 35 and Asp 52 it was confirmed that the protonation states of the two catalytic residues at pH 4.7 are protonated and deprotonated respectively, in agreement with the proposed mechanism[221].

2.4.2 Endothiapepsin

The aspartic proteinases constitute a family of enzymes that are involved in many important biological processes. For example, the retroviral aspartic proteinases, such as the HIV proteinase, are essential for maturation of the virus particle, and inhibitors have a proven therapeutic record in the treatment of AIDS. X-ray structures have been used to propose various mechanisms for the catalysis process, however, the positions of protons on the catalytic aspartic acid residues, and the ligand in these complexes, have not been determined with certainty. Thus the aim of the neutron study[224] was to locate these crucial hydrogen atom positions at the active site of an inhibitor complex. Crystals of endothiapepsin complexed with H261 (the inhibitor) were grown and then prepared for the neutron data collection by exchange of hydrogen for deuterium in the crystals *via* vapour diffusion. Two crystals, each of a volume ~3 mm^3 were used to collect the neutron data on LADI, at room temperature. The wavelength range used was 2.7 to 3.6 Å and data extended to a nominal resolution of 1.95 Å. The neutron structure was refined to 2.1 Å resolution and indicated that ~50% of the main-chain nitrogen atoms had exchanged their hydrogen atoms for deuterium atoms. The majority of residues that were resistant to exchange were found to be positioned within core *β*-sheet regions.

Most importantly, the data provided convincing evidence that Asp 215 is protonated and that Asp 32 is the negatively charged residue in the transition state complex. 256 solvent sites were modelled, with 41 of these modelled as D_2O. The majority of the well-defined water molecules are found close to the protein surface, and some are partially buried.
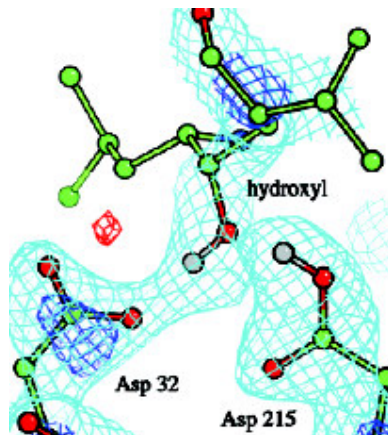
Figure 41.   Neutron density at the catalytic centre of endothiapepsin[224]. Asp 32 and Asp 215 are shown along with the transition state isostere. Original difference density (F_o - F_c) in dark blue and red lines which correspond to 2.5 and -2.5 σ, respectively. The $2F_o$ - F_c density is shown in cyan contoured at 1.2 r.m.s. The regions of positive density between the aspartic acid residues strongly suggests that the inhibitor hydroxyl group has a D-substituent oriented towards Asp 32 and that Asp 215 is deuterated on its outer carboxyl oxygen.

The endothiapepsin neutron study represents the largest protein structure for which a complete high-resolution neutron crystal structure analysis has been published. The space group for these co-crystals is $P2_1$ with unit cell dimensions: a = 43.1 Å, b = 75.7 Å, c = 42.9 Å, and β = 97.0°. As mentioned earlier, problems of spot-spot overlap can occur and become more crucial as the unit cell size increases. The narrow bandpass ($\delta\lambda/\lambda$ = 25%) using a Ni/Ti multilayer wavelength selector, was chosen in order to alleviate this problem and aid the signal-to-noise ratio.
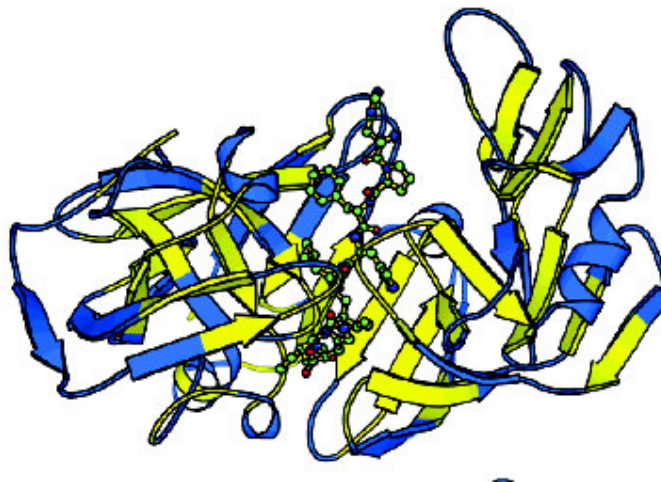
Figure 42.   Extent of main-chain deuteration in endothiapepsin[224]. Yellow shows
                   the regions that have not exchanged hydrogen for deuterium, whereas
                   those regions shown in blue are segments where amino acids have
                   been deuterated in the main-chain. The inhibitor (H261) is seen
                   occupying the cleft in the active-site.

2.4.3 Myoglobin

The only way to exchange all the hydrogen atoms for deuterium within protein
crystals is by the synthesis of fully deuterated protein. The first full report of the
synthesis and neutron crystal structure of a fully deuterated (perdeuterated) protein
was made for sperm whale met-myoglobin (met-Mb) at 2 Å resolution[225]. Neutron
data were collected[226] on a large single crystal (volume = 2.5 mm^3) on the H3A
station at the High Flux Beam Reactor at Brookhaven National Laboratory.

The structure revealed a wealth of chemical information about myoglobin, including
the protonation states of all the histidine residues, the geometry of hydrogen bonding
and the location and geometry of water molecules at the surface of the protein.
Although neutron diffraction on unlabelled myoglobin soaked in D_2O mother liquor
was able to determine the protonation states of exposed histidine residues, its ability
to determine the states of buried histidine residues, for which exchange is restricted,

is limited by the uncertainty in interpreting the negative density of the hydrogen atoms.

(A)   (B)   (C)
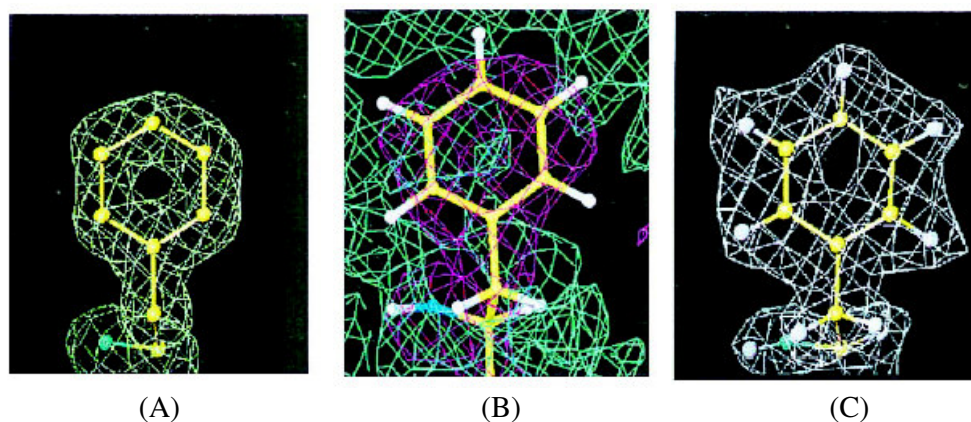
Figure 43. Deuterium (hydrogen) atoms can be located directly as positive peaks in $2F_o$ - F_c maps as illustrated by residue Phe 43 in Mb[225]. (A) The $2F_o$ - F_c X-ray map of fully deuterated Mb using 6.0 to 1.5 Å data, contoured at 1.0 r.m.s (0.84 e⁻/Å³) of the map. (B) A $2F_o$ - F_c neutron map on unlabelled Mb calculated to 2.0 Å resolution, with the pink map contoured at 1.0 r.m.s.and the blue map contoured at -1.0 r.m.s. (C) The $2F_o$ - F_c neutron map of fully deuterated Mb using 6.0 to 2.0 Å data, contoured at 1.0 r.m.s (1.03fermi/Å³).

The study demonstrated that neutron density Fourier maps from a fully deuterated sample are clearly superior to those obtained by previous methods such as vapour diffusion and soaking in D_2O buffers. One major limitation of the analysis of a hydrogenated or partially deuterated sample is density cancellation between atoms bonded to each other, which have opposite sign scattering lengths *e.g.* C-H, O-H bonds. This is overcome when a fully deuterated sample can be grown. Thus deuteration of proteins should greatly expand the applicability of neutron crystallography and make possible work on proteins for which it is impossible to produce large crystals.

2.4.4 Summary

The identification of hydrogen or deuterium atoms by high-resolution neutron crystallography has several consequences[146].

1. The determination of the protonation state of a particular group can be very important in understanding the function of a protein, as was the case in trypsin[187,192], crambin[200], myoglobin[172,225], lysozyme[210] and endothiapepsin[224].

2. The orientation of the amide groups ($CONH_2$) in asparagine and glutamine side-chains can be unambiguously determined: in X-ray diffraction the O and N atoms are very similar in terms of scattering length, whereas in neutron diffraction the entire NH_2 (or ND_2) group is visible, and thus is easily distinguishable from the single oxygen atom of the carbonyl group.

3. The identification of all three atoms in a water molecule leads to the determination of the orientation of that molecule. The location of water molecules by neutron diffraction is easier than by X-ray diffraction because each D_2O has three atoms of roughly similar neutron scattering power, whereas in X-ray diffraction the only effective scattering atom is the oxygen. Particular examples of ordered solvent have been reported for concanavalin A[51,52], oxymyoglobin[172], insulin[213] and vitamin B_{12}[139,217].

4. Neutron diffraction data can provide some information on protein dynamics. For example, the preferred orientation of terminal methyl groups was shown to be largely staggered in the case of crambin[199,200].

The fact that neutron diffraction can identify key hydrogen atoms within protein crystals is clear, however, due to the difficulties in performing a successful neutron experiment only a few studies have been undertaken so far. If we examine the protein data bank (http://www.rcsb.org/pdb) we see that there are only 12 entries listing neutron diffraction as the experimental technique with only 8 different proteins studied. The main draw back of using neutron diffraction in protein structure determination has been the requirement for relatively large protein crystals of at least $1mm^3$ and the long data acquisition times of up to several weeks per data

set owing to the low flux of neutrons at even the most powerful reactor sources. The large crystal size required for a neutron diffraction experiment of around $1mm^3$ is about 1000 times larger in volume than a protein crystal used in a typical X-ray diffraction experiment. It has been estimated that the volume required can be reduced by a factor of 10 if a perdeuterated protein crystal can be grown, thereby increasing the number of feasibly successful experiments that can be achieved with neutrons. Furthermore, advances in instrument design and detector technology, combined with use of the Laue or quasi-Laue technique have speeded up data collection times significantly and made many experiments that were considered too expensive (in terms of the amount of beam time needed) become reality.

Chapter 3

The peanut agglutinin/T-antigen complex

3.1   Summary

Thomsen-Friedenreich antigen, more generally know as T-antigen (Galβ1-3Gal$N$Ac), is a chemically well-defined tumor-associated antigen of non-oncofetal origin with a link to malignancy in man. Peanut agglutinin (PNA) is the most widely used probe for recognizing T-antigen.   Unlike other anti-T probes such as amaranthin and jacalin, it does not bind the more abundant cryptic T- and Tn-antigens (Gal$N$Acα-O-Ser/Thr) which are the sialylated derivatives of the T- and Tn- antigens respectively[227-229].   It is for this exclusive specificity of PNA and its consequent usefulness as a diagnostic tool that its importance has not diminished even after the advent of other T-antigen specific proteins[230]. The specificity of PNA is exploited widely for monitoring the differential expression of T-antigen for both the prognosis and diagnosis of malignancies[231-234].   PNA has found extensive use in the early detection of T-polyagglutinability which has proven to be life-saving in numerous instances. Ability of PNA to differentiate between immature thymocytes and their mature counterparts has found applications in bone-marrow transplantations[235].

The three-dimensional structure of the complex between PNA and T-antigen at room temperature has been determined to 2.25 Å resolution employing synchrotron radiation. The final model has good geometry, an R-factor of 16.5% for 59006 unique reflections and an R-free value of 19.1%. This represents an improvement from 2.5Å for the previous structure of this complex, which was undertaken by use of a CuKα rotating anode X-ray source[143]. Hence, the 2.25 Å resolution structure of the PNA/T-antigen complex now determined, provides us first of all with a higher resolution structure of the complex and furthermore, increases the ensemble of structures of PNA.

The asymmetric unit contains a tetramer, which binds one carbohydrate moiety per sub-unit (each of 236 amino acid residues) *via* direct and water-mediated hydrogen bonds and van der Waals interactions. The T-antigen is identified at the saccharide binding site and is bound directly to the protein through hydrogen bonds mediated by residues Asp 80, Asp 83, Gly 104, Asn 127, Ser 211 and Gly 213. The T-antigen is

also bound to the protein indirectly through water bridges mediated by residues Asn 41, Ile 101, Gly 104, Glu 129, (Tyr 130; two waters create the bridge), and Leu 212. Comparison of the structure with that of the corresponding complex with lactose reveals the specificity of peanut agglutinin for T-antigen is generated primarily by three specific water-mediated interactions, as previously proposed[143].

The number of water oxygen atoms modelled around the tetramer was 441. Of these, 319 agree within 1 Å of the water oxygen positions determined in the 2.25 Å resolution structure of the PNA/lactose complex[58] and 210 agree within 1 Å of the water oxygen positions determined in the 2.5 Å resolution structure of the PNA/T-antigen complex[143]. A total of 47 water molecules are found to be invariant in position between all four sub-units when the hydration shells are compared. These waters are mainly involved in stabilizing the loop sections of the protein, especially those which make up the saccharide binding region and those involved in binding of the monomers to form the tetrameric assembly. From analysis of the water structure of the PNA/T-antigen complex and other models of PNA, new insight into the positions and function of the invariant water molecules is gained. Particular emphasis is paid to those waters in the saccharide binding region and those surrounding it. The detailed study of the water structure of PNA illustrates that the bound water structure of any protein should be considered as an integral part of the structure, as well as highlighting its importance in structure based molecular drug design considerations.

3.2 Peanut agglutinin

Banerjee *et al.*[71] determined the orthorhombic crystal form of PNA to 2.95 Å resolution by using the multiple isomorphous replacement method. PNA is a 110 kDa homotetrameric non-glycosylated legume lectin. PNA belongs to the Gal/Gal*N*Ac-specific family of legume lectins. Each sub-unit in PNA was found to have essentially the same characteristic tertiary fold that is found in other legume lectins except in the loop sections.
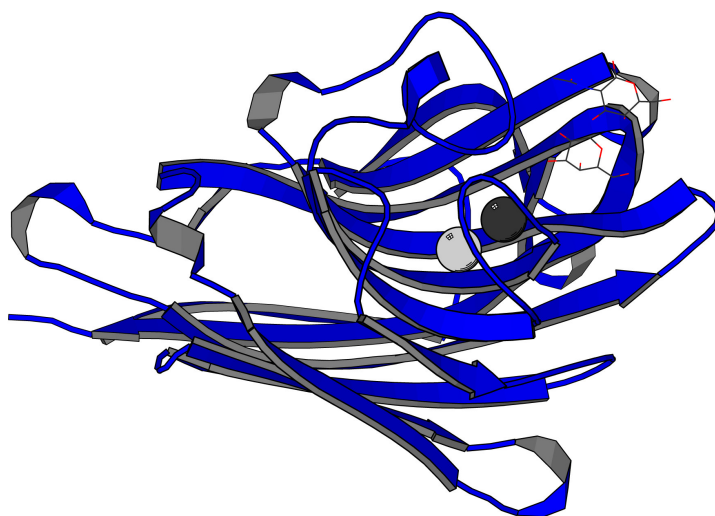
Figure 44. Ribbon diagram of a sub-unit in PNA, illustrating the legume lectin fold. The metal ions are depicted as spheres and the T-antigen is shown in stick-form at the saccharide binding site.

However, interestingly, PNA was found to have an unusual 'open' quaternary arrangement of sub-units, which, unlike other well-characterised tetrameric proteins with identical sub-units, has neither 222 (D_2) nor four-fold (C_4) symmetry[71]. Half of the tetramer, comprising sub-units A and D, is related to the other half sub-units B and C, by a molecular dyad, P. Sub-units A and D are related to each other by a local two-fold axis, R1. Sub-units B and C are related by another, different, local two-fold axis, R2. R1 and R2 are inclined with respect to each other by 73° and -73°

respectively and are skewed by 12 Å on either side of P. The combination of R1, R2 and P gives rise to an irrational screw axis perpendicular to and passing through all three. It relates sub-units C and A by 146° and 24 Å and sub-units D and B by -146° and -24 Å.
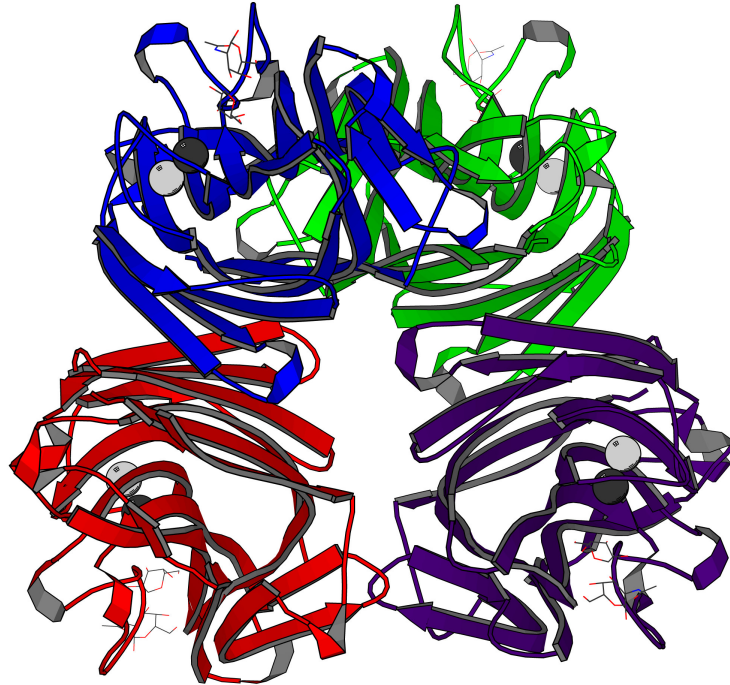
Figure 45.   A ribbon diagram illustrating the 'open' arrangement of sub-units in the tetramer of PNA.

Therefore, unlike any other legume lectin, PNA is best described as a dimer of two back-to-back dimers each two-fold symmetric[58,71]. Although the two dimers, *viz.* sub-units (A and D) and (B and C) in the tetramer, are also related by a dyad, the different dyads neither intersect nor are they mutually perpendicular. Consequently, the molecule has an open quaternary arrangement.

3.2.1 Dimerization

Based on structural arguments, it was concluded that sub-units A and D, and sub-units B and C constitute the natural dimers in the tetramer[71]. This was then confirmed by an analysis of inter-sub-unit interactions of the higher resolution PNA/lactose complex[58]. The formation of the PNA dimer involves the back-to-back association of monomers through the flat 6-stranded β-sheet. The β-sheets from the two monomers are inclined with respect to each other by 86 °.
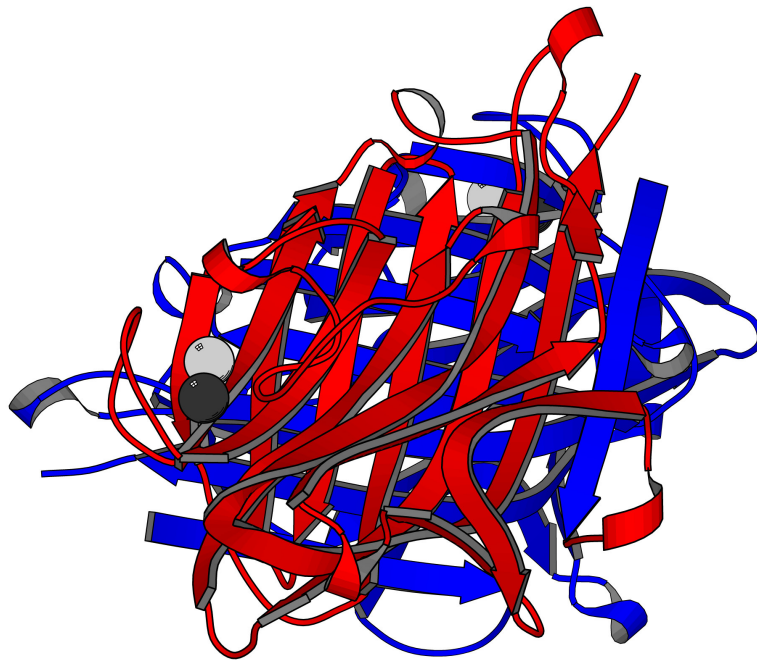
Figure 46.   The disposition of the two flat β-sheets in the inter-sub-unit interface in the PNA dimer. Sub-unit A is coloured blue, sub-unit D is coloured red.

There are a large number of van der Waals interactions between the two sub-units, involving residues distributed over all 6 strands of the flat sheet. There are also protein-protein hydrogen bonds and water bridges. Over half of the residues

involved have small hydrophilic side-chains *e.g.* 8 threonine residues, 2 serine residues and 2 aspartic acid residues. Therefore, residues with relatively short side-chains dominate the dimer interface.

3.2.2 Dimer association in the tetramer

Two interfaces are involved in the assembly of the tetramer from the two dimers, namely the interface between sub-units A and B and the interface between C and D. The A-B interface is formed by the side-by-side alignment of the two flat 6-stranded β-sheets. There are less inter-sub-unit contacts than at the dimer interface and also less hydrogen bonds. However, there are a larger number of water bridges in the A-B interface, which help tether the two sub-units A and B together. The association of the two sub-units A and B in PNA, is very similar to the dimer association found in concanavalin A, as both are formed by the side-by-side alignment of the two flat 6-stranded β-sheets. However, in concanavalin A the two β-sheets are held together by main-chain to main-chain hydrogen bonds, whereas in PNA water molecules are utilized to hold the two sub-units together.

The other interface involved in the tetrameric assembly is that between sub-units C and D. The C-D interface brings together the top regions of sub-units C and D and is made up of several residues. The majority of the residues involved in inter-sub-unit contacts at this interface are found within the small 5-stranded β-sheet, sheet 3. Like the other interfaces in PNA, water bridges help link the two monomers in the C-D interface.
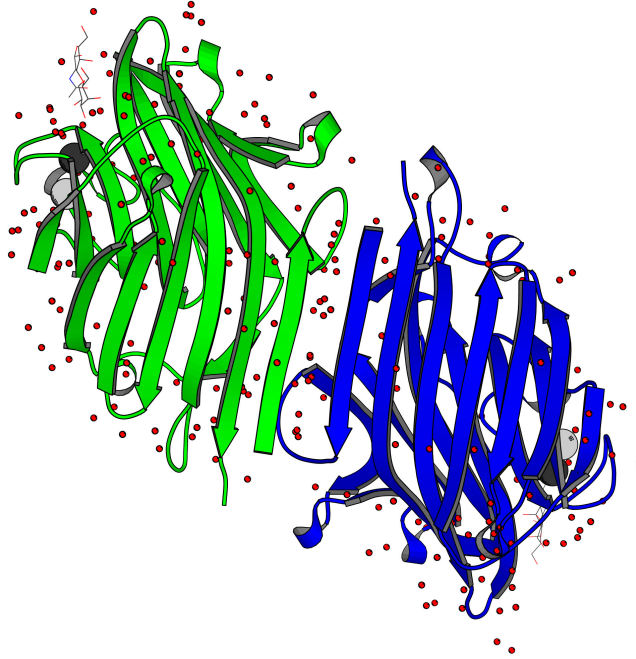
Figure 47.   A-B interface in the interaction between two dimers in the PNA tetramer. This is 'looser' than in the concanavalin A dimer *i.e.* joined by waters (see **Figure 50**).
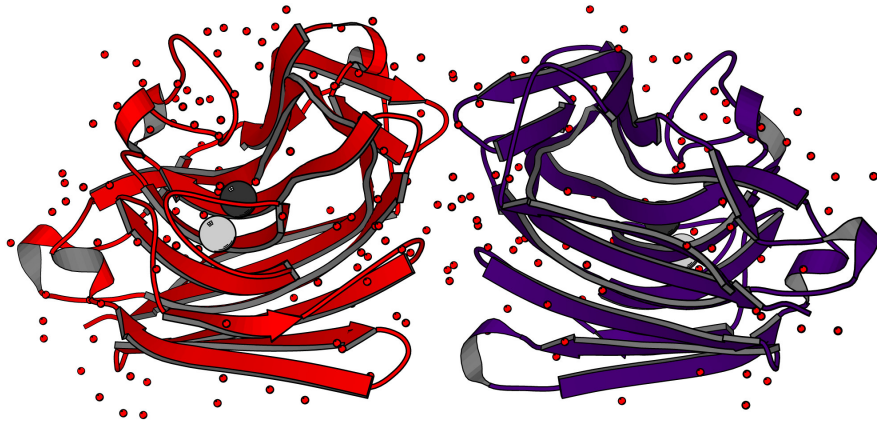
Figure 48.   The sub-units C and D, and their association to form the C-D interface.

3.2.3 Variability in quaternary structure of legume lectins

Legume lectins constitute a family of proteins in which small alterations arising from sequence variations in essentially the same tertiary structure lead to large changes in quaternary association[236]. All legume lectins studied up to now have been dimers or tetramers made up of dimers. The oligomeric state of pea lectin, favin, LOLI, lentil lectin, GSIV, EcorL, and WBAI is as a dimer, whereas concanavalin A, con Br, PNA, SBA, DGL and PHA-L are tetrameric. Therefore, the first or only step in sub-unit association is dimerization. There have been only a few different modes of dimerization observed within the lectins so far. The first mode is that seen in the majority of the lectins and involves the side-by-side alignment of the flat 6-stranded β-sheets from the two monomers. This method of dimerization is seen for con A, pea lectin, favin, LOLI, lentil lectin, PHA-L and soybean agglutinin. The two flat sheets join in these lectins to form a 12-stranded contiguous anti-parallel sheet, with main-chain to main-chain hydrogen bonds occurring between the two sheets. The dyad axes are perpendicular to the β-sheet and the interface may be described as an extended (II-type) interface[237]. Interestingly in all the legume lectin dimers which associate in this manner, the 12-stranded β-sheet presents a slightly concave surface.
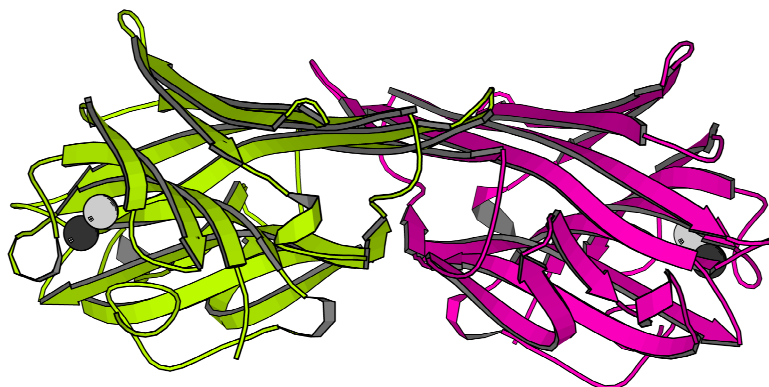
Figure 49.   The dimer in concanavalin A illustrates the side-by-side alignment of the two back β-sheets which forms a slightly concave 12-stranded surface.

Although the most dominant feature of the side-by-side association is the 12-stranded *β*-sheet encompassing both the sub-units, several other hydrophobic and polar interactions are also involved in stabilizing the arrangement. In addition to the back *β*-sheet, the other secondary structural elements that take part in inter-sub-unit interactions are (1) the loop that connects the first strand of the back *β*-sheet with the first strand of the curved front *β*-sheet, and (2) the loop between the second strand of the curved front *β*-sheet and the third strand of the back *β*-sheet[236].
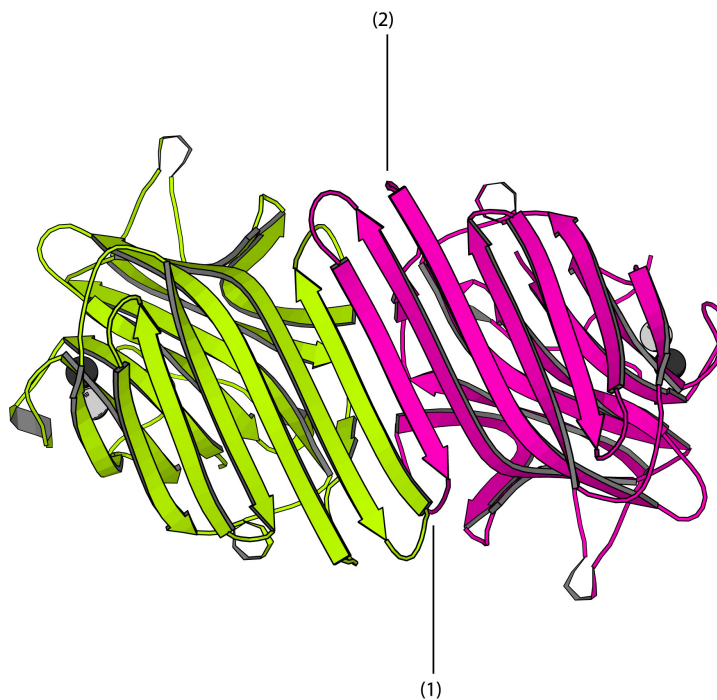
Figure 50.   The concanavalin A dimer, illustrates the side-by-side alignment of the 6-stranded back *β*-sheets.

A different mode of dimerization was observed in *Griffonia simplicifolia* lectin IV (GSIV)[108]. GSIV is a molecular dimer with a molecular weight of 56 kDa. Each sub-unit consists of 243 residues and contains one Ca^{2+} ion and one Mn^{2+} ion. Both sub-units have carbohydrate covalently bound, one has carbohydrate attached to both Asn 5 and Asn 18 whereas the other has carbohydrate linked only to Asn 18. GSIV has essentially the same tertiary structure as in other legume lectins, however, the monomers associate in a different manner. The GSIV mode of dimerization involves

the back-to-back association of the two flat *β*-sheets. In order to explain this difference in dimerization it was suggested that GSIV does not form the 'canonical dimer' observed in the majority of the legume lectins because this would lead to the burial of a charged residue, Glu 58, at the sub-unit interface. An alternative explanation for the non-occurrence of the canonical dimer was attributed to interactions involving the covalently bound carbohydrate[108]. Yet another mode of dimerization was observed for *Erythrina corallodendron* lectin (EcorL)[65]. EcorL is a dimer of a 30 kDa sub-unit, with a heptasaccharide *N*-linked to Asn 17 in each sub-unit. It was suggested that in this case the non-occurrence of the canonical dimer was due to interference by the bulky heptasaccharide, which forces the EcorL dimer into a different quaternary structure. Therefore, as both of these lectins have covalently bound sugar it was generally accepted that the reason for the non-occurrence of the dimerization mode of the others was due in someway to the glycosylation of these proteins.
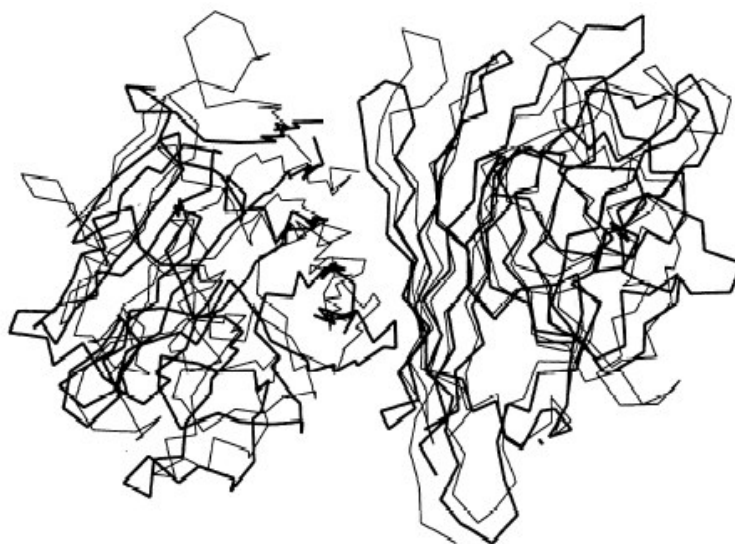
Figure 51.   *α*-carbon backbone superpositions of the PNA dimer (sub-units A and D in black) with the GSIV dimer[71] in grey.

The basic lectin from winged beans (WBAI) has 63% sequence identity with EcorL. However, unlike EcorL, the glycosylation sites in WBAI are far away from the

region that would constitute the interface in the canonical dimer and, hence, do not prevent the formation of such a dimer. Yet WBAI forms EcorL-type dimers, demonstrating that the mode of dimerization is primarily dictated by factors intrinsic to the protein itself[73]. Furthermore, PNA is not a glycoprotein and yet its mode of dimerization is very similar to that in GSIV *i.e.* involves the back-to-back association of the two flat *β*-sheets. Moreover, the A-B interface in PNA on tetramer formation is very similar to the dimer interface in con A. The main difference between the A-B interface and the dimer interface seen in con A, is that instead of the two flat sheets forming one 12 stranded contiguous sheet as in con A, the two sheets do not join close enough to form main-chain to main-chain hydrogen bonds. Replacing these main-chain to main-chain hydrogen bonds are water bridges that connect and help hold together the two sheets.
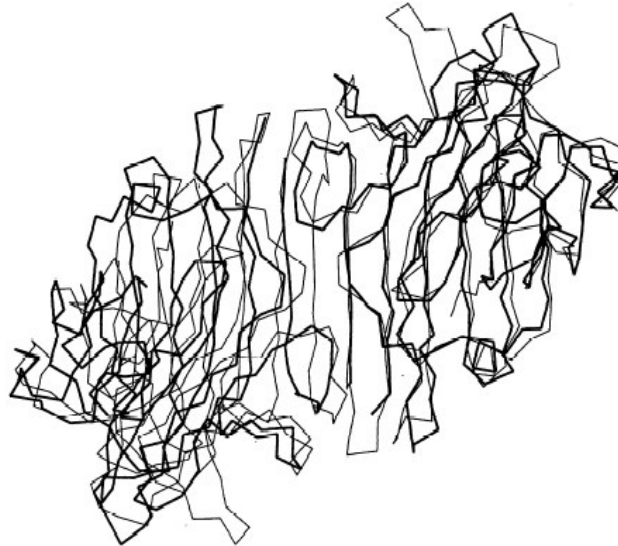
Figure 52.   *α*-carbon backbone superpositions of sub-units A and B in PNA (in black) with the con A dimer[71] in grey.

In PNA there is also the existence of another interface-the C-D interface (as described earlier). This type of interface has not yet been observed in any other legume lectin and is believed to be a fortuitous consequence of the two back-to-back and one side-by-side interfaces[58].

In summary, the occurrence in PNA of a GSIV-type dimerization and the presence of an interface of the type that occurs in the canonical legume lectin dimers, coupled with the fact that WBAI lectin forms an EcorL-type dimer even though its glycosylation sites are far removed from the interface, demonstrates that the variability in quaternary structure is not necessarily caused by interactions involving covalently bound sugar but more likely to be due to the variability in the protein structure itself.

Attempts[58,236] to rationalize this variability have been given in terms of the amount of non-polar surface area buried on association, interaction energy, and shape complementarity, by constructing energy-minimised models, in each of which, the sub-unit of one legume lectin is fitted into the quaternary structure of another. The results indicate that all the three indices favour and, thus, provide a rationale for the observed arrangements. However, the discrimination provided by buried surface area is marginal in a few instances. The same is true to a lesser extent, about that provided by shape complementarity. The relative values of interaction energy turn out to be a better discriminator than the other two indices.

3.2.4 Effect of pH on oligomeric equilibrium and saccharide binding properties

As with concanvalin A and wheat germ agglutinin (WGA), the conformation and saccharide binding properties of PNA depend on pH. PNA is tetrameric in neutral solution and dissociates reversibly into dimers below pH 5.1. Below pH 3.4, the lectin is totally dimeric. Lowering of the pH induces reversible changes in the tertiary and secondary structures of PNA. For solutions of dimeric PNA, containing only minor amounts of tetramers, the binding capacity was the same as for tetrameric PNA, binding sugar with the same stoichiometry as the tetramer (one binding site per protomer), but with an association constant one order of magnitude lower than that for the tetramer[238,239]. At physiological pH, however, the molecule is entirely tetrameric, with no evidence of association-dissociation[238].

The lectin crystallizes in four forms, one orthorhombic, two monoclinic and one triclinic, each containing a tetramer in the asymmetric unit[240,241]. The orthorhombic form crystallizes at neutral pH, whereas the other three crystal forms are grown at pH 4.6. The solvent contents of each crystal form range from 49% for the triclinic form to 57% for the orthorhombic form[241].

3.3   PNA-disaccharide interactions

3.3.1 The PNA/lactose complex at 2.25 Å resolution

The 2.95 Å PNA structure determined was the basis of the high-resolution refinement of the PNA/lactose complex at 2.25 Å resolution[58,71]. On binding of lactose it was shown that there were no changes in the overall tertiary structure of the protein. Each sub-unit was found to have the typical legume lectin fold as with all lectins studied so far, with root-mean-square deviation in α-carbon positions between pairs of sub-units ranging from 0.24 to 0.41 Å.

The carbohydrate binding region was found to be generated by residues in four loops, 91 to 106, 125 to 135, 75 to 83 and 211 to 216, at one edge of the sub-unit[58]. The binding site of legume lectins can be described as a shallow depression on loops associated with the concave face of the 7-stranded curved β-sheet.
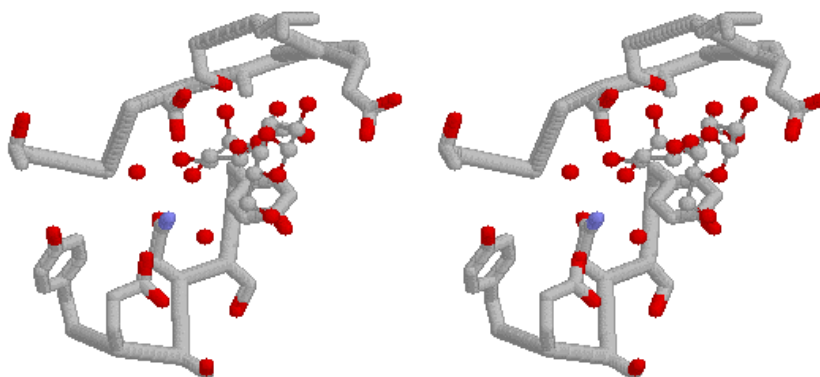
Figure 53.   Stereo view of the binding site in PNA. Lactose (Galβ1-4Glc) is shown at the binding site, as well as the two water molecules that link the lactose to the protein through water-mediated interactions.

The 2.25 Å model revealed that the lactose is bound to the protein through both direct hydrogen bonds and *via* water bridges. The protein residues involved in direct hydrogen bonds to the lactose include those invariant to all legume lectin-galactose complexes (Asp 83 OD1/2, Gly 104 N and Asn 127 ND2, in the PNA numbering scheme) and those that are specific to PNA itself (Ser 211 OG, Asp 80 OD2 and Gly 213 N).

The four invariant hydrogen bonds found in other legume lectin-galactose complexes are Asp 83 OD1-galactose 03, Gly 104 N-galactose 03, Asn127 ND2-galactose 03 and Asp 83 OD2-galactose 04. The mutual disposition of these invariant saccharide binding site residues in legume lectins were shown to be remarkably similar in both mannose/glucose binding (*e.g.* pea lectin, con A, favin, lentil lectin and LOLI) and galactose binding lectins (*e.g.* PNA, EcorL, GSIV and SBA). For example, the constellation of the four atoms superpose with a root-mean-square deviation of 0.23 Å for pea lectin and PNA, and 0.31 Å for EcorL and PNA. Identical spatial distribution of all these residues has been attributed to the constraints imposed by proximal coordination of the Ca^{2+} and Mn^{2+} ions. However, from a more detailed comparison of the binding site topologies of both pea lectin with PNA and EcorL with PNA, it is clear that except for the invariant features mentioned above, the similarity between the geometry of lectin-carbohydrate interactions in Gal-binding lectins and that in Man/Glc-binding lectins are much less striking; the position of the sugar ring itself for example, is substantially different. On the other hand, there are large similarities in binding site topology in lectins from the same group, although differences do exist. These differences are mainly confined to the fourth sugar-binding loop, which is suggested to fine-tune the specificity of each lectin[24,248,249]. Indeed this fourth loop is highly variable in terms of length, sequence and conformation.

The lectin-carbohydrate interactions that are specific to PNA were shown to be interactions between Gal O4 and Ser 211 OG, Gal O6 and Asp 80 OD2, Glc O3 and Ser 211 OG, and finally Glc O3 and Gly 213 N. As can be seen three out of four of these interactions involve residues from the fourth sugar-binding loop. The other

invariant features of legume lectin-carbohydrate interactions, namely the stacking of an aromatic residue (Tyr 125 in PNA) against the galactose ring and the proximity of an invariant alanine residue (Ala 82 in PNA) are also observed in this complex.

A total of 563 water oxygen atoms were modelled around the tetramer in this complex. The number of water molecules in the hydration shell of each sub-unit varied from 120 to 156. A total of 45 water oxygen atoms were found to be invariant between sub-units when the hydration shells of the four sub-units are compared, including the six found to exist in all legume lectin structures by Loris *et al*., (1994)[140]. The majority of the invariant water positions were shown to be involved in stabilising loops. Two water molecules are involved in bridging the lactose to the protein (see **Figure 53**). One water molecule connects the Gal 02 atom to Glu 129 OE2, while the other water molecule links the Gal 02 atom to Gly 104 N. The 2.25 Å resolution PNA/lactose complex also revealed that the 'open' quaternary structure is stabilized by hydrophobic, hydrogen-bonded and water-mediated interactions as described earlier (see section **3.2.1/3.2.2**).

3.3.2 The PNA/T-antigen complex at 2.5 Å

Thomsen-Friedenreich antigen, more generally know as T-antigen (Galβ1-3Gal$N$Ac), is a chemically well-defined tumor-associated antigen of non-oncofetal origin with a link to malignancy in man. This structure is generally expressed as O-linked glycans (Galβ1-3Gal$N$Ac-αSer/Thr), prominently so in more than 85% of human carcinomas[242-246]. It is however, cryptic or absent in normal cells. PNA is the most widely used probe for recognizing T-antigen.

The tertiary and quaternary structure of the lectin in the lectin/T-antigen complex are the same as that in its complex with lactose. The same is true about the location and geometry of the carbohydrate binding site. The four hydrogen bonds found in other legume lectin-galactose complexes occur in this complex also *i.e.* Asp 83 OD1-galactose 03, Gly 104 N-galactose 03, Asn127 ND2-galactose 03 and Asp 83

OD2-galactose 04[143]. The stacking of the aromatic residue (Tyr 125) against the galactose ring is also observed, as well as the close proximity of Ala 82[247].
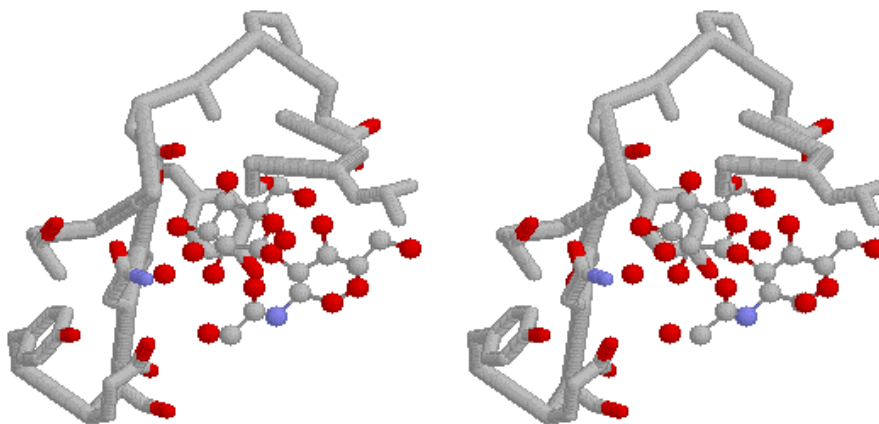
Figure 54.   Stereo view of the binding site in PNA. T-antigen (Galβ(1-3)Gal*N*Ac) is shown at the binding site, as well as the four water molecules that link the T-antigen to the protein through water-mediated interactions.

The interactions specific to PNA itself in binding T-antigen were identified. Both lactose and T-antigen are disaccharides with galactose as the first residue, and it was found that the hydrogen bonds to the galactose moiety are the same in both complexes. The second hexapyranose is β1-4 linked glucose in the case of lactose, while it is β1-3 linked *N*-acetylgalactosamine in T-antigen. The differences in the linkage and the conformation of the two disaccharides are such that O4 of the second ring in T-antigen occupies the same position as O3 of the second ring of lactose. Therefore, in the T-antigen complex, O4 of *N*-acetylgalactosamine is hydrogen bonded to Ser 211 OG and Gly 213 N. The number and the nature of non-polar interactions are also nearly the same in both complexes. PNA, however, binds T-antigen 20 times more strongly than it binds lactose[242]. The only additional interactions T-antigen has with PNA are two water-bridges involving the carbonyl oxygen of the acetoamido group. One molecule links Ile 101, while the other links both Asn 41 and Leu 212 of PNA with the carbonyl oxygen of the acetoamido group

of the second hexapyranose ring of the sugar. These two water molecules exist in all four sub-units in the PNA/lactose complex also, however, they interact only with the protein atoms in it, as there is no sugar atom in their immediate vicinity. Thus, the higher specificity of PNA for T-antigen and its 20-fold higher affinity for this disaccharide than that for lactose, were shown to be generated essentially by these two water-bridges[143]. This was perhaps the first time that water bridges were demonstrated to be largely responsible for generating specificity in protein-carbohydrate interactions.

In order to improve the level of detail of this very important and interesting protein, synchrotron radiation has been employed. Hence, the 2.25 Å resolution structure of the PNA/T-antigen complex now determined, provides us first of all with a higher resolution structure of the complex than previously available, and furthermore, increases the ensemble of structures of PNA. From analysis of the water structure of the PNA/T-antigen complex and other models of PNA, new insight into the positions and function of the invariant water molecules is gained. Particular emphasis is paid to those waters in the saccharide binding region and those surrounding it. The detailed study of the water structure of PNA illustrates that the bound water structure of any protein should be considered as an integral part of the structure, as well as highlighting its importance in structure based molecular drug design considerations.

3.4   Experimental

3.4.1 Crystallization of the PNA/T-antigen complex

The protein was prepared by affinity chromatography on cross-linked arabinogalactan[250]. The orthorhombic crystals of the T-antigen complex, space group $P2_12_12$ with $a = 128.89$ Å, $b = 125.67$ Å and $c = 75.97$ Å, were grown by A. Surolia, from a hanging drop of 5 mg/ml protein in 0.05 M sodium phosphate buffer (pH 7.0), containing 0.2 M sodium chloride, 0.02% sodium azide, 1.5 mM Galβ(1-3)Gal*N*Ac and 12% (w/v) PEG 8000 in the same buffer.

3.4.2 X-ray data collection, scaling and reduction

The PNA/T-antigen complex crystal data were collected from a single crystal at room temperature using synchrotron radiation of wavelength 0.9 Å. The data were collected on PX station 9.5 (fitted with a CCD detector) at the Daresbury Laboratory, U.K. The crystal to detector distance was 150 mm. A total of 100 images were collected in intervals of 1° per image, giving 511,163 observations and 83,198 unique reflections to a nominal resolution of 2.0 Å. The data were cut off at 2.25 Å due to large R_{merge} values and low I/sigma(I) values. Therefore, the final data set to 2.25 Å comprised a total of 59160 unique reflections, of which 59006 were used in the refinement.

Previously[58] intensity data were collected on a Siemens area detector system mounted on a GX 20 rotating anode generator. Two data sets were collected from two different crystals and the data were merged. The combined data gave rise to 173,209 observations and 50,564 unique reflections, of which 45510 reflections were used in the refinement. Therefore, it can be seen that a substantial increase in the number of unique reflections was attained for the new data *via* the use of synchrotron radiation.

The new data were indexed, integrated and scaled using MOSFLM [251]. SCALA[251] was used to scale batches of data together from processed images. The program applies scale and B-factors, adds together partially recorded reflections, monitors and rejects bad agreements between repeated measurements or symmetry equivalents and averages them for output. Using TRUNCATE[251] the intensities were converted to structure factor amplitudes.

An R-free factor[252] was calculated by excluding a randomly chosen set of reflections (~2.5%) from the refinement – a special case of the technique of cross-validation[253]. The agreement between F_{obs} and F_{calc} is independent of the refinement procedure. R-free flags were added and the dataset was completed to the high resolution limit by running the UNIQUEIFY[252] script. Below is a summary of the quality of the data collected.

Resolution (Å)	R-merge	I/sigma(I)	Number of reflections measured	Number of unique reflections	Multiplicity	% completeness
7.12	0.023	25.0	7565	2042	3.7	99.0
5.03	0.028	23.5	13897	3546	3.9	100
4.11	0.031	23.3	18002	4522	4.0	100
3.56	0.040	19.7	21130	5289	4.0	100
3.18	0.059	16.1	23790	5939	4.0	99.9
2.90	0.098	11.8	26337	6573	4.0	99.9
2.69	0.163	8.2	27974	7100	3.9	99.9
2.52	0.218	6.5	29554	7605	3.9	99.8
2.37	0.328	4.8	30609	8005	3.8	99.6
2.25	0.374	4.2	32192	8539	3.8	99.6
Overall	0.067	11.8	231050	59160	3.9	99.8

Table 2.   Summary of the PNA/T-antigen crystal complex X-ray data quality.

Resolution (Å)	R-merge	Number of reflections measured	Number of unique reflections	% completeness
3.99	0.0528	57970	10874	96.9
3.17	0.0725	48800	10275	94.8
2.77	0.1109	27178	9252	85.7
2.52	0.1266	17935	8388	78.2
2.34	0.1508	14219	7467	69.8
2.25	0.1512	7107	4508	42.3
Overall	0.0790	173209	50564	88.0

Table 3.   Summary of the PNA/lactose crystal complex X-ray data quality[58] for comparison to the data presented here.

Resolution (Å)	PNA/T-antigen data		PNA/lactose data[58]	
	Number of reflections	$<F/sigma(F)>$	Number of reflections	$<F/sigma(F)>$
7.12	1944	48.29	1239	57.26
5.03	3537	44.94	3446	71.90
4.11	4516	44.51	4411	70.25
3.56	5281	37.24	5089	54.64
3.18	5937	30.17	5504	41.80
2.90	6569	21.79	5471	28.00
2.69	7097	14.88	5283	22.37
2.52	7599	11.84	5250	18.10
2.37	7999	8.79	4794	13.99
2.25	8527	7.76	4023	11.23
Overall	59006	22.11	44510	36.29

Table 4.   Comparison of the PNA/T-antigen data and the PNA/lactose data[58] using the CCP4 program WILSON[251].

3.4.3 Refinement

The coordinates of the 2.25 Å PNA/lactose model[58] submitted to the Protein Data Bank[254] with code 2PEL were used as the starting model for the refinement. Water and lactose molecules were removed from the model and all B-factors reset to 30 Å^2. REFMAC-5[251] was used for the refinement against the new synchrotron radiation 2.25 Å resolution data. First, 10 cycles of rigid body refinement were run, after which the R-factor was 29.2%. LSQKAB[251], a program that employs a least-squares fit of two sets of co-ordinates was used to translate/rotate each of the four chains, *i.e.* the four sub-units, A, B, C, and D onto each other. This was done so as to identify the areas of each chain with high root-mean-square (r.m.s.) deviation. From this information, non-crystallographic symmetry restraints were introduced. Positional and B-factor refinement were carried out, which improved the R-factor to 20.6% (R-free = 22.0%).

$F_o - F_c$ maps were then checked using the graphics program 'O'[255] and clearly showed density for disaccharide in each of the four sub-units. Originally it was thought that the crystal that we collected synchrotron intensity data from was a crystal of the PNA/lactose complex. Therefore, each lactose molecule was built into the density and added to the model. After further cycles of refinement $2F_o - F_c$ maps were checked and minor positional adjustments were made to the protein and lactose molecules. However, it became apparent after modelling the lactose molecules into the $2F_o - F_c$ maps that something was clearly wrong. Although the fit to the electron density for the galactose sub-unit was excellent, the fit to the electron density of the glucose sub-unit was poor, with large conformational shifts required in order to fit the density satisfactorily. Therefore, the coordinates of the 2.5 Å PNA/T-antigen complex were downloaded from the protein data bank and using the CCP4 program LSQKAB were superimposed against the newly developed model (this work). From the fit of the T-antigen disaccharide to the electron density for the new data it was clear that in fact the data were indeed from a crystal of the PNA/T-antigen complex and not as first thought from a PNA/lactose complex. This was also verified by downloading the structure factors for both the 2.5 Å PNA/T-antigen complex[143] and

the 2.25 Å PNA/lactose complex[58] and comparing the maps with those calculated from the new data. Therefore, the lactose molecules were removed and T-antigen molecules built into the density.

Water molecules were identified and added using the program ARP_WARP[251], using $F_o - F_c$ maps to identify possible waters and $2F_o - F_c$ maps to remove them. The density of the water oxygens must be spherical in shape and within 3.3 Å of a nitrogen or oxygen atom. Waters closer than 2.2 Å to protein atoms were removed and/or if the peak height for the potential water was below 1.5 r.m.s. in the corresponding $2F_o - F_c$ map. Waters were added into the model if the peak height for the potential water was greater than 3.5 σ in the $F_o - F_c$ map. After 30 cycles of ARP_WARP, 503 waters were identified.

$2F_o - F_c$ and $F_o - F_c$ maps were then rechecked after further cycles of restrained refinement and waters removed or added where necessary. Adjustments to the model were made and alternate conformations of residues were also identified. Omit maps were used to verify the conformation at residues alanine-82 and aspartate-83, and confirmed the existence of the *cis*-peptide necessary for the binding of sugars. Positive and negative $F_o - F_c$ maps were checked for the existence of any remaining significant peaks (> 2 σ) and further adjustments made until no further peaks were evident.

The final R-factor for the new model including 6972 protein atoms, 4 Mn^{2+} ions, 4 Ca^{2+} ions, 104 carbohydrate atoms and 441 solvent atoms was 16.5% (R-free = 19.1%) for 59006 unique reflections to 2.25 Å resolution. Previously[58], the final R-factor was 16.4%, including 6976 protein atoms, $4Mn^{2+}$ ions, 4 Ca^{2+} ions, 92 carbohydrate atoms and 563 solvent atoms for 44510 unique reflections.

3.4.4 Refinement summary

Below is a summary of the refinement procedure that was followed.

Refinement program	Number of cycles/adjustments made to the model	R-factor (%)	R-free (%)
REFMAC5 (rigid body)	10	29.2	28.2
REFMAC5 (restrained)	10	20.6	22.0
OOPS2	Used to speed up and streamline model rebuilding.		
O (version 7)	Checked ; (a) protein main-chain and side-chains in density correctly. (b) all rotamers for each side-chain. (c) difference density for T-antigen in each sub-unit.		
ARP_WARP	30 cycles 503 waters added	18.0	19.8
O (version 7)	Checked (and removed where necessary) waters are spherical and oxygen atoms in density at > 1.5 r.m.s in $2F_o – F_c$ maps. Verified that hydrogen bond distances are reasonable (wat-wat & wat-protein). Examined any remaining significant $F_o – F_c$ peaks *i.e.* > 3 sigma.		
REFMAC5 (restrained)	10	16.5	19.1
O (version 7)	Checked protein, sugar and waters for good geometry and fit to density and made adjustments where necessary. Modelled the alternate conformations.		

3.5   Analysis of the 2.25 Å X-ray model

The final structure contains a tetramer in the asymmetric unit, of molecular weight 110 kDa. Each monomer of the tetramer is made from a single polypeptide chain of 236 amino acid residues, 232 of which are modelled (4 residues at the C-terminus are without density) and contains one Ca^{2+} ion and one Mn^{2+} ion. A total of 441 water molecules were modelled around the tetramer. The space group is $P2_12_12$ (orthorhombic) with unit cell dimensions ; a = 128.89 Å, b = 125.67 Å, c = 75.97 Å, $\alpha = \beta = \gamma = 90°$. Previously the unit-cell dimensions for the PNA/lactose model[58] were; a = 129.30 Å, b = 126.90 Å, c = 76.90 Å, $\alpha = \beta = \gamma = 90°$.

Total no. of atoms = 7525
Total no. of protein atoms = 6972
Total no. of sugar atoms = 104
Total no. of water molecules = 441
Others = 8 metal ions (4 Ca^{2+} ions and 4 Mn^{2+} ions)
Resolution range = ∞ to 2.25 Å
R-factor = 16.5%
R-free = 19.1%
Bond distance r.m.s. deviation = 0.010 Å
Bond angle r.m.s. deviation = 1.381°
Ramachandran core/additional (%) = 91.9/7.9
Mean B-factor for all protein atoms ($Å^2$) = 29.5
Mean B-factor for all side-chain atoms ($Å^2$) = 30.1
Mean B-factor for all main-chain atoms = ($Å^2$) = 29.0
Mean B-factor for all sugar atoms ($Å^2$) = 42.4
Mean B-factor for all water molecules ($Å^2$) = 38.2
Mean B-factor for all side-chains and waters ($Å^2$) = 31.4
Mean B-factor for all atoms ($Å^2$) = 30.2

Table 5.   Summary of the 2.25 Å PNA/T-antigen model.

Total no. of atoms = 7639

Total no. of protein atoms = 6976

Total no. of sugar atoms = 92

Total no. of water molecules = 563

Others = 8 metal ions (4 Ca^{2+} ions and 4 Mn^{2+} ions)

Resolution range = 10 to 2.25 Å

R-factor = 16.4%

R-free = N/A

Bond distance r.m.s. deviation = 0.010 Å

Bond angle distance r.m.s. deviation = 0.038 Å

Ramachandran core/additional (%) = 90.1/9.8

Mean B-factor for all protein atoms ($Å^2$) = 21.1

Mean B-factor for all side-chain atoms ($Å^2$) = 21.8

Mean B-factor for all main-chain atoms = ($Å^2$) = 20.5

Mean B-factor for all sugar atoms ($Å^2$) = 36.5

Mean B-factor for all water molecules ($Å^2$) = 36.8

Mean B-factor for all side-chains and waters ($Å^2$) = 24.3

Mean B-factor for all atoms ($Å^2$) = 22.5

Table 6.   Summary of the 2.25 Å PNA/lactose model[58].

The Ramachandran plot[256] for the final model is shown below and was calculated using PROCHECK[257]. There are no residues in the disallowed regions. 735 amino acids (91.9%) are found within the most favoured regions (A, B, L) and 63 residues (7.9%) are found within the additionally allowed regions (a, b, l, p) illustrating that the model has good geometry. There are no distorted main-chain bond lengths, angles or planar groups. Two residues are labelled on the plot (Asp 78 of chain A and B) as these residues are in the generously allowed regions. When comparing the geometry of the model with that of the starting model, we see that the model now has improved geometry with more residues within the core region of the plot (starting model 90.1% core, final model 91.9% core).

In addition, previously there were several distorted main-chain bond lengths and angles (30 and 31 respectively) whereas now there are none. Moreover, some of

these distorted residues identified in the starting model (Asn 41, Leu 212, Gly 213) are found within the saccharide binding site.
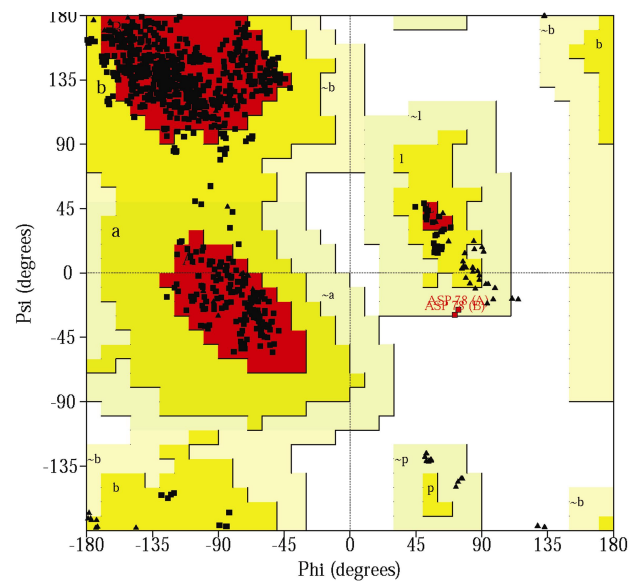
Figure 55.   The Ramachandran plot for the 2.25 Å PNA/T-antigen complex.

The overall root-mean-square deviations (*x, y, z*) between the 2.25 Å PNA/lactose structure starting model[58] (PDB code = 2PEL) and the final model were calculated as follows:

1. Between all protein atoms, r.m.s. deviation = 0.42 Å

2. Between all main-chain protein atoms, r.m.s. deviation = 0.35 Å

3. Between all side-chain protein atoms, r.m.s. deviation = 0.48 Å

4. Between all α-carbon protein atoms, r.m.s. deviation = 0.32 Å

The overall root-mean-square deviations (*x, y, z*) between the 2.5 Å PNA/T-antigen structure[143] (PDB code = 2TEP) and the final model were calculated as follows:

1. Between all protein atoms, r.m.s. deviation = 0.52 Å

2. Between all main-chain protein atoms, r.m.s. deviation = 0.31 Å

3. Between all side-chain protein atoms, r.m.s. deviation = 0.69 Å

4. Between all α-carbon protein atoms, r.m.s. deviation = 0.27 Å

The residues that exhibit the largest differences between the starting model (2PEL) and the final 2.25 Å PNA/T-antigen model are listed here, those given have root-mean-square deviations greater than 1 Å.

Main-chain large r.m.s. deviations;

Chain A – A60(Gly), A232(Thr).

Chain B – B99(Gly).

Chain D – D13(Glu), D60(Gly).

Side-chain large r.m.s. deviations;

Chain A – A59(Thr), A181(Asp), A232(Thr).

Chain B – B98(Ala), B112(Lys), B154(Asn).

Chain C – C93(Asp), C112(Lys).

Chain D – D12(Ser), D40(Val), D53(Arg), D59(Thr), D231(Thr), D232(Thr).

The amino acid residues above are all found within loop structures and on the surface of the protein, therefore, they have greater conformational freedom than those residues which are buried or involved in one of the 3 β-sheets, which make up the tertiary structure of PNA. The residue exhibiting the largest r.m.s. deviation is A232 (Thr 232). This residue is near the end of the polypeptide chain and therefore, is more flexible than those residues 'within' the core of the protein. Threonine 232 is also a highly exposed surface residue with an overall average accessible area value of 18.3 Å^2. The $2F_o - F_c$ electron density for this residue was poor and so the modelling of even the main-chain atoms of the residue proved problematic. As can be seen from the B-factor versus residue graph, for chain A (see **Figure 71**), this residue also has the highest B-factor (avB$_{A232}$ = 54.6 Å^2) of all the amino acids in chain A.

When the previous 2.5 Å PNA/T-antigen model (2TEP) and the newly refined model are superimposed the T-antigen molecules in each sub-unit exhibit small deviations. Below are given the r.m.s. deviation in atom positions for both the galactose and *N*-acetylgalactosamine sub-units of the T-antigen molecules. Note that the letters E

to H in parentheses below represent the chain identifier for the T-antigen molecules in the PDB file *e.g.* the T-antigen bound to chain A has a chain identifier of E.

Chain A (E) T-antigen r.m.s. deviation (Gal) = 0.42 Å

r.m.s. deviation (Gal*N*Ac) = 0.65 Å

Chain B (F) T-antigen r.m.s. deviation (Gal) = 0.26 Å

r.m.s. deviation (Gal*N*Ac) = 0.53 Å

Chain C (G) T-antigen r.m.s. deviation (Gal) = 0.30 Å

r.m.s. deviation (Gal*N*Ac) = 0.72 Å

Chain D (H) T-antigen r.m.s. deviation (Gal) = 0.31 Å

r.m.s. deviation (Gal*N*Ac) = 0.83 Å

The positions of the atoms in the galactose sub-unit of each T-antigen show excellent agreement with those of the 2TEP model, with slightly larger deviations for the second hexapyranose ring, Gal*N*Ac. This is unsurprising as the first hexapyranose ring makes more hydrogen bonds to the protein and has more van der Waals interactions with the protein.

The four sub-units in the molecule have essentially the same tertiary structure. The root-mean-square (r.m.s.) deviation in α-carbon positions between pairs of sub-units when they are superimposed range from 0.18 to 0.27 Å. The corresponding values for all protein atoms range from 0.48 and 0.67 Å. The smallest r.m.s. deviation is between sub-units A and B (0.18 Å for α-carbon positions) with the largest r.m.s. deviation between sub-units B and C (0.27 Å for α-carbon positions).

3.6   Metal sites

As is the case in other legume lectins, each monomer of peanut agglutinin contains one Ca^{2+} ion and one transition metal (in this case Mn^{2+} ion). The details of metal co-ordination are almost identical in the four sub-units. Below are the manganese to ligand distances.

Mn^{2+} ligand distances (Å)	Chain A	Chain B	Chain C	Chain D
Glu121 OE2	2.29 (2.32)	2.28 (2.33)	2.31 (2.32)	2.28 (2.32)
Asp123 OD2	2.27 (2.31)	2.30 (2.34)	2.26 (2.32)	2.29 (2.33)
Asp132 OD1	2.28 (2.31)	2.30 (2.34)	2.30 (2.32)	2.30 (2.33)
His137 NE2 (axial)	2.32 (2.36)	2.31 (2.36)	2.30 (2.34)	2.31 (2.34)
Water molecule (axial)	2.31 W76 (2.33)	2.33 W94 (2.34)	2.32 W56 (2.33)	2.33 W96 (2.34)
Water molecule	2.29 W449 (2.33)	2.29 W95 (2.32)	2.28 W504 (2.31)	2.29 W450 (2.33)

Table 7.   The manganese ligand distances in the four chains of the PNA tetramer (values from Banerjee *et al*., (1996)[58] are given in parentheses). The waters listed are numbered as in the PDB file.

The data agrees with that from the PNA/lactose model, in that Glu121 OE2, Asp123 OD2, Asp132 OD1, His137 NE2 and two water molecules co-ordinate to the manganese ion. When comparing the distances for chains A, B, C, and D, there is excellent agreement between all four; the average manganese-ligand distances being 2.29 Å, 2.30 Å, 2.30 Å, and 2.30 Å for A, B, C, and D respectively. The average manganese-ligand distances in the PNA/lactose model are 2.32 Å, 2.34 Å, 2.32 Å, and 2.33 Å for A, B, C, and D respectively.

The calcium-ligand distances are given (see **Table 8**). Asp123 OD1, Asp123 OD2, Tyr125 O, Asn127 OD1, Asp132 OD2 and two water molecules co-ordinate to the calcium ion. When comparing the distances of chains A, B, C, and D, there is excellent agreement between all four; the average calcium-ligand distances being 2.41 Å, 2.41 Å, 2.42 Å, and 2.40 Å for A, B, C, and D respectively. The average calcium-ligand distances in the PNA/lactose model are 2.42 Å for all four chains. The ligands at the calcium binding site form a pentagonal bipyramidal arrangement, with Asp 132 OD2, Tyr 125 O, and the two water molecules forming the square-planar ligands and Asp 123 OD1/2 and Asn 127 OD1 positioned at the axial positions. The metal to metal distances for each of the four sub-units are given. The calcium-manganese distances in the PNA/lactose complex varied between 4.13 and 4.39 Å in the four sub-units. The calcium-manganese distances in the PNA/T-antigen model reported here vary between 4.09 and 4.18 Å in the four sub-units.

Ca^{2+}-ligand distances (Å)	Chain A	Chain B	Chain C	Chain D
Asp123 OD1	2.50 (2.45)	2.46 (2.44)	2.50 (2.44)	2.44 (2.43)
Asp123 OD2	2.42 (2.46)	2.42 (2.42)	2.43 (2.44)	2.44 (2.44)
Asn127 OD1	2.42 (2.40)	2.42 (2.43)	2.46 (2.45)	2.44 (2.43)
Asp132 OD2	2.35 (2.39)	2.35 (2.39)	2.38 (2.41)	2.37 (2.41)
Tyr125 O	2.37 (2.40)	2.36 (2.40)	2.34 (2.38)	2.36 (2.39)
Water molecule	2.43 W114 (2.42)	2.43 W38 (2.43)	2.39 W34 (2.43)	2.38 W32 (2.40)
Water molecule	2.39 W27 (2.42)	2.41 W57 (2.43)	2.45 W177 (2.42)	2.39 W140 (2.41)

Table 8.   The calcium ligand distances in the four chains of the PNA tetramer (values from Banerjee *et al.*, (1996)[58] are given in parentheses).

If we compare the B-factors of the atoms chelated to the metals in each chain, there is good agreement in values between chain A and B, for both the calcium and manganese sites. The metal ligands in these two chains appear to be the most ordered *i.e.* have the lowest B-factors. This is also reflected in the B-factors of the metals themselves. The metal ligands in chain C appear to be the least ordered as both the manganese and calcium sites have the highest average B-factors. This is again reflected in the B-factors of the metals themselves. Chain D metal sites lie intermediate in B-value with respect to the other chains. Below are given the B-factors of all the metal ligand atoms.

Mn^{2+} ligand B-factors($Å^2$)	Chain A	Chain B	Chain C	Chain D
Glu121 OE2	25.5 (20.6)	24.5 (17.8)	29.3 (24.5)	24.5 (17.4)
Asp123 OD2	25.0 (19.4)	24.6 (17.5)	28.5 (25.3)	25.3 (19.6)
Asp132 OD1	27.3 (23.4)	26.1 (19.9)	33.4 (23.3)	30.1 (15.4)
His137 NE2	26.1 (20.9)	27.3 (19.8)	33.2 (24.9)	27.1 (22.0)
Water molecule	30.4 (26.3) W449 (W551)	29.4 (16.5) W94 (W554)	35.9 (23.6) W56 (W558)	31.7 (21.1) W96 (W562)
Water molecule	29.9 (23.9) W76 (W550)	31.2 (22.0) W95 (W555)	36.7 (28.8) W504 (W559)	33.2 (25.5) W450 (W563)

Table 9.   The manganese ligand B-factors in the four chains of the PNA tetramer (values from Banerjee *et al.*, (1996)[58] are given in parentheses).

If we compare the B-values of the manganese ligand atoms against each other they all are in the range 24 to 37 $Å^2$. If we compare the B-values of the calcium ligand atoms against each other they all lie in the range 22 to 32 $Å^2$. Some of the metal ligand atoms are part of the seven-stranded β-sheet structure of the PNA monomer

(Glu 121/Asp 123/His 137), and therefore, it is unsurprising that they exhibit low B-factors. However, even those that belong to loop structures (Tyr 125/Asn 127/Asp 132) also exhibit low B-factors.

Ca^{2+}-ligand B-factors($\mathring{A}^2$)	Chain A	Chain B	Chain C	Chain D
Asp123 OD1	25.2 (16.6)	25.1 (17.5)	28.5 (23.6)	25.9 (15.6)
Asp123 OD2	25.0 (19.4)	24.6 (17.5)	28.5 (25.3)	25.3 (19.6)
Asn127 OD1	25.1 (19.0)	23.8 (19.8)	30.6 (23.8)	27.6 (24.6)
Asp132 OD2	26.8 (18.1)	24.9 (16.8)	30.4 (22.8)	27.2 (20.0)
Tyr125 O	26.0 (18.4)	26.1 (17.0)	31.7 (23.7)	29.0 (17.7)
Water molecule	23.1 (15.4) W114 (W548)	22.1 (15.8) W38 (W552)	28.2 (18.3) W34 (W557)	25.0 (20.2) W32 (W561)
Water molecule	24.1 (17.4) W27 (W549)	24.1 (11.7) W57 (W553)	28.2 (20.1) W177 (W556)	25.9 (13.7) W140 (W560)

Table 10.   The calcium ligand B-factors in the four chains of the PNA tetramer (values from Banerjee *et al*., (1996)[58] are given in parentheses).

On the next page are the average B-factors for the metal ligands of each chain and the average B-factor for the metal ligands over all the four chains. Overall these atoms are well-ordered as they exhibit B-values which are less than the main-chain protein atoms (average B_{mc}= 29.0 $\mathring{A}^2$ compared to average $B_{Mn\ ligands}$= 29.0 $\mathring{A}^2$ and average $B_{Ca\ ligands}$= 26.4 $\mathring{A}^2$).

Average B-value of ligands to each metal ($\mathring{A}^2$)	Chain A	Chain B	Chain C	Chain D	All chains
Mn^{2+}	27.4 (22.4)	27.2 (18.9)	32.8 (25.1)	28.7 (20.2)	29.0 (21.6)
Ca^{2+}	25.0 (17.8)	24.4 (16.6)	29.4 (22.5)	26.6 (18.8)	26.4 (18.9)

Table 11. The average B-factors for the metal ligands of each chain and the average B-factor for the metal ligands over all the four chains (Banerjee *et al*., (1996)[58] values in parentheses).

Below are the B-factors for the manganese and calcium in all four chains.

Metal B-value($\mathring{A}^2$)	Chain A	Chain B	Chain C	Chain D
Mn^{2+}	33.2 (35.6)	33.21 (34.8)	38.3 (37.7)	35.0 (35.2)
Ca^{2+}	24.9 (16.1)	23.9 (16.8)	29.0 (18.7)	26.0 (17.7)

Table 12. B-factors for the manganese and calcium in all four chains.

If we compare the B-factors of the previous model with those here, we see that those from the PNA/lactose model are lower than the values calculated by REFMAC-5 for the PNA/T-antigen model, however, each structure shows approximately the same trend between chains. Possibly packing effects can explain the apparent trend in stability between chains, chain B being involved in a greater amount of intermolecular contacts thereby aiding the stability of the chain.

3.7   Sugar binding

Residues in four loops, 91 to 106, 125 to 135, 75 to 83 and 211 to 216, make up the carbohydrate binding pocket of the sub-unit in PNA. The invariant lectin-galactose interactions involving Asp 83, Gly 104 and Asn 127, are observed here, as in previous studies[58,143]. These interactions are given in **Table 14** and **Figure 57**. The stacking interaction between an aromatic residue (Tyr 125) and the galactose sugar ring is also identified as well as the close proximity of Ala 82.
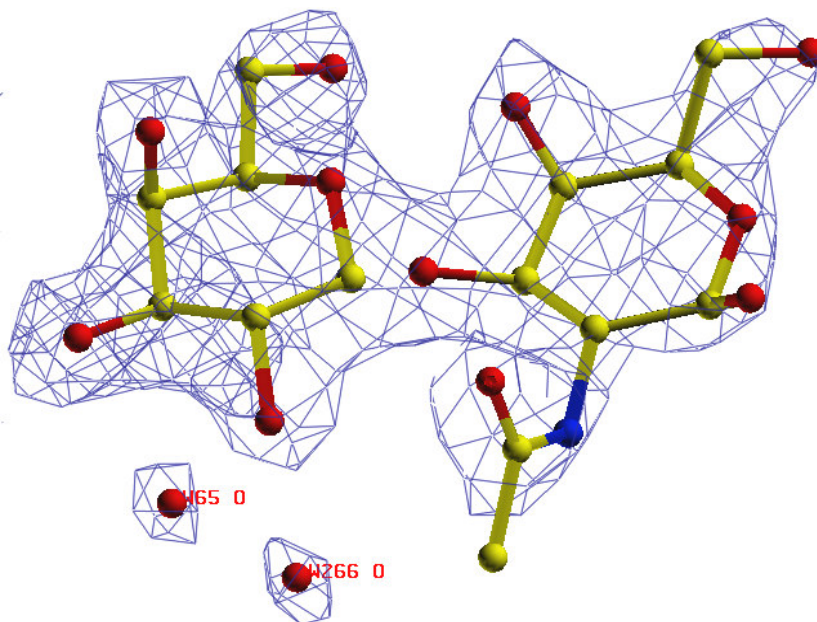
Figure 56.   The $2F_o - F_c$ electron density for the T-antigen molecule identified at the saccharide binding site of chain B in PNA.

All of the invariant residues are within 4 Å of the T-antigen molecule. The full list of the residues within 4 Å of the sugar are given here;

Asp 80, Ala 82, Asp 83, Gly 103, Gly 104, Tyr 125, Asn 127, Ser 211, Leu 212, Gly 213 and Gly 214.

In addition to the invariant hydrogen bonds, the galactose ring in T-antigen is involved in three more hydrogen bonds with PNA. The sugar atoms O4 and O5 interact with Ser 211 OG and O6 interacts with Asp 80 OD2. Serine 211 OG interacts with O3 and O4 of the Gal*N*Ac ring also. Yet another interaction with this ring involves O4 and Gly 213 N. These interactions have been highlighted in the previous PNA/T-antigen structure[143] at lower resolution and have been identified within the model presented here.
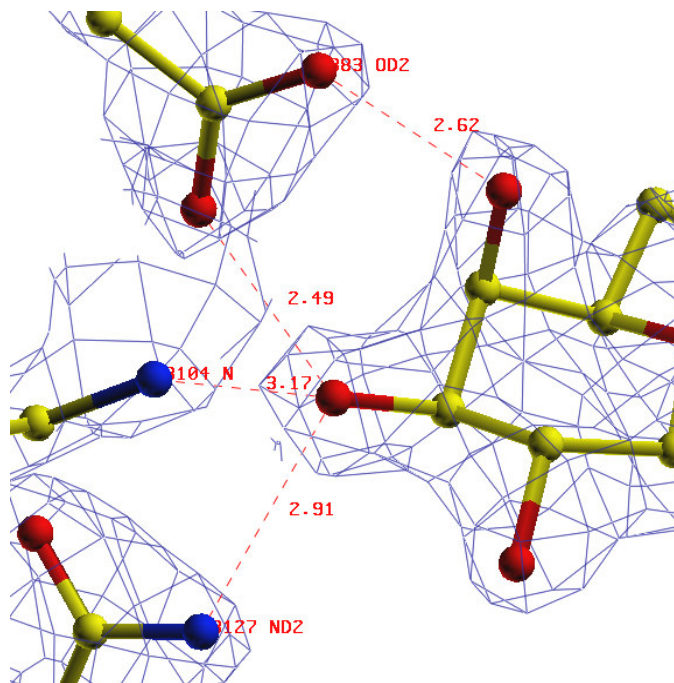
Figure 57.   Illustrates the four invariant hydrogen bonds between the galactose sugar ring atoms O3/O4 and the protein residues Asp 83, Gly 104 and Asn 127.
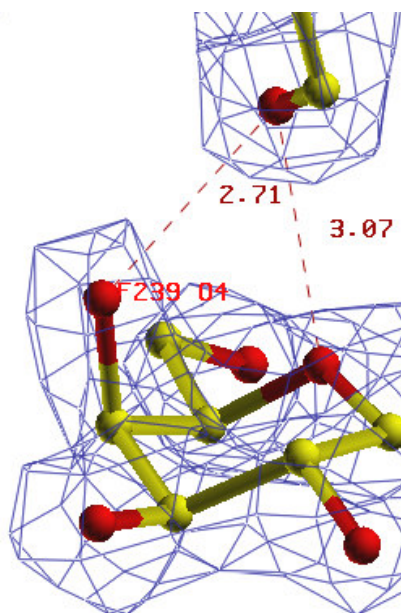
Figure 58.   Two hydrogen bonds specific to the PNA/T-antigen interaction, between the galactose sugar ring atoms O4/O5 and the side chain of Ser 211.
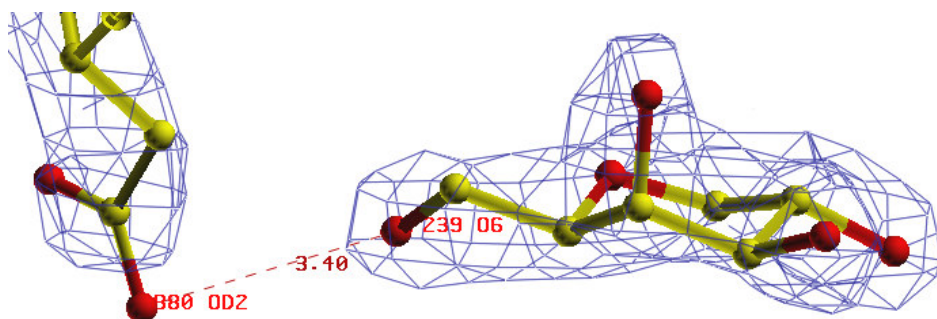
Figure 59.   Another hydrogen bond specific to the PNA/T-antigen interaction between the galactose sugar ring atom O6 and the side chain of Asp 80 OD2.
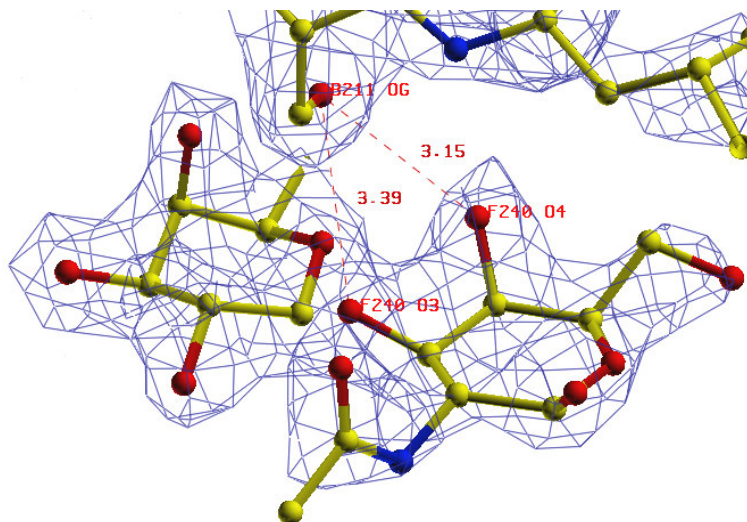
Figure 60.   Two hydrogen bonds specific to the PNA/T-antigen interaction between the Gal*N*Ac sugar ring atoms O3/O4 and the side chain of Ser 211.
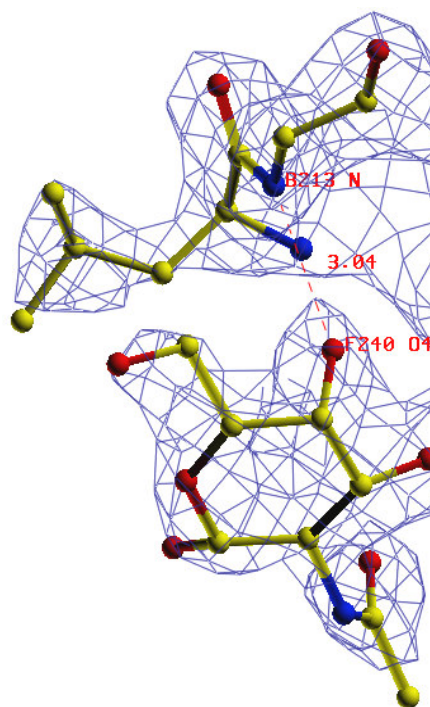
Figure 61.   A hydrogen bonds specific to the PNA/T-antigen interaction between the Gal*N*Ac sugar ring atom O4 and the main chain N of Gly 213.
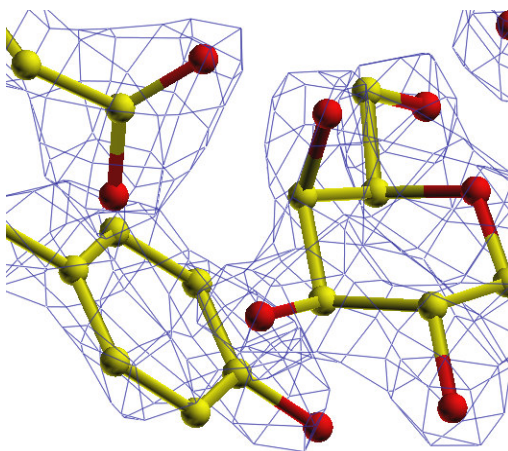
Figure 62.   Illustrates the stacking interaction between an aromatic residue
                 (Tyr 125) and the hydrophobic face of the galactose sub-unit of the
                 T-antigen molecule.

Water bridges are also involved in linking the T-antigen to the protein. In all four
chains of the PNA tetramer, there are four water molecules found within the
saccharide binding region at identical positions in each. The four equivalent waters
of each chain are given in **Table 13**.

	Chain A	Chain B	Chain C	Chain D
Water 1.	W379	W266	W517	W138
Water 2.	W125	W65	W232	W207
Water 3.	W147	W133	W184	W152
Water 4.	W136	W74	W255	W105

Table 13.   The four water molecules found at equivalent positions in all four
                 chains of the PNA tetramer (water numbers are those in the PDB file).

Water 1. Links Gal O2 to Glu 129 OE2. The water oxygen is hydrogen bonded to
the sugar atom and then to the oxygen atom OE2 of the glutamic acid side-chain, and
therefore connects the T-antigen molecule to the protein *i.e.* Gal O2 → water '1'
→ Glu 129 OE2.

Water 2. Creates another water-mediated hydrogen bond, this time between Gal O2 and Gly 104 N. *i.e.* Gal O2 → water '2' → Gly 104 N.

These two water-mediated interactions described above also exist in the PNA/lactose complex, however, the following water-mediated interactions are not found in the PNA/lactose complex.   These extra water-mediated interactions appear to be responsible for the twenty-fold increase in specificity of PNA towards T-antigen compared with lactose. The third water-mediated hydrogen bond occurs between Gal*N*Ac O7 and Ile 101 O. *i.e.* Gal*N*Ac O7 → water '3' → Ile 101 O. The fourth water is involved in two hydrogen bonds, that between Gal*N*Ac O7 and Asn 41 ND2, and that between Leu 212 N and Gal*N*Ac O7.
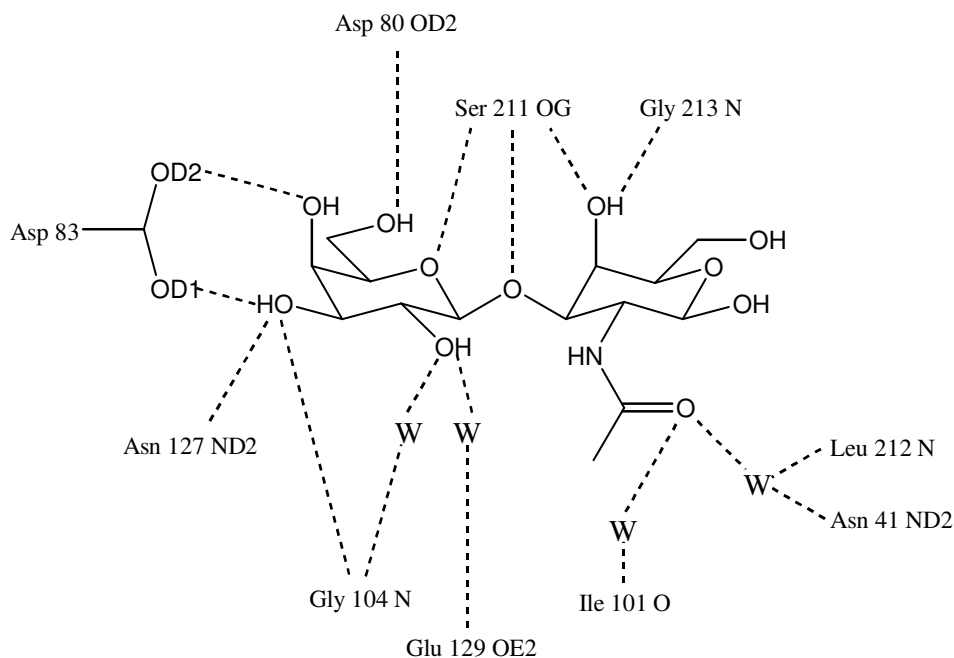
Figure 63.   Schematic of the hydrogen-bond interactions between the T- antigen molecule and the amino acid residues at the saccharide binding-site in PNA.
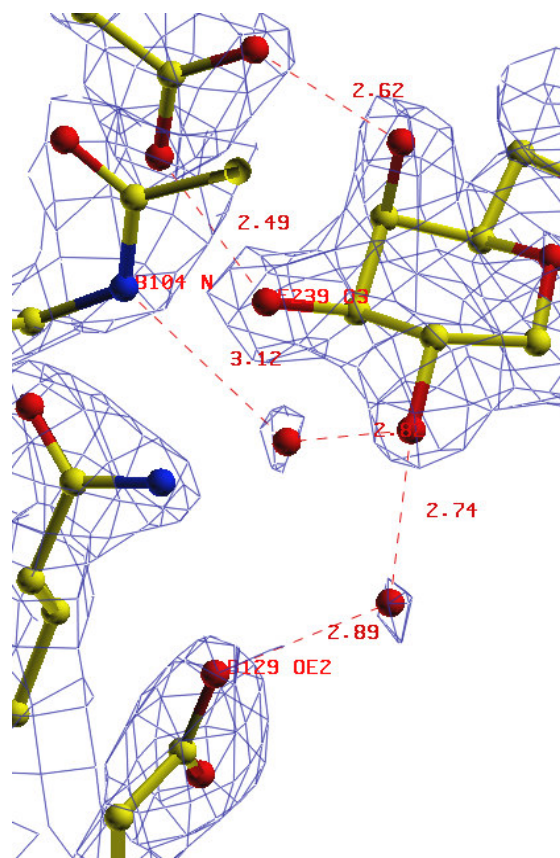
Figure 64.   Depicts two of the water-mediated hydrogen-bond interactions. One between the O2 of the galactose and Glu 129 OE2 and the second between the O2 of the galactose and Gly 104 N.
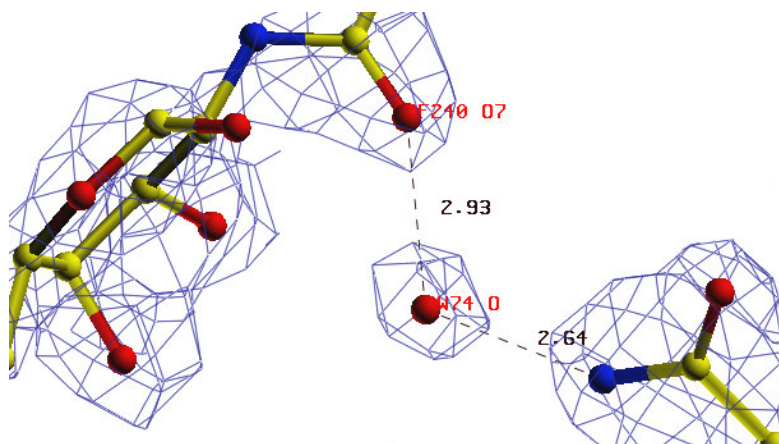
Figure 65.   Depicts the water-mediated hydrogen bond between the Gal*N*Ac O7 atom and Asn 41 ND2.
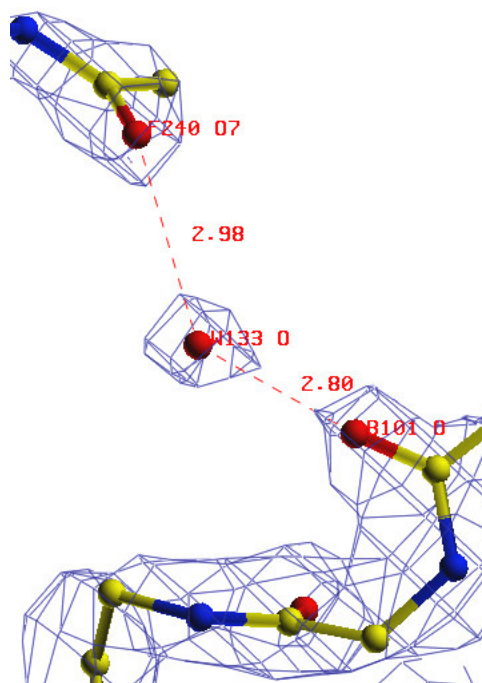
Figure 66.   Depicts the water-mediated hydrogen bond between the Gal*N*Ac O7 atom and Ile 101 O atom.
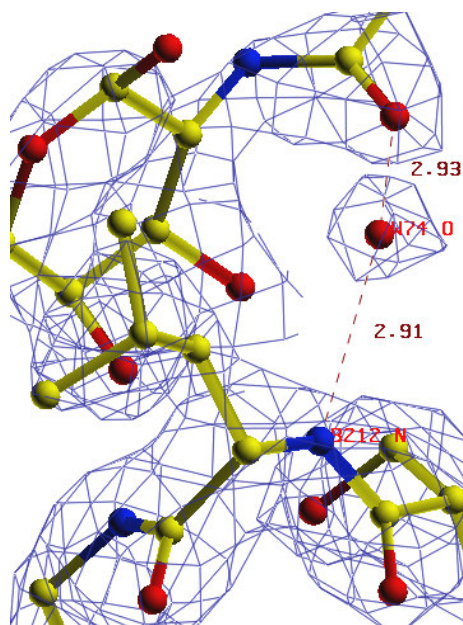
Figure 67.   Depicts the water-mediated hydrogen bond between the O7 atom of Gal*N*Ac and Leu 212 N.

	Chain A	Chain B	Chain C	Chain D
Invariant hydrogen bonds (Å)				
Asp 83 OD1 to Gal O3	2.50 (2.67)	2.49 (2.43)	2.46 (2.58)	2.62 (2.49)
Gly 104 N to Gal O3	3.06 (3.08)	3.17 (2.92)	2.98 (3.07)	3.06 (2.81)
Asn 127 ND2 to Gal O3	2.95 (2.86)	2.91 (3.00)	3.12 (3.12)	2.90 (3.02)
Asp 83 OD2 to Gal O4	2.58 (2.59)	2.62 (2.68)	2.76 (2.64)	2.68 (2.55)
PNA specific hydrogen bonds				
Ser 211 OG to Gal O4	2.66 (2.62)	2.71 (2.82)	2.79 (2.66)	2.79 (2.76)
Ser 211 OG to Gal O5	2.89 (3.12)	3.07 (3.34)	2.98 (3.09)	3.06 (3.16)
Asp 80 OD2 to Gal O6	3.09 (3.33)	3.40 (3.39)	3.50 (OD1) (3.36)	2.71 (2.98)
Ser 211 OG to Gal*N*Ac O3	3.18	3.39	3.23	3.32
Gly 213 N to Gal*N*Ac O4	2.91	3.04	2.95	3.00
Ser 211 OG to Gal*N*Ac O4	3.09	3.15	3.31	3.32
Water bridges				

Gal O2 to wat 1	3.05 (W379)	2.74 (W266)	2.58 (W517)	2.70 (W138)
Wat 1 to Glu 129 OE2	2.75	2.89	2.91	2.68
Gal O2 to Wat 2	2.80 (W125)	2.82 (W65)	2.86 (W232)	2.75 (W207)
Wat 2 to Gly 104 N	3.07	3.12	3.06	3.10
GalNAc O7 to Wat 3	3.12 (W147)	2.98 (W133)	3.19 (W184)	3.16 (W152)
Wat 3 to Ile 101 O	2.65	2.80	2.68	2.67
Wat 2 to wat 3	2.64	2.73	2.60	2.77
GalNAc O7 to Wat 4	2.77 (W136)	2.93 (W74)	2.90 (W255)	2.78 (W105)
Wat 4 to Asn 41 ND2	2.73	2.64	2.70	2.85
Wat 4 to Leu 212 N	2.99	2.91	3.08	2.96
Gal O2 to wat 5	4.58 (W205)	4.46 (W63)	4.24 (W284)	4.24 (W163)
Wat 5 to wat 2	3.00	2.71	2.64	2.56
Wat 5 to Glu 129 OE2	2.67	2.58	2.84	2.94
Wat 5 to Tyr 130 OH	2.93	2.84	2.94	2.84

Table 14. Hydrogen bonds and water-mediated interactions at the saccharide binding site in each of the four sub-units of the PNA tetramer.

131

The density for the T-antigen molecule was clearly identifiable in the maps. The galactose monomer of the T-antigen sits nearer to the protein residues than the Gal$N$Ac monomer. It therefore has more interactions with the protein (7 direct hydrogen bonds vs. 3 direct hydrogen bonds) and thus, is more tightly bound than the Gal$N$Ac hexapyranose ring. This is evident from the density and B-factors for the two sub-units of T-antigen *e.g.* chain B (F) T-antigen disaccharide, average B-factor$_{\mathbf{Gal}}$ = 14.0 Å^2 and average B-factor$_{\mathbf{Gal\textit{N}Ac}}$ = 38.1 Å^2.

The same trend in stability at the metal sites is observed for the T-antigen molecules in each chain when comparing B-factors, *i.e.* B, A, D, C. The average B-factors for the four T-antigen molecules in each chain being 33.0 Å^2, 27.9 Å^2, 59.0 Å^2, and 49.4 Å^2, for chains A, B, C, and D respectively. The values for the Ravishankar *et al.,* (1997)[143] data are 11.7 Å^2, 12.5 Å^2, 24.2 Å^2, and 21.9 Å^2 for the T-antigen molecules of chains A, B, C, and D respectively.

Below are given the B-factors of the sugar ligand atoms that make direct hydrogen bonds with the T-antigen molecule. The average B-factors for the seven sugar ligand atoms in each chain are 28.3 Å^2, 27.1 Å^2, 36.6 Å^2, and 31.1 Å^2 for chains A, B, C, and D respectively.

Sugar ligand atom	Chain A	Chain B	Chain C	Chain D
Asp 83 OD1	23.2	23.8	29.9	24.5
Gly 104 N	24.8	23.0	34.6	26.6
Asn 127 ND2	26.5	25.0	33.2	30.9
Asp 83 OD2	27.8	23.4	30.9	26.1
Ser 211 OG	26.8	25.7	36.7	31.3
Asp 80 OD2	42.3	43.2	51.5	44.7
Gly 213 N	26.5	25.3	39.6	33.1

Table 15.   B-factors (Å^2) of the sugar ligand atoms that make direct hydrogen bonds with the T-antigen molecule.

Even though these residues are all part of loop sections their B-factors are low when compared to main-chain protein atoms. The exception to this is Asp 80 OD2, which is positioned in the middle of a loop section and therefore, is in the most flexible position. From analysis of the average accessible area of Asp 80 we see that the side-chain atoms are highly exposed (side-chain value, chain A = 18.2 Å^2 *cf.* main-chain value, chain A = 0.18 Å^2). The main-chain atoms are involved in two hydrogen bonds, (A215 N→A80 O, A213 O→ A80 N) however, the only stabilising interaction concerning the side-chain is a single hydrogen bond between A80 OD2 and O6 of the galactose sub-unit. Previous PNA studies also displayed a high B-factor for Asp 80 OD2 with respect to the other sugar ligands[58,143].

3.8   Dimerization

An analysis of the inter-sub-unit interactions at the dimer interface was performed and compared against the contacts described for the PNA/lactose complex. There are a large number of interactions between the two sub-units *i.e.* A and D, B and C. 22 residues are involved in protein-protein interactions of less than 4 Å, while another 3 residues that do not make direct protein-protein interactions are also involved *via* water bridges. Therefore, overall 25 residues are involved in holding the two sub-units together, these are;

1, 3, 5 (water-bridge only), 64, 65, 66, 149, 164, 166, 168, 169 (water-bridge only), 170, 171, 173 (water-bridge only), 175, 177, 179, 183, 184, 185, 187, 188, 189, 226 and 228.

In the PNA/lactose complex 22 residues are involved in protein-protein interactions of less than 4 Å, with two other residues, 169 and 170 assisting through water-mediated interactions. Therefore, overall 24 residues were shown to be involved in holding the two sub-units together in the PNA/lactose complex. The residues involved are the same as those for the PNA/T-antigen complex with the omission of residue 5. Residue 5 is connected to residue 183 through a water bridge involving water 358. In the PNA/lactose complex there is no water at the equivalent position to this water, and so the residue is not seen to be involved in linking the two sub-units.

The residues involved are distributed over all 6 strands of the flat sheet and make ~105 inter-atomic contacts of less than 4 Å (107 in the A-D interface, 103 in the B-C interface) in the PNA/T-antigen complex dimer interface. In the PNA/lactose complex there are around 90 inter-atomic contacts of less than 4 Å (89 in the A-D interface, 91 in the B-C interface) involved in the dimer interface. Of these 11 can be described as hydrogen bonds in both the PNA/lactose complex and the PNA/T-antigen complex:

Thr 3 OG1----Asp 184 OD1

Ser 64 OG----Thr 187 OG1

Thr 187 OG1----Asp 168 OD2

Ile 185 N----Thr 228 OG1

Ile 188 N----Asp 168 OD1, their two-fold equivalents, and

Ser 175 OG----Ser 175 OG.

Water bridges also occur at the dimer interface. In the A-D interface in the PNA/T-antigen complex, 9 waters were shown to interact with oxygen or nitrogen atoms in both sub-units at distances of less than 3.5 Å. These waters are W9, W24, W25, W62, W340, W358, W373, W428 and W515. There are also 6 other waters (W317, W299, W58, W29, W180, W408) that are within 4 Å of both sub-units, making a total of 15 waters identified at the A-D interface.
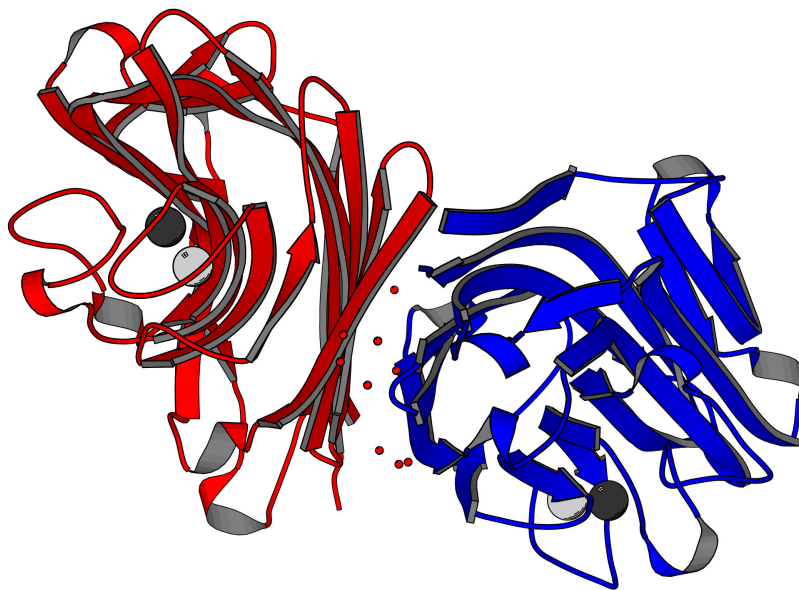
Figure 68. The A-D interface in the PNA dimer. The 9 waters that link the sub-units through water-mediated hydrogen bonds are also shown.

In the A-D interface in the PNA/lactose complex, 6 waters were shown to interact with oxygen or nitrogen atoms in both sub-units at distances of less than 3.5 Å. These waters are W14, W25, W41, W343, W402 and W470. There are also 11 other waters (W26, W47, W77, W99, W176, W205, W333, W339, W351, W391, W461) that are within 4 Å of both sub-units, making a total of 17 waters identified at the A-D interface.

In the B-C interface in the PNA/T-antigen complex a total of 14 waters were identified, 9 of which interact with oxygen or nitrogen atoms in both sub-units at a distance of less than 3.5 Å. These waters are W2, W17, W50, W196, W327, W424, W36, W129 and W324. The other five waters, which are within 4 Å of both sub-units are W148, W211, W313, W369 and W374. In the B-C interface in the PNA/lactose complex a total of 19 waters were identified, 9 of which interact with oxygen or nitrogen atoms in both sub-units at a distance of less than 3.5 Å. These waters are W42, W70, W400, W193, W158, W28, W440, W259 and W547. The other 10 waters are within 4 Å of both sub-units and are given also for completeness; W155, W295, W412, W436, W445, W463, W469, W482, W488 and W526.

Comparing the position of the waters at each interface, (A-D and B-C), between the two complexes, shows there is good agreement between the two. When comparing the positions of the 15 waters at the A-D interface in the T-antigen complex with the 17 waters at the A-D interface in the lactose complex, we see that 12 of these match within 1 Å. The equivalent waters are given below;

PNA/T-antigen waters at A-D interface		Corresponding water in PNA/lactose complex
W9	=	W14
W24	=	W402
W25	=	W41
W62	=	W25
W340	=	W343
W373	=	W470
W58	=	W47
W29	=	W26
W180	=	W77
W408	=	W351
W317	=	W231
W299	=	W333

Table 16.   Water molecules found at equivalent positions at the A-D interface in both the PNA/T-antigen complex and the PNA/lactose complex. The numbers of the waters are those from the PDB files.

When comparing the positions of the 19 waters identified at the B-C interface in the PNA/lactose complex with the 14 waters identified at the B-C interface in the PNA/T-antigen complex, 12 are seen to agree in position within 1 Å. The equivalent waters at this interface are given below;

PNA/T-antigen waters at B-C interface		Corresponding water in PNA/lactose complex
W2	=	W42
W17	=	W70
W50	=	W400
W196	=	W193
W327	=	W158
W36	=	W28
W129	=	W440
W324	=	W259
W211	=	W60
W148	=	W436
W313	=	W338
W374	=	W526

Table 17.   Water molecules found at equivalent positions at the B-C interface in both the PNA/T-antigen complex and the PNA/lactose complex. The numbers of the waters are those from the PDB files.

Calculation of the buried surface area on dimerization shows that 1920 Å^2 of surface area is buried on dimerization, of which 71% is found to be non-polar. Therefore, it can be seen that dimerization involves substantial hydrophobic interactions, hydrogen bonds and water bridges.

3.9   Dimer association in the tetramer

Two interfaces are involved in the assembly of the tetramer from the two dimers, namely the interface between sub-units A and B and the interface between C and D. The A-B interface is formed by the side-by-side alignment of the two flat 6-stranded β-sheets. There are 72 inter-sub-unit contacts of less than 4 Å in the PNA/T-antigen complex and 65 inter-sub-unit contacts of less than 4 Å in the PNA/lactose complex. The residues involved in direct protein-protein interactions in the A-B interface of the PNA/T-antigen complex and the PNA/lactose complex being residue 2 from the first strand of the flat β-sheet, residue 231 from the second strand of the flat β-sheet, residue 48 from the second strand of the 7-stranded curved β-sheet, and residues 12 to 17, 50, 51, 53 and 201 from the loops that connect the two β-sheets. They include side-chain to side-chain hydrogen bonds (6 hydrogen bonds altogether) between Ser 12 and Glu 2, and between Arg 53 and Glu 13 in both the T-antigen and lactose complexes *i.e.*

Glu 2 OE1-----Ser 12 OG

Glu 13 O -----Arg 53 NH1

Glu 13 O -----Arg 53 NH2 and their two-fold equivalents

There are therefore, less inter-sub-unit contacts than at the dimer interface and also less hydrogen bonds. However, there are a larger number of water bridges in the A-B interface, which help tether the two sub-units A and B together. In total there are 21 waters located at the A-B interface in the PNA/T-antigen complex, remarkably more than at the dimer interface. Of these 21 water oxygen atoms, 15 are seen to interact with oxygen or nitrogen atoms in both sub-units at a distance of less than 3.5 Å, while 6 waters are within 4 Å of both sub-units.
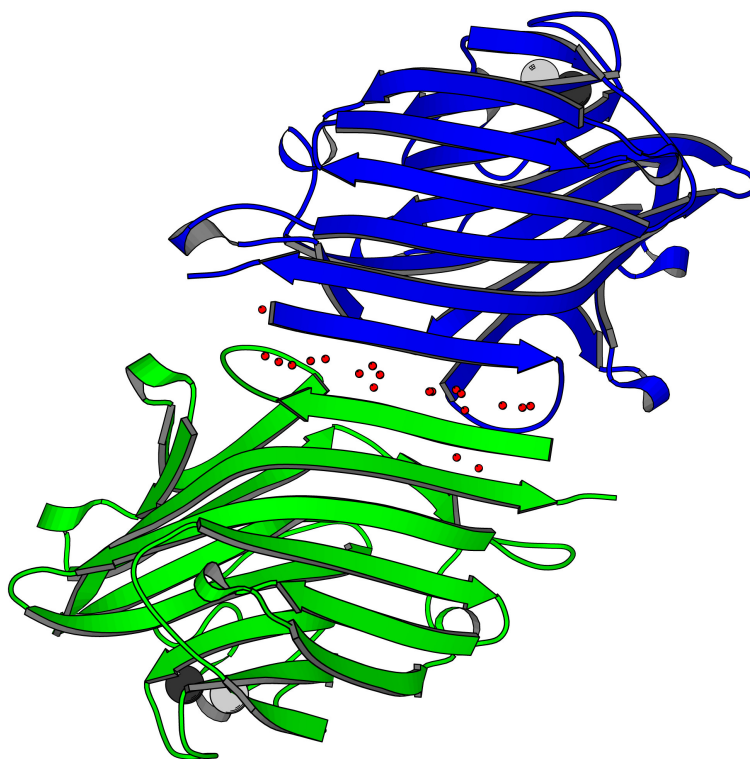
Figure 69.   The A-B interface in the PNA tetramer.   The waters that make hydrogen bonds to both sub-units at the interface are shown.

The PNA/lactose complex also has substantially more water bridges at the A-B interface than at the dimer interface, with 26 waters involved in the A-B interface. Of these 26 waters identified in the lactose complex, 18 are seen to interact with oxygen or nitrogen atoms in both sub-units at a distance of less than 3.5 Å, while 8 waters are within 4 Å of both sub-units.

In the PNA/T-antigen complex all the peptide carbonyl and amide groups of residues 3, 5 and 7 in the first strand of the flat sheet are involved in these bridges as well as residues in the 9-17 loop (9, 10, 12, 14, 15) and also residues 1, 2, 50, 53, and 201.

In the PNA/lactose complex all the peptide carbonyl and amide groups of residues 3, 5, and 7 in the first strand of the flat sheet are involved in these bridges as well as residues in the 9-17 loop and residues 51, 53, 201 and 231. Of the 15 waters at the

A-B interface in the PNA/T-antigen complex which are seen to make hydrogen bonds to both sub-units, 13 are seen to have waters at equivalent positions in the PNA/lactose complex. Of the 21 waters identified overall at the A-B interface in the PNA/T-antigen complex, 18 are seen to agree with the positions of the water oxygen atoms in the PNA/lactose complex. The matches between the two complexes are given here;

PNA/T-antigen waters at A-B interface		Corresponding waters in PNA/lactose complex
W40	=	W230
W188	=	W92
W201	=	W254
W214	=	W217
W222	=	W172
W319	=	W313
W202	=	W131
W221	=	W79
W233	=	W287
W303	=	W453
W343	=	W102
W389	=	W500
W406	=	W290
W55	=	W38
W238	=	W345
W181	=	W44
W501	=	W146
W355	=	W429

Table 18.   Water molecules found at equivalent positions at the A-B interface in both the PNA/T-antigen complex and the PNA/lactose complex. The numbers of the waters are those from the PDB files.

Calculation of the amount of buried surface area at the A-B interface is quite low at 1385 Å^2 with a non-polar component of 50%.

The other interface involved in the tetrameric assembly is that between sub-units C and D. The C-D interface brings together the top regions of sub-units C and D and is made up of several residues. The majority of the residues involved in inter-sub-unit contacts at this interface are found within the small 5-stranded β-sheet, sheet 3. The residues are: 27, 31, 33, 72, 74, 160, 217, 219. Two residues 28 and 29, positioned in a turn that connects two adjacent strands in this sheet also make inter-sub-unit contacts, as well as several other residues, 9, 10, 37, 158 and 221, from loops at the top of the sub-unit. In the PNA/lactose complex residue 77 is also involved in protein-protein interactions as well as those mentioned above. Residue 77 is a lysine and therefore, has a long side-chain with large conformational freedom. In the PNA/T-antigen complex the side-chain is slightly shifted away from the interface and so does not make an interaction with the other sub-unit. There are 65 inter-atomic contacts of less than 4 Å in the PNA/T-antigen complex compared to 58 in the PNA/lactose complex, with 12 of these being identified as hydrogen bonds in both complexes. Five out of six of the hydrogen bonds involve either of the two residues Lys 74 and Arg 221. The hydrogen bonds are listed here:

Lys 74 NZ----Asn 9 O
Gln 33 NE2----Ser 28 OG
Lys 74 NZ----Asn 29 O
Arg 221 NH2----Glu 72 OE1
Arg 221 NH1----Glu 72 OE1
Arg 221 NH2----Gly 158 O and their two-fold equivalents

There are also 13 water oxygen atoms identified at the C-D interface in the PNA/T-antigen complex, seven of these water atoms are at distances less than 3.5 Å from an oxygen or nitrogen in both sub-units, while 6 are at distances of less than 4 Å from both sub-units. The seven waters which make hydrogen bonds to both sub-units are W31, W509, W73, W97, W194, W483 and W513.
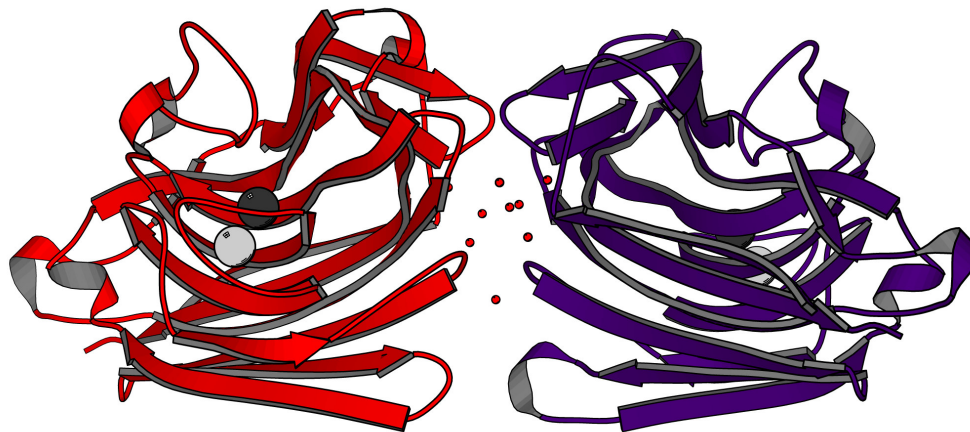
Figure 70.   The C-D interface in the PNA tetramer. with the waters that bridge the
                two monomers shown.

In the PNA/lactose complex, 9 waters (W18, W89, W119, W143, W256, W282,
W326, W362, W405) were identified at the C-D interface at a distance of less than
3.5 Å from an oxygen or nitrogen in both sub-units, while 6 waters are at a distance
of less than 4 Å from both sub-units. Of the seven waters that make hydrogen bonds
to both sub-units in the C-D interface of the PNA/T-antigen complex, six are found
at identical positions in the PNA/lactose complex. Of the 13 waters identified
overall at this interface in the PNA/T-antigen complex, 10 are seen at identical
positions in the PNA/lactose complex. The matches are shown once again over the
page;

The surface area buried in the association between the two sub-units is low at
1246 $Å^2$ of which only 36% is non-polar.

PNA/T-antigen waters at C-D interface		Corresponding water in PNA/lactose complex
W31	=	W18
W509	=	W256
W73	=	W119
W97	=	W89
W194	=	W326
W513	=	W237
W414	=	W405
W209	=	W316
W236	=	W419
W271	=	W105

Table 19.   Water molecules found at equivalent positions at the C-D interface in both the PNA/T-antigen complex and the PNA/lactose complex. The numbers of the waters are those from the PDB files.

3.10 Temperature-factor analysis

Given here are the temperature-factor versus residue number plots for each of the protein polypeptide chains A, B, C and D. It can be seen from the plots that there is close similarity in the shape of the curves between chains A and B. There is also good agreement between all four chains for some areas of the chain *i.e.* regions of high and low B–factors correspond well between A and B, with some areas correlating in all four.
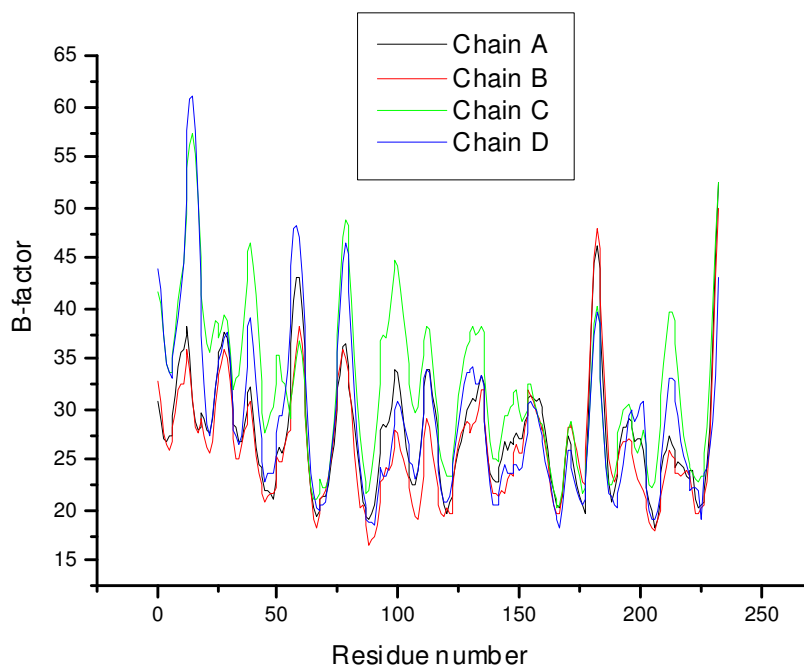
Figure 71.   B-factor versus residue number graph for chains A, B, C and D of the
                   PNA/T-antigen complex tetramer.

As with other legume lectins, the framework of the molecule consists of three *β*-sheets, a six-stranded flat sheet (residues 2 to 8, 64 to 70, 162 to 168, 173 to 179, 186 to 192 and 222 to 229) at the back of the sub-unit (sheet 1), a seven-stranded curved sheet (18 to 23, 42 to 49, 84 to 90, 117 to 124, 136 to 143, 149 to 153 and 203 to 210) towards the front of the sub-unit (sheet 2) and a small five-stranded sheet (25 to 27, 31 to 34, 71 to 74, 160 to 161 and 217 to 220), sheet 3, which has a major role

in holding the two larger sheets together[58]. Loops make up 54% of the structure. If we examine the areas with high B-factor they correspond well to loop areas in between the β-sheet sections. For both chains A and B the C-terminus (residues 230-232) exhibits the highest B-factors. The density at the C-terminus is very weak and so the atoms cannot be modelled easily. The loop region 180 to 185 in chains A and B has residues with high B-factors also, averaging around 45 Å^2. The loop region including residues 50 to 63 also exhibits high B-values in chains A and B, mainly residues 57-59. These loops have weak density and are more disordered than the stable secondary structures such as the β-sheet.
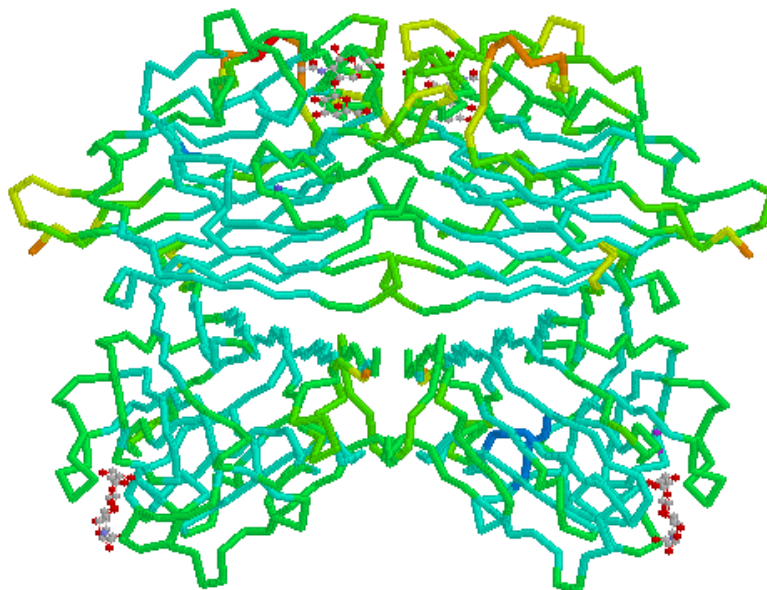
Figure 72.   The protein chain backbone of the PNA tetramer, coloured according to temperature factor of the residues. The colour coding is as follows; blue↔v. low B-factor, green↔low B-factor, yellow↔medium B-factor, orange↔high B-factor, red↔v. high B-factor. The T-antigen molecules are also shown coloured according to the Corey, Pauling and Koltun (CPK) colour scheme, where carbon is grey, oxygen is red and nitrogen is blue.

The region of highest mobility in chains C and D is for residues 9 to 17, which again, is an interconnecting loop region. This loop region is one of the most flexible areas of the whole molecule as it does not contain many stabilising hydrogen bonds. The C-terminus of chains C and D also exhibit high B-factors as in chains A and B. Yet another region with high B-factors in chains C and D is the loop region involving residues 75 to 83. Overall, it can be seen from the plots that chains A and B, in general have lower B-factors than chains C and D, and with certain areas of the chain that are disordered/ordered in all four. Chain C has the highest B-factors of all the chains. The average B-factors for each chain are given below (Banerjee *et al.,* (1996)[58] values in parentheses);

Average B-factor ($\mathring{A}^2$) for ;	Chain A	Chain B	Chain C	Chain D	Overall
Main-chain	27.8 (19.5)	26.4 (18.5)	32.4 (22.9)	29.5 (21.0)	29.0 (20.5)
Side-chain	28.9 (20.9)	27.6 (19.7)	33.3 (24.1)	30.6 (22.6)	30.1 (21.8)
Whole chain	28.3 (20.2)	26.9 (19.1)	32.9 (23.5)	30.0 (21.8)	29.5 (21.1)

Table 20.   The average B-factors for main-chain atoms, side-chain atoms and the whole chain for each sub-unit (Banerjee *et al.,* (1996)[58] values in parentheses).

As expected, the average values for all the side-chains of each sub-unit are higher than those for the main-chain atoms, due to the fact that, the side-chains are more flexible with less constraints on their conformation. Main-chain atoms are involved in many more stabilising hydrogen bonds than side-chains and therefore, show less mobility. The same trend is seen again in both sets of data whereby chain B has the lowest B-factors and chain C has the highest.

3.11 Water structure analysis

The final refined model contains 441 water molecule oxygen atoms. These water oxygen atoms are all present in the final $2F_o$ - F_c electron density maps at greater than 1.5 r.m.s. The average B-factor of all the water oxygen atoms is 38.2 Å^2. This value compares well with that of the PNA/lactose study at 2.25 Å where the 563 waters identified had an average B-factor of 36.8 Å^2.
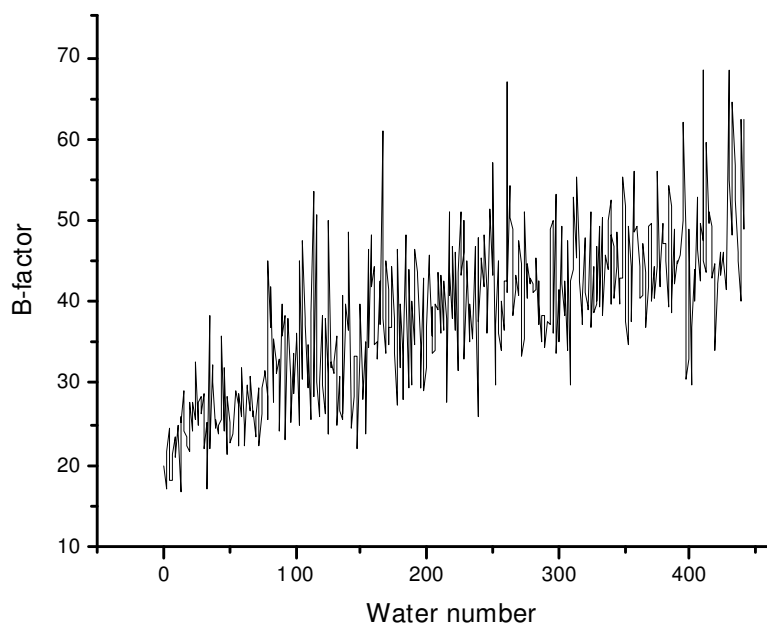
Figure 73.   The B-factor versus water residue number plot for the 441 water oxygen atoms identified in the refined model.

An analysis of the position of these waters within the tetramer was carried out. Of the 441, 120 are seen to be completely buried within the protein while 321 are surface waters. As would be expected the average B-factor of the buried waters is considerably less than those on the surface (averageB$_{120\text{-buried}}$ = 31.3 Å^2 and averageB$_{321\text{-surface}}$ = 40.7 Å^2). The buried waters achieve many more hydrogen bonds to the protein and so exhibit greater stability.

It is of interest to compare the hydration shells of the four chemically equivalent, but crystallographically independent, sub-units in the tetramer in the asymmetric unit. In particular it is instructive to examine the extent to which the hydration shells of individual sub-units reflect the molecular and local dyads that relate the sub-units themselves. The number of water molecules in the hydration shell of chains A, B, C and D are 110, 131, 92, and 108 respectively.

The average B-factor of the waters in each chain were calculated, the values are as follows;

Chain A→110 waters, average B-factor = 37.5 Å^2

Chain B→131 waters, average B-factor = 36.5 Å^2

Chain C→ 92 waters, average B-factor = 41.2 Å^2

Chain D→108 waters, average B-factor = 38.3 Å^2

The 441 waters were also analysed in terms of the number of bonds made from each water molecule to protein atoms. In the table below it can be seen that 120 waters interact only with either other waters and/or sugar atoms. As is expected these waters exhibit the highest B-factors as they are more mobile due to having less stabilising interactions.

Number of bonds per water oxygen	Number of waters	Average B-factor (Å^2)
0	120	41.9
1	196	39.6
2	100	33.5
3	23	27.4
4	2	32.3

Table 21.   Table to show the number of bonds made from each water molecule to protein atoms and their average B-factors.

The total number of water contacts was also calculated overall and for each chain, with the values given in the table below:

Chain	Total number of water contacts
A	119
B	130
C	113
D	126
ALL	488

Table 22.   The total number of water contacts calculated overall and for each chain in the PNA tetramer.

The waters in Chain B are seen to make the most water contacts of any of the chains- this is unsurprising as chain B contains the most water oxygen positions.

By superimposing the starting model (2PEL) with the refined model using LSQKAB, an analysis of the number of waters that correspond *i.e.* are equivalent, between the two structures was achieved.   Waters were seen to be equivalent if after superposition of the two sets of co-ordinates the waters are within 1 Å of each other. Of the 441 water oxygen positions, 319 are within 1 Å of the 563 waters in the starting model.   The average B-factor of these 319 waters is 35.6 Å^2, with the average B-factor of the 319 that agree in the starting model (2PEL) being 31.2 Å^2. Of the 120 buried waters in the final refined model, 111 of these match the positions of the water oxygen atoms in the starting model.   Of the 321 surface waters in the final refined model, 208 of these match the positions of the water oxygen atoms in the starting model.   This illustrates that a greater percentage of the buried waters agree with the starting model than the surface waters (92.5% buried agree *cf.* 64.8% surface agree) as would be expected due to the increased mobility in general of surface waters which make less interactions with the protein.

The variation in the number of water molecules in the four hydration shells itself indicates their substantial non-equivalence, however, a more rigorous analysis is necessary to quantify the extent of the variation. To determine the number of invariant water molecules between different sub-units, a water molecule in the hydration shell of one sub-unit and one in the hydration shell of another were considered invariant if they interact with at least one common protein atom and if the distance between the two is less than 1 Å when the two sub-units along with their hydration shells are superimposed on each other. When this criterion is used the number of invariant water molecules between pairs of sub-units were as follows:

Pair of sub-units	Number of invariant water oxygen atoms
A-B	83
B-C	63
B-D	66
C-A	59
C-D	59
A-D	66
ALL FOUR	47

Table 23. The number of invariant water molecules between pairs of sub-units.

As can be seen from the table above the greatest number of invariant water positions between sub-units occurs between chains A and B. This is probably due to the large number of water bridges (19) that connect the two sub-units. The number of invariant water molecules with respect to all four sub-units is 47. These include the six water oxygen positions identified by Loris *et al.,* (1994)[140] to exist in the structures of all legume lectins and their complexes. Four of them are involved in metal co-ordination, two of which also interact with protein atoms that hydrogen bond to the T-antigen molecule. Of the remaining two, one connects the 144 to 148 loop and the C-terminus of the 136 to 143 strand, while the other stabilises the *β*-bulge contiguous to sheet 2 between loops 50 to 63 and 193 to 202. Approximately half (22/47) of the invariant waters are found to be buried, this at first

may seem to be a rather low number, but when we examine the positions of the invariant waters, many of them are found by loop sections and make hydrogen bonds to residues within these loop sections. Therefore, it seems probable that the function of these invariant surface waters is to ensure the correct structural relations among the loops and with the rest of the molecule.

The average B-factors of the 47 invariant water oxygen atoms in each chain was also calculated as well as the B-factor of all 188 invariant water oxygen atoms. They are given below:

CHAIN	B-factor of invariant water oxygen atoms(Å^2)
A	31.3
B	29.5
C	36.9
D	33.3
ALL CHAINS	32.7

Table 24.   The average B-factors of the 47 invariant water oxygens in each chain and for all four chains.

The 47 invariant water oxygen positions in each chain were then compared against both the water positions in the PNA/lactose study at 2.25 Å and those in the PNA/T-antigen study at 2.5 Å. Overall 163 of the 188 (47 x 4) invariant water positions match within 1 Å of the water positions in the PNA/lactose study at 2.25 Å and 123/188 agree within 1 Å of the water positions in the PNA/T-antigen study at 2.5 Å. It is clear from these results that there is better agreement between water oxygen positions between the two higher resolution structures *i.e.* the PNA/lactose study at 2.25 Å and the PNA/T-antigen study at 2.25 Å presented here.

CHAIN	2PEL model	2TEP model
A	43/47	33/47
B	41/47	32/47
C	39/47	28/47
D	40/47	30/47

Table 25.   Gives the number of water positions that agree within 1 Å of the 47 invariant water positions identified in each monomer of the tetramer between the refined PNA/T-antigen model and the models 2PEL and 2TEP respectively.

The position of the 47 invariant water oxygen atoms were analysed. Of the 47, 21 are seen to make hydrogen bonds with atoms only in loop sections. Four of the waters are found within the saccharide binding site, being those which create the water bridges between the T-antigen and the protein. Four of the invariant waters are the water molecules linked to the metals, two to each metal. Five waters are associated solely with sheet 1 and three with sheet 2. Six waters are found to make hydrogen bonds to both loop sections and sheet 2, one water interacts with sheet 1 and loop sections and two waters interact with sheet 3 and loop sections. One of the invariant waters makes only links to other waters.

Therefore, overall it can be seen that the majority of the invariant waters are involved in stabilising the loop sections. As the β-sheet regions are stable secondary motifs with main-chain to main-chain hydrogen bonds occurring between chains these areas do not require the extra stabilization given by the water-mediated hydrogen bonds. On the other hand the loop sections have much greater mobility and less protein-protein interactions, therefore, it is unsurprising that interactions involving water molecules are utilized in order to stabilize their overall conformations, especially when these loop sections are a crucial part of the binding site region as in legume lectins.   Further analysis was done to see which loops are stabilized by these waters

and which sections are not. When the number of hydrogen bonds from the invariant waters to each loop section are tallied, those that are found to have the most interactions with the invariant waters are the loop sections listed below;

Loop section	Number of H-bonds from the invariant waters to the loop
91-116	13
35-41	9
75-83	7
144-148	7
211-216	7
125-135	6
154-159	6

Table 26.   Gives the number of hydrogen bonds from the invariant waters to each loop section.

As can be seen from the list above, the loop sections which have the most interactions with the invariant waters are those that make up the binding pocket in PNA. There are 21/47 waters that interact with at least one of the four binding region loops. All 21 of these waters are present within the 2.25 Å structure of the PNA/lactose complex. Seventeen of these 21 waters are also present in the 2.5 Å structure of the PNA/T-antigen complex.

3.12 Alternate conformations

Altogether a total of 13 alternate conformations were modelled for the tetramer. The number of alternate conformations modelled for each chain being 3, 4, 3, and 3 for chains A, B, C and D respectively. In all four sub-units the residues threonine 25 and serine 68 exhibit large F_o - F_c peaks at their side-chains indicative of multiple occupancies. Both of these amino acid residues contain a hydroxyl group which is free to rotate and therefore, these residues are commonly identified in more than one conformation.

The other residues identified with multiple occupancies are threonine 111 and valine 148. The multiple occupancies appear to be identified at residues where the alternate conformation is able to form a hydrogen bond to a proximal water molecule, this is illustrated below in **Figure 74**. All of the residues that are identified with alternate conformations are found exclusively on the surface of the protein, and are well distributed around the surface. Therefore, there is no evidence in PNA of cooperation between the alternate conformations. It is unsurprising that the residues within the core of the protein do not exhibit differing conformations as these residues are generally involved in β-sheet secondary structures and so are more likely to be locked in a particular conformation. Given here are all the multiple occupancies identified.

Chain	Alternate conformations
A	Thr 25, Ser 68, Val 148
B	Thr 25, Ser 68, Thr 111, Val 148
C	Thr 25, Ser 68, Val 148
D	Thr 25, Ser 68, Thr 111

Table 27.   Gives the number of multiple occupancies modelled for each of the four chains of the PNA tetramer.
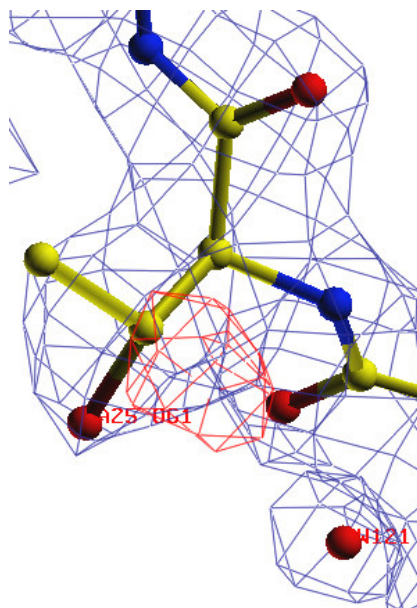
Figure 74.   The two alternate conformations modelled for threonine 25 of chain A. The $2F_o - F_c$ electron density is at 2 r.m.s. in blue, and the $F_o - F_c$ density is shown in red at 3σ. As can be seen the alternate conformation is stabilized by forming a hydrogen bond with a water molecule.
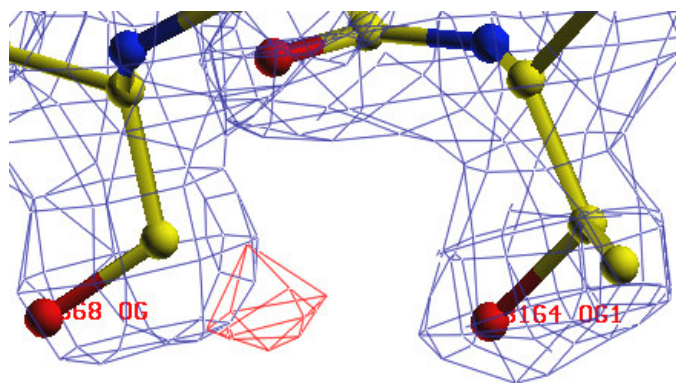
Figure 75.   The $2F_o - F_c/F_o - F_c$ maps indicate the two alternate conformations for Ser 68 of chain B.

3.13 Concluding remarks

The three-dimensional structure of the complex between PNA and T-antigen at room temperature has been determined to 2.25 Å resolution employing synchrotron radiation. This represents an improvement from 2.5Å resolution for the previous structure of this complex, which was undertaken by use of a CuK$_{\alpha}$ rotating anode X-ray source[143]. Hence, the 2.25 Å resolution structure of the PNA/T-antigen complex now determined, provides us first of all with a higher resolution structure of the complex and furthermore, increases the ensemble of structures of PNA. This has enabled us to confirm the results of the lower resolution study, in that the enhanced affinity for T-antigen over lactose is generated by three extra water mediated interactions, involving two waters. Furthermore, from analysis of the water structure in general, it has been shown how important the bound waters are for the effective functioning of the protein. The waters are involved in binding the sugar directly through water mediated hydrogen bonds and indirectly by stabilising the metal ions, the *cis*-peptide and the loops of the saccharide binding site. They are also involved in subunit association, in which waters are used to help tether monomers together correctly. This study then has increased the knowledge of an important protein widely useful as a diagnostic tool, and in more general terms has provided specific details on the bound water structure of proteins.

Chapter 4

Flash-cooling large protein-crystals

4.1   Introduction

Although many structures can be solved from small crystals with X-ray synchrotron radiation, it is far easier to solve and refine structures when strong data is recorded from large crystals. Better data will be recorded from a large crystal because the standard deviation of the data is controlled by counting statistics that are related to the square root of the total number of counts in the reflection and its background. Thus, a larger crystal scatters more X-rays relative to its background and generally has lower sigma values. This can be important at the early stages of structural determination, since the sigma values control the quality of the small differences inherent in multiple wavelength measurements that are now used for most phasing experiments[284]. Large crystals are also preferred for structure elucidation of viruses due to the larger molecular size of viruses compared to proteins and for weakly diffracting proteins with high solvent content. Furthermore, due to the inherent low flux of neutron sources, large crystals are at present essential for high-resolution neutron crystallographic studies.

It is well accepted however, that the bottleneck of any crystal structure determination is growing crystals of any size[285,286]. However, if large crystals can be grown, in order to harness the many advantages gained from cryocrystallography, it is essential that efficient cooling of the crystals is achieved. It has been found however, that successful flash-cooling of large crystals (by large crystals we mean volumes of $\sim 1mm^3$) is more difficult than with smaller crystals. Flash-cooling of macromolecular crystals has been found to extend the lifetime of the crystals in intense X-ray beams. Cooling the crystal, generally to around 100K, reduces the X-ray radiation damage to the crystal. Many studies have investigated the effects of radiation damage on macromolecular crystals in both real and reciprocal-space[287-299]. Flash-cooling the crystal can provide the opportunity to potentially collect a complete data set from a single crystal, thereby eradicating problems of non-isomorphism and saving valuable crystals. By collecting the data at cryo-temperatures thermal disorder can be reduced, manifested by reduction in B-values and enhanced density definition. Furthermore, it can also reduce background

scattering (diffuse scattering), and therefore, can aid the signal-to-noise ratio and hence, the resolution limit. Cryocrystallography also opens up the possibility of freeze trapping of intermediates in biological reactions[300,301] and on a practical note it simplifies sample handling in the experiment.

There are however, several problems that can occur by cooling the crystal. Cooling can create disorder, which reduces data quality, degrades the resolution limit and therefore, reduces accuracy of electron density maps. In extreme cases, cooling the crystal can cause cracking and dislocations, making data collection impossible. A simple description of the imperfections of protein crystals was given by Nave[302] following on from the work of Helliwell[303]. The imperfections in a crystal can be described by a simple mosaic-block model that describes three characteristic parameters. These parameters are the size s of the mosaic blocks, the angular spread ω of the blocks and the variation in cell dimensions δa between blocks. A common effect observed by cooling macromolecular crystals is an increase in mosaicity (lattice orientations). Small crystals at room temperature generally have very small mosaicities ($<0.02°$), with this value being increased by the characteristics of the X-ray beam *i.e.* the incident beam divergence[304,305]. When cooled, protein crystals generally have mosaicities of $0.2°$ or more. As large crystals in general have larger mosaic spreads at room temperature, this effect is amplified on cooling, so that very large mosaicities are observed.

4.2   Experimental

When using neutron data to refine a crystal structure, it is common to use an X-ray structure as the starting model for the refinement. Ideally the X-ray structure should also be derived from the same crystal as the neutron data, in order to eliminate (or at least reduce) problems with non-isomorphism. Using the same crystal for both neutron and X-ray data collections is possible due to the fact that neutrons are non-perturbing and do not cause any radiation damage to the crystal. Therefore, after successfully collecting neutron diffraction data from a large concanavalin A crystal (1.8 x 1.3 x 0.7mm) at 12K to 2.5 Å resolution (see **Chapter 5** for details), we then collected an X-ray data set at 100K on the same crystal to 1.65 Å (see **Chapter 5** for details) in order to accurately determine the positions of the heavy atoms.

The original flash-cooling of the large concanavalin A crystal was achieved as follows; first the crystal was soaked in a 50:50 (v/v) solution of the D_2O mother liquor and perdeuterated 2-methyl-2,4-pentanediol (MPD) for 2 hours, then the crystal (mounted in a loop) was plunged directly into liquid nitrogen (77K). After the neutron data collection on LADI at the ILL, the crystal was kept at liquid nitrogen temperatures (77K) and immediately mounted on a goniometer of a rotating anode X-ray diffractometer, where it was directly placed in a cryostream, maintained at 100K. Prior to collecting the X-ray data set, a step-scan across the crystal was performed in order to collect the diffraction data from the best possible position of the crystal. This step-scan produced extremely different results in terms of the quality of diffraction observed at different positions along the crystal. The area of the crystal which displayed the strongest diffraction was identified and an X-ray data set to 1.65Å resolution was collected and used as the starting model for the neutron refinement. However, due to the remarkable differences in diffraction observed from the step-scan, this led us to investigate this further. Each image was indexed and integrated to investigate the fall-off in resolution limit and the changes in mosaic spread at the different positions. The orientation and exposure time was kept constant for each image.

4.3   Results and discussion

Below are shown the diffraction images observed from the step-scan. The images were taken at $\Phi = 252°$ (arbitrary), with $\Delta\Phi = 0.05°$ for an exposure time of 300 seconds, on a rotating anode X-ray diffractometer fitted with a Mar detector. Image number 1 was taken at a position nearest to the pin of the loop, and then each successive image was taken at points gradually further away from the pin by translating the crystal position (see **Figure 77**). The distance between successive images was ~0.25mm.
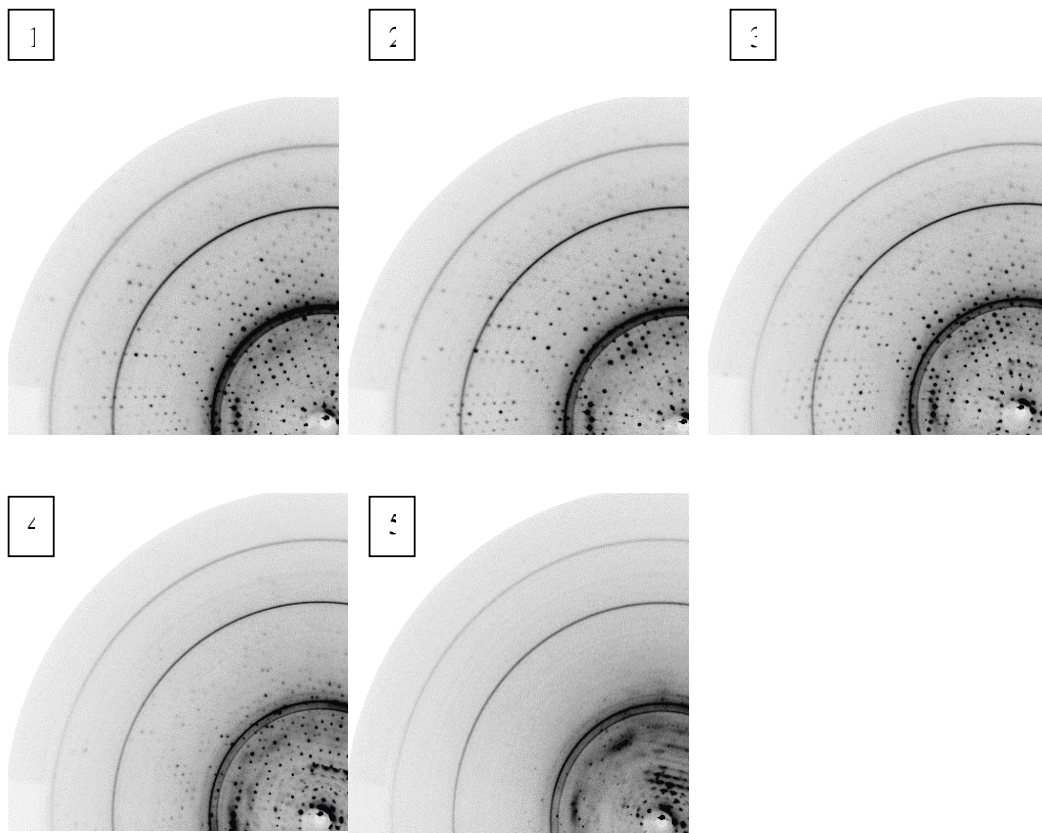
Figure 76.   Quadrants of the five diffraction patterns produced from the step-scan across the large concanavalin A crystal. Image numbers 1 (nearest to the pin) to 5 (furthest from the pin) are given.

From visual inspection of the diffraction patterns we can see that a dramatic reduction in diffraction was observed on moving to the end of the crystal furthest from the pin (images 4 and 5). In fact, the last image (image 5) exhibits such extreme loss of diffraction that this image could not be indexed. Furthermore, from examination of the outer spots on the patterns it appears that the two first images (those nearest to the pin) are the best, with the third image slightly worse. Therefore, there appears to be a progressive loss in diffraction as we move away from the pin.

By indexing and integrating the first four images (1-4), we can illustrate semi-quantitatively the difference in diffraction quality. A plot of the I/sigma(I) values versus resolution illustrates the differences observed (**Figure 78**).
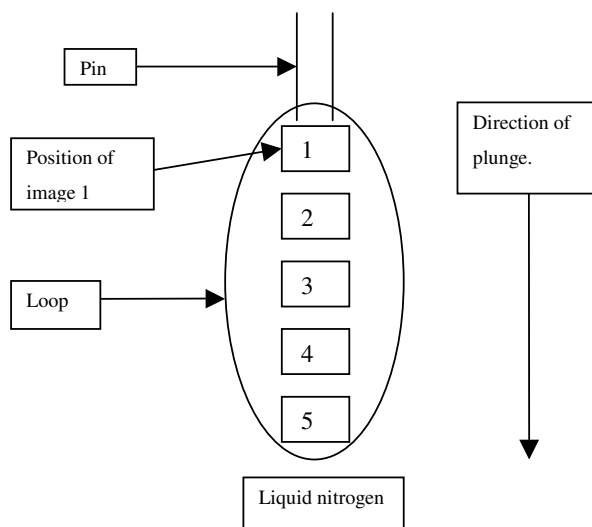
Figure 77.   Schematic of the crystal in the loop, with the positions at which the patterns were taken.

The plot of the I/sigma(I) values does not give completely consistent results to the visual inspection, as it indicates that image 2 has the best overall I/sigma(I) values, however, the values calculated can only be taken as semi-quantitative as they are each derived from only one image and so may include large errors. Nevertheless, the plot does indicate that the images that were taken nearest to the pin (1-3) display the best diffraction. In this way, both the analyses are consistent.
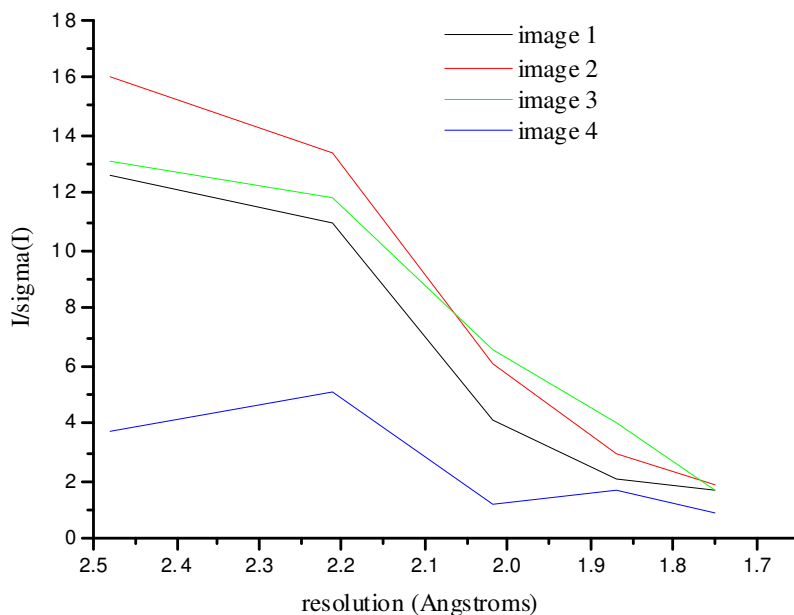
Figure 78.   Plot of I/sigma(I) versus resolution for each of the four indexed images.

The large concanavalin A crystal used for this work was plunged into the liquid nitrogen with the end furthest from the pin first, so that the last part of the crystal to be immersed in the liquid nitrogen was the area closest to the pin (as shown in **Figure 77**). Considering the effects of cooling on the solvent in the crystal, we would expect the area of the crystal which is cooled at the fastest rate to be the least damaged *i.e.* the end furthest from the pin. This is because in order to cool successfully without harming the crystal we need to avoid ice formation within the crystal by rapidly vitrifying the solvent to an amorphous phase. However, from the results, it appears that the area cooled fastest exhibits the greatest degree of disorder. It is not clear why this is. One possibility is that the high turbulence at the end plunged first has damaged the crystal whereas at the other end it is less turbulent, and so the diffraction is better. There are many factors involved in the success of cryo-cooling protein crystals, such as the size of the crystal, the cooling rate, the cooling medium, solvent content, solvent composition (cryoprotectants and buffers) and the protein itself. It is clear from the results of the X-ray step-scan that in order to

successfully cryo-cool large protein crystals consistently well, further investigations are essential to shed more light on the processes that dominate the cooling.

At LADI, the first really large, neutron sized protein crystal to be cryo-cooled successfully was this concanavalin A crystal. This was a breakthrough. Consequently, at LADI, a large rubredoxin crystal was frozen. These are the largest protein crystals frozen to date to my/our knowledge.

Chapter 5

Saccharide-free concanavalin A

5.1 Summary

The neutron crystal structure of the saccharide-free form of concanavalin A and its bound water at 12K have been determined to 2.5 Å resolution. The final model is in fact based on an X-ray refinement of the non-deuterium atoms to 1.65 Å resolution using an identically prepared crystal, data measured at 100K, and with an R-factor of 19.5% for 26234 unique reflections and an R-free value of 22.0%. The final neutron data based model, including deuterium atoms with the heavier atoms held fixed, still has good geometry, an R-factor of 25.1% for 6462 unique reflections and an R-free value of 32.7% for the neutron data.

The bound water structure has been determined, with 167 waters modelled around the monomer as D_2O and 60 waters modelled as oxygen only. Overall, there are more water molecules identified as D_2O at 12K (167) than at room temperature (88). Moreover, the water molecules at 12K on average have lower B-factors than those at room temperature, 17.6 Å^2 and 37.8 Å^2 respectively.

29 water oxygen atom positions (24 D_2O, 5 Oxygen) were found to agree within 1 Å between the 12K and room temperature[51] neutron studies, 25 (22 D_2O, 3 Oxygen) of these also agree within 1 Å of the 0.94 Å ultra-high resolution X-ray water oxygen atom positions[50]. The water molecules that exhibit the best agreement in position and orientation between the two neutron structures and the ultra-high resolution X-ray structure, are those that play some functional role for the protein; either in stabilizing the overall structure (*e.g. β*-bulge water, *β*-hairpin water) or as an integral part of the saccharide-binding mechanism (*e.g.* saccharide binding site waters and metal-coordinated waters). As expected, the majority of these waters exhibit low B-factors and good 'boomerang' type neutron density.

5 water molecules found within the saccharide binding site were identified and the position of their D-atoms determined. The improved definition of the nuclear density at the saccharide binding site at 12K compared to room temperature enabled a

hydrogen bonding description to be suggested between key amino acid residues and the water molecules.

In contrast to the saccharide binding site, increased disorder at the metal sites was observed at 12K compared to the room temperature data. This observation agrees with current results from EPR experiments in which the local symmetry at the Mn^{2+} site deviates from the crystal symmetry at low temperatures (4K, 120K and 135K)[258,307,308].

5.2   Saccharide-free concanavalin A

Concanavalin A associates as a dimer of dimers to form a tetramer of 100 kDa total molecular weight. In the saccharide-free crystal form, space group *I*222, a single monomer of 25 kDa constitutes the asymmetric unit. The concanavalin A monomer contains 237 amino acid residues, and binds a Ca^{2+} ion and a transition-metal ion (usually a Mn^{2+} ion) which are both necessary for saccharide binding[61,145,259]. The transition-metal binding site (*S*1) has been shown to have affinity for Ni, Co, Zn, Mn, or Cd and the calcium-binding site (*S*2) was found to be specific for Ca or Cd ions[260,261]. As isolated, the native protein contains a mixture of metals at the transition-metal site, predominantly manganese.
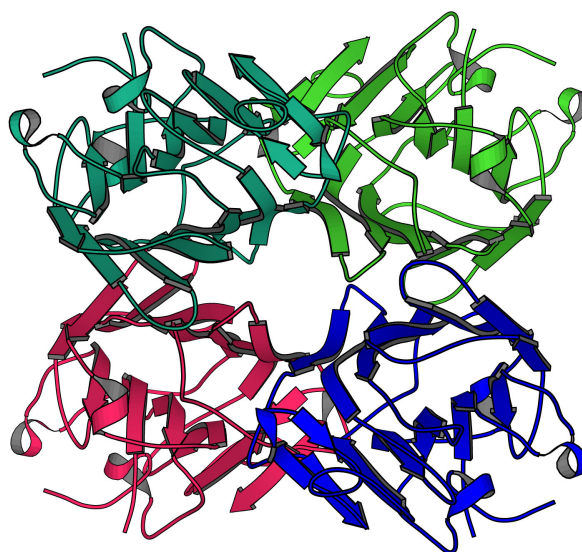
Figure 79.   The biological unit in concanavalin A is a classical *I*222 tetramer. The asymmetric unit contains a single monomer, each of which are shown here in different colours.

The basic X-ray crystal structure of the sugar-free form of the protein was determined[61,259] in the 1970's and more recently X-ray crystal structures of the protein and its complexes with sugars have revealed the intricacies of many aspects of the protein structure, conformational changes and ligand recognition. The X-ray

studies culminated with an ultra-high resolution crystal structure of the sugar-free form at 0.94 Å resolution, with data collected at 110K[50]. This X-ray structure has the precision and accuracy equivalent to those of small molecule crystal structures. The high precision of the atomic coordinates and the direct observation of certain hydrogen atom positions, on both the protein and the bound solvent, provided a new insight into the protein structure. However, despite such high resolution, only 12 water molecules with both hydrogen atoms visible *i.e.* as H-O-H, were identified.
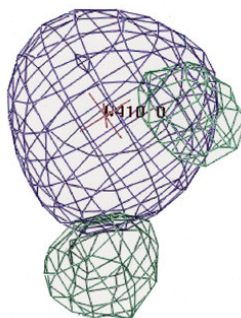
Figure 80.   A water molecule with both the hydrogen atoms observed directly, from the UHR X-ray structure of saccharide-free concanavalin A[50].
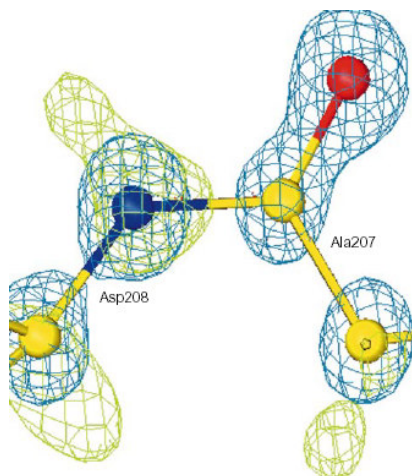
Figure 81.   The *cis*-peptide between Ala 207 and Asp 208 in concanavalin A with the position of the amide hydrogen atom clearly visible[50]. The $2F_o$ - F_c map is shown in blue (at 3 r.m.s.), and the $F_o – F_c$ map in yellow-green (at 1.75 σ).

5.2.1 Neutron studies of the saccharide-free form

Using the neutron quasi-Laue method, Habash *et al.,* (1997)[52], studied the saccharide-free form of concanavalin A at room temperature and obtained data to 2.75 Å using LADI, ILL. The unit cell for saccharide-free concanavalin A was reasonably large ($a \approx 89$ Å, $b \approx 86$ Å, $c \approx 62$ Å) compared to high-resolution protein neutron diffraction studies undertaken previously, yet due to the recent improvements in data acquisition times, the data collection was only 11 days. Partial H/D exchange was attempted by vapour diffusion. However, the bound solvent was found to be invisible, as would be the case if it were still H_2O rather than D_2O, where at 2.75 Å resolution the two hydrogen atoms would effectively cancel the positive nuclear density for the oxygen.

By utilizing a more effective D_2O/H_2O exchange method, involving liquid dialysis over a 4 months period, Habash *et al.,* (2000)[51], improved on their previous study and determined the structure of the deuterated, saccharide-free form of concanavalin A to 2.4 Å at room temperature. The higher resolution is itself indicative of reduced hydrogen content since the incoherent background scattering is thereby reduced. This neutron Laue study showed the nature of the metal ligand regions of concanavalin A and revealed directly that the four bound waters' D-atoms are indeed pointing away from the Mn^{2+} and Ca^{2+} ions. The study also revealed the positions and orientations of the bound water D-atoms at the saccharide binding site although the clarity of the nuclear density at the individual waters was poor *i.e.* not revealing them singly but rather as a cluster. If the definition of the nuclear density for these water molecules is improved by lowering the temperature to 12K so that individual water molecules are observed, more confidence can be placed on the refined D-atom positions. Such details one can imagine are important for accurate molecular modelling and dynamics calculations to provide a new level of detail for the comprehension of molecular recognition processes.
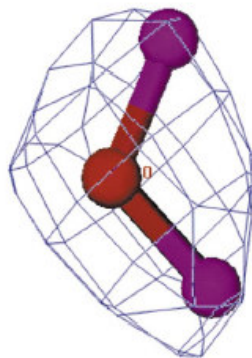
Figure 82.   A water molecule in the room temperature neutron study[51] showing the orientation of the D_2O nuclear density. The $2F_o - F_c$ map is shown (at 1.5 r.m.s.).

Comparing results from the two previous room temperature neutron studies[51,52] it is clear that the bound solvent in the 2.4 Å neutron structure is much richer in detail[51]. Whereas before the bound water molecules had essentially zero density at 2.75 Å resolution[52], at 2.4 Å resolution, 88 out of 148 water molecules were assigned as D_2O. The extensive direct soaking of the crystal in D_2O yielded the D-atom positions of the bound water molecules. The reduction of the number of hydrogen atoms *via* this D_2O soaking method also enhanced the neutron diffraction resolution limit, presumably *via* a reduction in the hydrogen background scattering. Habash *et al.,* (2000) concluded that neutron Laue diffraction of the bound water D_2O molecules, even at medium resolution is seen to be more attractive than ultra-high SR X-ray crystallography in assigning many more co-ordinates of the D-atom positions of the bound D_2O. Some 62 of the 88 D_2O molecules that were revealed (visible) at room temperature in the neutron Laue diffraction case[51] where found to agree within 1 Å of the water oxygen atom positions identified in the ultra-high SR X-ray crystallography case[50]. These water positions that agree can be seen as common waters between the two structures and so it is clear that the neutron data is favourable to the ultra-high SR X-ray crystallography case[50] in which only 12 H_2O molecules were revealed, even at the lower temperature of 110K. However, it must be pointed out that the water molecules that are visible in the ultra-high SR X-ray crystallography study are well determined *i.e.* have accurately defined oxygen atom

positions, and so these can be confidently combined with those from the neutron Laue diffraction study. Thus, the complementarity of synchrotron radiation X-ray protein crystallography and neutron Laue protein crystallography is ideally suited for the direct observation of hydrogen atoms in protein structure and function studies and in the detailed study of the bound solvent (as D_2O).
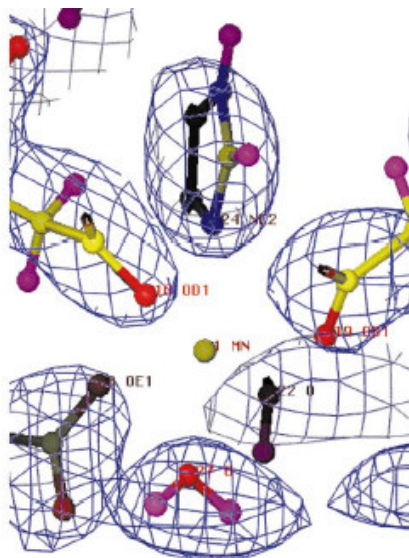
Figure 83. The $2F_o - F_c$ neutron density map at room temperature around the Mn^{2+} coordination sphere of ligands[51].
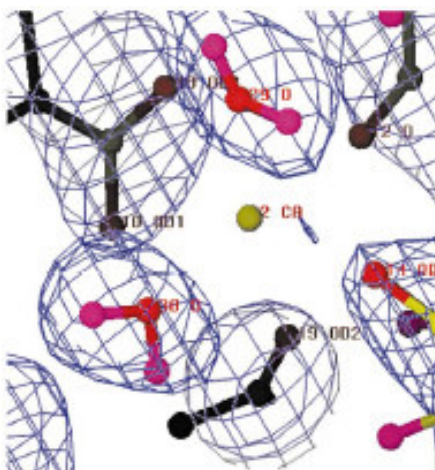
Figure 84. The $2F_o - F_c$ neutron density map at room temperature around the Ca^{2+} coordination sphere of ligands[51].

To continue from this previous neutron study at room temperature, we have performed a neutron quasi-Laue study at cryo-temperature (12K) on the saccharide-free form of concanavalin A. This neutron diffraction study at 12K aims to examine the bound water deuterium atom definition in the neutron maps, at low temperature versus the existing results at room temperature. In particular, the bound waters at the saccharide binding site are displaced by the sugar-ligand on binding and the orientation of these waters to the approach of the ligand is important for accurate thermodynamics and modelling studies. Furthermore, the overall bound water structure at cryotemperature versus room temperature is of interest as it will be one of the first such dual temperature neutron protein crystallographic studies. By comparing the positions of the water molecules identified at both room and cryotemperatures from both X-ray and neutron models of concanavalin A, it is possible to identify those at invariant positions crucial for the efficient functioning of the protein.

The comparison of the B-factors of the common waters at three temperatures (12K neutron at 2.4 Å; 100K UHR X-ray at 0.94 Å; 293K neutron at 2.4 Å) will also reveal insights into their dynamic behaviour. In particular there is a glass transition in proteins at approximately 180K where a sudden change in vibrational amplitudes occurs. This has been studied by neutron inelastic scattering and X-ray crystallography based structures at different temperatures but has not included neutron protein crystal structures either side of the glass transition temperature due to the previously insurmounted hurdle of freezing neutron-sized crystals. The concanavalin A neutron 12K structure can now be compared with the neutron 293K structure, including bound water deuterium atoms, the first such protein structure comparison.

5.2.2 The low-temperature anomaly

EPR studies at room temperature of the Mn^{2+} binding site of single crystals of concanavalin A are completely consistent with site symmetry implied by the space group, *I*222. This has been verified in CW EPR studies on single crystals published more than 25 years ago[263]. In current pulsed EPR studies[258,307,308], however, the local symmetry at the Mn^{2+} site deviates from the crystal symmetry at low temperatures (4K, 120K and 135K), but returns to the expected *I*222 room temperature symmetry when the crystal is warmed to 250K. Nevertheless, at all temperatures studied so far by both X-ray and neutron crystallographic techniques (110K to room temperature) the crystal remains *I*222.

This has led to the hypothesis that there may be two or more Mn^{2+} site structures that average out in the crystallographic experiments at all temperatures but, on the ESR time scale, averaging only occurs at temperatures above 250K but not at 135K and below. Possible candidates for such isomerization are the two water molecules bound to the Mn^{2+} ion. If, for example, a water molecule can rotate about the bisector of the H-O-H bond, the orientation of the unshared-pair electrons may have a profound effect on the electronic properties of the Mn^{2+} ion. Only low temperature neutron diffraction has the potential to examine the orientation of the water molecules by locating the water hydrogen atoms (as deuterium) and thus compare directly with the EPR liquid helium temperature measurements.

5.3   Experimental

5.3.1   Crystallization of saccharide-free concanavalin A

Large crystals of concanavalin A, grown by Dr A. J. Kalb (Gilboa) by batch dialysis as described by Greer *et al*., (1970)[264] were selected for the neutron study. Crystals were grown at room temperature from protein solutions in dialysis bags immersed for several weeks in 0.1 M NaNO$_3$, 0.05 M Tris-acetate, 1 mM MnCl$_2$, 1 mM CaCl$_2$, pH 6.5. In previous neutron studies on concanavalin A[51,52] crystals have been grown from H$_2$O and then soaked in D$_2$O for H/D exchange. A more efficient exchange method has been used in this study by growing the crystals directly from D$_2$O.

Figure 85.   A crystal of saccharide free concanavalin A. The length can be up to 2 mm, and therefore, are suitable for neutron protein crystallogaphy studies, as well as SR protein crystallography.

5.3.2 Neutron data collection, scaling and reduction

Two neutron data sets were collected (each of 17 images) from two separate crystals of dimensions 2.6 x 1.8 x 1.2 mm and 1.8 x 1.3 x 0.7 mm respectively. The crystals were first soaked in a 50:50 (*v/v*) mixture of D_2O mother liquor and perdeuterated 2-methyl-2,4-pentanediol (d-MPD) for 2 hours, before flash-cooling by plunging into liquid N_2 (77K). The use of perdeuterated MPD as the cryoprotectant was used in order to further minimize incoherent scattering from hydrogen atoms. The crystal in the loop was mounted on the cold head of a displex cryostat (pre-cooled to 12K) on LADI. Having already maximized neutron flux, and taken test shots to align the crystal, data were collected on LADI using the neutron cold beam. A narrow bandpass ($\delta\lambda/\lambda = 25\%$) of wavelength range 2.9 to 3.7 Å was obtained using a multi-mirror bandpass filter comprising 40 Si-crystal mirrors each with 748 alternating 74-90 Å thick Ti and Ni layers. The exposure time per image was 24 h; 5 min was needed to scan the image plate and 1 min to erase it. The pixel size was 0.2 x 0.2 mm. For the eight 200 x 200 mm plates, a full, scanned, pattern comprised 4000 x 2000 pixels. The neutron Laue data were processed using the Daresbury Laboratory software (LAUEGEN and LSCALE) developed for synchrotron X-ray Laue data processing[265,266], but modified for neutron Laue data with a cylindrical detector geometry and the polarization correction removed[267]. The orientation of each crystal was determined by auto-indexing using the LAUEGEN[267] program. The unit cell was refined in LAUEGEN to; $a = 89.40$ Å, $b = 86.71$ Å, $c = 62.07$ Å for set 1 and to; $a = 89.61$ Å, $b = 86.58$ Å, $c = 62.03$ Å for set 2. Integration of each Laue reflection was also performed. The program LSCALE[268] was used to derive the wavelength-normalization curve using the intensities of symmetry-equivalent reflections measured at different wavelengths. During this procedure there is usually no need to account for degradation of the crystal during exposure as it will not suffer from radiation damage. Moreover, the absorption effects are fairly small for neutrons even at long wavelengths, and a partial correction is also obtained *via* the scaling of symmetry related reflections. The internal agreement factor from LSCALE for the intensities of reflections with wavelength differences of less than 0.1 Å was 13.9% for set 1 and 14.6% for set 2. The two data sets were seen to be

sufficiently isomorphous (by comparison with SCALEIT[251], R(F) = 26%, weighted R(F) = 15.3%) to combine and merge them together to produce a single data set of 6462 unique reflections to 2.5 Å resolution (76.0% complete with R_{merge} (I) of 0.235 from SCALA[251]).
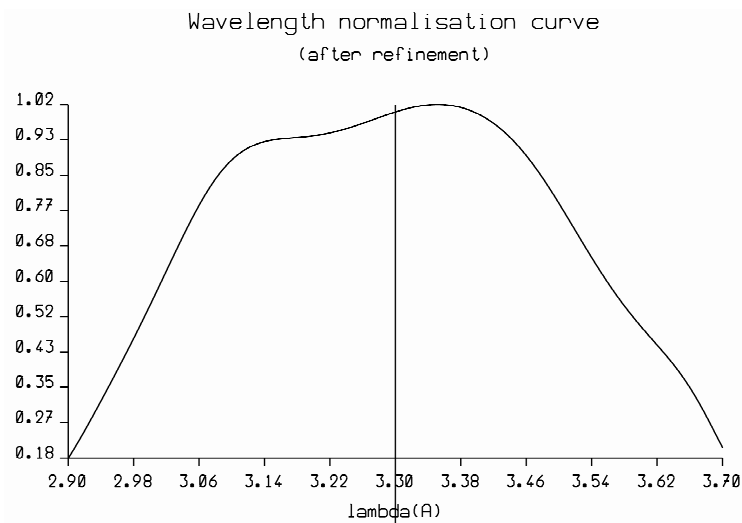
Figure 86.   Wavelength normalization curve for set one of the 12K neutron data, taken from LSCALE[268].
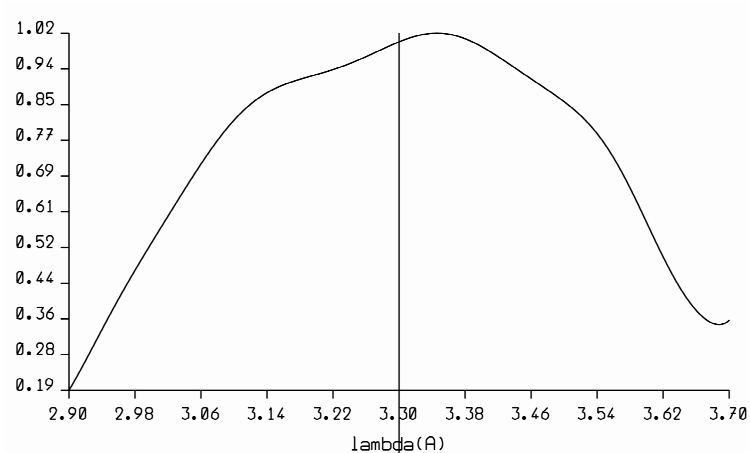
Figure 87.   Wavelength normalization curve for set two of the 12K neutron data, taken from LSCALE[268].

Resolution (Å)	Number of reflections	Completeness % (shell)	I/sigma(I)	R-merge (on Intensity)	<F/sigma(F)>	Mlplcty
7.91	244	82.5	8.4	0.094	15.80	2.2
5.59	464	90.8	9.0	0.120	16.60	3.1
4.56	586	89.7	9.7	0.136	17.88	3.8
3.95	660	86.8	9.2	0.151	17.17	3.9
3.54	699	82.4	8.2	0.152	15.10	3.5
3.23	703	74.7	6.8	0.159	12.78	3.1
2.99	623	62.3	5.8	0.171	10.52	2.9
2.80	530	49.9	4.2	0.156	7.66	2.4
2.64	495	42.5	3.5	0.165	6.42	2.2
2.50	451	37.7	3.0	0.153	5.43	2.0
Overall	5455	64.3	6.9	0.145	12.68	3.0

Table 28.   Summary of set one 2.5 Å resolution concanavalin A neutron data quality.

Resolution (Å)	Number of reflections	Completeness % (shell)	I/sigma(I)	R-merge (on Intensity)	<F/sigma(F)>	Mlplcty
7.91	237	80.5	9.2	0.070	17.35	2.4
5.59	441	87.3	9.2	0.120	17.10	3.1
4.56	546	83.8	9.7	0.147	18.11	3.7
3.95	620	81.8	8.5	0.161	15.71	3.6
3.54	644	76.7	7.0	0.159	12.83	3.1
3.23	612	65.5	5.8	0.151	10.75	2.7
2.99	573	57.3	4.9	0.164	9.01	2.5
2.80	483	45.7	3.8	0.160	7.00	2.2
2.64	464	39.8	3.3	0.148	5.95	2.0
2.50	435	36.2	3.1	0.158	5.71	2.0
Overall	5055	59.8	6.4	0.143	11.85	2.8

Table 29.   Summary of set two concanavalin A neutron data quality.

179
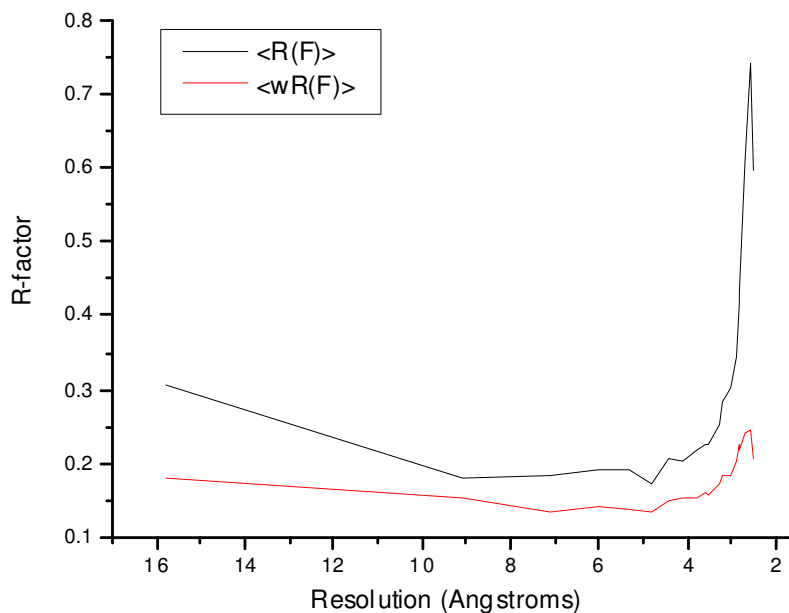
Figure 88.   Plot of the R-factor (on F) against the resolution of the data from
SCALEIT for the combined two neutron data sets at 12K.

As can be seen both wavelength normalization curves are very similar, with a slight
dip in the curve centered around $\lambda = 3.22$ Å. This dip is caused by another machine
positioned upstream at the ILL, which operates on the same beamline (H142) as
LADI. Throughout each data collection the other machine was operating at a
constant wavelength and thus the loss of neutrons due to this other machine was
constant over the data collection period.

Resolution (Å)	Number of reflections	Completeness % (shell)	I/sigma(I)	R-merge (on Intensity)	<F/sigma(F)>	Mlplcty
7.91	283	93.5	7.7	0.143	13.83	4.0
5.59	494	95.6	8.3	0.183	15.20	5.7
4.56	616	93.6	9.0	0.202	16.54	7.0
3.95	709	93.1	8.0	0.241	14.62	6.8
3.54	763	89.5	6.8	0.248	12.47	5.8
3.23	799	84.7	5.5	0.257	10.06	4.7
2.99	752	74.9	4.5	0.289	8.02	4.2
2.80	705	66.3	3.2	0.295	5.69	3.1
2.64	690	59.1	2.7	0.313	4.81	2.8
2.50	651	54.2	2.3	0.358	4.26	2.6
Overall	6462	76.0	5.6	0.235	10.16	4.7

Table 30.   Summary of the combined 2.5 Å resolution concanavalin A neutron data quality.

5.3.3 Sources of error: assessment of the neutron data quality

To recap the neutron final data set is a merged set of two data sets each from one crystal; grown in the same batch. Data runs on LADI were 2 months apart. During this wait time for the second crystal it was simply sitting in its mother liquor. The unit cell parameters for the two crystals are:-

	a	*b*	*c*
neutron 12K crystal 1	89.40	86.71	62.07
neutron 12K crystal 2	89.61	86.58	62.03
X-ray 100K crystal 2	89.16	86.129	61.594

Table 31.   The unit cell parameters for the two crystals used for the neutron and X-ray data collections at 12K and 100K respectively.

The most precise, and indeed accurate set is from the CuK_α crystal 2 determination, since the wavelength is known absolutely. For the LADI sets, and Laue data in general, each reflection is known precisely in theta angle alone and the *d* spacing is derived approximately, albeit quite well. So, one sees that for crystal 2 there is a fixed ratio between the two values for *a*, for *b* and for *c* (1.005, 1.005, 1.007 respectively).   The deviation between the LADI data sets cell parameters is respectively 1.002, 0.999, 0.999 and thus very similar to crystal 1. Thus at worst we can say that the errors of positioning atoms in the unit cell due to this effect would be at about +/- 0.15%, less than due to the resolution limit effect of 2.5 Å (neutron) with 1.65 Å (X-ray) from the curvature of the nuclear/electron density.

Sources of error for the diffraction data structure factor amplitudes lie within the individual LADI data sets and the quality of the process of merging them together. Within crystal 1 LADI data SCALA merge stats are 14.5% overall on I and 15.3% on I in the highest resolution shell. The data were 64.3% complete overall and 37.7% complete in the highest shell. Within crystal 2 LADI data the SCALA merge stats are 14.3% overall on I and 15.8% on I in the highest resolution shell. The data were

59.8% complete overall and 36.2% complete in the highest shell. The two LSCALE determinations of the LADI lambda curve is a check on the stability of the incident neutron beam spectral intensity curve and of the efficacy of the wavelength normalization algorithm/software. They are closely similar curves. The SCALEIT agreement between set 1 and set 2 have <R(F)> of 26%, <wR(F)> of 15.3% overall; <R(F)> on the strongest reflections of 13.6% and <wR(F)> of 9.6%. The lower wR values show that the data is affected by being weak. The stronger band values show that the data can be merged reasonably well although they are values higher than typical for two X-ray data sets from similarly prepared crystals at similar resolution (2.5 Å) *i.e.* where <R(F)> would be ~4%. In the SCALA run bringing set 1 and set 2 together shows similar values <R(I)> on the strongest data of 19.5% *i.e.* ~8.75% on F. A final check on merged data quality is the <F/sig(F)> which is overall 10.2 and outer shell 4.3 (from TRUNCATE). To sum up, the quality of the neutron LADI data from these two merged sets is worse than what one would expect from X-ray data at equivalent resolution but nevertheless reasonable *i.e.* better than the comparison between a native and a heavy atom derivative, and approximately as good as a native compared with an inhibitor soak. For further comparison, two random *i.e.* totally different data sets would have <R(F)> of ~60% *i.e.* much worse than the two neutron LADI sets agreement. Finally, there are structure to structure validation cross checks that are obviously possible *e.g.* 12K 2.5 Å neutron model and 100K 1.65 Å X-ray model-the current work, compared with the UHR 0.94 Å X-ray structure of Deacon *et al.,* (1997)[50] and compared with the 293K 2.4 Å neutron and 1.8 Å X-ray structures of Habash *et al.,* (2000)[51]. These structure to structure comparisons are discussed later in this chapter in **sections 5.4** and **5.5**.

5.3.4 X-ray data collection, scaling and reduction

As a starting model for the neutron structure refinement of concanavalin A and its bound water molecules at 12K, the atomic positions of an X-ray structure at cryo-temperature (100K) were determined. Due to the lack of radiation damage to the crystal in a neutron diffraction experiment it was possible to use the same crystal (maintained at cryo-temperature) for the X-ray data collection also. Therefore, we were able to collect a complete X-ray data set to 1.65 Å at 100K on the second crystal that we had used for the neutron experiment (we were unable to recover and transport the first crystal successfully).

A total of 260 images were collected in intervals of 1° per image. 140 images were collected at a crystal-to-detector distance of 116 mm and 120 images at a crystal-to-detector distance of 82 mm. The data were collected using a rotating-anode X-ray source fitted with a 'Mar30cm' detector, and processed in DENZO[269] and merged in SCALEPACK[269]. The refined unit-cell parameters were $a = 89.16$ Å, $b = 86.129$ Å, $c = 61.594$ Å. 186,877 observations were recorded which reduced to 26234 unique reflections. The 26234 unique reflections to a resolution of 1.65 Å were 91% complete and had an overall I/sigma(I) of 15.9. The mosaicity of the crystal was 1.13°, a high value, but not surprising for a large, cryocooled crystal of dimensions 1.8 x 1.3 x 0.7 mm^3.

Resolution shell (Å)	Number of unique reflections	Linear R-factor (see below for definition)	Square R-factor (see below for definition)	Completeness (%)	Chi^2 value (see below for definition)
3.55	2526	0.058	0.111	82.7	1.25
2.82	2839	0.064	0.099	97.0	1.34
2.46	2807	0.079	0.177	97.5	1.78
2.24	2464	0.086	0.095	85.1	1.74
2.08	2425	0.104	0.145	84.5	1.61
1.96	2745	0.120	0.222	95.5	1.49
1.86	2474	0.174	0.245	86.3	1.47
1.78	2662	0.180	0.163	94.0	1.17
1.71	2676	0.162	0.167	93.5	1.04
1.65	2616	0.206	0.186	92.4	0.92
Overall	26234	0.076	0.116	90.8	1.44

Table 32. Summary of 1.65 Å concanavalin A X-ray data quality.

Linear R-factor = SUM [ABS(I - <I>)] / SUM (I)

Square R-factor = SUM [(I - <I>)^2] / SUM (I^2)

Chi^2 = SUM [(I - <I>)^2] / (Error^2.N/ (N-1))

1.65 Å X-ray data collected at 100K		
Resolution (Å)	Number of reflections	<F/sigma(F)>
5.22	874	56.47
3.69	1533	62.47
3.01	1903	64.19
2.61	2539	71.92
2.33	2819	67.60
2.13	2425	47.91
1.97	3313	46.53
1.84	3336	28.63
1.74	3689	21.90
1.65	3838	13.95
Overall	26269	43.42

Table 33. Summary of 1.65 Å X-ray data quality calculated using the CCP4 program WILSON[251].

5.3.5 X-ray structure refinement

As the starting structure, the co-ordinates for the 0.94 Å saccharide-free structure[50] (PDB code =1NLS) were refined by REFMAC-5[251] against all the X-ray diffraction data, with water molecules removed and the B-factors reset to 30 Å^2. Rigid body refinement was first employed, then positional and isotropic B-factor refinement, after which the R-factor was 26.4% and the R-free was 28.2%. At this stage ARP_WARP[270] was run in order to identify water oxygen atom positions. 212 such waters were found, including those found as ligands to the S1 (Mn^{2+}) and S2 (Ca^{2+}) metal sites, and those that are expelled on binding of saccharide in the saccharide binding site. The R-factor at this stage was 19.4% and R-free was 22%. Using the $F_o - F_c$ and $2F_o - F_c$ electron density maps the waters were analysed and removed/added where necessary. Indeed from the analysis it was clear that some of the water molecules identified by ARP_WARP were in fact carbon and oxygen atoms belonging to a molecule of the cryoprotectant, 2-methyl-2,4-pentanediol (MPD). Therefore, MPD was built into that density and added to the model.

Further analysis indicated that one of the water oxygen atoms identified by ARP_WARP was found to be at the same position (almost identical coordinates) as a third metal site (known as S3) previously located by a Patterson analysis of cadmium-substituted concanavalin A[271]. This site had also been identified[272] as the binding site of several of the heavy metal ions originally used to solve the protein structure, and furthermore, in the structure of the cadmium-substituted concanavalin A structure[273]. The metal and its symmetry equivalent are located in the cleft between the monomers which form the conventional dimer. When this site was modelled as a water oxygen atom, a very large (15 σ) positive $F_o - F_c$ peak was observed at this site and the temperature factor of this 'water' was 2 Å^2 (the minimum allowed). The site was also found to have octahedral co-ordination geometry, suggesting that this is occupied by a third metal (*e.g.* Mn^{2+}) and not a water molecule.
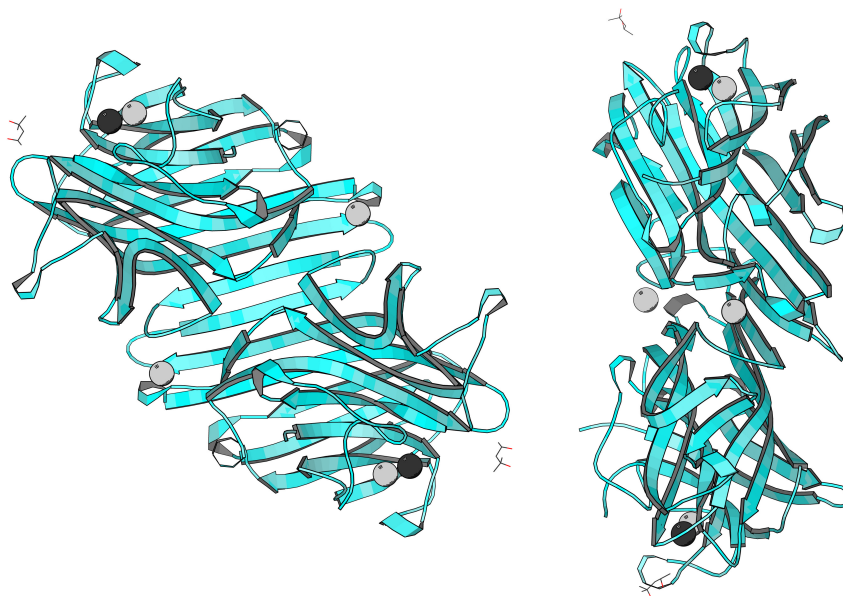
Figure 89.   Two views of the conventional dimer of concanavalin A. The two symmetry related $S3$ site metals are in the cleft (shown as grey spheres) between the two monomers. Also shown are the other two metal sites per monomer ($S1$ and $S2$) which are coloured grey and black respectively.

The metals able to occupy the other two metal sites, $S1$ and $S2$ were tested at the $S3$ site. Each metal, in turn, was included in the model at the $S3$ position, and positional and B-factor refinement employed. After 15 cycles of refinement, $2F_o – F_c$ and $F_o – F_c$ maps were calculated and the size of the peaks at the $S3$ site were checked. **Table 34** below shows the results of these refinements and indicates that the site is possibly occupied by a manganese ion, this having the smallest $F_o – F_c$ peak (4 σ) and a reasonable B-factor of 16.5 $Å^2$. The assignment of the third metal site was also tested by an omit-map calculation. All three metals were omitted from the structure and 15 cycles of positional and B-factor refinement were run to reduce model bias. The three highest peaks in the resulting $F_o - F_c$ map are shown in **Table 35** and corresponded to the three metal sites.

The final X-ray model contained 1809 protein atoms, 3 metal ions ($2Mn^{2+}$ ions and a Ca^{2+} ion), 8 MPD atoms and 217 water oxygen positions. The final R-factor was 19.5% and R-free was 22.0%. Of the 217 water sites identified, 181 were found to be within 1 Å of the 319 water sites located in the 0.94 Å X-ray model at 110K[50].

Atom type	R-factor (%)	R-free (%)	Positive $F_o - F_c$ peak size (σ)	Negative $F_o - F_c$ peak size (σ)	Positive $2F_o - F_c$ peak size (r.m.s.)	B-factor (Å^2)
Mn^{2+}	19.5	22.0	4	0	19	16.5
Ca^{2+}	19.6	22.0	7	0	18	12.9
Cd^{2+}	19.5	22.1	0	-7	20	27.5
Ni^{2+}	19.4	21.9	4	0	19	18.9
Water oxygen	19.7	22.1	15	0	17	2.0

Table 34.   Atom types tested at the octahedral coordinated site, *S*3 and the results of the separate refinements.

Metal site	$F_o - F_c$ map peak size (σ)	B-factor of metal when site occupied (Å^2)
*S*1	53	8.4
*S*2	24	14.3
*S*3	32	16.5

Table 35.   Omit $F_o - F_c$ map peak size for the three metal sites.

5.3.6 X-ray refinement summary

Below is a summary of the refinement procedure that was followed.

Refinement program	Number of cycles/adjustments made to the model	R-factor (%)	R-free (%)
REFMAC5 (rigid body)	10	28.6	28.2
REFMAC5 (restrained)	10	26.4	28.2
O (version 7)	Checked ; (a) protein main chain and side chains in density correctly. (b) all rotamers for each side chain.		
ARP_WARP	30 cycles 212 waters added		
O (version 7)	Checked $2F_o - F_c$ and $F_o - F_c$ electron density maps and removed/added water oxygen atoms where necessary. [Water density is spherical and oxygen atoms in density at > 2 r.m.s. in $2F_o - F_c$ maps.] Checked hydrogen bond distances (water-water & water-protein). Replaced certain water positions for; an MPD molecule and Mn^{2+} ion at *S*3 metal site.		
REFMAC5 (restrained)	10	19.5	22.0
O (version 7)	Checked protein, metals, MPD and waters for good geometry and fit to density and made adjustments where necessary. Modelled the alternate conformations.		

5.3.7 Neutron structure refinement

The refined X-ray structure at 1.65 Å was used as the starting model for the neutron refinement using CNS[274]. In this 1.65 Å X-ray model, hydrogen atoms were generated for the protein and the bound water oxygen atoms according to an algorithm in the CNS program, in preparation for the neutron refinement. All non-exchangeable hydrogen atoms *i.e.* aliphatic hydrogen atoms, were given the neutron scattering length for H (b_H = -0.374), while all the exchangeable hydrogen atoms *i.e.* those attached to O or N, were given the neutron scattering length for D (b_D = 0.6674). The first step was a 20-cycle rigid body refinement of the protein and bound water, using all the neutron data to 2.5 Å. Next, the exchangeable D-atom positions of the hydroxyl groups of Ser, Thr, and Tyr protein residues, and of the bound water molecules were refined (positional and B-factor refinement) with the rest of the model fixed. Analysis of the model was used to check the progress of the refinement. In a few cases (seven) the fit to the density for the bound water atoms was correct *e.g.* the D-atoms of the water molecules moved into the neutron density while the water oxygen atoms were held at their X-ray oxygen atom positions (**Figure 90**).

However, in order to satisfactorily fit the density for all the bound waters, the positions of certain water oxygen atoms determined by the X-ray structure refinement were required to be shifted (**Figure 91**). Therefore, positional and B-factor refinement for the waters was repeated, but this time including the water oxygen atoms. This greatly improved the agreement between the model and the nuclear density. In order to assess if the individual water sites in the two models correspond, analysis of the hydrogen bond interactions was made. If the two waters being compared interact with the same protein residues/waters then they were deemed to correspond. Of the 217 waters identified from the X-ray data, 77 could be modelled as D_2O, 20 were modelled as oxygen alone and 120 had no visible nuclear density. The average shift in oxygen atom positions for the 97 waters that correspond between X-ray and neutron models was 1.06 Å.
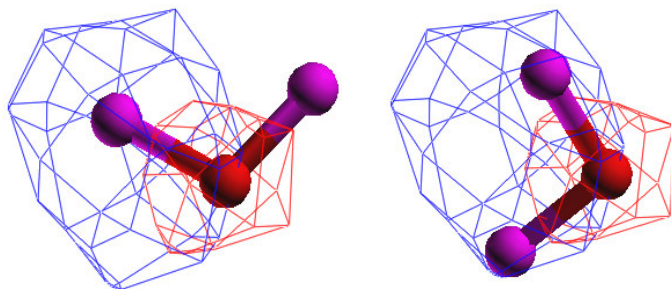
Figure 90.   Refinement of the water molecule D-atoms with the oxygen atoms fixed. *Left*, a water molecule (W22 in neutron file, W24 in X-ray file) with the oxygen atom at the position identified from the X-ray refinement and the D-atoms at the calculated positions. The fit to the $2F_o - F_c$ neutron density in blue (at 1.6 r.m.s.) is incorrect. *Right,* the same water molecule after refinement of the D-atoms with the oxygen atom fixed at the X-ray determined position. The fit is better (albeit not perfect) with the D-atoms moving into the nuclear density.
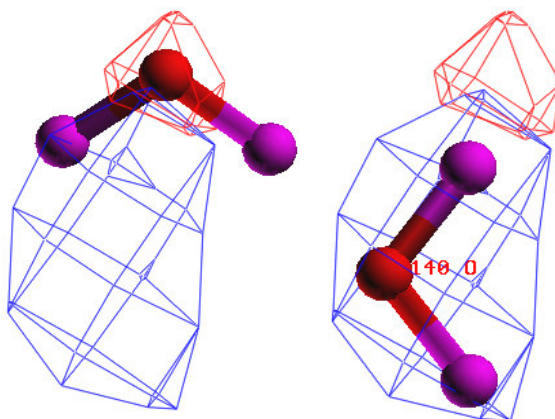
Figure 91.   Refinement of D_2O as a whole molecule *i.e.* refining both the D- and O-positions for each water molecule. *Left*, the position of the water when the oxygen is at the X-ray determined position and the D-atoms are at the refined positions. As can be seen, the fit to the nuclear density is incorrect.   *Right*, the position and orientation after refinement of the water as D_2O is now correct in relation to the neutron data.

The positions of the exchangeable D-atoms of the hydroxyl groups of protein residues were examined before and after refinement of their positions. Most D-atoms were seen to move into the nuclear density for the model indicating the refinement strategy was working (**Figures 92 and 93**). Below is an example of a threonine residue in which the exchangeable side-chain D-atom position has been refined, improving the fit to the density.
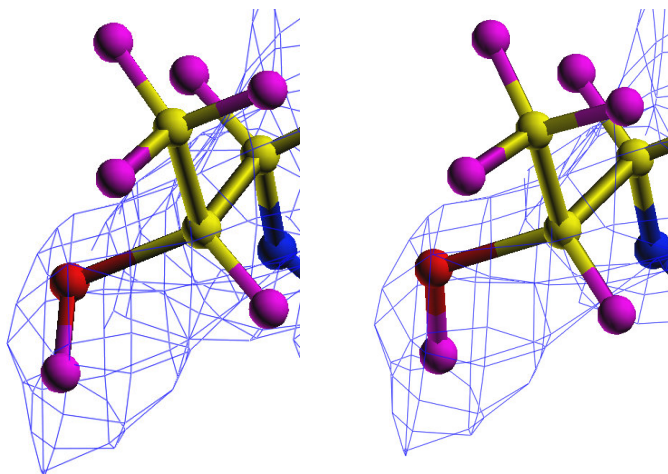
Figure 92.   The $2F_o - F_c$ neutron density at 1.6 r.m.s. for the side-chain of threonine 49. *Left*, before refinement of the hydroxyl D-atom of this residue it can be seen that the deuterium atom sits slightly out of the positive nuclear density, while the nuclear density for the methyl group of threonine is invisible due to cancellation of the positive and negative density for the carbon and hydrogen atoms respectively. *Right*, after refinement of the D-atom position the atom now sits in the nuclear density.

After modelling the position and orientation of the waters identified from the X-ray data, further waters were added to the model using the neutron data alone. $F_o - F_c$ maps were calculated and used to identify possible water peaks using the water-picking procedure in CNS. Water oxygen atoms were identified on the basis of satisfying certain distance restraints and having positive $F_o - F_c$ map σ values of greater than 3.5. H-atoms were once again generated for these new water oxygen positions and given the scattering length for deuterium. Positional and B-factor

refinement of the original water molecules plus the 130 extra added waters (as D_2O) was performed, with the rest of the model fixed. $2F_o - F_c$ and $F_o - F_c$ maps were used to check the fit to the density. Of the 130 extra added waters, 90 could be modelled satisfactorily as D_2O while 40 were modelled as oxygen alone.
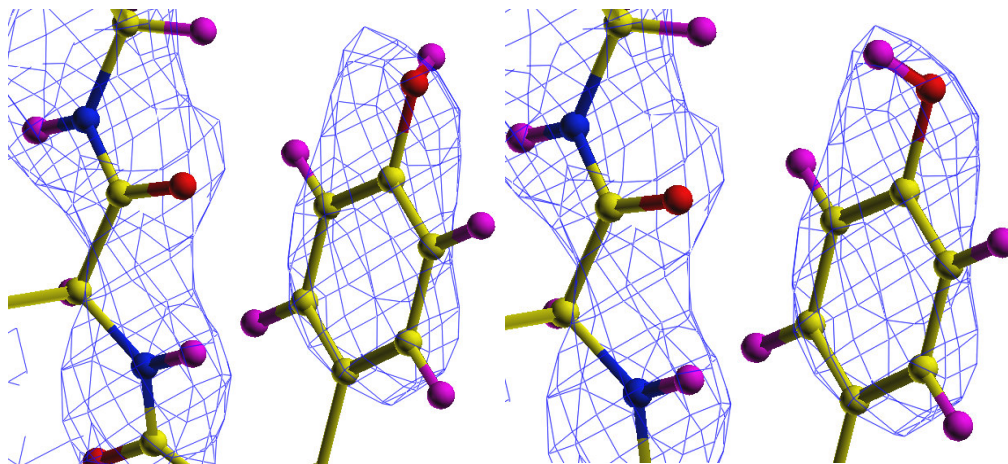
Figure 93.   The $2F_o - F_c$ neutron density at 1.5 r.m.s. for tyrosine 54. *Left*, before refinement of the hydroxyl D-atom of this residue it can be seen that the deuterium atom sits slightly out of the positive nuclear density. *Right*, after refinement of the D-atom position the atom now sits in the positive nuclear density. Also shown is part of the main chain where the hydrogen atoms of the amide group have exchanged for deuterium atoms and thus are in positive nuclear density whereas those hydrogen atoms attached to carbon atoms are unexchanged and thus have negative scattering nuclear density (not shown).

5.3.8 Tests with the neutron data alone

It should be noted that a full refinement of all atomic positions of the protein led to strongly diverging R-values (R-factor and R-free). The reason for this is partly the large scattering density of the solvent, which is more than twice as large as the scattering density of the protein, and partly the high number of parameters to be refined. While the X-ray model contains a total of 2037 atoms (1809 protein atoms, 3 metals, 8 MPD atoms and 217 water oxygen atoms), approximately another 2000 H/D atoms have to be included into the model for the refinement of the neutron data. With ~4000 atoms in the neutron structure and each atom having 4 parameters to refine (x, y, z, and B), this gives a total of ~16000 parameters to refine the whole model. The neutron data set contains 6462 unique reflections, and even though there are bond distance and angle restraints to help, trying to refine the whole model gives a data-to-parameter ratio that is very poor. The risk of over-refinement is therefore inherent at this resolution and so could lead to inaccurate results. Indeed the X-ray analysis will always give a better structure for the non-H-atoms, simply because the ratio of the number of reflections to the number of parameters is so much better (26234 unique X-ray reflections, 8148 parameters, data-to-parameter ratio of ~3 without including stereochemical restraints). The usefulness of the neutron data is in elucidating that which cannot be seen with X-rays, and any full refinement of all the atoms using the neutron data will lead to a less precise structure. Therefore, the non-hydrogen atoms of the protein were kept at the positions from the X-ray refinement (apart from rigid body refinement) and only B-factors of these atoms were refined. Even though the protein atom positions are not refined by the neutron data, the agreement between the model and the neutron density is excellent (**Figure 94**). By only refining the exchangeable D-atoms of the protein (399 atoms) and the bound water atoms, we reduce the number of parameters dramatically and can therefore refine these atoms with much less risk of over-refinement.
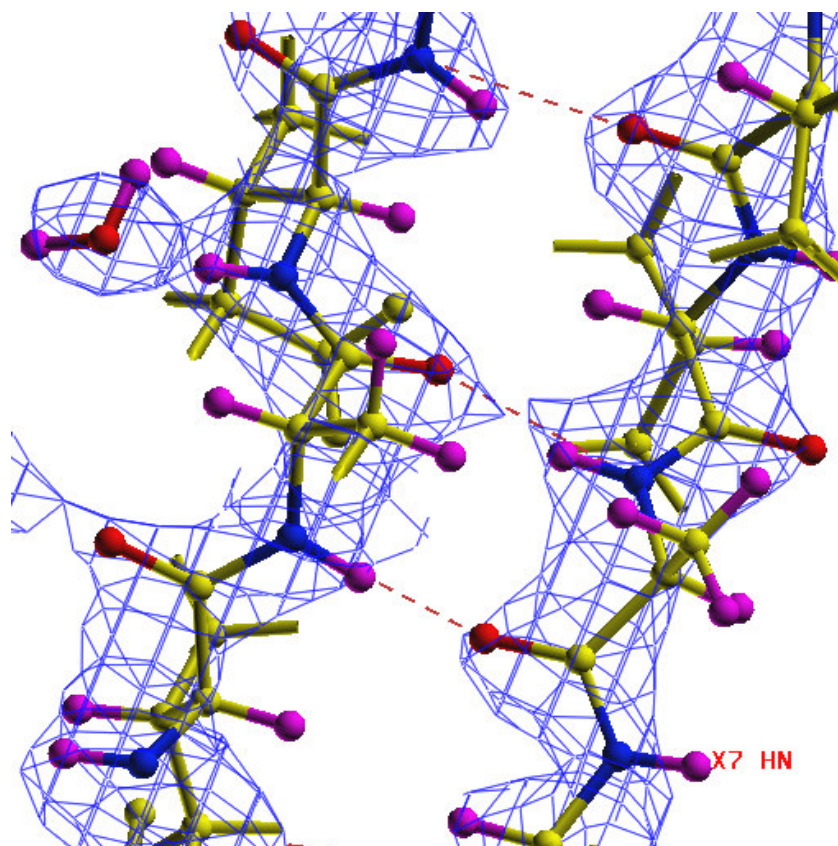
Figure 94.   Two sections of the seven stranded β-sheet with the $2F_o - F_c$ neutron
density (at 1.6 r.m.s.) overlaid.   Even though the protein atom
positions are not refined by the neutron data, the agreement between
the model and the neutron density is excellent.

5.4   Analysis of the X-ray model

The final structure contains a monomer in the asymmetric unit, of molecular weight 25 kDa. Each monomer of the classical *I*222 tetramer is made from a single polypeptide chain of 237 amino acid residues, all of which are modelled, and contains one Ca^{2+} ion and two Mn^{2+} ions. A total of 217 water molecules were modelled around the monomer. The unit cell dimensions are; a = 89.16 Å, b = 86.129 Å, c = 61.594 Å, α = β = γ = 90°.

Total no. of atoms = 2037
Total no. of protein atoms = 1809
Total no. of MPD atoms = 8
Total no. of water molecules = 217
Others = 3 metal ions (1 Ca^{2+} ion and 2 Mn^{2+} ions)
Resolution range = ∞ to 1.65 Å
R-factor = 19.5%
R-free = 22.0%
Bond distance r.m.s. deviation = 0.007 Å
Bond angle r.m.s. deviation = 1.292°
Ramachandran core/additional (%) = 85.6/13.0
Mean B-factor for all protein atoms ($Å^2$) = 19.3
Mean B-factor for all side chain atoms ($Å^2$) = 20.1
Mean B-factor for all main chain atoms = ($Å^2$) = 18.7
Mean B-factor for all MPD atoms ($Å^2$) = 33.5
Mean B-factor for all water molecules ($Å^2$) = 28.8
Mean B-factor for all side chains and waters ($Å^2$) = 21.9
Mean B-factor for all atoms ($Å^2$) = 20.4

Table 36.   Summary of the 1.65 Å resolution X-ray model of saccharide-free concanavalin A at 100K.

The Ramachandran plot[256] for the X-ray saccharide-free concanavalin A model is shown in **Figure 95** and was calculated using PROCHECK[257]. 178 amino acids (85.6%) are found within the most favoured regions (A, B, L) and 27 residues

(13.0%) are found within the additionally allowed regions (a, b, l, p) illustrating that the model has good geometry. There are no distorted main chain bond lengths, angles or planar groups. 3 residues are labelled on the plot, 2 of these residues are in the generously allowed regions (Thr 120 and Glu 122) and one residue is in the disallowed region (His 121). When comparing the geometry of the model with that of the starting model, we see that previously there were a few distorted main chain bond lengths, angles and planar groups (5, 3, and 7 respectively) whereas now there are none. Furthermore, in the starting model there were also two extra *cis*-peptide linkages (120-121, 121-122) which were modelled due to the disorder of this loop section, whereas they are now identified in the normal *-trans* configuration. However, beforehand there were more residues in the most favoured regions than in the 1.65 Å D_2O soaked concanavalin A model at 100K presented here, 185 (88.9%) and 178 (85.6%) respectively.
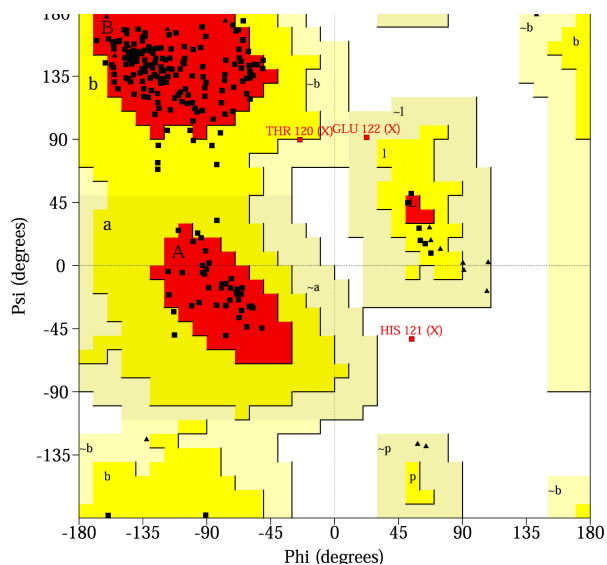
Figure 95.   The Ramachandran plot for the 1.65 Å resolution X-ray model of saccharide-free concanavalin A.

The overall root-mean-square deviations (*x, y, z*) between the 0.94 Å concanavalin A starting model (PDB code = 1NLS)[50] and the final X-ray model were calculated as follows:

1. Between all protein atoms, r.m.s. deviation = 0.405 Å

2. Between all main chain protein atoms, r.m.s. deviation = 0.182 Å

3. Between all side chain protein atoms, r.m.s. deviation = 0.545 Å

4. Between all α-carbon protein atoms, r.m.s. deviation = 0.146 Å

The residues that exhibit the largest differences between the starting model (1NLS) and the final 1.65 Å X-ray model of saccharide-free concanavalin A are listed, those given have root-mean-square deviations greater than 1 Å.

Main chain large r.m.s. deviations;

His 121, Glu 122

Side chain large r.m.s. deviations;

Val 75, His 121, Lys 135

The amino acid residues above are all found within loop structures and on the surface of the protein, therefore, they have greater conformational freedom than those residues which are buried or involved in one of the 3 β-sheets. Valine 75 exhibits a large side-chain r.m.s. deviation as this residue has been modelled in an alternate conformation to that of the starting model. However, as shown in **Figures 96/113** the residue appears to take two favoured conformations and therefore, could be modelled in the same conformation as the starting model. Lysine 135 exhibits a large side-chain shift from the starting model. This is unsurprising as the long side chain is very flexible and can occupy many different low energy conformations. The residue exhibiting the largest r.m.s. deviation is Histidine 121 (1.431 Å). The $2F_o - F_c$ electron density for this residue was poor and so the modelling of even the main-chain atoms of the residue proved problematic. As can be seen from the B-factor versus residue graph for the protein chain (see **Figure 97**), this residue also

has the highest B-factor (average B-factor$_{121}$ = 42.0 Å^2) of all the amino acids in the monomer.
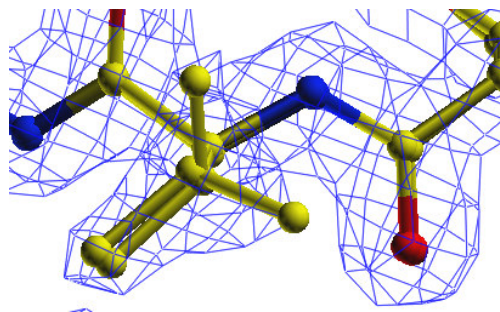
Figure 96.   The two alternate conformations of the side-chain of valine 75 from the starting model (1NLS) and the 1.65 Å model presented here.
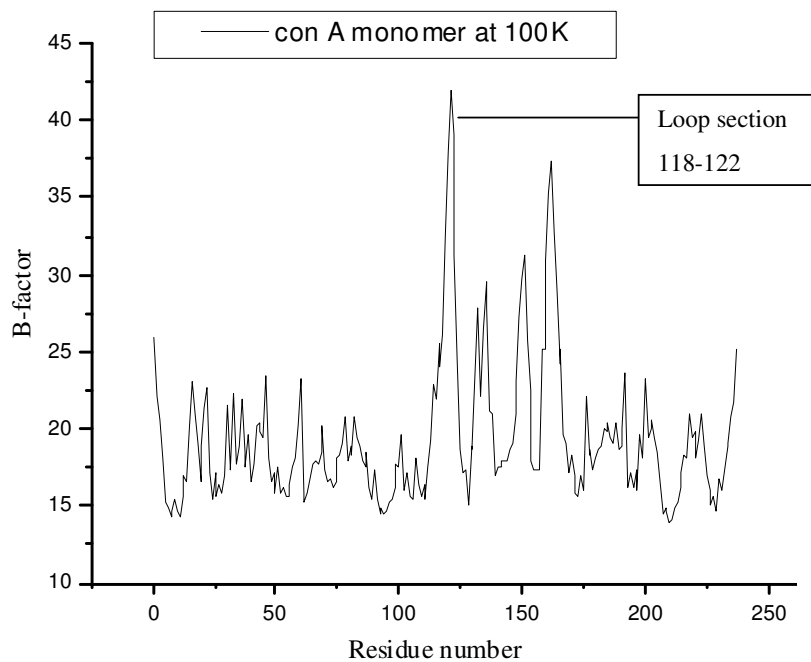
Figure 97.   The B-factor versus residue plot for the monomer in saccharide-free concanavalin A.

5.4.1 Metal sites

The manganese to ligand distances at the *S*1 transition-metal site are given (see **Table 37**).   The data agrees with that from previous models of saccharide-free concanavalin A, in that Glu 8 OE2, Asp 10 OD2, Asp 19 OD1, His 24 NE2 and two water molecules co-ordinate to the manganese ion. When comparing the distances for the starting model (1NLS) and the refined model there is excellent agreement between them, the largest difference being 0.037 Å. The two waters which act as ligands to the Mn^{2+} ion appear as well-ordered, spherical electron density. Moreover, they have reasonably low B-factors (W2 B-factor = 12.3 $Å^2$, W6 B-factor = 13.8 $Å^2$) as expected for these conserved water sites.

The calcium-ligand distances are given (see **Table 38**). Asp 10 OD1, Asp 10 OD2, Tyr 12 O, Asn 14 OD1, Asp 19 OD2 and two water molecules co-ordinate to the calcium ion. There is excellent agreement in ligand distances between the starting and refined models. The water oxygen atoms identified as ligands to the Ca^{2+} ion at the *S*2 site also exhibit strong spherical density and have reasonably low B-factors (W3 B-factor = 15.1 $Å^2$, W19 B-factor = 15.3 $Å^2$).

The manganese to ligand distances at the *S*3 metal binding site are given (see **Table 39**). For comparison, the ligand distances from the Cd-concanavalin A structure[273] are given also. In the Cd-concanavalin A structure an aspartic acid side chain (Asp 136 OD1/2) from a symmetry related subunit helps chelate the Cd ion. The Cd ion was found to be pseudo-octahedrally coordinated. In the refined X-ray structure presented here this was found not to be the case with the aspartic acid side chain observed in a different conformation (see **Figure 101**). The Mn^{2+} ion at the *S*3 site is found to be octahedrally coordinated with a water molecule (W101 in the PDB file) occupying the position at which the Asp 136 side chain is observed in the Cd-concanavalin A structure.

The Ca^{2+} to Mn^{2+} distances for Deacon *et al.,*[50] and the refined model are 4.20 Å and 4.22 Å respectively.
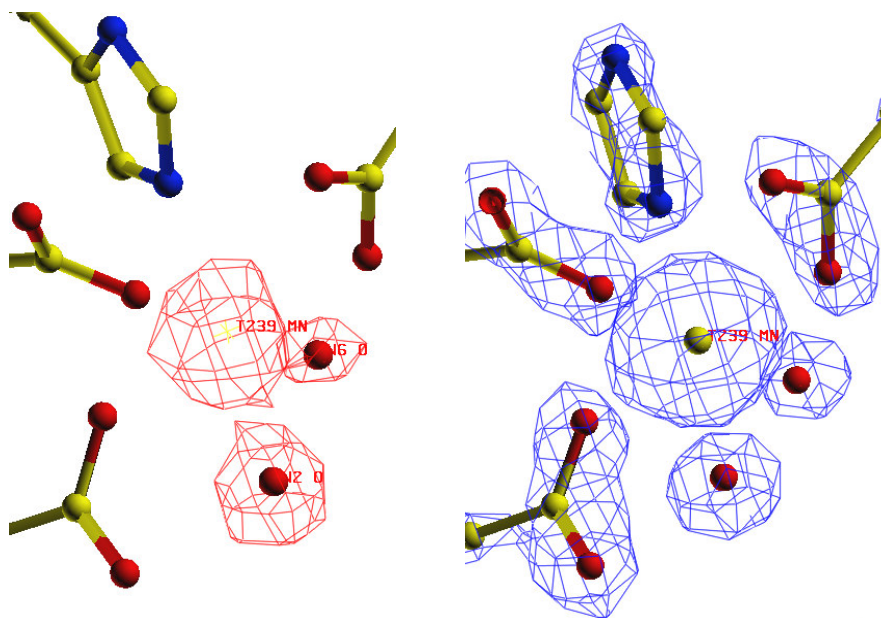
Figure 98.   The $S1$ site in concanavalin A at 1.65 Å resolution. *Left,* the $F_o - F_c$ omit map (at 6 σ) for the two waters at the $S1$ site. There is also a residual peak for the manganese ion. *Right,* the $2F_o - F_c$ electron density map (at 2.6 r.m.s.) is shown, as well as the water oxygen atoms and protein residues that coordinate to the Mn^{2+} ion.

Mn^{2+} ligand distances at the $S1$ site(Å)	1.65 Å X-ray con A at 100K	Deacon *et al.,* (1997)
Glu 8 OE2	2.18	2.166
Asp 10 OD2	2.19	2.153
Asp 19 OD1	2.21	2.192
His 24 NE2	2.27	2.233
Water molecule	2.28 (W2)	2.261 (W412)
Water molecule	2.19 (W6)	2.182 (W413)

Table 37.   The manganese ligand distances at the $S1$ site in the monomer of concanavalin A. Values from Deacon *et al.*, (1997)[50] are given for comparison. The waters listed are numbered as in the PDB file.

201

Ca^{2+}-ligand distances at the S2 site (Å)	1.65 Å X-ray con A at 100K	Deacon *et al.*, (1997)[50]
Asp 10 OD1	2.47	2.469
Asp 10 OD2	2.51	2.513
Tyr 12 O	2.37	2.383
Asp 19 OD2	2.40	2.405
Asn 14 OD1	2.34	2.350
Water molecule	W3 2.36	W410 2.360
Water molecule	W19 2.43	W411 2.428

Table 38.   The calcium ligand distances at the *S*2 site in the concanavalin A monomer.   Values from Deacon *et al.*, (1997)[50] are given for comparison.

Figure 99.   The *S*2 site in concanavalin A at 1.65 Å resolution.   The $2F_o - F_c$ electron density map (at 2.6 r.m.s.) is shown in blue, as well as the water oxygen atoms and protein residues that coordinate to the Ca^{2+} ion and their ligand distances.

Mn^{2+} ligand distances at the *S*3 site (Å)	1.65 Å X-ray con A at 100K	Cd-substituted concanavalin A structure[273]
Glu 183 OE1	2.80	2.49
Glu 183 OE2	2.25	2.37
Glu 87 OE2	2.28	2.49
Water molecule	2.32 (W98)	2.71 (W125)
Water molecule	2.57 (W224)	2.80 (W104)
Water molecule	2.25 (W101)	
Asp 136 OD2*		2.36
Asp 136 OD1*		2.91
*Residue from a symmetry related subunit		

Table 39.   The manganese ligand distances at the *S*3 site. Ligands and distances from the Cd-concanavalin A structure[273] are given also.
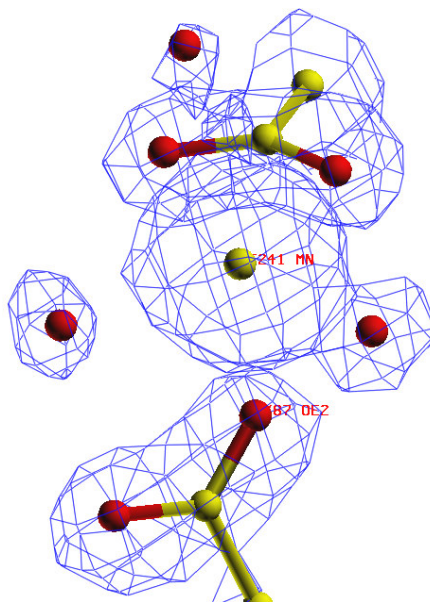
Figure 100.   The *S*3 site in concanavalin A at 1.65 Å resolution. The $2F_o - F_c$ electron density map (at 1.4 r.m.s.) is shown in blue, as well as the water oxygen atoms and protein residues that coordinate to the Mn^{2+} ion.
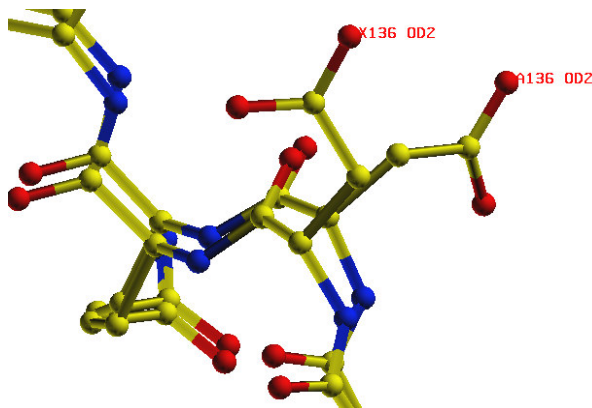
Figure 101.   The two different conformations taken by the aspartic acid side chain of residue 136. In the 1.65 Å resolution X-ray model here the Asp 136 side chain is labelled as X136 OD2, whereas the conformation taken in the Cd-concanavalin A structure is labelled as A136 OD2.

If we compare the B-factors of the atoms chelated to the metals, there is good agreement in values between the starting model (1NLS) and the refined model, for both the calcium and manganese sites *i.e.* the metal ligand atoms for both models have low B-factors of less than 20 Å^2 and exhibit a small spread in values within each set. Below are given the B-factors of all the metal ligand atoms. If we compare the B-values of the manganese ligand atoms at the *S*1 site in the 1.65 Å structure against each other they all are in the range 12.3 to 14.7 Å^2. If we compare the B-values of the calcium ligand atoms at the *S*2 site against each other they are in the range 14.7 to 15.8 Å^2. Some of the metal ligand atoms are part of the seven-stranded *β*-sheet structure of the concanavalin A monomer (Glu 8, Asp 10, Tyr 12, His 24), and therefore, it is unsurprising that they exhibit low B-factors. However, even those that belong to loop structures (Asn 14, Asp 19) also exhibit low B-factors.

Mn^{2+} ligand B-factors at $S1$ ($Å^2$)	1.65 Å X-ray con A at 100K	Deacon *et al.,* (1997)
Glu 8 OE2	14.2	6.81
Asp 10 OD2	14.7	6.50
Asp 19 OD1	14.5	7.33
His 24 NE2	14.2	6.72
Water molecule	W2 12.3	W412 7.41
Water molecule	W6 13.8	W413 8.68

Table 40.   The manganese ligand B-factors at $S1$ in concanavalin A.   Values from Deacon *et al.*, (1997)[50] are given for comparison.

Ca^{2+}-ligand B-factors at $S2$ ($Å^2$)	1.65 Å X-ray con A at 100K	Deacon *et al.,* (1997)
Asp 10 OD1	14.7	7.10
Asp 10 OD2	14.7	6.50
Tyr 12 O	15.8	7.85
Asp 19 OD2	15.4	7.28
Asn 14 OD1	15.8	7.89
Water molecule	W3 15.1	W410 7.69
Water molecule	W19 15.3	W411 8.23

Table 41.   The calcium ligand B-factors at the $S2$ site in the concanavalin A monomer.   Values from Deacon *et al.*, (1997)[50] are given for comparison.
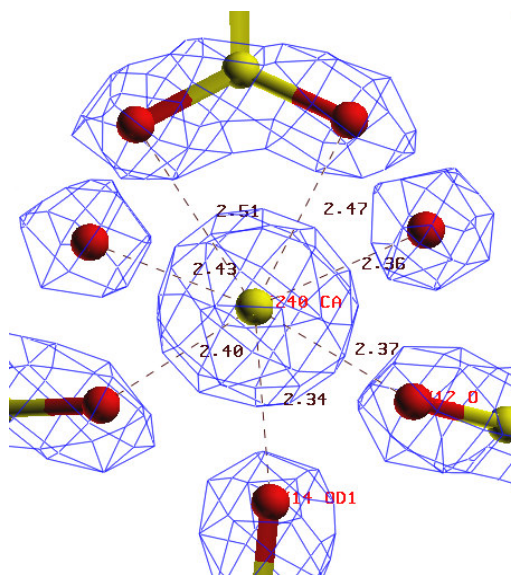
Mn^{2+} ligand B-factors at the $S3$ site ($Å^2$)	1.65 Å X-ray con A at 100K	Cd-substituted concanavalin A structure[273]
Glu 183 OE1	21.7	31
Glu 183 OE2	21.8	24
Glu 87 OE2	19.0	20
Water molecule	27.4 (W98)	30 (W125)
Water molecule	35.6 (W224)	34 (W104)
Water molecule	36.8 (W101)	n/a
Asp 136 OD2*	n/a	44
Asp 136 OD1*	n/a	47
*Residue from a symmetry related subunit		

Table 42.   The manganese ligand B-factors at the $S3$ site. Ligands and B-factors from the Cd-concanavalin A structure[273] are given also.

Below are the B-factors for the manganese and calcium ions.

Metal B-value ($Å^2$)	1.65 Å X-ray con A at 100K	Deacon *et al.,* (1997)
$S1$ (Mn^{2+})	8.4	6.45
$S2$ (Ca^{2+})	14.3	6.37
$S3$ (Mn^{2+})	16.5	n/a

Table 43.   B-factors for the two manganese and one calcium ion in the X-ray model of saccharide-free concanavalin A.

5.4.2 The saccharide binding site at 100K

Residues in four loops, 11 to 23, 97 to 104, 201 to 208 and 217 to 237, make up the carbohydrate binding pocket in concanavalin A. The positions of the waters identified within the saccharide binding site of the 1.65 Å structure were compared with previous X-ray models[50,51]. Five waters were identified, three of which (W4, W9, W36) were found to agree (within 1 Å) with those found in both the 0.94 Å UHR X-ray structure at 100K and the 1.8 Å D_2O soaked X-ray structure at room temperature. As can be seen from the figure below, four of the five waters identified (W4, W9, W47, W68) are found positioned at very similar positions to the oxygen atoms of the hydroxyl groups of the bound glucoside.

Figure 102.   The $2F_o - F_c$ electron density (at 2 r.m.s.) for the five waters identified in the monosaccharide binding site, with the positions of methyl-glucoside sugar atoms from the methyl-glucoside/concanavalin A complex[8] superimposed.

The waters are listed in the following table. The water labelled as W4 is found at hydrogen bonding distance from Asp 208 OD1, Asn 14 ND2, and W36. The water labelled as W9 is found at hydrogen bonding distance from Asp 208 OD2, W36 and

W68. The water labelled as W36 is found at hydrogen bonding distance from W4, W9 and W105. The water labelled as W47 is found at hydrogen bonding distance from W68. The water labelled as W68 is found at hydrogen bonding distance from Leu 99 N, W9, W12 and W47. Therefore, the waters form a network of hydrogen bonds between each other and three key residues (Asn 14, Leu 99, Asp 208) in the saccharide binding site.

Saccharide binding site waters	100K 1.65 Å X-ray structure	Room temperature 1.8 Å X-ray structure[51]	110K UHR X-ray structure[50]
WI	W4 (16.0 Å^2)	W13 (14.5 Å^2)	W252 (9.3 Å^2)
WII	W9 (17.5 Å^2)	W89 (17.7 Å^2)	W264 (10.3 Å^2)
WIII	W36 (17.3 Å^2)	W6 (32.4 Å^2)	W375 (12.2 Å^2)
WIV	W47 (18.5 Å^2)		
WV	W68 (16.8 Å^2)		W269 (11.2 Å^2)

Table 44.   The five water oxygen atom sites identified at the monosaccharide binding site in the 1.65 Å resolution model of saccharide-free concanavalin A, compared against those identified in the monosaccharide binding site in previous models.
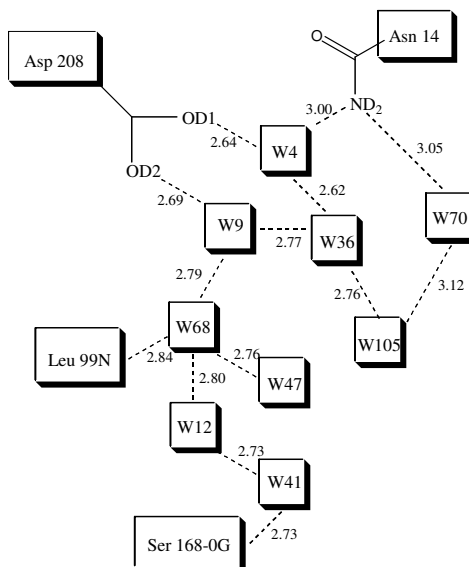
Figure 103.   Schematic of the X-ray hydrogen bonding interactions with distances.
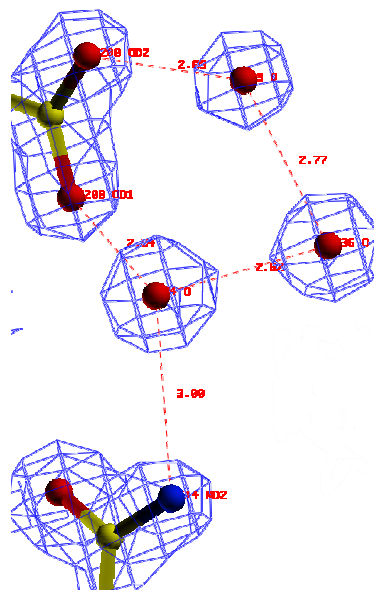
Figure 104.   The 3 invariant waters (W4, W9, W36) identified in the monosaccharide binding site, with the $2F_o - F_c$ electron density shown at 2 r.m.s.
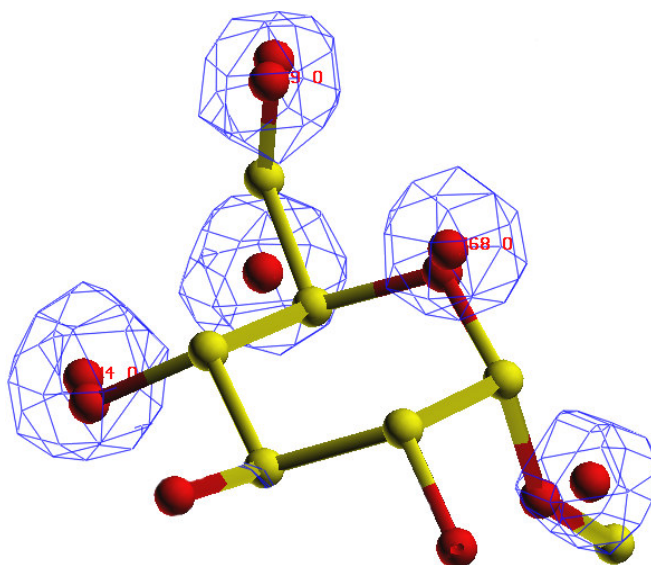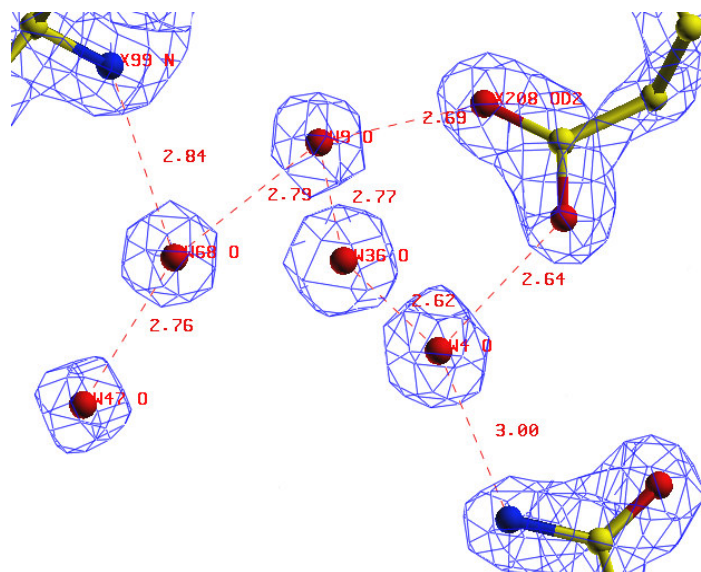

Figure 105.   The $2F_o - F_c$ electron density shown at 2 r.m.s. for the five waters identified at the monosaccharide binding site and their respective hydrogen bonding interactions and distances.

Analysis of water sites in the 'extended region' of the saccharide binding site, indicated that another two water sites (W41 and W105 in X-ray PDB file at 100K) agree within 1 Å in all three models. These are given in the table below. W41 is stabilized through hydrogen bonding interactions with the side chain of a serine residue (Ser 168 OG) and two waters (W12 and W134). W105 is stabilized through hydrogen bonds made to W36 and W70.
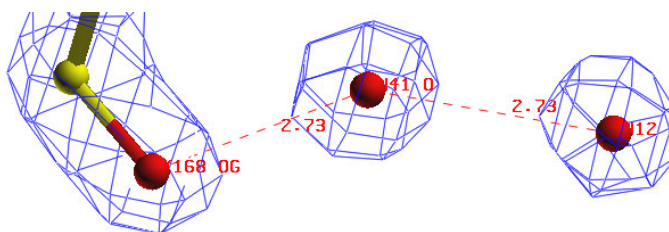
Figure 106.   The water numbered 41 in the 100K 1.65 Å X-ray structure is located in the extended region of the saccharide binding site.

Saccharide binding site waters	100K X-ray structure	Room temperature X-ray structure[51]	110K UHR X-ray structure[50]
WVI	W12 (19.5 Å^2)		
WVII	W41 (21.6 Å^2)	W126 (37.7 Å^2)	W296 (16.4 Å^2)
WVIII	W70 (20.2 Å^2)		W279 (14.3 Å^2)
WIX	W105 (26.8 Å^2)	W7 (52.0 Å^2)	W348 (15.5 Å^2)

Table 45.   The four water oxygen atom sites identified in the 'extended region' of the saccharide binding site in the 1.65 Å resolution model of saccharide-free concanavalin A, compared against those identified in the 'extended region' of the saccharide binding site in previous models.
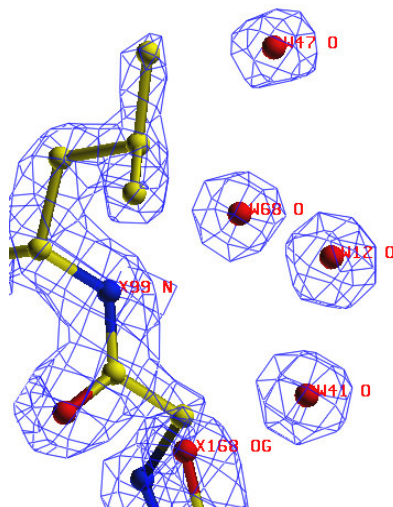
Figure 107.   Four of the water oxygen atoms identified in the saccharide binding site region. Two of the waters (W47 and W68) are found within the monosaccharide binding site, while the other two waters (W12 and W41) are found within the extended region of the saccharide binding site.
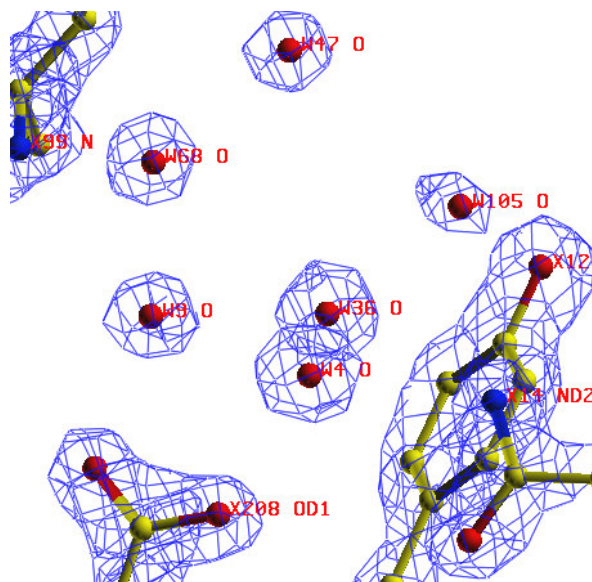
Figure 108.   Six of the water oxygen atoms identified in the saccharide binding site region. Five of the waters (W4, W9, W36, W47, W68) are found within the monosaccharide binding site, while the other water (W105) is found within the 'extended region' of the saccharide binding site.

5.4.3 Temperature-factor analysis

By examination of the B-factor versus residue plot (**Figure 97**) for the saccharide-free monomer of concanavalin A, the areas with high B-factor correspond well to loop areas in between the *β*-sheet sections. In particular, four loop sections; 118-122, 133-136, 150-151, and 160-164 display the highest B-factors. Residues belonging to these four loop sections are labelled in **Figure 109**.
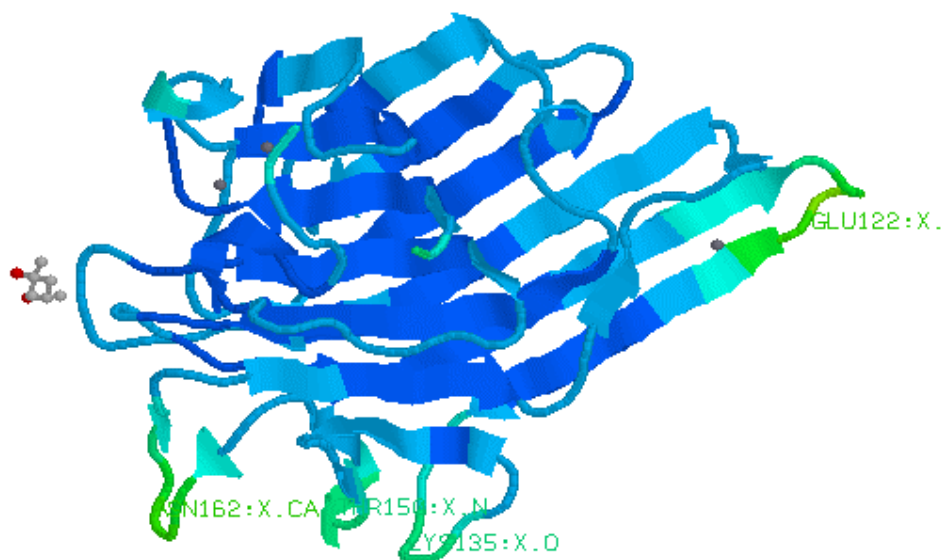
Figure 109.   The protein chain backbone of the con A monomer, coloured according to temperature factor of the residues. The colour coding is as follows; blue↔v.low B-factor, green↔low B factor, yellow↔medium B-factor, orange↔high B-factor, red↔v. high B-factor. Also shown, coloured according to the CPK definition is the molecule of cryoprotectant (MPD) identified.

As can be seen from the figure above, the areas with low B-factors (coloured blue) correspond to the *β*-sheet sections. Overall, it can be seen that the structure is extremely well ordered with no regions of the protein displaying high B-factors (as indicated by the lack of areas coloured orange or red). The majority of the protein atoms (1330/1809) are found to have associated B-factors of less than 20 Å^2, with

only a few (11/1809) having B-factors greater than 40 Å^2. The majority (96/217) of the water oxygen atoms have B-factors in the range 20 to 30 Å^2, with all of them having B-factors less than 60 Å^2. All three of the metal atoms have B-factors less than 20 Å^2, indicating their stability. In particular, the Mn^{2+} ion at the S1 site has the lowest B-factor of all the atoms (8.4 Å^2). The MPD atoms indicate they have reasonable mobility as all the atoms lie in the range 30 to 40 Å^2. The MPD molecule identified in the model is at a reasonable distance (~4 Å) from the protein surface and makes only a few stabilising interactions with other atoms. Interestingly the MPD molecule is positioned at the entrance to the saccharide binding site with the nearest residues being Leu 99 and Tyr 100 as shown in **Figure 110**.
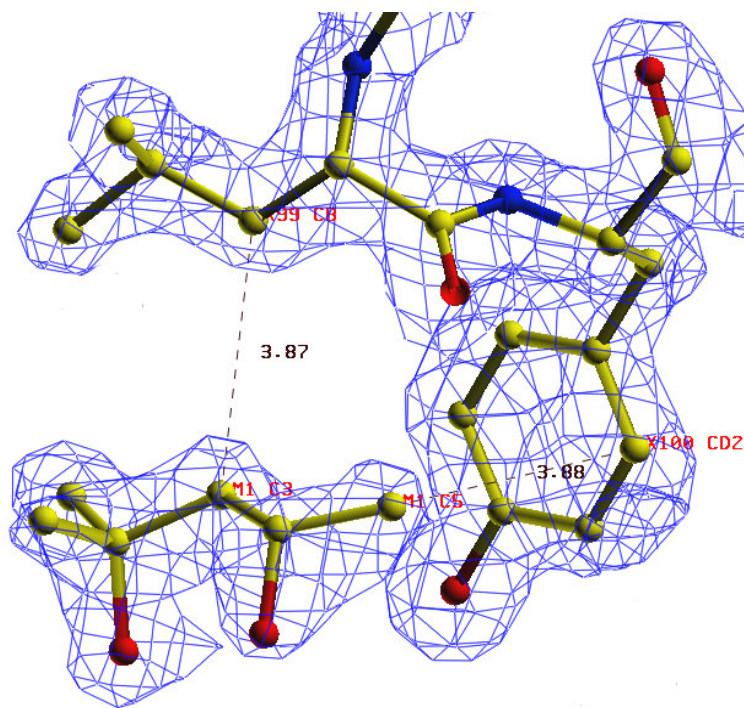
Figure 110.   The $2F_o - F_c$ electron density shown at 1.5 r.m.s. for the molecule of cryoprotectant, MPD. Also shown are the nearest protein residues (Leu 99 and Tyr 100) which are at the entrance to the saccharide binding site.

The average B-factors for main chain, side chain and all the atoms are given below .
Deacon *et al.,* (1997)[50] values are given also.

Average B-factor ($\mathring{A}^2$) for ;	1.65 $\mathring{A}$ X-ray con A at 100K	Deacon *et al.,* (1997)
Main chain	18.7	10.6
Side chain	20.1	15.5
Whole chain	19.3	13.0

Table 46.   The average B-factors for main chain atoms, side chain atoms and the whole chain for concanavalin A. Deacon *et al.,* (1997)[50] values are given also.

The average value for all the side chains is higher than that for the main chain atoms due to the fact that the side chains are more flexible with less constraints on their conformation. Main chain atoms are involved in many more stabilising hydrogen bonds than side chains and therefore, show less mobility.

5.4.4 Water structure analysis

The final refined model contains 217 water molecule oxygen atoms. These water oxygen atoms are all present in the final $2F_o$ - F_c electron density maps at greater than 2 r.m.s. of the map. The average B-factor of all the water oxygen atoms is 28.8 $\mathring{A}^2$. This value is similar to that of the UHR X-ray study at 110K at 0.94 $\mathring{A}$ resolution, where the 319 waters identified had an average B-factor of 27.2 $\mathring{A}^2$.

An analysis of the position of these waters within the monomer was carried out. Of the 217, 39 are seen to be completely buried within the protein while 178 are surface waters.   As would be expected the average B-factor of the buried waters is considerably less than those on the surface (averageB**buried** = 22.0 $\mathring{A}^2$ and

averageB$_{surface}$ = 30.3 Å^2). The buried waters achieve many more hydrogen bonds to the protein and so exhibit greater stability.

The 217 waters were also analysed in terms of the number of bonds made from each water molecule to protein atoms. In the table below it can be seen that 65 waters interact only with either other waters and/or MPD atoms. As is expected these waters exhibit the highest B-factors as they are more mobile due to having less stabilising interactions.

Number of bonds per water oxygen to protein atoms	Number of waters	Average B-factor (Å^2)
0	65	32.3
1	110	28.5
2	32	26.0
3	10	19.4

Table 47.   Table to show the number of bonds made from each water molecule to protein atoms and their average B-factors.

The total number of water contacts was also calculated.

Chain	Total number of water contacts
A	204

Table 48.   The total number of water contacts in the concanavalin A monomer.

By superimposing the starting model (1NLS) with the refined model using LSQKAB, an analysis of the number of waters that correspond *i.e.* are equivalent, between the two structures was achieved. Waters were seen to be equivalent if, after least-squares superposition of the two sets of co-ordinates the waters are within 1 Å of each other. Of the 217 water oxygen positions, 181 are within 1 Å of the 319

waters in the starting model. The average B-factor of these 181 waters is 27.4 $\mathring{A}^2$, with the average B-factor of the 181 that agree in the starting model (1NLS) being 21.6 $\mathring{A}^2$. Of the 39 buried waters in the final refined model, all 39 of these match the positions of the water oxygen atoms in the starting model. Of the 178 surface waters in the final refined model, 142 of these match the positions of the water oxygen atoms in the starting model. Therefore, a greater percentage of the buried waters agree with the starting model than the surface waters (100% buried agree *cf.* 80% surface agree) as would be expected due to the increased mobility of surface waters which make less interactions with the protein.

Approximately 20% (39/181) of the invariant waters are found to be buried, this at first may seem to be a rather low number, but when we examine the positions of the invariant waters, many of them are found by loop sections and make hydrogen bonds to residues within these loop sections. Therefore, it seems probable that the function of these invariant surface waters is to ensure the correct structural relations among the loops and with the rest of the molecule.

The position of the 181 invariant water oxygen atoms were analysed. Of the 181, 94 are seen to make hydrogen bonds only with atoms in loop sections. 25 waters are associated solely with the 6-stranded β-sheet (sheet 1), 11 with the 7-stranded β-sheet (sheet 2) and 3 with the 5-stranded β-sheet (sheet 3). 15 waters are found to make hydrogen bonds to both loop sections and sheet 2, including those coordinated to the metals at the *S*1 and *S*2 sites. 7 waters interact with sheet 1 and loop sections and 5 waters interact with sheet 3 and loop sections. One water oxygen position interacts with both sheets 1 and 2. The remaining 20 interact only with other water oxygen atoms. Five of the invariant waters are found within the saccharide binding site, being those which are expelled on binding saccharide. Four of these (W4, W9, W47, W68) are found bound to loop sections. The other water (W36) in the saccharide binding site is located above the Tyr 12 aromatic ring side-chain as noted by Deacon *et al.,* (1997)[50].
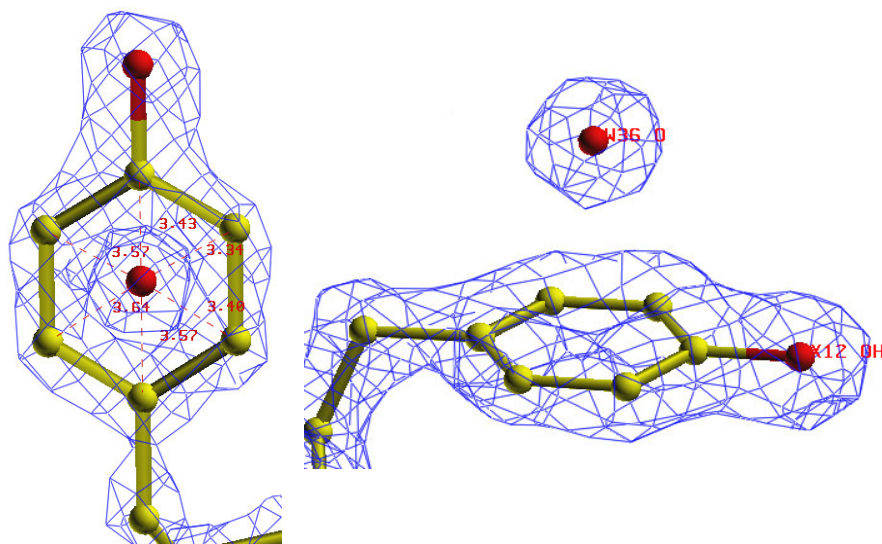
Figure 111.   Two views of a water molecule (W36 in the PDB file) in the
saccharide-binding site located directly above the aromatic ring of Tyr
12. The water molecule is located approximately equidistant from the
six carbon atoms of the ring. $2F_o - F_c$ electron density (*left,* at 2.5
r.m.s. and *right,* at 1.6 r.m.s.) is shown also.

Overall, it can be seen that the majority of the invariant waters are involved in
stabilising the loop sections. As the β-sheet regions are stable secondary motifs with
main-chain to main-chain hydrogen bonds occurring between chains these areas do
not require the extra stabilization given by the water mediated hydrogen bonds. On
the other hand the loop sections have much greater mobility and less protein-protein
interactions, therefore, it is unsurprising that interactions involving water molecules
are utilized in order to stabilize their overall conformations, especially when these
loop sections are a crucial part of the binding site region as in legume lectins.
Further analysis was done to see which loops are stabilized by these waters and
which sections are not. When the number of hydrogen bonds from the invariant
waters to each loop section are tallied, those that are found to have the most
interactions with the invariant waters are the loop sections listed below in **Table 49**.

Loop section	Number of H-bonds from the invariant waters to the loop
1-3	6
11-23 (sbs loop)	16
30-35	13
40-46	13
56-59	3
67-72	10
79-87	12
97-104 (sbs loop)	7
118-122	4
131-139	7
145-146	3
149-152	3
156-169	9
176-185	7
201-208 (sbs loop)	8
217-237 (sbs loop)	27

Table 49.   Gives the number of hydrogen bonds from the invariant waters to each loop section. In the table above sbs = saccharide binding site.

As can be seen from the list above, the loop sections make a large number of hydrogen bond interactions with the invariant waters. The four loops which create the carbohydrate binding pocket in legume lectins are given above. In particular, it can be seen that the two loops 11-23 and 217-237 have the largest number of hydrogen bond interactions with the invariant waters. This is unsurprising as it is necessary that these loops are in the correct spatial positions relative to each other in order to define the lectins specificity and affinity for certain sugars, in this case mannose and glucose. The other loops which are seen to make a large number of hydrogen bond interactions with the invariant waters are loops 30-35, 40-46 and 79-87. Loop 30-35 connects the first and second strands of the 7-stranded curved β-sheet (sheet 2), loop 40-46 connects the 7th strand of sheet 2 with the 3rd strand of

sheet 1 and loop 79-87 connects the 1st strand of sheet 1 with the 5th strand of sheet 2. There are 59/217 waters and 50/181 invariant waters that interact with at least one of the four binding region loops.
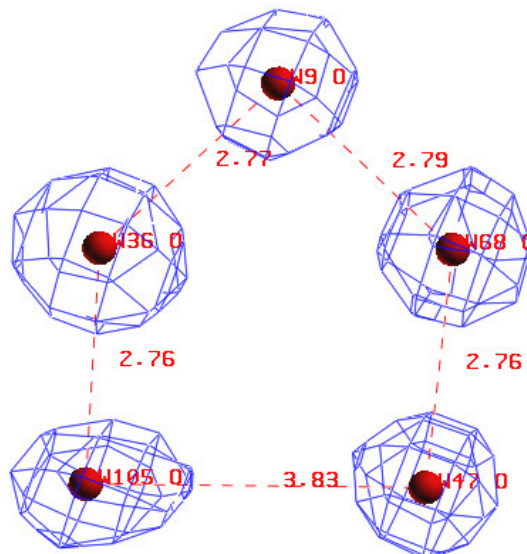
Figure 112. A pentagon-like arrangement of five water oxygen atoms identified in the saccharide binding site region. Four of the five waters are found within the saccharide binding site (W9, W36, W47, W68) and one (W105) in the extended region of the saccharide binding site.

5.4.5 Alternate conformations

Five alternate conformations were identified in the 1.65 Å X-ray structure, three of these are valine residues (Val 65, 75, 188), one a serine (Ser 113) and one an isoleucine (Ile 25). With the exception of Val 75, these were all identified as having alternate conformations in the UHR X-ray structure[50].
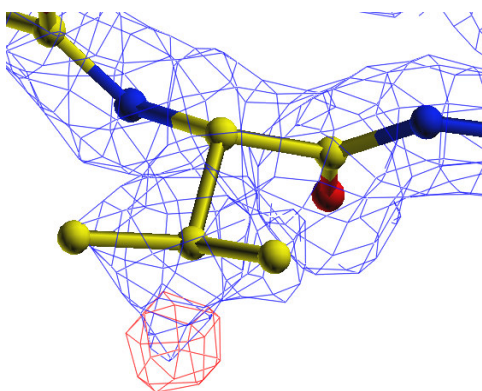
Figure 113.   The side chain of valine 75. The $2F_o - F_c$ electron density is shown in blue (at 1.5 r.m.s.) and the $F_o - F_c$ electron density which indicates the alternate conformation is shown in red (at 4 σ)
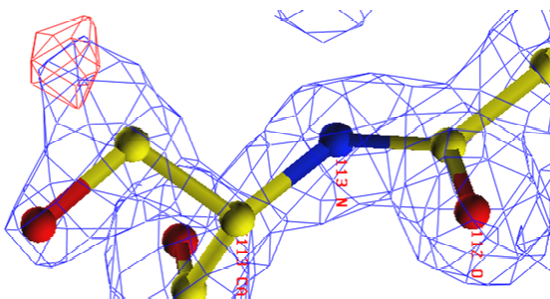
Figure 114.   The side chain of serine 113. The $2F_o - F_c$ electron density is shown in blue (at 1.5 r.m.s.) and the $F_o - F_c$ electron density which indicates the alternate conformation is shown in red (at 4 σ)
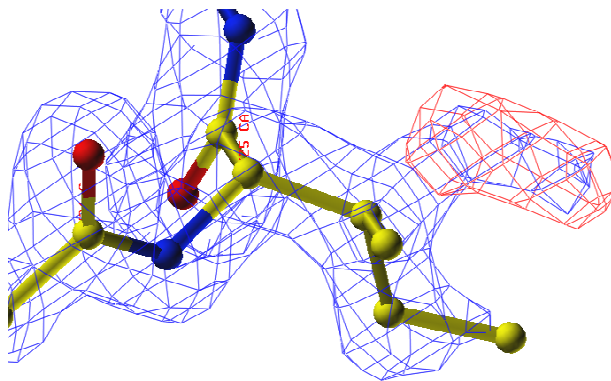
Figure 115.   The side chain of isoleucine 25.   The $2F_o - F_c$ electron density is shown in blue (at 1.5 r.m.s.) and the $F_o - F_c$ electron density which indicates the alternate conformation is shown in red (at 4 σ)

5.5   Analysis of the neutron model

The final structure contains a monomer in the asymmetric unit, of molecular weight 25 kDa.   Each monomer of the classical *I*222 tetramer is made from a single polypeptide chain of 237 amino acid residues, all of which are modelled, and contains one Mn^{2+} ion at the *S*1 site and one Ca^{2+} ion at the *S*2 site.   No metal ion was identified at the *S*3 site in the neutron model.   A total of 227 water molecules were modelled around the monomer.   The unit cell dimensions are; a = 89.16 Å, b = 86.129 Å, c = 61.594 Å, $α$= $β$ = $γ$ = 90°.   At room temperature[51] the unit cell dimensions were found to be; a = 89.11 Å, b = 87.58 Å, c = 63.26 Å, $α$= $β$ = $γ$ = 90°.

Total no. of atoms = 4127

Total no. of protein atoms = 3564

Total no. of water molecules = 227

Total no. of D_2O molecules = 167

Total no. of O-only water sites = 60

Others = 2 metal ions (one Ca^{2+} ion and one Mn^{2+} ion)

Resolution range = ∞ to 2.5 Å

R-factor = 25.1%

R-free = 32.7%

Ramachandran core/additional (%) = 85.6/13.9

r.m.s. B-factor for 237 main-chain residues = 0.89

Mean B-factor for all protein atoms ($Å^2$) = 24.9

Mean B-factor for all side chain atoms ($Å^2$) = 25.0

Mean B-factor for all main chain atoms = ($Å^2$) = 24.6

Mean B-factor for all water molecules ($Å^2$) = 19.2

Mean B-factor for 167 D_2O molecules ($Å^2$) = 17.6

Mean B-factor for 60 O-only water sites ($Å^2$) = 32.2

Mean B-factor for all side chains and waters ($Å^2$) = 23.8

Mean B-factor for all atoms ($Å^2$) = 24.1

Table 50.   Summary of the 2.5 Å resolution neutron model of saccharide-free concanavalin A at 12K.

Total no. of atoms = 4010

Total no. of protein atoms = 3564

Total no. of water molecules = 148

Total no. of D_2O molecules = 88

Total no. of O-only or D-O water sites = 60

Others = 2 metal ions (one Ca^{2+} ion and one Mn^{2+} ion)

Resolution range = ∞ to 2.4 Å

R-factor = 27.0%

R-free = 30.1%

Bond distance r.m.s. deviation = 0.018 Å

Bond angle r.m.s. deviation = 2.2°

Ramachandran core/additional (%) = 87.5/12.5

r.m.s. B-factor for 237 main-chain residues = 3.44

Mean B-factor for all protein atoms ($Å^2$) = 20.4

Mean B-factor for all side chain atoms ($Å^2$) = 21.5

Mean B-factor for all main chain atoms = ($Å^2$) = 17.5

Mean B-factor for all water molecules ($Å^2$) = 43.0

Mean B-factor for 88 D_2O molecules ($Å^2$) = 37.8

Mean B-factor for 60 O-only water sites ($Å^2$) = 50.6

Mean B-factor for all side chains and waters ($Å^2$) = 24.6

Mean B-factor for all atoms ($Å^2$) = 22.9

Table 51.   Summary of the 2.4 Å resolution neutron model of saccharide-free concanavalin A at 293K.

The Ramachandran plot[256] for the neutron saccharide-free concanavalin A model at 12K is shown in **Figure 116** and was calculated using PROCHECK[257]. There is one residue, His 121, in the disallowed region and this is labelled on the plot. His 121, as already noted from the X-ray model refinement, is located in an extremely mobile loop section (118-122) and is found to be one of the most disordered residues modelled in all crystal structures of concanavalin A, as illustrated by the residue's high B-factor (average B_{121} = 58.4 $Å^2$). Furthermore, His 121 is the only residue with atoms of B-factor greater than 60 $Å^2$. 178 amino acids (85.6%) are found within the most favoured regions (A, B, L) and 29 residues (13.9%) are found within the additionally allowed regions (a, b, l, p) illustrating that the model has good

geometry. There are no distorted main chain bond lengths or planar groups. When comparing the geometry of the model with that of the room temperature neutron model, we see that the room temperature model has slightly better geometry, with more residues within the core region of the plot (room temperature neutron model 87.5% core, 12K neutron model 85.6% core). However, the room temperature model contained two distorted main-chain bond angles (N-CA-C for Leu 9 and Phe 212), whereas now the only distorted main-chain bond angle is for Glu 122 which is part of the disordered loop 118-122.
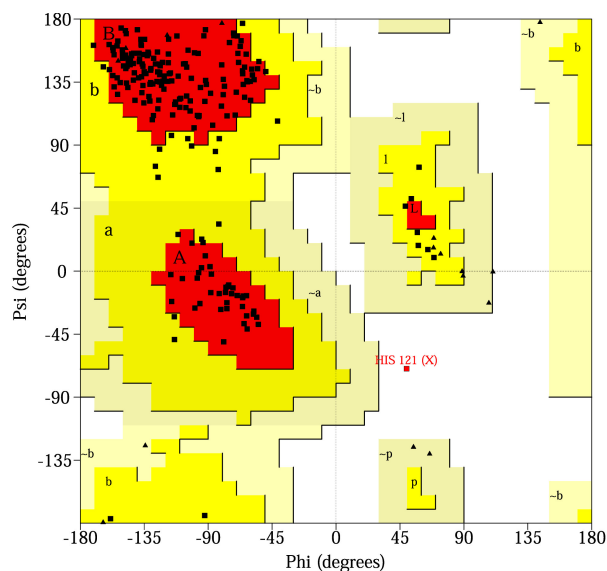
Figure 116.   The Ramachandran plot for the 2.5 Å resolution neutron model of saccharide-free concanavalin A at 12K.

The overall root-mean-square deviations (*x, y, z*) between the room temperature neutron model and the 12K neutron model were calculated as follows:

1. Between all protein atoms, r.m.s. deviation = 0.694 Å

2. Between all main chain protein atoms, r.m.s. deviation = 0.315 Å

3. Between all side chain protein atoms, r.m.s. deviation = 0.876 Å

4. Between all α-carbon protein atoms, r.m.s. deviation = 0.279 Å

The residues that exhibit the largest differences between the 2.4 Å room temperature neutron model and the final 2.5 Å neutron model at 12K of saccharide-free concanavalin A are listed, those given have root-mean-square deviations greater than 1 Å.

Main chain large r.m.s. deviations;

Ala 1, Asn 237

Side chain large r.m.s. deviations;

Ala 1, Asp 16, Lys 36, Met 42, Lys 46, Arg 60, Asn 69, Thr 74, Asp 82, Ser 113, Ser 117, Asn 118, Thr 120, Asn 124, Met 129, Asn 131, Gln 132, Ser 134, Asp 151, Arg 158, Ser 161, Asn 162, Ser 164, Val 188, Ser 223, Asn 237

The amino acid residues above are all found within loop structures and/or on the surface of the protein, therefore, they have greater conformational freedom than those residues which are buried and/or involved in one of the 3 β-sheets. Half of the large side-chain deviations are from serine or asparagine residues (13/26, six for each). For the serine residues this is because the hydroxyl group is free to rotate and can take up many conformations. As the hydroxyl group contains an exchangeable hydrogen atom, the position of this atom has been refined against the neutron data and thus the large number of serine residues exhibiting large side-chain deviations is unsurprising. For the asparagine residues, the neutron density can aid in the correct assignment of the side chain groups (C=O and ND_2). With X-rays, both these groups have similar scattering signals whereas with neutrons the ND_2 group has a much larger scattering signal than the carbonyl group due to the enhanced scattering from the two deuterium atoms as well as the nitrogen atom. Therefore, by examination of the neutron density it can be possible to reassign or flip the two groups of the asparagine side-chain in order to correctly fit the nuclear density (see **Figure 118**).
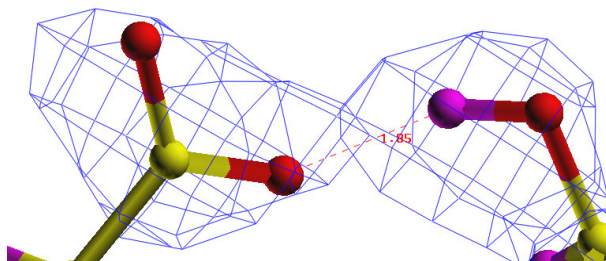
Figure 117.   The $2F_o - F_c$ neutron density (at 1.5 r.m.s) for the side-chain carboxyl group of Asp 28 and hydroxyl group of Ser 34. A hydrogen bond is formed between the D atom of the serine hydroxyl and an oxygen of the carboxyl group of the aspartic acid.
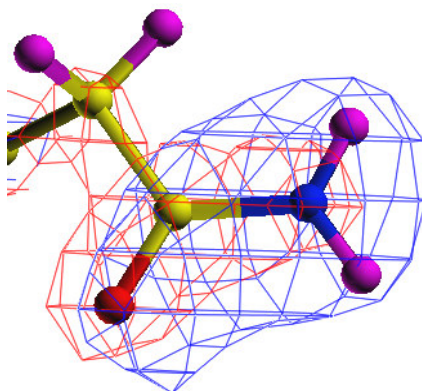
Figure 118.   The $2F_o - F_c$ neutron density and the $2F_o - F_c$ electron density (both at 1.5 r.m.s) for the side-chain of Asn 131. The correct orientation of the side-chain can be assigned more easily with neutrons than X-rays due to the larger scattering signal from the ND_2 group compared to the carbonyl group.

The residue exhibiting the largest r.m.s. deviation is Ala 1 (main-chain r.m.s. deviation = 1.015 Å, side-chain r.m.s. deviation = 2.779 Å), this residue is at the N-terminus of the polypeptide chain and therefore, is more flexible than those residues within the core of the protein. As can be seen from the B-factor versus residue graph, for the 12K neutron concanavalin A model (see **Figure 119**), this residue also has a high B-value (average $B_{Ala\ 1} = 37.6$ Å^2) with only lysine residues or those in the disordered loops having higher values.
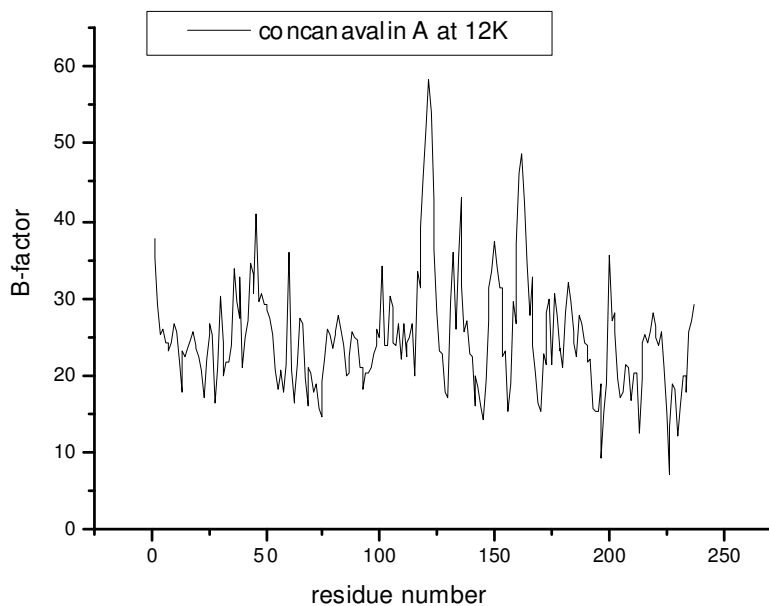
Figure 119.   The B-factor versus residue plot for the 2.5 Å saccharide-free form of concanavalin A at 12K.
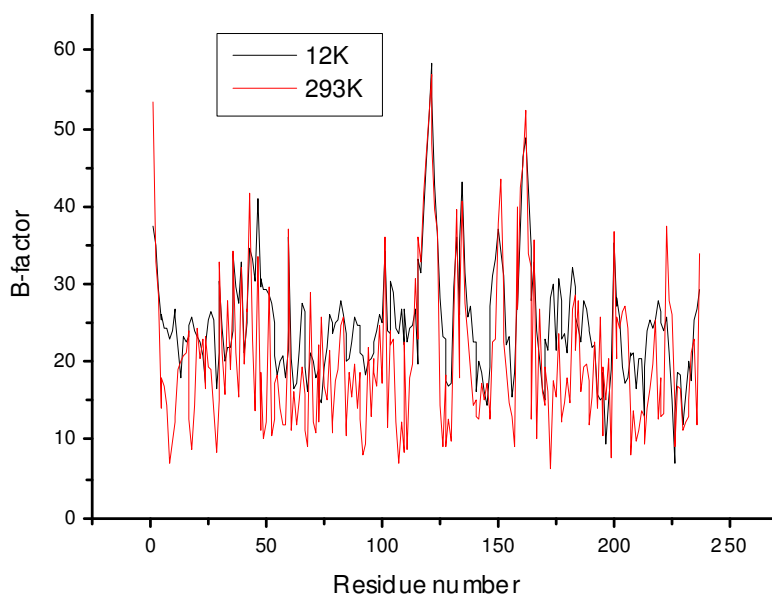
Figure 120.   The B-factor versus residue plot for the neutron models of saccharide-free concanavalin A at 12 K and 293K.

In agreement with the 1.65 Å X-ray structure, the areas which are found to have high B-factors are the loop sections. In particular, residues 120-122 and 162 exhibit the highest B-factors of above 50 $\mathring{A}^2$ and these are labelled in **Figure 121**.
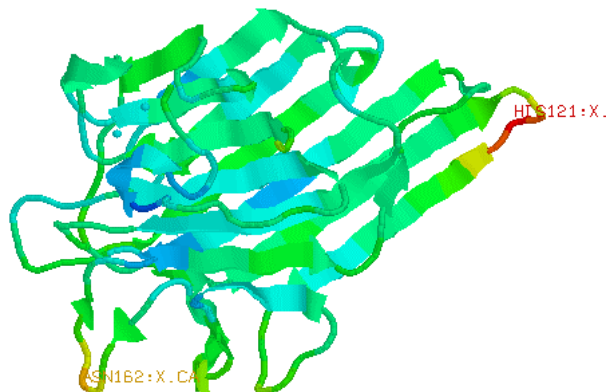
Figure 121.　The protein chain backbone of the con A monomer, coloured according to temperature factor of the residues. The colour coding is as in **Figure 109**.

Certain lysine residues exhibit reasonably high B-factors (Lys 39, 46, 101, 135, 200) of above 40 $\mathring{A}^2$, due to the flexible, long side-chain having larger mobility and thus is generally more disordered.
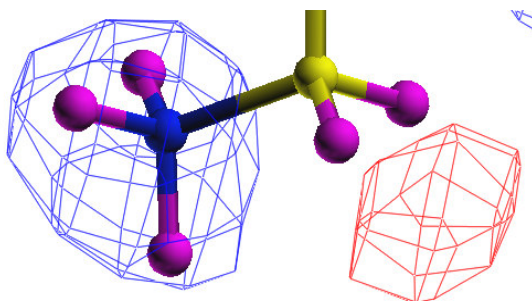
Figure 122.　The $2F_o - F_c$ neutron density in blue at 2 r.m.s. and the $2F_o - F_c$ neutron density in red at -2 r.m.s. for part of the side chain of lysine 35. The negative scattering signal for hydrogen is observed for the aliphatic hydrogen atoms.
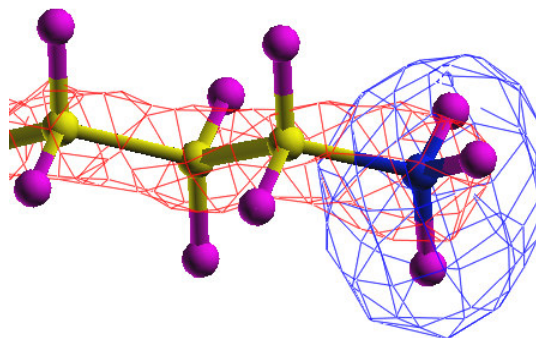
Figure 123.   The $2F_o - F_c$ neutron density in blue and the $2F_o - F_c$ electron density
in red (both at 2 r.m.s.) for the side chain of lysine 35. The ND_3
group has a large scattering signal with neutrons and thus exhibits
strong nuclear density, in contrast to the long aliphatic side chain
nuclear density which has been cancelled out by contributions from
the positive and negative nuclear scattering signals for C and H
respectively. The complementarity of the two techniques is illustrated
by the excellent agreement between the maps.

The protein residues with the lowest B-factors are found to be two Phe residues (197
and 213) and Thr 226. All three of these residues are buried and involved in several
hydrogen bond interactions with other residues.
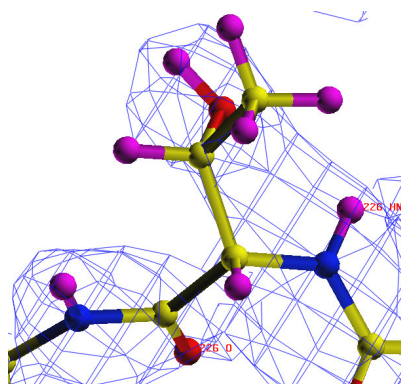
Figure 124.   The $2F_o - F_c$ neutron density for threonine 226 at 2 r.m.s. illustrates
that the side-chain hydroxyl group hydrogen atom has exchanged for
deuterium, as have the main chain amide hydrogen atoms.
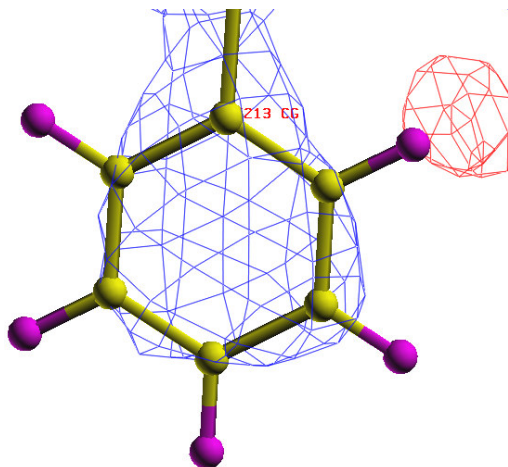
Figure 125. The $2F_o - F_c$ neutron density in blue at 1.8 r.m.s. and the $2F_o - F_c$ neutron density in red at -2.5 r.m.s. for phenylalanine 213. The negative nuclear density for one of the aliphatic hydrogen atoms is visible.
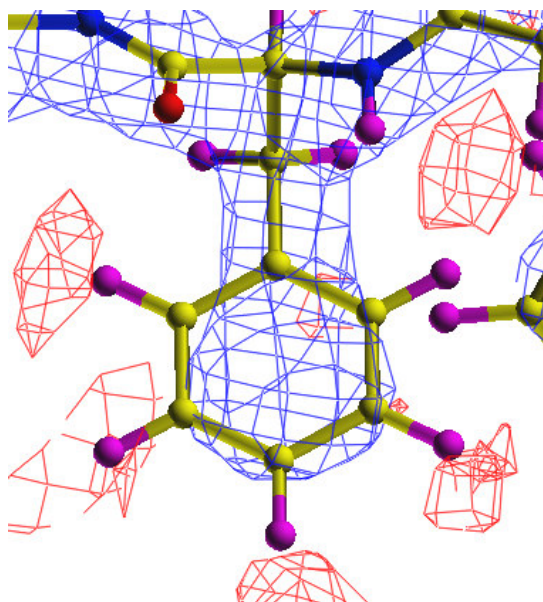
Figure 126. The $2F_o - F_c$ neutron density in blue at 1.8 r.m.s. and the $2F_o - F_c$ neutron density in red at -2.5 r.m.s. for phenylalanine 197. The negative nuclear density for four of the aliphatic hydrogen atoms is visible.

In general the trend in B-factors per protein residue is similar between the two neutron structures, even though they are at different temperatures. In each model the surface exposed loops have the highest B-factors, and buried residues involved in several hydrogen bonding interactions have the lowest. The room temperature structure shows larger variations in B-factor across the protein chain than the 12K structure, as seen from the r.m.s. in B-values for protein residues for the two structures (r.m.s. B for 237 main-chain residues; 12K = 0.89 and 293K = 3.43). In both the 12K and 293K neutron models, the residues Ala 1, Lys 135 and Asn 162 appear to be highly disordered as they have high B-factors at both temperatures. Moreover, loop 118-122 exhibits the highest B-factors in both.

Figure 127 plots the average B-factors for protein and water atoms in different models of concanavalin A at various temperatures. The downward trend in B-factors for the waters, at ever lower temperatures, makes physical sense. This can be attributed to the reduction in dynamic disorder as the temperature is decreased. The final upward trend for the B-factors of the protein atoms between 100K and 12K must be due to the fact that the protein atom coordinates were not refined at 12K (B-factors only) against the neutron data alone, whereas the waters (coordinates and B-factors) were.
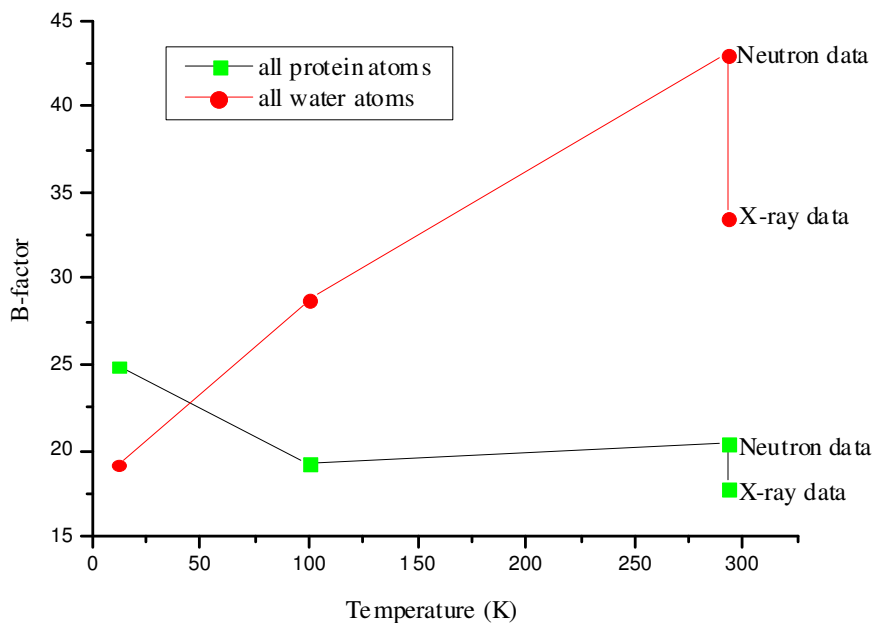
Figure 127.   Plot of the average B-factors for all protein atoms and all water atoms in saccharide-free concanavalin A at various temperatures. The four models used are the 12K 2.5 Å neutron model, the 100K 1.65 Å X-ray model, the 293K 2.4 Å neutron model and the 1.8 Å X-ray model.

The neutron model B-factors are overall slightly higher than those calculated for the refined X-ray model, nevertheless the same overall shape to the curve is seen for both.
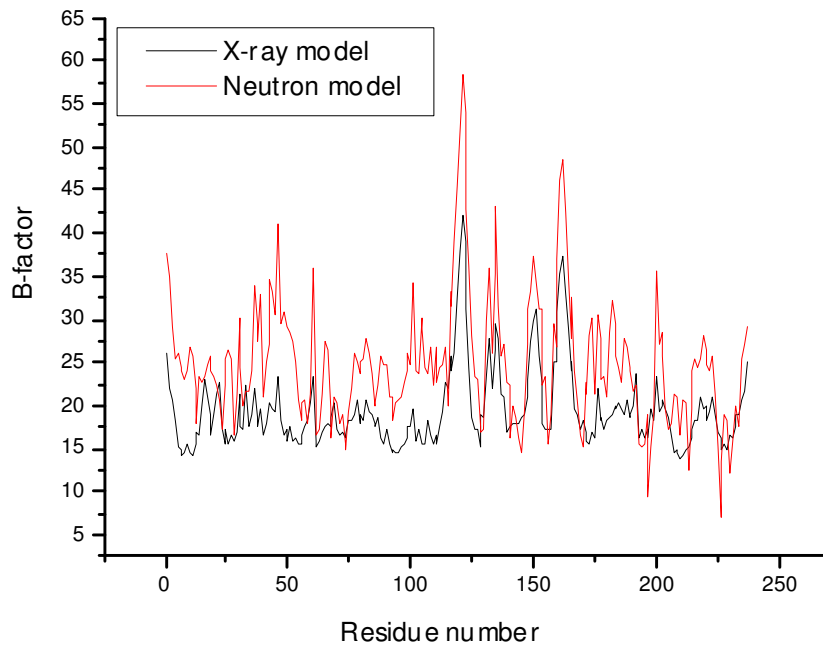
Figure 128.   B-factor versus residue plot for the X-ray and neutron models of saccharide-free form of concanavalin A at 100K and 12K respectively.

5.5.1 The saccharide binding site at 12K

The room temperature neutron study revealed the positions and orientations for three bound waters (as D_2O) at the saccharide binding site. However, the clarity of the nuclear density at these three individual waters was poor *i.e.* not revealing them singly but rather as a cluster (see **Figure 129**). One of the main aims of this study was to improve the definition of the nuclear density for these water molecules. If the water molecules are revealed with individual, distinct nuclear density, more confidence can be placed on the refined D-atom positions. These details can then be incorporated into concanavalin A models in the future, giving rise to more accurate molecular modelling and dynamics calculations.
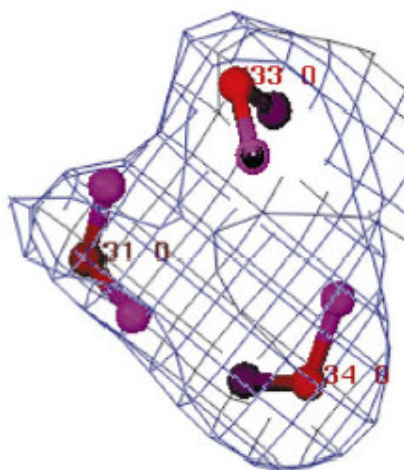
Figure 129.   The $2F_o$ - F_c neutron map (at 1.5 r.m.s.) showing the three waters at the saccharide-binding site at room temperature[51]. The waters are not observed as distinct, single water molecules.

The five water molecules identified in the 1.65 Å X-ray model have been observed in the 2.5 Å neutron model also. However, unlike the X-ray determination in which only the oxygen atoms were revealed, the neutron data has revealed the positions and orientations of the water D-atoms as well.

Saccharide binding site waters	2.5 Å neutron structure at 12K	1.65 Å X-ray structure at 100K	1.8 Å X-ray structure at 293K	2.4 A neutron structure at 293K
WI	W2 (1.2 Å^2)	W4 (14.0 Å^2)	W13 (14.5 Å^2)	W33 (29.8 Å^2)
WII	W32 (1.0 Å^2)	W9 (15.5 Å^2)	W89 (17.7 Å^2)	W34 (20.5 Å^2)
WIII	W124 (4.6 Å^2)	W36 (15.2 Å^2)	W6 (32.4 Å^2)	W31 (23.0 Å^2)
WIV	W65 (3.9 Å^2)	W47 (16.2 Å^2)		
WV	W49 (1.0 Å^2)	W68 (14.3 Å^2)		

Table 52.   The B-factors for the five water sites identified in the monosaccharide-binding site from the room temperature X-ray and neutron models and the cryo-temperature X-ray and neutron models.

Table 52 above gives the B-factors of the waters identified at the saccharide binding site, in neutron and X-ray models, both at cryo-and room temperatures. The B-factors for the five waters have clearly been reduced by collecting the data at lower temperatures. For the five waters at 12K, their B-factors are all less than 5 Å^2, whereas at room temperature, the three waters identified have B-factors all greater than 20 Å^2. This reduction in B-factor indicates a reduction in dynamic disorder. This has helped improve the definition of the nuclear density of the waters, thereby aiding the elucidation of an accurate hydrogen bonding scheme for these waters.

From analysis of the positions and orientations of the five water molecules in the 12K model, it has been possible to suggest a hydrogen bonding network between the water molecules themselves and between key protein residues and the waters. Below is a schematic of the hydrogen bonding network identified at 12K.
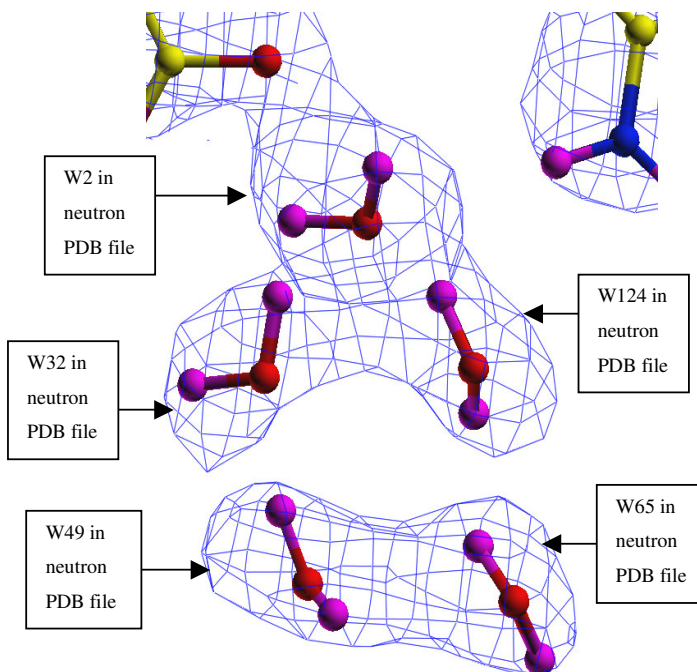
Figure 130.   The $2F_o - F_c$ neutron density (at 1.5 r.m.s) for the five water molecules identified at the saccharide binding site in the neutron model of concanavalin A at 12K.
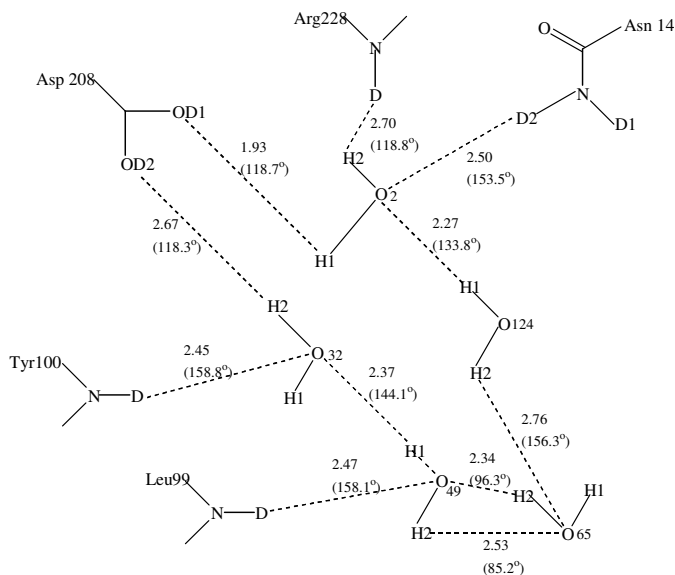
**Figure 131.**   Schematic of the neutron data derived hydrogen bonding network.
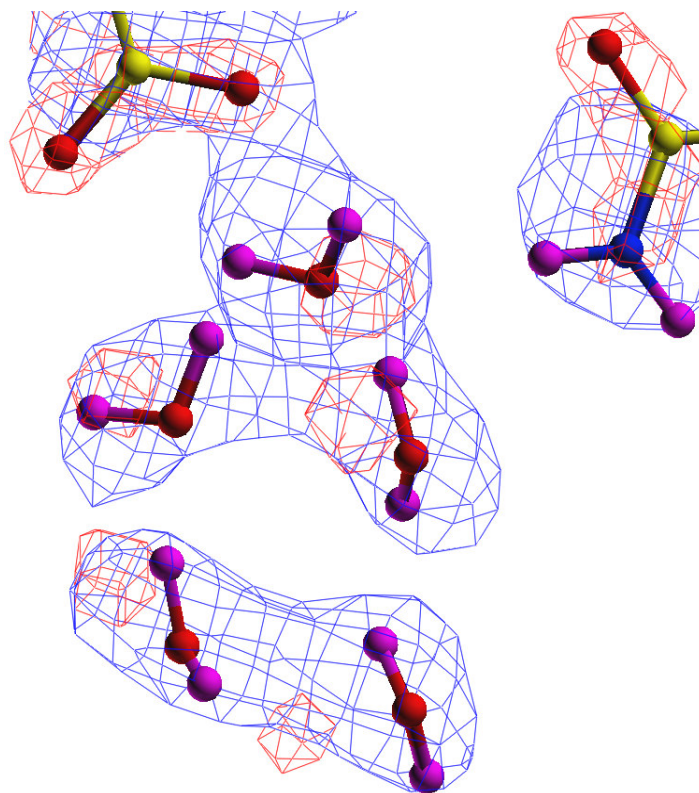
Figure 132.   The $2F_o - F_c$ neutron density in blue (at 1.8 r.m.s) and the $2F_o - F_c$ electron density in red (at 2.5 r.m.s.) for the 5 waters identified in the 2.5 Å neutron and 1.65 Å X-ray models of saccharide-free concanavalin A presented here. In the neutron model at 12K they are observed directly as D_2O, while only their oxygen atom positions are revealed visible from the X-ray data.

The figure above illustrates the complementarity between the two diffraction techniques employed. For well-ordered waters such as the five in the saccharide binding site, there is good agreement in position between the neutron and X-ray data. In particular, the water at the top of this figure (W2 in the neutron PDB file) exhibits excellent agreement in oxygen atom position between the X-ray and neutron data with a shift of only 0.437 Å. The shifts for all five of the waters are given in **Table 53**.

Saccharide binding site waters	Oxygen atom shift between X-ray and neutron models (Å)
WI	0.437
WII	0.747
WIII	0.748
WIV	1.289
WV	1.087

Table 53.   Oxygen atom shifts between the 1.65 Å X-ray model at 100K and the 2.5 Å neutron model at 12K for the five waters identified in the saccharide binding site.
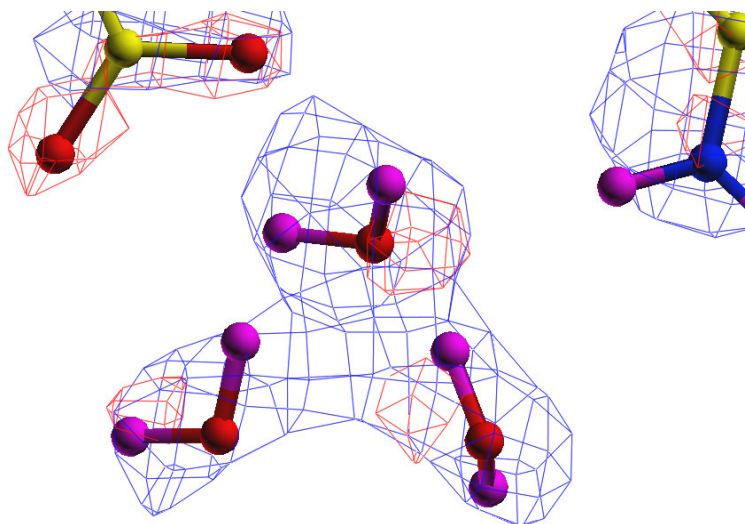
Figure 133.   The $2F_o - F_c$ neutron density in blue (at 1.8 r.m.s) and the $2F_o - F_c$ electron density in red (at 2.5 r.m.s.) for the 3 invariant waters identified at the saccharide binding site in the neutron and X-ray models of saccharide-free concanavalin A at cryotemperature (12K and 100K respectively). The carboxylic acid group of Asp 208 and the amino group of Asn 14 are also shown as they are involved in hydrogen bonds with the saccharide binding site waters.

5.5.2 Metal sites

One specific area of interest for the 12K neutron study was to investigate the orientation of the water molecules that co-ordinate to the Mn^{2+} and Ca^{2+} ions. At room temperature, the D-atoms of the water molecules were shown to be pointing away from the metal atoms[51] and this provided crucial information to aid the interpretation of spectra from an EPR spectroscopic study of the Mn^{2+} binding site in saccharide-free concanavalin A crystals. EPR studies at room temperature of the Mn^{2+} binding site of single crystals of concanavalin A are completely consistent with site symmetry implied by the space group, *I*222. However, in current pulsed EPR studies[258,307,308], the local symmetry at the Mn^{2+} site deviates from the crystal symmetry at low temperatures (4K, 120K and 135K), but returns to the expected *I*222 room temperature symmetry when the crystal is warmed to 250K. It has been suggested that if a water molecule can rotate about the bisector of the H-O-H bond, the orientation of the unshared-pair electrons may have a profound effect on the electronic properties of the Mn^{2+} ion. It has been suggested that this may be the cause of the observed differences in EPR spectra for the Mn^{2+} site at low temperatures. Only low temperature neutron diffraction has the potential to examine the orientation of the water molecules by locating the water hydrogen atoms (as deuterium).

As the four metal site water molecules are crucial for the efficient functioning of the protein and are found to be conserved in not only all concanavalin A crystal structures but also all legume lectin crystal structures in general, one would perhaps expect these waters to be in well-defined orientations and positions in the neutron structure at 12K. However, from analysis of the $2F_o – F_c$ and $F_o – F_c$ neutron density maps at 12K, two of the four metal site water molecules displayed poor nuclear density and the orientation of them could not be defined accurately. The disorder of these specific metal-site waters is illustrated by their associated B-values which are both above 40 $Å^2$. W39 is one of the two water molecules found coordinated to the Ca^{2+} ion and has a B-factor of 40.4 $Å^2$ and W147 is one of the two water molecules found coordinated to the Mn^{2+} ion and has a B-factor of 48.6 $Å^2$. The other water

coordinated to the Mn^{2+} ion (W129) displays strong positive nuclear density and a low B-factor of 11.3 $\mathring{A}^2$, while the other water coordinated to the Ca^{2+} ion (W20) displays only reasonable density, indicated by its mid-range B-factor of 27.3 $\mathring{A}^2$ (see **Figures 134/135**). If we compare the B-factors of the metal site waters at 12K with those for the metal site waters at room temperature, we see that the same two water sites at each temperature have the highest B-factors. However, at room temperature all of the metal site water B-factors are less than 35 $\mathring{A}^2$, and the waters appear to be well-ordered. They all display strong positive nuclear density and are in well-defined positions and orientations (see **Figures 83/84**). The definition of the nuclear density at 12K for the bound water in general appears to have been improved relative to the room temperature data. This is illustrated by the average B-values for all the waters at each temperature (B-factor$_{12K\ Waters}$ = 19.2 $\mathring{A}^2$ and B-factor$_{293K\ Waters}$ = 43.0 $\mathring{A}^2$) and by the improved definition of the saccharide binding site waters at 12K. Yet this improved definition does not occur at the metal sites. The current EPR results show that at low temperatures the site-symmetry at the Mn^{2+} site breaks down and it has been suggested that this is possibly caused by the water molecules rotating in some way. If the water molecules are more mobile and have weaker stabilizing interactions with their surroundings they would then appear in neutron maps as disordered waters with high B-factors and with a lack of good quality boomerang shaped neutron density. In summary, the neutron data at 12K was unable to accurately define the positions of two of the four water molecules at the metal sites, in contrast to the room temperature data, in which the positions and orientations of all four waters were defined. This cannot be used as evidence of symmetry breakdown at the metal sites at low temperatures, however, it does indicate that a certain degree of disorder is evident at the metal sites in the large cryocooled single crystals of concanavalin A at 12K and thereby is consistent with the results from EPR experiments. However, the 100K 1.65 $\mathring{A}$ X-ray results presented in this thesis, and at 0.94 $\mathring{A}$ resolution by Deacon *et al.,* (1997)[50] do not show disorder in these two waters whereas the EPR already gives that indication at 120K and 135K. This is then an inconsistency between the EPR results and the crystallography.
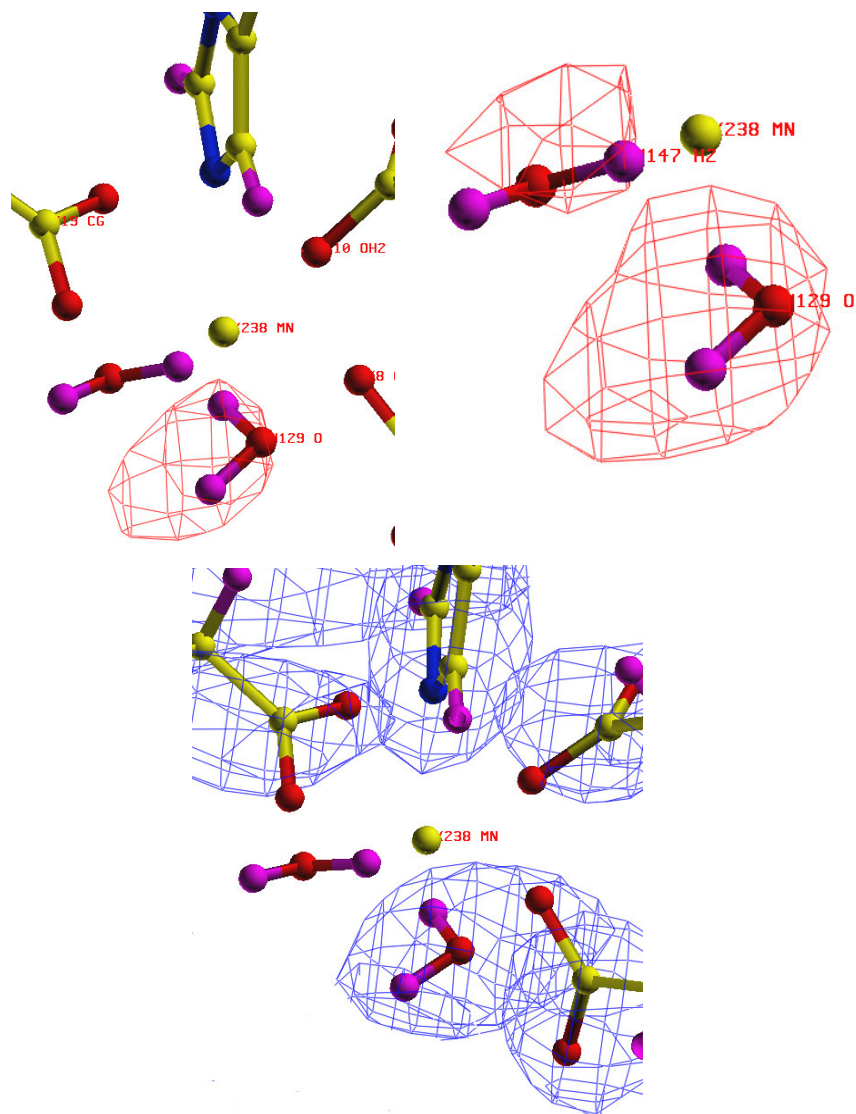
Figure 134. The Mn^{2+} binding site (*S*1) at 12K in concanavalin A. *Top Left,* the $F_o - F_c$ neutron omit map (at 2.5 r.m.s.) for the two waters at the *S*1 site. *Top Right,* the $F_o - F_c$ neutron omit map (at the lower level of 2.0 r.m.s.) for the two waters at the *S*1 site. As can be seen at this level of the map, some nuclear density is visible for W147, however, it is still very weak indicating this site is highly disordered at 12K. *Bottom,* the $2F_o - F_c$ neutron density (at 1.5 r.m.s) is shown. As can be seen, there is no visible density for one of the two water molecules (W147 in the PDB file).
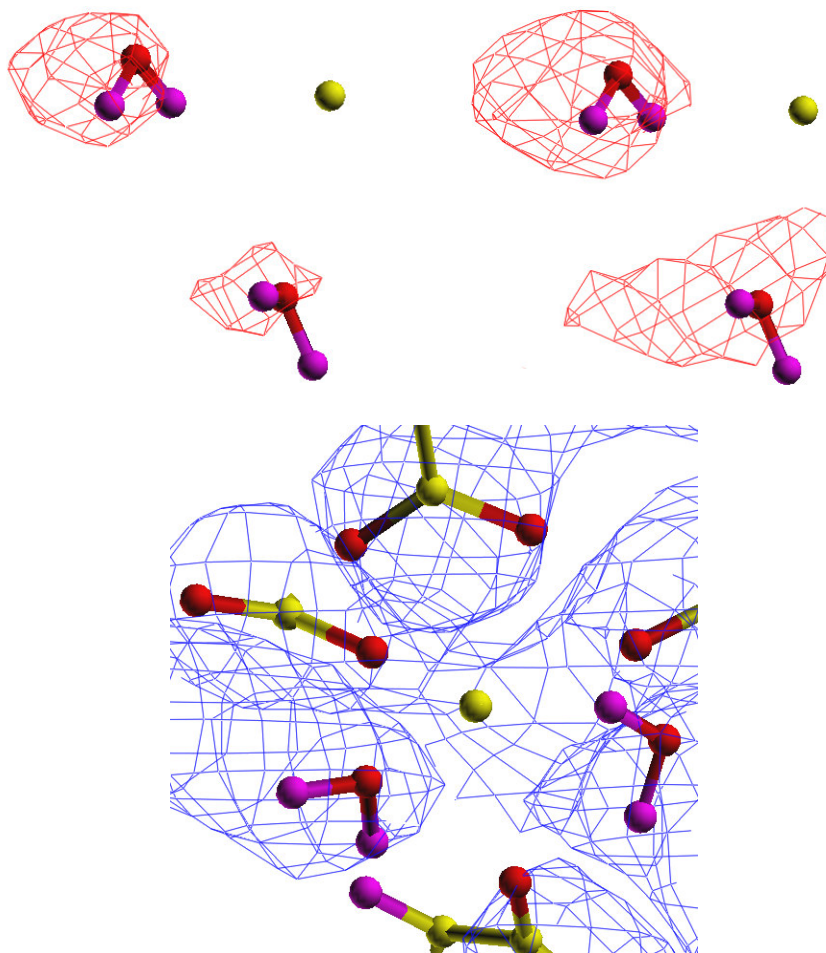
Figure 135. The Ca^{2+} binding site (*S*2) at 12K in concanavalin A. *Top Left,* the F_o – F_c neutron omit map (at 2.5 r.m.s.) for the two waters at the *S*2 site. *Top Right,* the F_o – F_c neutron omit map (at the lower level of 2.0 r.m.s.) for the two waters at the *S*1 site. *Bottom,* The $2F_o$ – F_c neutron density (at 1.2 r.m.s) is shown. As can be seen, there is poor visible density for one of the two water molecules (W39 in the PDB file) and only reasonable density for the other water molecule (W20).

Metal site waters	2.5 Å neutron structure at 12K	1.65 Å X-ray structure at 100K	1.8 Å X-ray structure at 293K (PDB code = 1QNY)[51]	2.4 Å neutron structure at 293K (PDB code = 1C57)[51]
(Mn^{2+}) WA	W129(11.3 Å^2)	W2 (9.7 Å^2)	W8 (9.9 Å^2)	W27 (22.0 Å^2)
(Mn^{2+}) WB	W147 (48.6 Å^2)	W6 (11.0 Å^2)	W9 (13.0 Å^2)	W22 (34.0 Å^2)
(Ca^{2+}) WC	W39 (40.4 Å^2)	W3 (12.9 Å^2)	W10 (9.6 Å^2)	W29 (33.0 Å^2)
(Ca^{2+}) WD	W20 (27.3 Å^2)	W19 (13.0 Å^2)	W11 (13.4 Å^2)	W30 (9.8 Å^2)

Table 54.   The water numbers (as given in the PDB files) and their associated B-factors, for the water molecules coordinated to the metal ions in the X-ray and neutron models of concanavalin A at both room and cryotemperatures.

S1 Mn^{2+} ligand distances (Å)	This study	Habash *et al.,* (2000)[51]
Glu 8 OE2	2.18	2.19
Asp 10 OD2	2.19	2.17
Asp 19 OD1	2.21	2.24
His 24 NE2 (axial)	2.27	2.28
Water molecule (metal-to-oxygen distance) (axial)	W129 2.71	W27 2.28
Water molecule (metal-to-oxygen distance)	W147 2.47	W22 2.17

Table 55.   The manganese ligand distances in the monomer of concanavalin A. Values from Habash *et al*., are given for comparison. The waters listed are numbered as in the PDB file.

The data agrees with that from the 293K neutron model of saccharide-free concanavalin A, in that Glu 8 OE2, Asp 10 OD2, Asp 19 OD1, His 24 NE2 and two

water molecules co-ordinate to the manganese ion. The protein to Mn^{2+} ligand distances agree well between the two neutron studies. These are, in any case, the X-ray determined positions. However, when comparing the positions for the water molecules at the metal sites, there are large differences between the two structures. The 12K neutron model gives Mn^{2+} to water oxygen distances that are significantly longer (by ~0.35 Å) than in the 293K neutron model. An important point to make here is that the 12K water oxygen positions were refined against the neutron data unlike the room temperature study of Habash *et al.,* (2000)[51] in which the oxygen atoms were fixed at the X-ray positions.

The calcium-ligand distances are given (see **Table 56**). Asp 10 OD1/OD2, Tyr 12 O, Asp 19 OD2, Asn 14 OD1 and two water molecules co-ordinate to the calcium ion. When comparing the distances, once again the metal-to-protein ligand distances agree whereas the metal-to-water distances are longer at 12K than 293K (by ~0.4 Å).

$S2$ Ca^{2+}-ligand distances (Å)	This study	Habash *et al.,* (2000)[51]
Asp 10 OD1	2.47	2.48
Asp 10 OD2	2.51	2.47
Tyr 12 O	2.37	2.34
Asp 19 OD2	2.40	2.44
Asn 14 OD1	2.34	2.36
Water molecule (metal-to-oxygen distance)	W20 2.72	W30 2.38
Water molecule (metal-to-oxygen distance)	W39 2.93	W29 2.42

Table 56.   The calcium ligand distances in the concanavalin A monomer. Values from Habash *et al.,* are given for comparison.

If we compare the B-factors of the atoms chelated to the Mn^{2+} in the 12K structure, we see that the protein atoms all have B-values around 23 $\mathring{A}^2$ with very little range (22.8 to 23.8 $\mathring{A}^2$) between them. However, the B-factors of the two waters chelated to the Mn^{2+} ion do not exhibit similar B-values to the protein atoms, with one having a much larger B-factor (W147) and the other (W129) a lower B-factor. Below are given the B-factors of all the Mn^{2+} ligand atoms in both the cryo- and room temperature neutron structures. At room temperature the B-values for the water molecules are more consistent with the B-values for the protein atoms. There is however, more range in B-value between the protein atoms at room temperature.

$S1$ Mn^{2+} ligand B-factors($\mathring{A}^2$)	This study	Habash *et al.,* (2000)[51]
Glu 8 OE2	23.0	13.6
Asp 10 OD2	23.8	9.6
Asp 19 OD1	22.8	22.2
His 24 NE2 (axial)	23.4	35.4
Water molecule	W129 11.3	W27 22.0
Water molecule	W147 48.6	W22 34.0

Table 57.   The manganese ligand B-factors in concanavalin A. Values from Habash *et al.*, are given for comparison.

If we compare the B-factors of the atoms chelated to the Ca^{2+} in the 12K structure, we see that the protein atoms all have B-values around 24 $\mathring{A}^2$, except for Tyr 12 O which has a low B-factor of 13.1 $\mathring{A}^2$. The B-factors of the two waters chelated to the Ca^{2+} ion are higher than those for any of the protein atoms chelated to the Ca^{2+} ion. On the next page are given the B-factors of all the Ca^{2+} ligand atoms, in both the cryo- and room temperature neutron structures. At room temperature the B-values for the water molecules are more consistent with the B-values for the protein atoms, although the water numbered W29 exhibits quite a high B-factor in the room temperature data.

Manganese, like hydrogen, has a negative and small scattering factor with neutrons ($b_H = -0.373$). Indeed, from analysis of the neutron maps, the negative peak from the manganese at 12K is weak and can only be observed at low sigma levels. The relative lack of negative nuclear density at the Mn^{2+} site is indicated partly by the high B-factor of Mn^{2+} of 52.7 $\mathring{A}^2$. Furthermore, the native concanavalin A protein crystals used for the data collection are known to have a mixed population of transition metals at the *S*1 site[276]. The other transition metals that are able to bind at the *S*1 site (*e.g.* Ni, Co, Zn) have positive neutron scattering lengths and would therefore, cancel out the weak scattering signal from manganese. Likewise the residual peak at the *S*1 site in the X-ray map may be due to the mixed population effect (see **Figure 98**).

*S*2 Ca^{2+}-ligand B-factors($\mathring{A}^2$)	This study	Habash *et al.,* (1997)[51]
Asp 10 OD1	24.2	2.0
Asp 10 OD2	23.8	9.6
Asp 19 OD2	22.7	9.3
Tyr 12 O	13.1	15.2
Asn 14 OD1	23.2	10.0
Water molecule	W20 27.3	W30 9.8
Water molecule	W39 40.4	W29 33.0

Table 58. The calcium ligand B-factors in the concanavalin A monomer. Values from Habash *et al.*, are given for comparison.

Metal B-value($\mathring{A}^2$)	This study	Habash *et al.,* (2000)[51]
*S*1 Mn^{2+}	52.7	27.7
*S*2 Ca^{2+}	28.7	2.0

Table 59. B-factors for the manganese and calcium ions in the neutron model of saccharide-free concanavalin A at 12K.

5.5.3 Water structure analysis

The final refined model contains 227 water molecules, 167 of these are modelled as D_2O and 60 are modelled as water oxygen atoms alone. These 227 water molecules are all present in the final $2F_o - F_c$ neutron density maps at greater than 1.5 r.m.s. The average B-factor of all the water atoms is 19.2 $Å^2$. As noted previously, this value is considerably less than that of the room temperature neutron study at 2.4 Å resolution where the 148 waters identified had an average B-factor of 43.0 $Å^2$. Furthermore, the average B-factor of the 167 D_2O molecules at 12K was 17.6 $Å^2$, compared to an average B-factor of 37.8 $Å^2$ for the 88 D_2O molecules at room temperature. The average B-factor for the waters at 12K that could be modelled as oxygen-only was 32.2 $Å^2$. As expected, this value is higher than for the 167 D_2O molecules, due to the weaker nuclear density for these sites. At room temperature, the 60 sites modelled as O or DO had an average B-factor of 50.6 $Å^2$.

At 12K only one of the water sites (W194) has a B-factor over 50 $Å^2$, this water sits at the surface of the monomer and could only be modelled as oxygen, as the density was spherical and weak. Thirteen further water sites have B-factors greater than 40 $Å^2$. All of these are modelled as oxygen alone, except for the two disordered waters (W39 and W147) coordinated to the metals (see **4.5.2**). 43 more waters have B-factors greater than 30 $Å^2$; seventeen of these are modelled as D_2O and 26 as oxygen only. 46 of the waters have very low B-factors of less than 10 $Å^2$; these are all modelled as D_2O.

An analysis of the position of the bound waters within the monomer was carried out. Of the 227, 16 (13 D_2O, 3 O-only) are seen to be completely buried within the protein while 211 (154 D_2O, 57 O-only) are surface waters. The buried waters include the four metal-site waters, a water molecule that stabilizes a β-hairpin structure (W16) and a water molecule (W22) that stabilizes a β-bulge motif. Normally one would expect the average B-factor of the buried waters to be considerably less than those on the surface. However, due to the disorder at the *S*1

and *S*2 metal sites, the buried water molecules in fact have a higher average B-value than the surface waters; (average-$B_{\mathbf{buried}}$ = 24.2 Å^2 and average-$B_{\mathbf{surface}}$ = 18.8 Å^2).
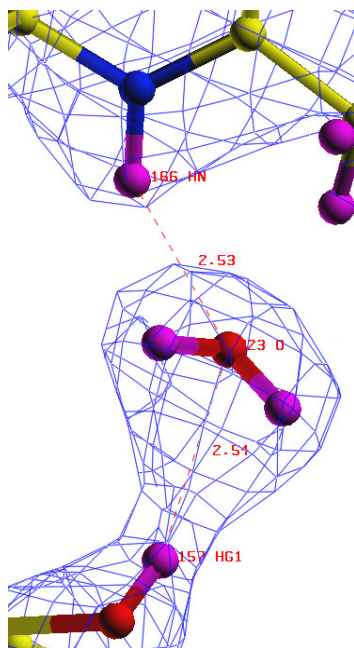
Figure 136. One of the 16 buried waters identified in the 2.5 Å neutron structure at 12K. The $2F_o - F_c$ neutron density is in blue (at 1.2 r.m.s.). The water molecule (W23 in the neutron PDB file) is hydrogen bonded to the main chain amide deuterium atom of Gln 166 and the side chain hydroxyl deuterium of Thr 157.
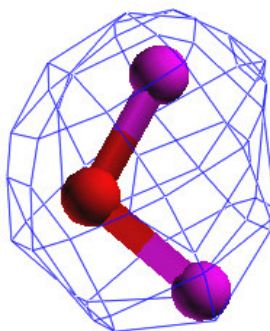
Figure 137. One of the 16 buried water molecules identified at 12K in the 2.5 Å neutron structure of concanavalin A. This water molecule (W1) has a low B-factor (3.2 Å^2) and strong positive nuclear density.

By superimposing the starting 1.65 Å X-ray model with the refined neutron model using LSQKAB, an analysis of the number of waters that correspond *i.e.* are equivalent, between the two structures was achieved. Waters were seen to be equivalent if after superposition of the two sets of co-ordinates the waters are within 1 Å of each other. Of the 227 water oxygen positions, 44 (37 as D_2O and 7 as O-only) are within 1 Å of the 217 waters in the starting model. The average B-factor of these 44 waters is 19.8 $Å^2$ (average B-factor for 37 D_2O = 19.0 $Å^2$, average B-factor for 7 O-only = 32.5 $Å^2$). The average B-factor of the 44 that agree in the starting model being 25.3 $Å^2$. Of the 16 buried waters in the final refined model, 9 of these match the positions of the water oxygen atoms in the starting model. Of the 211 surface waters in the final refined model, 35 of these match the positions of the water oxygen atoms in the 1.65 Å X-ray model. This illustrates that a greater percentage of the buried waters agree with the starting model than the surface waters (56.3% buried agree *cf.* 16.6% surface agree) as would be expected due to the increased mobility in general of surface waters which make less interactions with the protein.

Further comparisons of the water structure of the 12K neutron model were made against both the 2.4 Å neutron model at 293K and the UHR 0.94 Å X-ray model at 110K. Of the 227 water oxygen positions, 29 (24 as D_2O and 5 as O-only) are within 1 Å of the 148 waters in the 2.4 Å neutron model at 293K. The average B-factor of these 29 waters at 12K is 20.5 $Å^2$ (average B-factor for 24 D_2O = 19.5 $Å^2$, average B-factor for 5 O-only = 34.6 $Å^2$). The average B-factor of these 29 waters that correspond at 293K is 38.7 $Å^2$ (average B-factor for 21 D_2O = 37.1 $Å^2$, average B-factor for 8 O-only = 50.8 $Å^2$).

Of the 227 water oxygen positions, 48 (40 as D_2O and 8 as O-only) are within 1 Å of the 319 waters in the 0.94 Å X-ray model at 110K. The average B-factor of these 48 waters at 12K is 19.4 $Å^2$ (average B-factor for 40 D_2O = 18.3 $Å^2$, average B-factor for 8 O-only = 35.7 $Å^2$).

In order to identify the water sites that are conserved in the neutron models and the UHR X-ray structure, those that agree between the two neutron structures (29 water sites) were compared against the positions in the UHR X-ray structure. Of the 29 that agree, 25 of these also agree within 1 Å of the positions in the 0.94 Å X-ray structure. These water sites are therefore, highly conserved as they are present in all of the models tested. These conserved water sites were examined in order to discover where they are situated and what interactions they make with their surroundings. In **Figures 141**-**148**, some of these conserved water sites are shown for the two neutron structures (12K and 293K) and the UHR X-ray structure at 110K. When comparing the orientations of these conserved water molecules in the two neutron structures at 12K and 293K, some of them agree and others do not. For certain sites, the orientation of the water molecules at 12K and 293K are very different, and thus different hydrogen bond interactions are observed (*e.g.* **Figures 141/147**), while in other sites the orientation of the water molecules at 12K and 293K are the same, or at least very similar, in that reasonable hydrogen bonding interactions (angles and distances) with the same protein residues are made (*e.g.* **Figure 143**).

The seven waters located by R. Loris *et al.,* (1994)[140] have been identified in the 12K neutron model and are given in **Table 60**. There are 4 conserved waters at the metal sites, one in the saccharide binding site hydrogen-bonded to 228 N, 208 OD2 and Asn 14 ND2, one linked to Phe175 O, Ala 177 O and Val 91 N (β-bulge) and one linked to Arg 33 O, Ile 29 O and Ser 31 O (β-hairpin).

The seven waters conserved in all legume lectins.	12K neutron structure	100K X-ray structure	293K neutron structure
W1 (*S1*)	W129 (11.3 Å^2)	W2 (12.3 Å^2)	W27 (22.0 Å^2)
W2 (*S1*)	W147 (48.6 Å^2)	W6 (13.8 Å^2)	W22 (34.0 Å^2)
W3 (*S2*)	W39 (40.4 Å^2)	W3 (15.1 Å^2)	W29 (33.0 Å^2)
W4 (*S2*)	W20 (27.3 Å^2)	W19 (15.3 Å^2)	W30 (9.8 Å^2)
W5 (sbs)	W2 (1.2 Å^2)	W4 (16.0 Å^2)	W33 (29.8 Å^2)
W6 (β-bulge)	W22 (30.6 Å^2)	W24 (14.7 Å^2)	W52 (47.4 Å^2)
W7 (β-hairpin)	W16 (37.2 Å^2)	W30 (18.1 Å^2)	W51 (41.0 Å^2)

Table 60.   The seven water sites identified as being conserved in all legume lectin crystal structures. The table gives the water number in the PDB file and the waters B-factor.
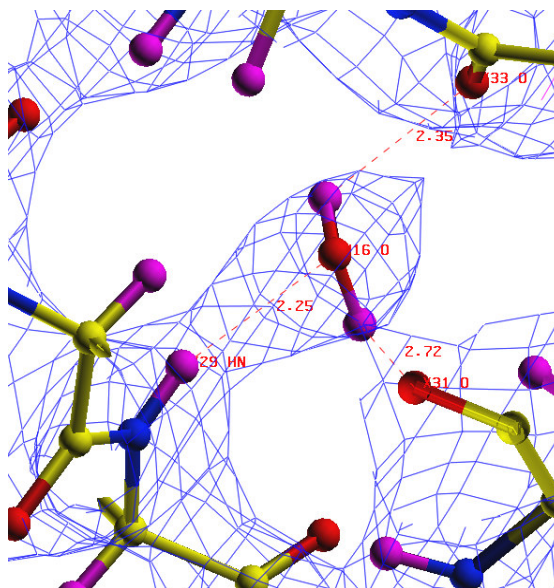
Figure 138. A water molecule (W16 in neutron model) identified from the 12K data found in all the crystal structures of saccharide-free concanavalin A. This water stabilizes a β-hairpin structure within the monomer.
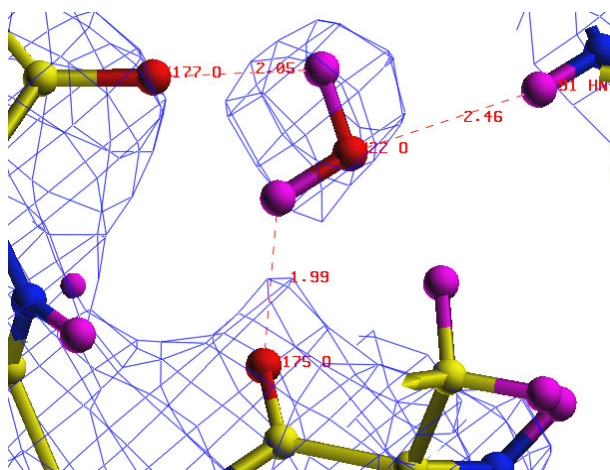
Figure 139. A water molecule (W22 in neutron model) identified from the 12K data found in all the crystal structures of saccharide-free concanavalin A. This water stabilizes a β-bulge structure within the monomer.
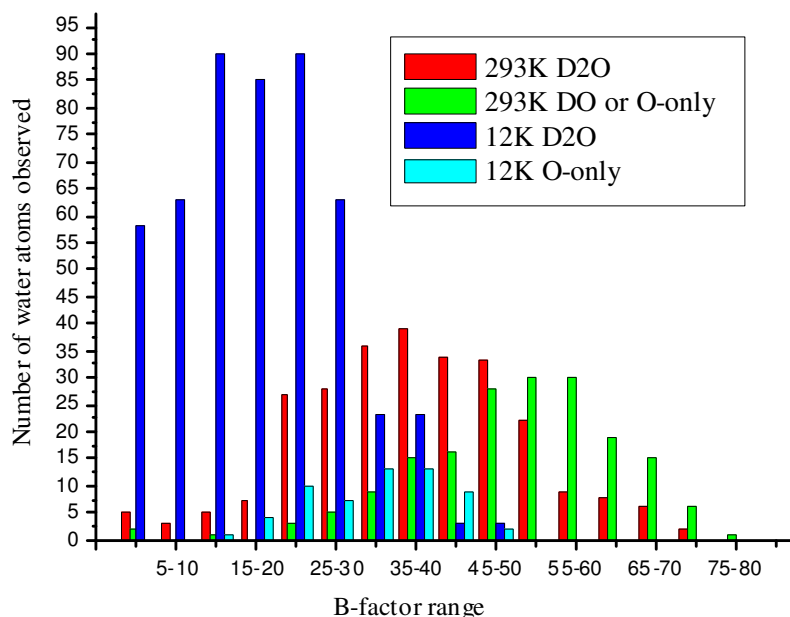
Figure 140.   The number of water atoms observed which have B-factors within a given range, at two different temperatures (12K and 293K)

As can be seen from the plot above (**Figure 140**) at 12K there are a large number of D_2O atoms observed with associated low B-factors. At 293K less D_2O atoms are observed and furthermore, those that are revealed have higher B-factors than at 12K. Therefore, the advantage of collecting the data at cryotemperatures rather than room temperature is that many more waters are observed as D_2O, thus gaining a more complete picture of the hydration shells of the protein. The reduction in dynamic disorder has clearly lowered the B-factors of the bound water and thereby enhanced the definition of the nuclear density. This is illustrated by the improved definition of the nuclear density for the three invariant water molecules in the saccharide binding site.

Figure 141.   *Top*, the $2F_o - F_c$ neutron density (at 2 r.m.s) at 12K for one of the water molecules (W1) identified at an identical position in 3 different models (PDB codes 1NLS, 1C57 and the neutron structure at 12K presented here) of saccharide-free concanavalin A. *Middle,* the $2F_o - F_c$ neutron density (at 2 r.m.s) at 293K for the same water position. The orientation of the water molecule is different at 12K. *Bottom,* the $2F_o - F_c$ X-ray density (at 3 r.m.s) at 110K for the same water position in the ultra-high resolution X-ray model at 0.94 Å resolution.
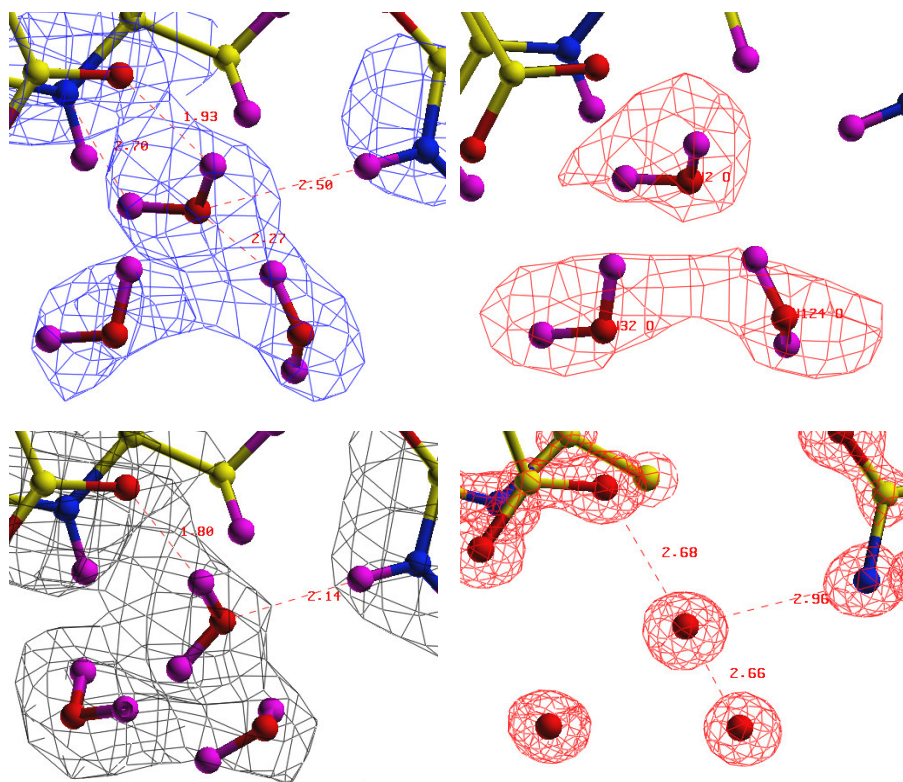
Figure 142.   *Top left*, the $2F_o - F_c$ neutron density (at 2 r.m.s) at 12K for three of the conserved water molecules (W2, W32 and W124) identified at similar positions in 3 different models of saccharide-free concanavalin A. *Top right,* the $F_o - F_c$ omit map (at 2 r.m.s.) for the three waters at 12K. *Bottom left,* the $2F_o - F_c$ neutron density (at 1.5 r.m.s) at 293K for the same 3 water sites. The orientation of the water molecules is different at 293K. *Bottom right,* the $2F_o - F_c$ X-ray density (at 4 r.m.s) at 110K for the same three water sites in the ultra-high resolution X-ray model at 0.94 Å resolution.
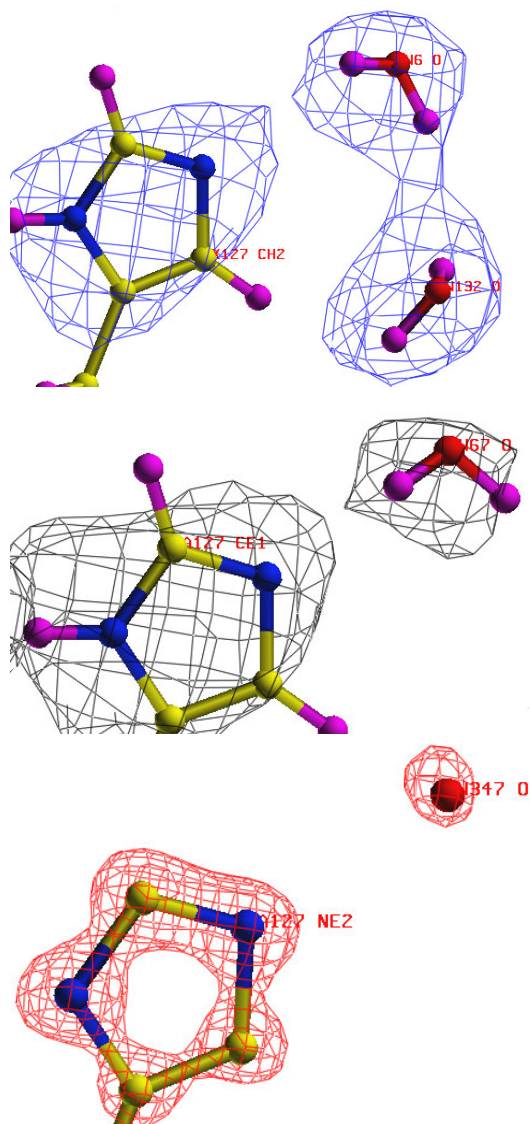
Figure 143.   *Top*, the $2F_o - F_c$ neutron density (at 1.5 r.m.s) at 12K for one of the water molecules (W6) identified at an identical position in 3 different models of saccharide-free concanavalin A.   *Middle,* the $2F_o - F_c$ neutron density (at 1 r.m.s) at 293K for the same water position.   The orientation of the water molecule is quite similar at 12K.   *Bottom,* the $2F_o - F_c$ X-ray density (at 3 r.m.s) at 110K for the same water position in the ultra-high resolution X-ray model at 0.94 Å resolution.
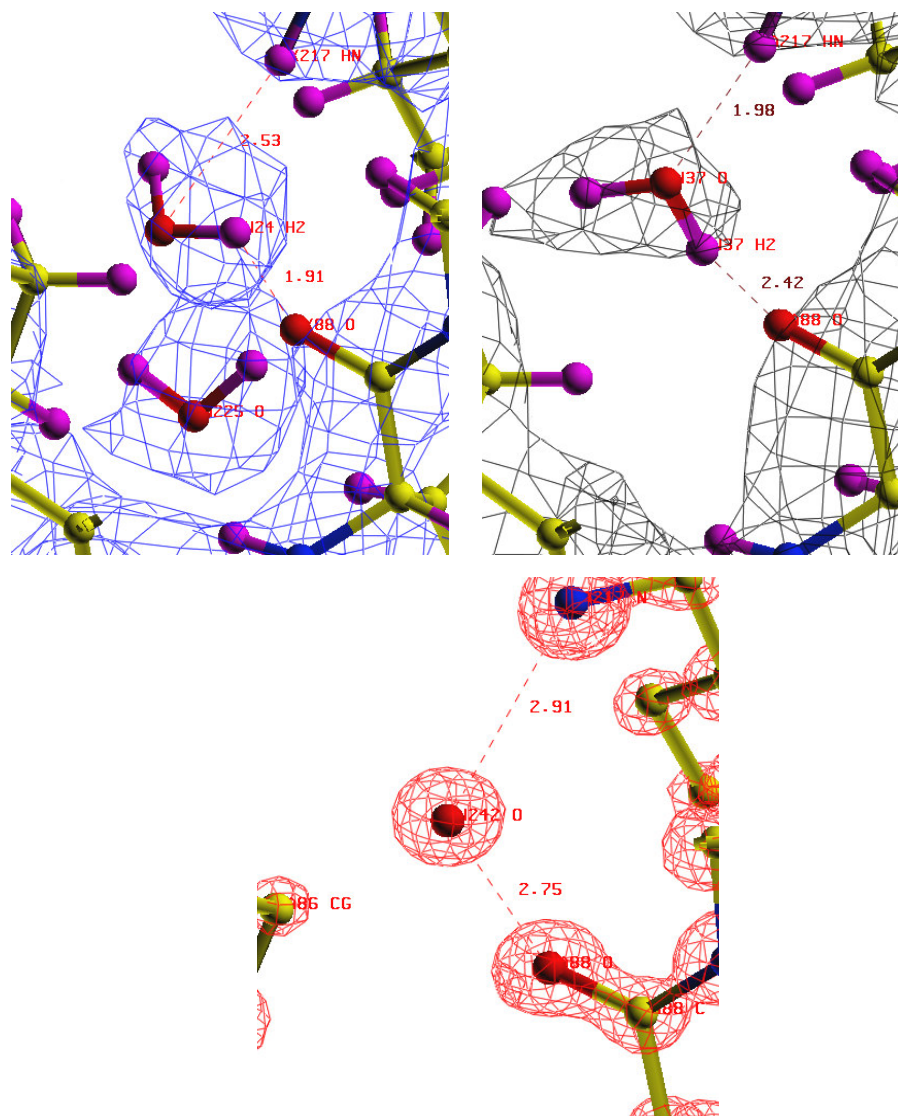
Figure 144.   *Top left*, the $2F_o - F_c$ neutron density (at 1.2 r.m.s) at 12K for one of the water molecules (W24) identified at an identical position in 3 different models of saccharide-free concanavalin A. *Top right,* the $2F_o - F_c$ neutron density (at 1.0 r.m.s) at 293K for the same water position. *Bottom,* the $2F_o - F_c$ X-ray density (at 3.5 r.m.s) at 110K for the same water position in the ultra-high resolution X-ray model at 0.94 Å resolution.
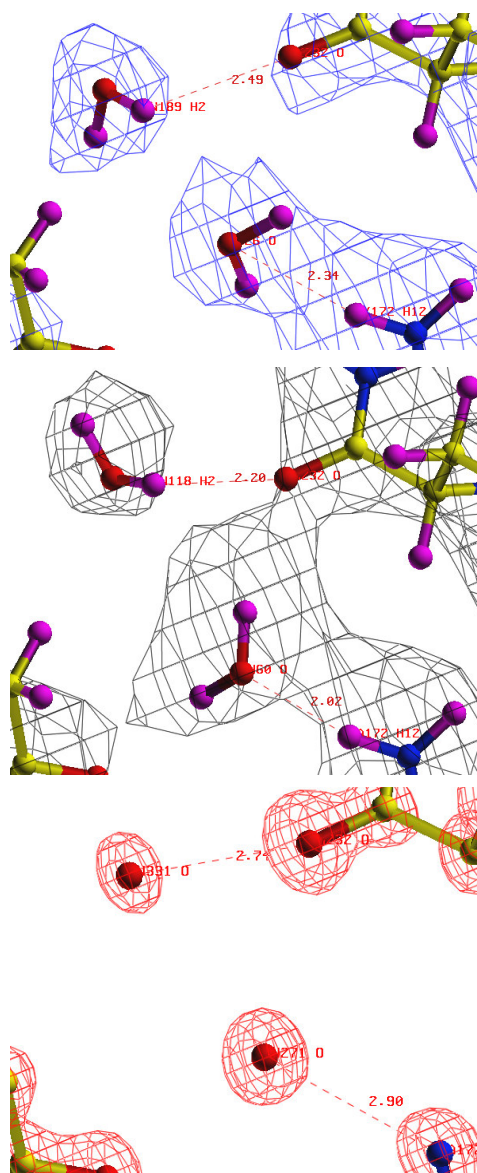
Figure 145. *Top*, the $2F_o - F_c$ neutron density (at 1.5 r.m.s) at 12K for two of the water molecules (W26 and W189) identified at identical positions in 3 different models of saccharide-free concanavalin A. *Middle,* the $2F_o - F_c$ neutron density (at 1.5 r.m.s) at 293K for the same water positions. The orientation of both water molecules is different at 12K. *Bottom,* the $2F_o - F_c$ X-ray density (at 3 r.m.s) at 110K for the same water positions in the ultra-high resolution X-ray model at 0.94 Å resolution.
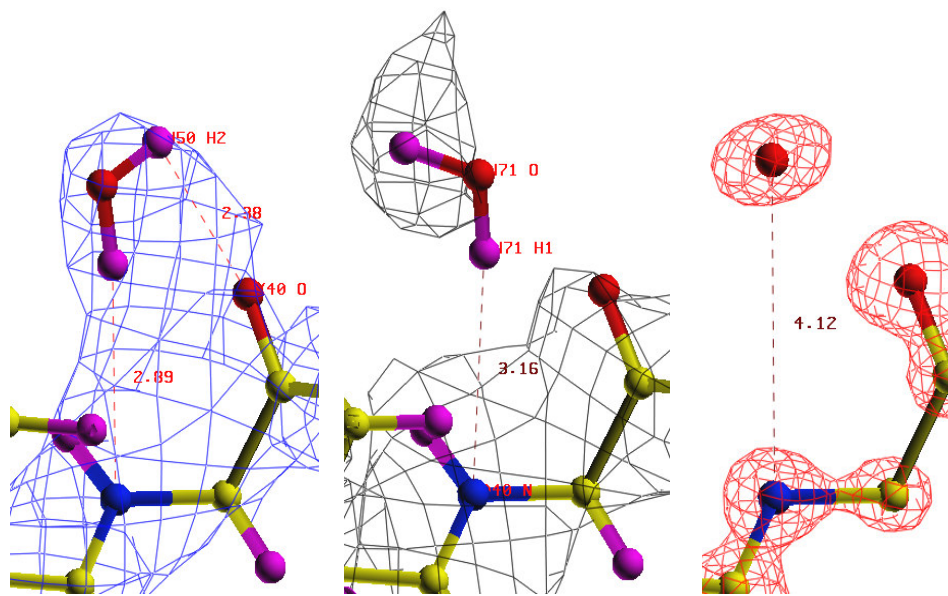
Figure 146.   *Left*, the $2F_o - F_c$ neutron density (at 1.5 r.m.s) at 12K for one of the water molecules (W50) identified at an identical position in 3 different models of saccharide-free concanavalin A.   *Middle,* the $2F_o - F_c$ neutron density (at 1.5 r.m.s) at 293K for the same water position.   The orientation of the water molecule is different at 12K. *Right,* the $2F_o - F_c$ X-ray density (at 3 r.m.s) at 110K for the same water position in the ultra-high resolution X-ray model at 0.94 Å resolution.
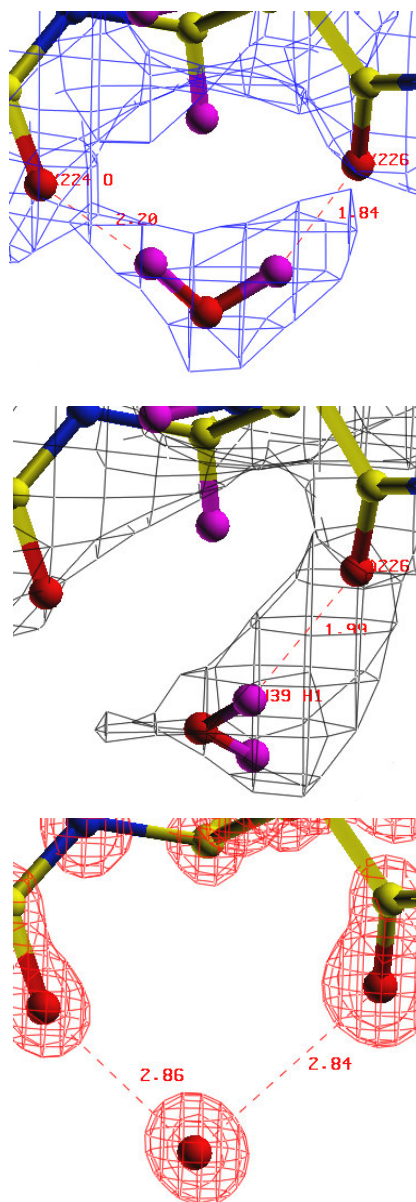
Figure 147.   *Top*, the $2F_o - F_c$ neutron density (at 1.2 r.m.s) at 12K for one of the water molecules (W75) identified at an identical position in 3 different models of saccharide-free concanavalin A.   *Middle,* the $2F_o - F_c$ neutron density (at 1 r.m.s) at 293K for the same water position.   *Bottom,* the $2F_o - F_c$ X-ray density (at 3 r.m.s) at 110K for the same water position in the ultra-high resolution X-ray model.
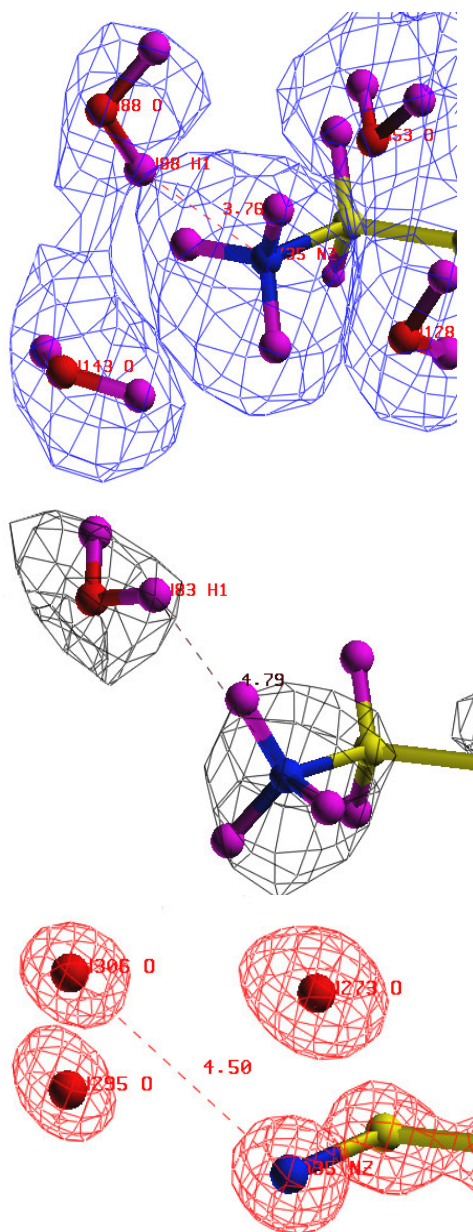
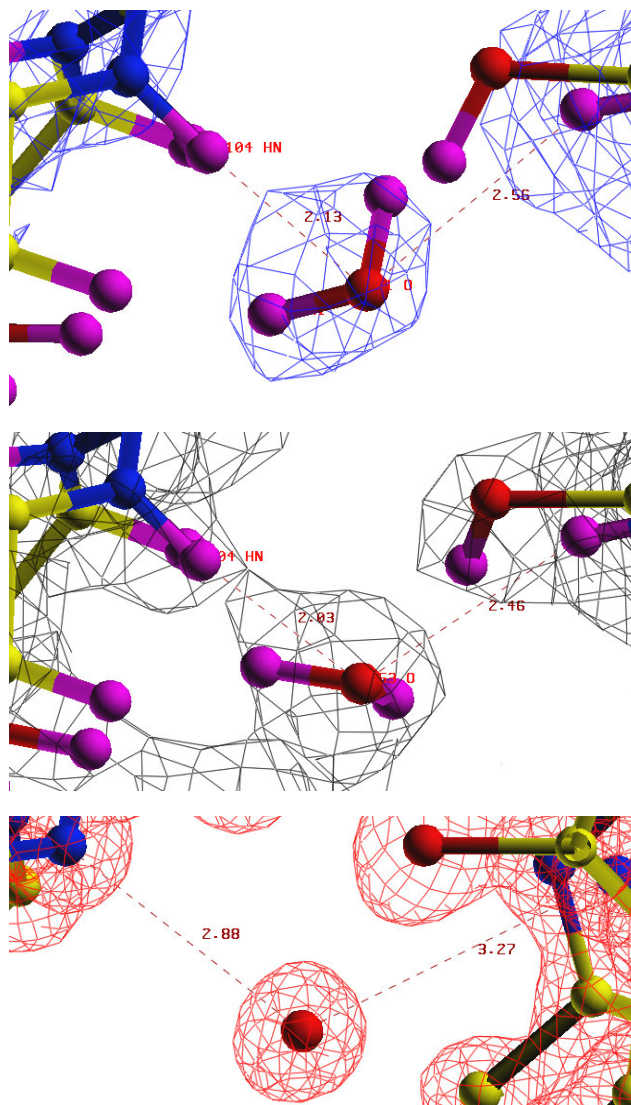**Figure 148.** *Top*, the $2F_o - F_c$ neutron density (at 2 r.m.s) at 12K for one of the water molecules (W88) identified at an identical position in 3 different models of saccharide-free concanavalin A. *Middle,* the $2F_o - F_c$ neutron density (at 1 r.m.s) at 293K for the same water position. *Bottom,* the $2F_o - F_c$ X-ray density (at 3 r.m.s) at 110K for the same water position in the ultra-high resolution X-ray model.

5.6   Concluding remarks

In this chapter, we have shown that it is possible to successfully cryo-cool a large protein crystal of dimensions (1.8 x 1.3 x 0.7mm) and collect both a neutron data set at 12K to 2.5 Å and an X-ray data set at 100K to 1.65 Å resolution. The 12K neutron structure has been compared to the 293K neutron structure[51]. Many more waters as D_2O are identified at 12K than at 293K. Moreover, the water molecules at 12K have lower average B-factors than those at room temperature.   Furthermore, by comparison of the positions of the D_2O waters at each temperature, those that are highly conserved at both temperatures have been identified. These conserved waters are also found to be those waters which have low B-factors, and are well-ordered displaying strong positive nuclear density. These well-ordered waters exhibit small shifts (~0.2 Å) in water oxygen atom positions between the X-ray (1.65 Å structure and 0.94 Å structure[50]) and neutron structures (2.5 Å 12K and the 2.4 Å 293K structures), while larger shifts (~1 Å) are observed for those waters that have high B-factors and weaker nuclear density. Thus the 'X then N' refinement strategy for the heavy atoms and then the exchangeable and water D-atoms has worked, as in the Habash *et al.,* (2000)[51] study. Clearly though, an 'X plus N' joint refinement would still be interesting in the future

More specifically, the nuclear definition of the waters bound in the saccharide binding site has been improved at 12K relative to 293K. At 293K the waters were identified as a cluster *i.e.* not singly, whereas at 12K the definition has been improved, and thus more confidence can be placed in the D-atom positions. This has enabled a hydrogen bonding description to be suggested between key amino acid residues in the saccharide binding site and the conserved water molecules.   In contrast to the saccharide binding site, increased disorder at the metal sites was observed at 12K compared to the room temperature data. This observation agrees with current results from EPR experiments in which the local symmetry at the Mn^{2+} site deviates from the crystal symmetry at low temperatures (4K, 120K and 135K)[258,307,308]. This was not supported by the 1.65 Å X-ray structure at 100K which still showed good density for all four waters.

Chapter 6

Saccharide-bound concanavalin A

6.1   Summary

The three-dimensional structure of the complex between methyl-α-D-glucopyranoside and D_2O cubic concanavalin A has been determined to 2.1 Å resolution. The structure was determined using the co-ordinates of the concanavalin A/methyl-α-D-glucopyranoside complex[8] at 2 Å resolution as the starting model. The final model, refined to 2.1 Å has good geometry, an R-factor of 18.2% and an R-free value of 20.1%. The number of D_2O water oxygen atoms identified is 266. An improvement in the structural precision has been achieved (more data) compared to previously[8]. A total of 43239 unique reflections (99.9% complete) to 2.1 Å resolution were collected compared to 34648 unique reflections (64% complete) to 2.0 Å resolution for the H_2O cubic concanavalin A structure.

The asymmetric unit contains a dimer, each sub-unit is made up of 237 residues which each binds one carbohydrate moiety, *via* hydrogen-bonds, van der Waals and water mediated interactions. The alpha-methyl glucoside is identified at the monosaccharide binding site and is bound by Tyr 100, Asp 208, Asn 14, Arg 228, and Leu 99. It is also bound to Thr 226 and Gly 224 by way of water mediated hydrogen bonds.

Comparisons of the H_2O and D_2O cubic concanavalin A structures were made and the two were found to be similar. Metal to ligand distances and sugar-lectin hydrogen bond interactions were compared and found to be the same within the precision of the structures. However, the B-factors of the D_2O structure are worse than for the H_2O structure and this warrants further study.

6.2   Saccharide-bound concanavalin A

The solution of the cubic crystal form of concanavalin A complexed with the monosaccharide methyl-α-D-glucopyranoside (glucoside) was determined by Harrop *et al.,* (1996)[262]. The crystallization of concanavalin A with glucoside was reported some years beforehand by Yariv *et al.,* (1987)[275] and yet the crystal structure solution of the cubic form proved difficult. Harrop *et al.,* found it necessary to use multiple isomorphous replacement to directly calculate phases and electron density maps for this crystal form, even though the structure of concanavalin A was already well known. Initial molecular replacement calculations in this high symmetry crystal proved impossible to interpret owing to the preponderance of symmetry related peaks in the rotation and translation maps. Crystallographic investigations using a K_2PtCl_4 heavy-atom derivative were misleading as they seemed to indicate that an interpretation of the resulting isomorphous difference Patterson in space group *I*23 was correct, rather than *I*$2_1$3. This also appeared to be consistent with iron uptake experiments in aqueous solution, as well as the high solvent content of these crystals, suggesting that concanavalin A might resemble ferritin, which is composed of 24 identical monomers arranged as a closed shell of 432 point symmetry[276]. In the end, however, the interpretation in space group *I*$2_1$3 was found to be correct. The derivitization involved replacing the transition metal, or the calcium ion, with cadmium, in two different combinations. Cadmium was used to replace cobalt at the transition-metal binding site (*S*1), and to replace calcium at the calcium binding site (*S*2). These exchanges gave relatively small isomorphous signals; 21 e^- for the replacement of cobalt, and 28 e^- for the replacement of calcium. However, these small differences yielded interpretable difference Patterson maps, due to the high quality of the crystals and their stability in the beam. The maps were calculated at 5.4 Å resolution, and were based on phases derived from single- and double-substituted cadmium differences. They revealed a classical concanavalin A tetramer of 222 point symmetry[262].

This crystal structure was therefore solved *via* specific metal-ion exchanges, based on the known behaviour of the concanavalin A molecule. The situation of using

small isomorphous differences in these metal-ion exchange data sets is analogous to synchrotron MAD data collection where small signals are extracted from highly redundant data sets.

Previous studies have addressed the binding of methyl-α-D-glucopyranoside and methyl-α-D-mannopyranoside to concanavalin A[8,49,144]. These two sugars are epimers of each other, differing in stereochemistry at the C2 hydroxyl. This small structural difference nevertheless provokes a completely different crystal space group ($P2_12_12_1$ versus $I2_13$) and crystal habit (plates versus rhombic dodecahedra).

The complex of concanavalin A and methyl-α-D-mannopyranoside (mannoside) was originally reported at 2.9 Å by Derewenda *et al*., (1989)[49] and subsequently at 2.0 Å by Naismith *et al*., (1994)[144]. The mannoside/concanavalin A complex contains an entire tetramer in the asymmetric unit. The structure was solved by molecular replacement using refined coordinates of an *I*222 crystal of saccharide-free concanavalin A. The 2.9 Å study by Derewenda *et al*., (1989)[49] revealed that the mannoside molecule is identified in a shallow crevice near the surface of the protein, bound in the C1 chair conformation. A network of hydrogen bonds were shown to connect the monosaccharide to the protein. The oxygen atoms O3, O4, O5 and O6 of the mannoside are hydrogen bonded to residues Asn 14, Leu 99, Tyr 100, Asp 208 and Arg 228. The oxygen atoms O2 and O1 were found to extend into the solvent, with O2 hydrogen bonded through a water molecule to an adjacent asymmetric unit in two distinct ways. In the first type, the water molecule forms a bridge to Ser 168 and Thr 226 of an adjacent tetramer; in the other it forms a link to Asp 71 of the adjacent tetramer.

$P2_12_12_1$ crystals dissolve if depleted of mannoside. Due to the identification of a water-mediated interaction between different asymmetric units, Derewenda *et al*., suggested that the problem of the crystals dissolving when depleted of mannoside is apparently because the C2 hydroxyl group stabilizes the crystals by cross-linking adjacent, symmetry-related subunits. Furthermore, as glucoside differs to mannoside only in its configuration at the C2 position, the observation that the crystals do not

266

survive exposure to glucoside was suggested to be due to the lack of this water mediated hydrogen bond interaction if glucoside were bound instead of mannoside *i.e.* the O2 of glucoside would not be well placed for cross-linking the asymmetric units in the $P2_12_12_1$ crystals.

Naismith *et al*., (1994) added further detail by improving the resolution limit of the mannoside/concanavalin A complex to 2 Å. It was shown that although the four subunits in the mannoside/concanavalin A complex are not crystallographically equivalent, the protein-saccharide interactions in all four are nearly identical. The differences that do occur between the four sites were shown to be in the water networks that surround each saccharide. These water networks were shown to be involved in crystal packing.

The structure was compared with a saccharide-free structure at 1.6 Å (cobalt substituted at the transition metal site)[145]. In the saccharide-free structure, the saccharide-binding site is occupied by three well-ordered water molecules and the side chain of Asp 71 from a neighbouring molecule. For the mannoside/concanavalin A complex it was shown that the oxygen-atom positions of the hydroxyl groups of the sugar occupy the same positions as the oxygen-atom positions of the well ordered water molecules which exist in the saccharide-free structure. Furthermore, it was shown that on binding of saccharide the side chain of Tyr 100 is directly affected whereby it rotates out of the saccharide binding site to avoid steric clashes with the sugar. Tyr 12 was found to also move slightly towards the saccharide site to form van der Waals interactions with the sugar.

The complex of concanavalin A and methyl-*α*-D-glucopyranoside was originally solved at 5 Å with refinement then carried out using higher resolution data to 3.6 Å[262]. The resolution of the cubic complex was later improved to 2.0 Å by Bradbrook *et al*., (1998)[8]. The glucoside was clearly identified in the saccharide binding site of each monomer. The sugar was found to adopt the same orientation and conformation as mannoside. The binding interactions were also very similar,

with the same hydrogen bonds and van der Waals interactions between the sugar and protein identified.

A wealth of calorimetric data exists for the binding of mannoside and glucoside to concanavalin A[277-279]. Mannoside is found to bind to concanavalin A with a stronger affinity than glucoside and yet from analysis of the crystal structures of both complexes it is unclear why this is the case. The most visible difference between the complexes where shown to involve the C2 hydroxyl. In the mannoside complex the C2 hydroxyl makes three additional van der Waals contacts and an interaction with Thr 226 *via* a water molecule (as described above). For mannoside, the water mediated interaction between Thr 226 and the C2 hydroxyl is part of an inter-tetramer interaction for two subunits. For the other two subunits, where this is not the case, the interaction distances are much longer. Hence, the importance of this interaction in solution as opposed to the crystal is not clear. The van der Waals interactions observed for the sugar C2 hydroxyl in the mannoside complex are absent in the glucoside complex, and could therefore, contribute to mannoside's increased binding affinity. However, summing over the whole binding site, it was found that there are no extra van der Waals interactions for mannoside compared with glucoside.

Bradbrook *et al*., used both static and dynamic modelling studies, in addition to structural comparisons in an attempt to understand these observed thermodynamic differences. The 'static' interaction energies of the substrate in each crystallographic subunit were calculated by molecular mechanics and found to be essentially the same for both sugars, suggesting that dynamic models are required in order to more accurately relate structure to thermodynamics. Therefore, solution models were developed using crystal structures as the starting point. These models were then used to examine features of dynamic protein-sugar binding that cannot be revealed by static structural studies and energy minimisation. Trends in the r.m.s. deviations on sugar atom positions in the molecular dynamics ensemble were compared with crystallographic B-factors. The glucoside r.m.s. values exhibited a fairly small spread, with C7 undergoing the largest average displacement, in agreement with B-

factors. Mannoside r.m.s. values were shown to have a fairly large spread, again in agreement with the crystallographic B-factors. These results were interpreted as being due to the mannoside trajectories accessing different binding sites within the complex, which aren't accessible to the glucoside complex.

Fluctuations in interaction distances between the C2 hydroxyl and residues in the active site within 7 Å were monitored. It was found that possible fluctuating hydrogen bonds in the molecular dynamics ensembles exist, that are not present in the crystallographic study. An important interaction of note was found to be that between Leu 99 N and the C2 hydroxyl. In the mannoside complex the distance of this interaction in the molecular dynamics ensemble was found to range from 2.4 to 5.8 Å, whereas in the glucoside complex this interaction varied from 4.8 to 7.5 Å. This suggests that the mannoside complex is able to form more temporary hydrogen bonds with residues within the binding site, providing a possible explanation of the observed thermodynamic differences. Furthermore, a calculated average interaction energy was found to be more favourable for mannoside than glucoside, by 4.9 ± 3.6 kcal mol^{-1} (comparable with the experimentally determined binding energy difference of 1.6 ± 0.3 kcal mol^{-1}). Hence, the sugar-concanavalin A interaction energies favour mannoside on averaging over a molecular dynamics ensemble, a conclusion not gained from the crystal structure ensembles. However, on consideration of all terms contributing to the binding enthalpy a difference was not found; the more favourable interaction energy term for mannoside (ΔH_{inter}) was found to be cancelled out by a more favourable internal energy term for the glucoside sugar, glucoside being favoured by 5.5 ± 3.1 kcal mol^{-1}.

This study[8] illustrates the difficulties involved in relating structure to thermodynamics and furthermore, highlights areas necessary of improvement when making such a comparison. In order to accurately model the sugar binding process and accurately calculate energy changes involved, we first require the correct positions of *all* the atoms involved in the system, both before and after binding. Therefore, accurate positional assignment of all water, protein and sugar atoms are essential, including hydrogen atoms positions, which are key atoms directly involved

in such processes. This requires the combined forces of X-ray and neutron diffraction techniques so that both the heavy and light atoms of the structure can be directly positioned without the need for inferring atomic positions. Moreover, it is clear that a static crystal structure does not represent the dynamical nature of proteins. In the sugar-binding process, the protein, its bound solvent and the sugar ligand are all moving to some extent and so to represent this more accurately a larger ensemble of crystal structures needs to be determined. Ideally structures from solution studies are required as we have seen that problems involving crystal packing effects can be misinterpreted and given a greater degree of importance than actually is the case in solution. Also, data used for comparison should be collected at physiological pH, temperature and pressure. It is clear therefore, that the process of binding is not a simple one and over-simplification of the process will no doubt lead to errors and possibly inaccurate interpretations and false conclusions.

In order to determine the hydrogen atom positions of the sugar, the protein and its associated bound solvent, a neutron structure of the complex needs to be derived. As it is the norm to deuterate the samples in a neutron diffraction experiment to reduce the incoherent scattering from hydrogen, it is first necessary to explore the validity of isotopic substitution to test whether the normal and isotopic enriched protein species differ significantly in their properties. It is generally assumed that the replacement of hydrogen by deuterium has little effect on the structure and biological activity of proteins.

The crystal structures of two different proteins, a small 149 amino-acid single domain *Staphylococcal* nuclease (Gamble *et al.,* 1994)[163] and the larger 393 amino-acid elongation factor Tu from *Escherichia coli* (Cooper *et al.,* 1998)[162] have been solved in both protiated and perdeuterated forms (96% and 95% replacement respectively). Perdeuteration was shown not to affect the overall fold of these proteins. However, limited results at present are available and so further comparisons are required. In the study performed here, a cubic crystal of concanavalin A complexed with glucoside that has been extensively soaked in D_2O, is refined using X-ray diffraction data and compared with the structure derived from

a 'normal' hydrogenated cubic crystal of the complex. An analysis of the structure of the protein, its solvent and the bound saccharide can provide us with information on the differences observed between the two. In the case that no structural differences are observed between the two structures, it then appears reasonable to add the positions of hydrogen atoms derived from neutron diffraction data of deuterated samples into thermodynamic calculations in the future. Once this information is incorporated, thermodynamic calculations can be repeated with a greater degree of confidence derived from the results.

The determination of the hydrogen atom positions for the 50 kDa dimer complex of concanavalin A with glucoside represents a real challenge for both SR X-ray and neutron techniques. At present, the largest published structure determined on LADI is the 33 kDa aspartic protease endothiapepsin. With large unit cells the crowding of spots on a fixed radius detector becomes limiting. On the current LADI instrument this can be overcome in part by using narrower band-pass filters (*e.g.* $\delta\lambda/\lambda = 10\%$, however, this would then increase data collection times) or by tuning to longer wavelengths. Larger unit-cell systems require yet larger crystals to compensate for their weak scattering signal, but then the increased spot size further exacerbates the problem of spatial overlap.

The practicality of measuring useful neutron diffraction data on biological molecules of a large unit cell is of considerable interest to the crystallographic community studying biochemical mechanisms involving hydrogen atoms. If we look at the molecular weight histogram for the yeast genome as an example, it peaks at 30 kDa and so at least half of all proteins in the genome are out of reach of current neutron protein crystallography capabilities even if big crystals can be grown. Therefore, in order to study successfully larger molecular weight/unit-cell problems certain design modifications need to be achieved. On LADI with a D_2O soaked crystal and a crystal size of 4 x 3 x 2 mm, the neutron diffraction resolution was at 3.5 Å and insufficient for atomic analysis. Therefore, a new detector configuration has been tested at the LADI instrument position. The new device provides a very large active area (800 x 800 mm) on a flat plate mount which allows the crystal to detector

distance to be varied and optimized to meet the requirements of the particular system. With an optimized detector distance with this new detector configuration we hope to improve the resolution limit and thus elucidate the positions of the sugar recognition hydrogen atoms (as deuterium).
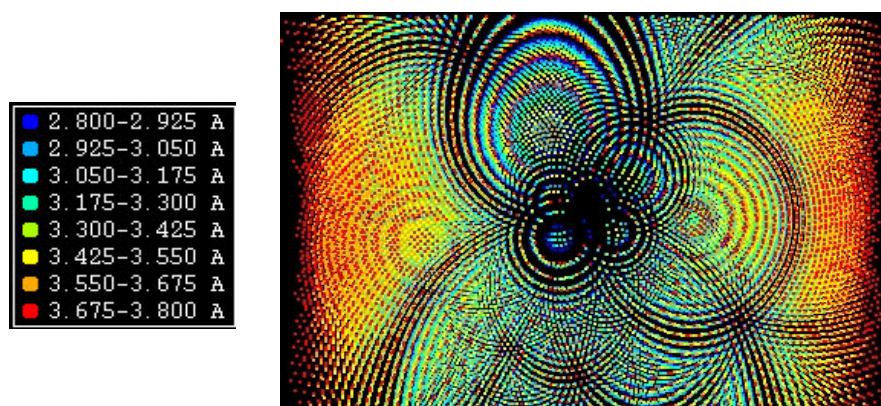
Figure 149. Laue simulation of the diffraction pattern from the concanavalin A/glucoside complex, on the LADI cylindrical detector, at a fixed distance of 159.2 mm. In this situation ~94% of reflections are spatially overlapped (using d_{min} = 2.5 Å and λ range 3.8-2.8 Å). The image is coloured according to the wavelength of the neutrons, as shown to the left of the simulated image.
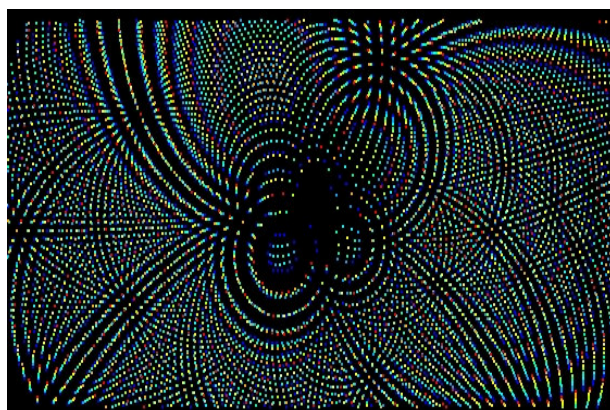
Figure 150. Laue simulation of the diffraction pattern from the concanavalin A/glucoside complex, this time on a flat detector, at a distance of 250 mm. In this situation ~41% of reflections are spatially overlapped. Coloured as in **Figure 149**.

6.3   Experimental

6.3.1 Crystallization of the concanavalin A methyl-
α-D-glucopyranoside complex

Attempts to crystallize concanavalin A with glucoside by soaking concanavalin A crystals in aqueous solutions of glucoside failed. It was found that the sugar binding site is blocked in the saccharide-free form and therefore, soaking of the sugar into the protein is not feasible. Therefore, co-crystallization was tried and proved successful[275] with crystals grown as regular rhombic dodecahedra. The crystal is formed only in the presence of glucoside and dissolves in the presence of mannoside.

Crystals of the complex were grown by dialysis by Dr A. J. Kalb (Gilboa). Portions of protein solution (0.2 ml) were equilibrated with 20 ml of 1.8 M sodium phosphate and 0.1 M methyl-*α*-D-glucopyranoside at pH 6.8. Crystals formed in a few days within a limited range of conditions; a ratio of NaH_2PO_4 to Na_2HPO_4 from 2 to 0.5 and a protein concentration of 120 mg/ml were used. The crystals were stable at glucoside concentrations above 1 mM but dissolved at lower concentrations of this saccharide, or in 0.1M methyl-*α*-D-mannopyranoside[275].

In the orthorhombic crystals of saccharide-free concanavalin A, the crystal volume per unit of protein molecular weight, V_m, is 2.35 A^3/dalton[128]. In the cubic crystal form of the saccharide complex, the extremely large value of V_m, 3.91 A^3/dalton[275], attests to a major alteration in the packing scheme and this is evident from the change in the space group. The crystals grow to a size of 1 mm^3 in a few days, and in some cases they can attain maximum dimensions of 5 mm. These large crystals naturally lend themselves to a neutron Laue diffraction study.

6.3.2 Deuteration of the concanavalin A crystals

Large crystals, grown in H_2O, selected for the X-ray study were transferred to 1 ml of crystallization buffer made up in 99.9% D_2O (Aldrich) with 0.1 M $NaNO_3$, 0.05 M tris-acetate, 1 mM $MnCl_2$, 1 mM $CaCl_2$, pH 6.5 in a tightly sealed test-tube for a 4 month soaking period. Attempts to grow cubic crystals directly with D_2O and the cubic crystal growth conditions, resulted in crystals of maximum dimensions of 1 mm[51].

6.3.3 X-ray data collection, scaling and reduction

SR X-ray diffraction data were collected at room temperature on the Quadriga undulator beamline fitted with a marCCD detector, at the ESRF by Dr S. Arzt. Diffraction data were collected from three crystals to 1.74 Å nominal resolution and a crystal translation of 0.2 mm was made after every 4 images, with a 1 oscillation per image. A total of 150 images were collected from the 3 crystals, however, the data obtained from the third crystal were discarded. The data were indexed, integrated and scaled using MOSFLM[251].

SCALA[251] was used to scale batches of data together from processed images. A free-R-factor[252] was calculated by excluding a randomly chosen set of reflections (~5%) from the refinement – a special case of the technique of cross-validation[253]. The agreement between F_{obs} and F_{calc} is independent of the refinement procedure. The data were cut-off at 2.1 Å resolution due to poor R-merge values and weak I/sigma(I) values. The quality of the data for the D_2O cubic concanavalin A were checked against the data used by Bradbrook *et al.,* (1998)[8], using SCALEIT[251]. The data are in good agreement, overall R(F) = 12% to 2.1 Å resolution, with the strongest reflections having R(F) = 8.8%. Below is a summary of the data processing statistics.

Resolution (Å)	Number of reflections	Completeness % (shell)	I/sigma(I)	R-merge (on Intensity)	<F/sigma(F)>	Mlplcty
5.51	2543	99.8	35.0	0.041	67.65	11.4
3.90	4649	99.5	34.0	0.046	65.05	11.6
3.18	6017	100.0	28.6	0.052	53.73	11.2
2.75	7091	100.0	19.9	0.076	36.74	11.0
2.46	8029	100.0	12.4	0.136	22.49	11.0
2.25	8846	100.0	8.7	0.212	15.58	10.9
2.08	9600	100.0	5.9	0.327	10.58	10.8
1.95	10283	99.8	3.7	0.543	6.96	10.8
1.84	10969	100.0	1.5	1.412	3.91	10.9
1.74	11228	99.3	0.5	3.342	2.68	6.8
Overall	79255	99.9	10.7	0.071	21.40	10.4

Table 61.   Summary of 2.1 Å D_2O cubic concanavalin A X-ray data quality

6.3.4 Refinement

The co-ordinates of the 2.0 Å concanavalin A/glucoside model[8] submitted to the Protein Data Bank[254] with code 1GIC were used as the starting model for the refinement. All sugar and water molecules were initially removed from the starting model and all the atomic B-factors were reset to 30 Å^2. REFMAC-5[251] was used for the refinement against the 2.1 Å resolution data. 10 cycles of rigid-body refinement reduced the R-factor to 24.8% (R-free = 23.9%). After restrained refinement the R-factor had reduced to 21.0%, with an associated R-free of 21.1%. $F_o - F_c$ maps were then checked and clearly showed density for the monosaccharide in each of the two subunits, therefore, glucoside molecules were built into the density. The model was checked for errors using the graphics program O[255]. After any improvements to the model were made by rebuilding, ARP_WARP[270] was run with water oxygen atoms identified and added to the model on the basis of difference density coupled with several distance constraints. Rejection of atoms is carried out on the basis of

the density interpolated at the atomic centre, the deviation of the density shape from sphericity and some distance criteria. A total of 251 bound D_2O sites were identified around the dimer from ARP_WARP, and as the water molecules contribute to the overall scattering, their inclusion reduces the R-factor. The model was compared with $2F_o$ - F_c/F_o – F_c electron density maps using the graphics program 'O'. Protein, saccharide and solvent model adjustments were made where appropriate. Certain waters were removed and others added to give a final number of water oxygen atom positions of 266. Omit maps were used to verify the conformation at residues Ala 207 and Asp 208, and confirmed the existence of the *cis*-peptide necessary for the binding of sugars. The final R-factor against the 2.1 Å data set was 18.2% for 43239 unique reflections, with R-free of 20.1% from 2315 reflections (5.1% of the unique reflections).

6.3.5 Refinement summary

Below is a summary of the refinement procedure that was followed.

Refinement program	Number of cycles/adjustments made to the model	R-factor (%)	R-free (%)
REFMAC5 (rigid body)	10	24.8	23.9
REFMAC5 (restrained)	10	21.0	21.1
O (version 7)	Checked ; (a) protein main chain and side chains in density correctly. (b) all rotamers for each side chain. Added difference density for glucoside in each sub-unit.		
ARP_WARP	30 cycles 251 waters added	20.4	22.0
O (version 7)	Checked (and removed where necessary) waters are spherical and oxygen atoms in density at > 1.0 sigma. H-bond distances (water-water & water-protein). Any remaining significant $Fo - Fc$ peaks.		
REFMAC5 (restrained)	10	18.2	20.1
O (version 7)	Checked protein, sugar and waters for good geometry and fit to density and made adjustments where necessary. Modelled the alternate conformations.		

6.4 Analysis of the 2.1 Å cubic concanavalin A model

The final structure contains a dimer in the asymmetric unit, of molecular weight ~50 kDa. Each monomer contains 237 amino acids, one Ca^{2+} ion and one Mn^{2+} ion and one glucoside bound at the monosaccharide binding site. 266 D_2O water molecules were modelled around the dimer. The space group is $I2_13$ with unit cell dimensions of $a = b = c = 167.747$ Å, $\alpha = \beta = \gamma = 90°$. Below is a summary of the structure and on the next page, a summary of the structure refined by Bradbrook *et al.,* (1998)[8] for comparison.

Total no. of atoms = 3914

Total no. of protein atoms = 3618

Total no. of sugar atoms = 26

Total no. of water molecules = 266

Total no. of metals = 4

Resolution range = ∞ – 2.1 Å

R-factor = 18.2%

R-free = 20.1%

Bond distance r.m.s. deviation = 0.009 Å

Bond angle r.m.s. deviation = 1.305 °

Ramachandran core/allowed/generously (%) = 88.2/11.5/0.2

Mean B-factor for all protein atoms ($Å^2$) = 38.8

Mean B-factor for side chain atoms ($Å^2$) = 39.5

Mean B-factor for main chain atoms ($Å^2$) = 38.1

Mean B-factor for sugar atoms ($Å^2$) = 36.4

Mean B-factor for 266 D_2O water molecules ($Å^2$) = 54.2

Mean B-factor for all side chains and waters ($Å^2$) = 41.4

Mean B-factor for 4 metals ($Å^2$) = 32.3

Mean B-factor for all atoms ($Å^2$) = 39.8

Correlation coefficient F_{obs} to F_{calc} = 0.965 [R-free = 0.960]

Table 62. Summary of the 2.1 Å D_2O cubic concanavalin A/methyl-α-D-glucoside complex.

Total no. of atoms = 4008

Total no. of protein atoms = 3618

Total no. of sugar atoms = 26

Total no. of water molecules = 360

Total no. of metals = 4

Resolution range = ∞ – 2.0 Å

R-factor = 17.0%

R-free = n/a

Bond distance r.m.s. deviation = 0.014 Å

Bond angle r.m.s. deviation = 3.2 °

Ramachandran core/allowed/generously (%) = 86.1/13.7/0.2

Mean B-factor for all protein atoms (Å^2) = 20.6

Mean B-factor for side chain atoms (Å^2) = 21.7

Mean B-factor for main chain atoms (Å^2) = 19.6

Mean B-factor for sugar atoms (Å^2) = 19.3

Mean B-factor for 360 H_2O water oxygen atoms (Å^2) = 43.6

Mean B-factor for all side chains and waters (Å^2) = 25.4

Mean B-factor for 4 metals (Å^2) = 13.6

Mean B-factor for all atoms (Å^2) = 22.6

Table 63.   Summary of the 2.0 Å concanavalin A/methyl-α-D-glucoside complex[8].

The Ramachandran plot[256] for the final model is given. The Ramachandran plot was calculated using PROCHECK[257]. There are no residues in the disallowed regions, with the majority of residues in the most favoured regions (88.2%). There are a few residues (11.5%) in the additional allowed regions and one residue (Asp203, chain A) in the generously allowed region. Asp203 is located in a loop region at the surface of the dimer, and has increased mobility relative to residues involved in stable secondary structure motifs.

If we compare the final model summary for the data presented here, against that for Bradrook *et al.,* (1998)[8], we see that there is an overall improvement in the percentage of residues in the core regions (88.2% vs. 86.1%), and also an

improvement in the percentage of main chain bond lengths/angles and planar groups within the given limits. There are also no distorted main chain bond distances, main chain bond angles or planar groups, whereas before there were 25 distorted main chain bonds, 39 distorted main chain angles and 5 distorted planar groups.

Figure 151.   The Ramachandran plot for the concanavalin A/ methyl-α-D-glucoside complex at 2.1 Å resolution.

The overall root-mean-square deviations (*x, y, z*) between the 2.0 Å concanavalin A/glucoside complex starting model (PDB code = 1GIC)[8] and the final 2.1 Å D$_2$O concanavalin A/glucoside complex model were calculated as follows:

1. Between all protein atoms, r.m.s. deviation = 0.331 Å

2. Between all main chain protein atoms, r.m.s. deviation = 0.218 Å

3. Between all side chain protein atoms, r.m.s. deviation = 0.422 Å

4. Between all α-carbon protein atoms, r.m.s. deviation = 0.182 Å

Overall then it can be seen that the two structures superimpose well with only small r.m.s. deviations.

The residues that exhibit the largest differences between the starting model (1GIC) and the final 2.1 Å X-ray model of glucoside bound concanavalin A are listed, those given have root-mean-square deviations greater than 1 Å.

Main chain large r.m.s. deviations;

Thr 150 (chain B)

Side chain large r.m.s. deviations;

Asn 69 (chains A and B), His 121 (B), Glu 122 (B), Lys 135 (A), Asp 151 (B), Arg 158 (A), Ser 161 (A), Ser 164 (B), Ser 184 (A), Lys 200 (B), Ser 204 (B)

Threonine 150 is the only residue to exhibit a large main chain r.m.s. deviation of greater than 1 Å. This is due to the necessity for the peptide unit to be flipped relative to the starting model, in order to correctly fit the electron density, as **Figure 152** illustrates. In the 2.0 Å resolution structure determined by Bradbrook *et al.,* (1998)[8] the oxygen of the peptide is not within the density, whereas if it were to be flipped then it would correctly fit the density, as in the 2.1 Å model presented here. Many of the residues that have large side chain shifts are found within specific loop sections (118-122, 159-165, 201-207) as noted previously. As expected the residues with the large r.m.s. shifts also have the highest associated B-factors. For example, like the saccharide-free structure, His 121 has the highest B-factor of all the residues in each of the two subunits (chain A; B-factor$_{121}$ = 67.4 Å^2, chain B; B-factor$_{121}$ = 69.1 Å^2). The high flexibility of lysine and conformational freedom of serine cause these residues to exhibit large r.m.s. deviations.
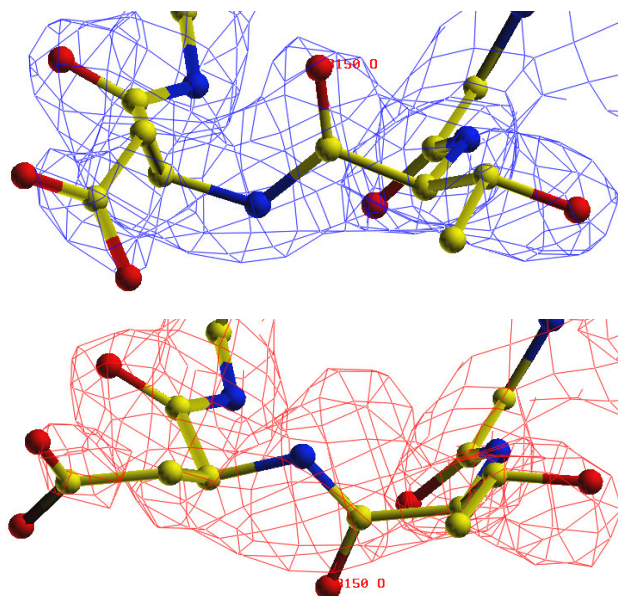
Figure 152.   *Top,* the $2F_o - F_c$ electron density for threonine 150 of chain B in the D_2O cubic concanavalin A model. *Bottom,* the $2F_o - F_c$ electron density for threonine 150 of chain B in the H_2O cubic concanavalin A model[8]. The peptide unit is in a different conformation in the two models.

6.5   Temperature-factor analysis

Given below is the temperature-factor versus residue number plots for both protein chains A and B in the D_2O cubic concanavalin A model. It can be seen from the shapes of the curves that there is very close similarity between chains A and B. Regions of high and low B-factors correspond well between the two chains, with the main regions of disorder being those noted in previous concanavalin A crystal structures[48-50,142,144] *i.e.* residues 118 to 122, 159 to 165, 201 to 207. Overall, chain B has a lower average B-factor than chain A, 36.8 Å^2 and 40.7 Å^2 respectively. This is also reflected in the average B-factors of the sugars bound to each subunit; chain A methyl-glucoside average B-factor = 42.8 Å^2 and chain B methyl-glucoside average B-factor = 30.1 Å^2.
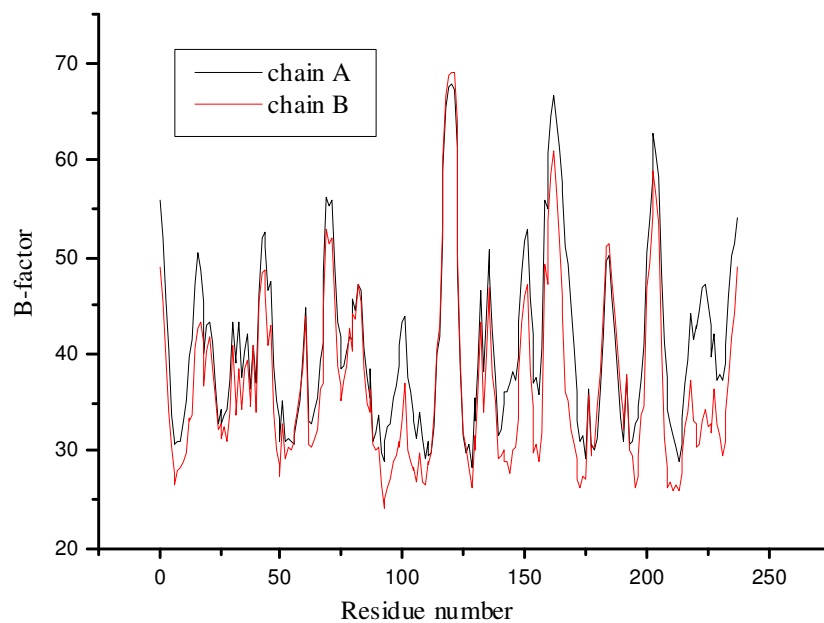
Figure 153. The B-factor versus residue number plot for chains A and B in D_2O cubic concanavalin A.
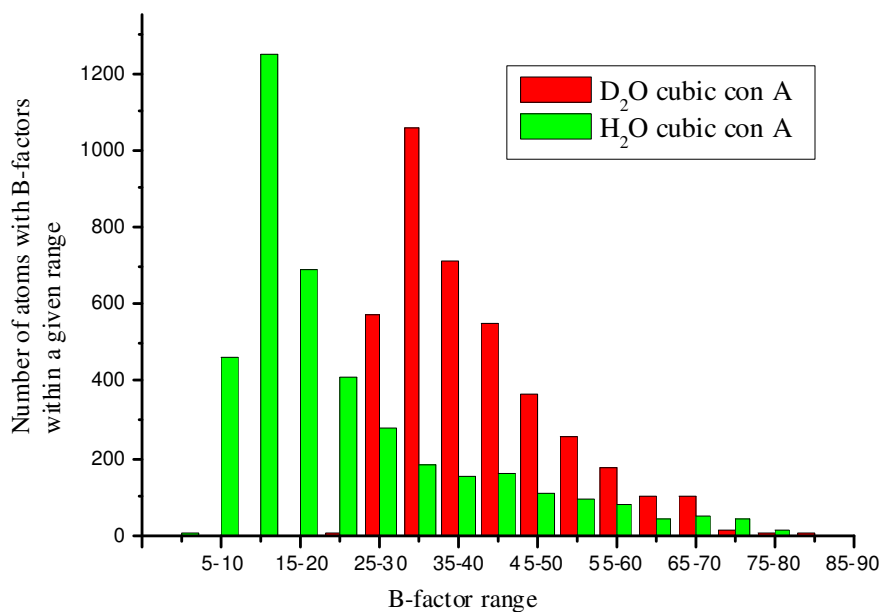
Figure 154. The number of atoms with B-factors within a given range in the 2.1 Å model of cubic concanavalin A.

From the plot above we can see that the majority of the atoms in the structure have B-factors in the range 30-35 Å^2. The only atoms with a B-factor greater than 80 Å^2 are four waters (W105, W139, W191, W227), while the only protein atoms with a B-factor greater than 70 Å^2 belong to His 121 of chain B. None of the atoms have B-factors of less than 20 Å^2. The residue which exhibits the lowest average B-factor is Gly 92 of chain B (average B-factor = 24.2 Å^2) with Leu 93, Ala 211 and Phe 213 also exhibiting low B-factors of ~25 Å^2. These residues all lie in close proximity to each other and are found buried within the core of the protein (see **Figure 155**). The data for the D_2O cubic concanavalin A were collected at room temperature, as unfortunately the crystals are not stable below room temperature and thus in this case the use of cryo-techniques was not possible. Therefore, as expected, the B-factors for atoms in this model are generally elevated compared to cryo-data sets, due to the increased dynamic disorder.
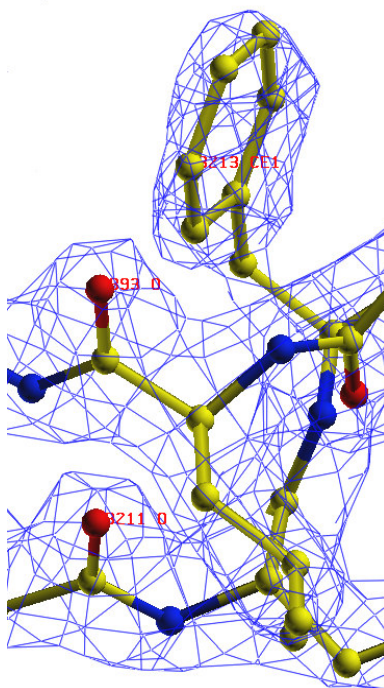
Figure 155.   The residues that exhibit the lowest B-factors of any of the protein residues in the D_2O cubic structure are all situated in close proximity to each other within the core of the protein.

From comparison of the B-factors for both the 2.0 Å H_2O and 2.1 Å D_2O models for cubic concanavalin A we see that the areas of high and low B-factors correspond between the two models. Furthermore, like the 2.1 Å D_2O cubic concanavalin A model, the 2.0 Å H_2O model has slightly higher average B-factors for chain A compared to chain B; 21.2 $Å^2$ and 20.0 $Å^2$ respectively. However, the D_2O cubic concanavalin A model has higher overall average B-factors than the H_2O cubic concanavalin A model. The average B-factor for all protein atoms in the 2.0 Å H_2O model is 20.6 $Å^2$, whereas it is almost double this in the 2.1 Å D_2O cubic concanavalin A model at 38.8 $Å^2$. Furthermore, the average B-factor for all water atoms in the 2.0 Å H_2O model is 43.6 $Å^2$, whereas it is 54.2 $Å^2$ in the 2.1 Å D_2O cubic concanavalin A model. Higher resolution structures generally give lower B-factors, however, the difference in resolution in this case is only small (0.1 Å) and so is possibly not the cause of the elevated B-factors. Perhaps a more likely reason for the elevated B-factors is that because the data were collected using synchrotron radiation at the ESRF at room temperature, some radiation damage may have occurred during data collection.
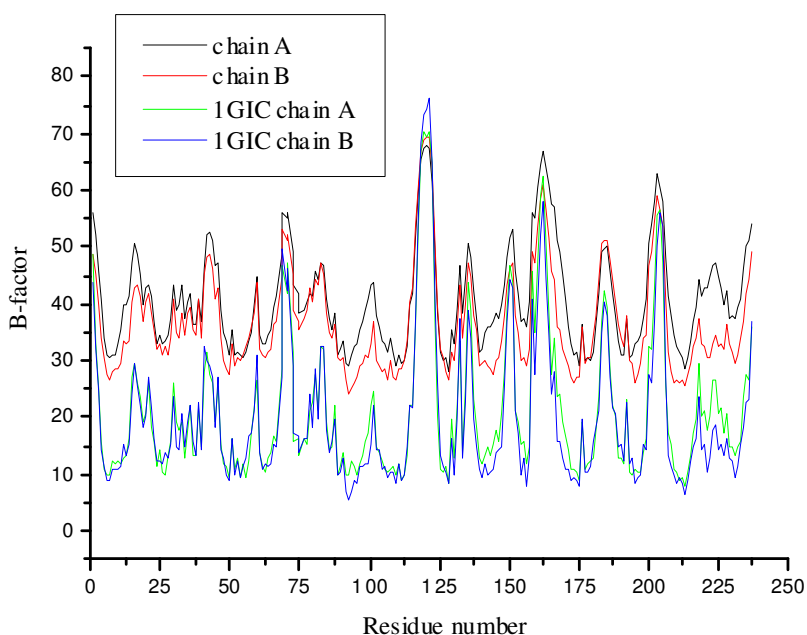
Figure 156.   The B-factor versus residue number plot for both the 2.1 Å D_2O and 2.0 Å H_2O cubic concanavalin A models.

285

6.6   Sugar binding

Clear difference electron density for methyl-α-D-glucoside was observed in the saccharide binding site of each subunit. The sugars O3, O4 and O6 all form co-operative hydrogen bonds. The sugar O4 and O6 form a bidentate hydrogen bond with Asp208. All the interactions noted previously[8] are observed with excellent agreement in hydrogen-bond distances in each instance. Networks of hydrogen bonds are formed, many mediated by water molecules. One interaction that was noted by Bradbrook *et al.,*(1998)[8] found to be present in both the sub-units of the concanavalin A/glucoside complex, is an interaction between the sugar C2 hydroxyl and a water, which binds to the oxygen OG1 of Threonine 226. The C2 oxygen, to water, to Thr 226 distances are given in the table below along with the other donor to acceptor hydrogen bond distances for each sub-unit (for some interactions the assignments of donor and acceptor are arbitrary).

The Thr226 – sugar O2 water mediated interaction was highlighted due to the fact that the differences in binding affinity observed between the mannoside and glucoside complexes may be in part due to the difference at C2[8]. The Thr226W – sugarO2 distance was ~2.7 Å for each sub-unit in the mannoside complex whereas in the glucoside complex it is shown to be ~4.0 Å. Bradbrook *et al.,* (1998)[8] suggested that this weaker interaction for the glucoside complex could partially explain the weaker binding affinity. In the data presented here for the glucoside complex, the Thr226 water is identified in chains A and B and is in good agreement with the distances from Bradbrook *et al.,*(1998).

Donor	Acceptor	Chain A	Chain B	(A) Bradbrook *et al.,*(1998)[8]	(B) Bradbrook *et al.,*(1998)[8]
Arg 228 N	Sug O3	2.97	2.92	3.02	3.00
Sug O4	Asp208 OD2	2.65	2.57	2.76	2.86
Asn 14 ND2	Sug O4	2.76	2.85	2.82	2.81
Arg 228 N	Sug O4	3.39	3.34	3.25	3.26
Leu 99 N	Sug O5	3.23	3.13	2.99	3.08
Sug O6	Asp208 OD1	2.72	2.82	2.73	2.99
Tyr 100 N	Sug O6	3.13	3.04	3.09	3.01
Leu 99 N	Sug O6	3.21	3.17	3.07	3.14
Thr 226 W	Sug O2	(W291) 4.48	(W194) 4.12	4.34	4.06
Thr 226 W	Thr 226 OG1	2.59	2.70	2.74	2.82

Table 64. The hydrogen bond interactions between the methyl-α-D-glucoside and the protein atoms in the cubic concanavalin A dimer in the refined model presented here and the model refined by Bradbrook *et al.,* (1998)[8].

Further water mediated interactions of note are found between the sugar atoms O2 and O3, and the main chain oxygen atoms of residues Gly 224 and Thr 226. Two water molecules are involved in bridging the protein and sugar as follows;
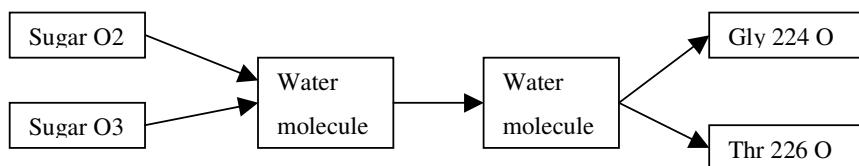
Figure 157. Schematic of water mediated interactions between the O2 and O3 atoms of the methyl-glucoside and the main chain oxygen atoms of residues 224 and 226 in the cubic concanavalin A dimer.
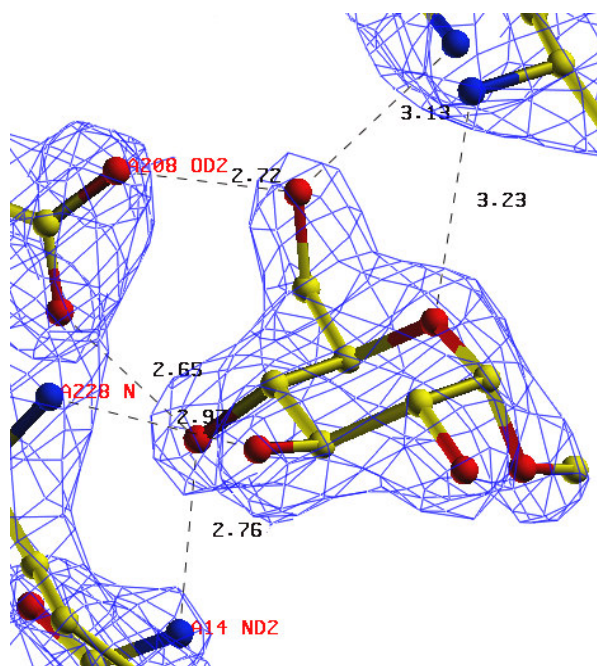
Figure 158.   The $2F_o - F_c$ electron density for the glucoside bound at the saccharide binding site of chain A. The sugar is in the chair conformation, with the protein side chain and main chain atoms which bind the sugar also shown.
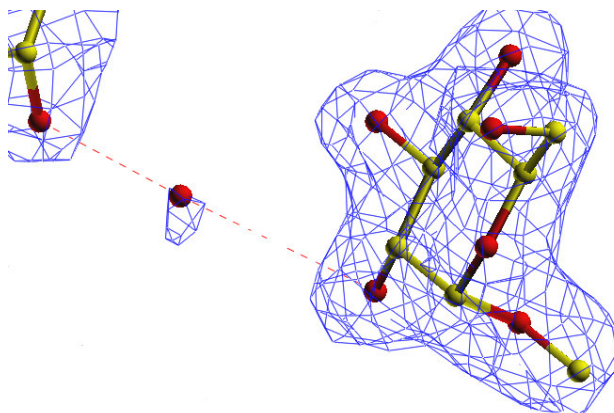
Figure 159.   A weak water mediated hydrogen bond between the side chain oxygen of Thr 226 and the glucoside in chain B. The water to protein distance is a reasonable length for a hydrogen bond (2.7 Å) whereas the water to glucoside distance is rather long for a hydrogen bond at 4.12 Å.

Atom	Atom	Chain A	Chain B
Sugar O3	Water oxygen	2.76 (W190)	2.65 (W68)
Sugar O2	Water oxygen	3.02 (W190)	3.26 (W68)
Water oxygen	Water oxygen	3.07 (W190-W113)	2.76 (W68-W99)
Water oxygen	Gly 224 O	2.96 (W113)	2.92 (W99)
Water oxygen	Thr 226 O	2.51 (W113)	2.54 (W99)

Table 65.　　　Water mediated interactions in the saccharide binding site other than those noted by Bradbrook *et al.,* (1998)[8]
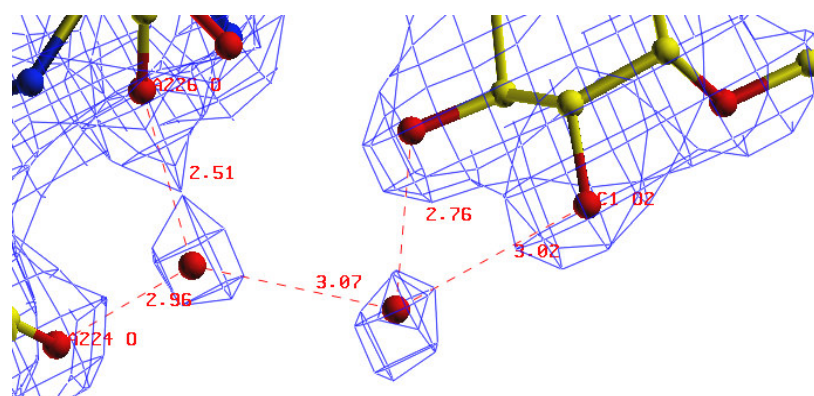
Figure 160.　　Two water mediated interactions between the methyl-glucoside sugar atoms O2 and O3, and the main chain oxygen atoms of Gly 224 and Thr 226 in chain A of the D_2O cubic concanavalin A dimer.

6.7   Van der Waals interactions

Van der Waals contacts are defined here as inter-atomic distances less than 4.1 Å. This was chosen as a cut-off for comparison with the earlier study by Bradbrook *et al.,* (1998)[8]. Van der Waals interactions within the binding site are the only ones listed. The same contacts are seen for both sub-units, however, there are subtle differences in distance as shown in the table below. The results also show that the same interactions can be seen for both the data here and for that of Bradbrook *et al.,* (1998).

Sugar atom	Protein atom	Chain A	Chain B	Bradbrook *et al.,*(A)[8]	Bradbrook *et al.,*(B)[8]
C-1	99 N	3.95	3.80	3.69	3.82
	99 CB	3.87	3.75	3.71	3.81
C-3	228 N	3.94	3.92	3.91	3.94
C-4	14 ND2	3.87	3.96	3.96	3.93
	208 CG	3.95	3.84	3.98	4.05
	208 OD1	3.69	3.53	3.73	3.64
	208 OD2	3.34	3.32	3.43	3.66
	227 CA	3.97	3.93	3.95	3.91
	228 N	3.74	3.77	3.66	3.73
C-5	208 OD1	4.02	4.03		
	12 CE1	4.05	4.02	3.94	4.01
	99 N			4.05	
C-6	12 CG	4.01	3.97	4.02	3.96
	12 CD1	3.77	3.76	3.65	3.79
	12 CD2				4.09
	12 CE1	3.83	3.79	3.71	3.74
	12 CE2				4.05
	12 CZ		4.01	3.99	3.86
	99 N			3.87	4.07
	100 N	3.95	3.90		3.90
	100 CD1			3.80	3.98

	100 CD2	3.74	4.10		
	100 CB				4.02
	207 CB	3.63	3.62	3.60	3.69
C-6	208 CG	3.90	3.92	3.89	4.09
	208 OD1	3.55	3.60	3.44	3.61
	208 OD2	3.45	3.46	3.64	3.82
C-7	99 CB		4.08		
	99 CD2		3.94		3.97
	12 OH				
O-2	98 CA				
	98 C				
	99 N				
	227 CA				
O-3	227 CA	3.60	3.56	3.68	3.47
	227 C	3.78	3.73	3.87	3.77
	228 N	2.97	2.92	3.02	3.00
	228 CA	3.88	3.81	3.88	3.92
	228 CB	3.59	3.47	3.49	3.53
	228 CG	3.56	3.96	3.38	3.91
	228 CD	3.54	3.45	3.63	3.33
	228 NE		4.03		3.86
O-4	12 CD1	3.85	3.86	3.87	3.81
	12 CD2				
	12 CE1				4.08
	12 CE2				
	14 CG	3.58	3.69	3.73	3.73
	14 OD1	3.59	3.72	3.74	3.72
	14 ND2	2.76	2.85	2.82	2.81
	208 CG	3.63	3.49	3.68	3.62
	208 OD1	3.84	3.66	3.88	3.66
	208 OD2	2.65	2.57	2.76	2.86
	227 C		4.04	4.10	4.02
	228 N	3.39	3.34	3.25	3.26
	228 CA	3.68	3.64	3.68	3.69
	228 CB	3.38	3.35	3.40	3.35
O-5	98 CA			4.06	
	98 C			3.97	4.08

	99 N	3.23	3.13	2.99	3.08
	99 CA	3.94	3.80	3.74	3.80
	99 CB	3.51	3.34	3.37	3.36
	100 N				
O-6	98 N	3.41	3.51	3.27	3.64
	98 CA	3.32	3.36	3.22	3.60
	98 C	3.57	3.57	3.43	3.65
	99 N	3.21	3.17	3.08	3.14
	99 CA		4.02	4.09	4.03
	99 C		3.99	4.03	3.96
	100 N	3.13	3.04	3.09	3.01
	100 CA	3.83	3.74	3.89	3.71
	100 C	3.95	3.92	3.98	3.97
	100 O	3.40	3.39	3.41	3.46
	100 CB	3.99	3.83	3.97	3.56
	207 CB	3.32	3.29	3.35	3.29
	208 CG	3.57	3.66	3.57	3.80
	208 OD1	2.72	2.82	2.73	2.99
	208 OD2	3.69	3.81	3.78	3.97
Total number		53	60	59	64
Total distance		191.28	218.55	210.54	236.64
Average distance		3.61	3.64	3.57	3.70

Table 66.   Van der Waals contacts between the methyl-α-D-glucoside and the protein atoms in the cubic concanavalin A dimer for the 2.1 Å D_2O model presented here and the model refined by Bradbrook *et al.,* (1998)[8].

We can see from the average distances of the van der Waals interactions that there is very good agreement between the two sub-units and the earlier study. Chain A in the D_2O cubic structure appears to have slightly less van der Waals interactions than chain B, indicating that the sugar in chain B is more tightly bound to the protein chain. This is seen in the B-factors of the two glucoside sugars;

Chain A; glucoside average B-factor = 42.8 Å^2

Chain B; glucoside average B-factor = 30.1 Å^2

6.8   Metal sites

The tables below give the metal to ligand distances for each chain. As can be seen there is excellent agreement in distances when comparing the two cubic structures. The largest difference for the manganese to ligand distances in the two structures being 0.04 Å. In all of the analyses of lectin metal sites in this thesis we have seen that the ligand positions and therefore, the metal to ligand distances, are highly conserved with excellent agreement between all the various models both saccharide-free and saccharide bound.

Ligand	Distance (Å)	Distance (Å) Bradrook *et al.*,(1998)[8]	Ligand B-factor (Å^2)
Asp 10 OD2	2.10	2.10	33.7
W206	2.08	2.09	32.6
Glu 8 OE2	2.31	2.31	32.4
W37	2.08	2.04	34.6
His 24 NE2	2.27	2.28	31.7
Asp 19 OD1	2.28	2.27	33.9

Table 67. The manganese-ligand distances in chain A of the cubic concanavalin A dimer. Distances from Bradrook *et al.* are given for comparison.

Ligand	Distance (Å)	Distance (Å) Bradbrook *et al.*,(1998)[8]	Ligand B-factor (Å^2)
Asp 10 OD2	2.13	2.14	28.2
W212	2.20	2.22	30.4
Glu 8 OE2	2.23	2.22	30.6
W219	2.09	2.10	30.7
His 24 NE2	2.30	2.31	32.7
Asp 19 OD1	2.26	2.26	31.6

Table 68. The manganese-ligand distances in chain B of the cubic concanavalin A dimer. Distances from Bradrook *et al.* are given for comparison.

Ligand	Distance (Å)	Distance (Å) Bradbrook *et al.,*(1998)[8]	Ligand B-factor (Å^2)
Asn 14 OD1	2.53	2.53	37.4
W26	2.49	2.47	34.2
Tyr 12 O	2.28	2.29	37.0
Asp 19 OD2	2.34	2.34	35.0
Asp 10 OD2	2.61	2.62	33.7
W220	2.20	2.22	35.0
Asp 10 OD1	2.35	2.35	34.9

Table 69.   The calcium-ligand distances in chain A of the cubic concanavalin A dimer. Distances from Bradrook *et al.* are given for comparison.

Ligand	Distance (Å)	Distance (Å) Bradbrook *et al.,*(1998)[8]	Ligand B-factor (Å^2)
Asn 14 OD1	2.47	2.48	30.5
W252	2.15	2.12	29.8
Tyr 12 O	2.33	2.34	31.4
Asp 19 OD2	2.37	2.37	30.3
Asp 10 OD2	2.48	2.48	28.2
W23	2.41	2.41	29.7
Asp 10 OD1	2.29	2.28	29.2

Table 70.   The calcium-ligand distances in chain B of the cubic concanavalin A dimer. Distances from Bradrook *et al.* are given for comparison.

The calcium to ligand distances also show little variation between the two structures, the largest difference being 0.03 Å. It can be seen therefore, that at the metal sites there is no apparent difference in structure between the hydrogenated and D_2O soaked cubic crystal structures.
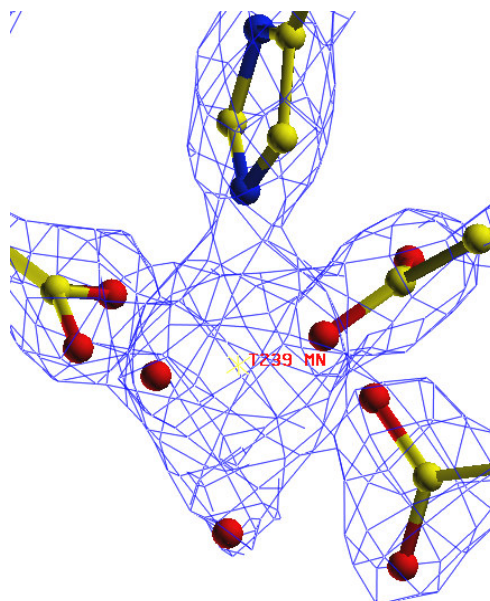
Figure 161.   The manganese binding site of chain B of the D_2O cubic concanavalin
                   A model at 2.1 Å resolution.
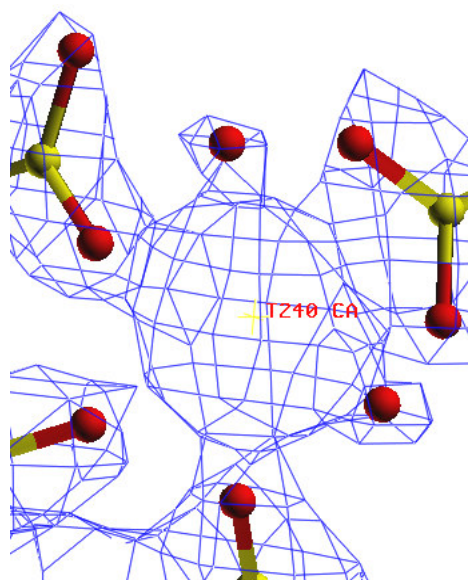
Figure 162.   The calcium binding site of chain B of the D_2O cubic concanavalin A
                   model at 2.1 Å resolution.

6.9   Water structure analysis: conserved waters

The 2.1 Å D_2O cubic concanavalin A model contains 266 water oxygen atoms with an average overall B-factor for the waters of 54.2 $Å^2$. Of these 266 waters, 32 are seen to be completely buried (B-factor$_{Buried\ waters}$ = 40.7 $Å^2$) while 234 are found on the surface (B-factor$_{Surface\ waters}$ = 56.1 $Å^2$). After superposition of the starting model dimer (1GIC) with the D_2O cubic concanavalin A model using LSQKAB[251], a hydration site was considered conserved if both the H_2O and D_2O oxygen atoms were found to be within a pre-defined distance of 1.0 Å. When this cut-off is used 181 waters are seen to be conserved. Of these 181, 30 are buried and 151 are positioned on the surface.

The number of waters that belong to each subunit of the dimer was calculated, 129 are associated with chain A (B-factor$_{Chain\ A\ waters}$ = 55.8 $Å^2$) and 137 with chain B (B-factor$_{Chain\ B\ waters}$ = 52.7 $Å^2$). The 266 waters were also analysed in terms of the number of bonds made from each water molecule to protein atoms. In the table below it can be seen that 70 waters interact only with either other waters and/or sugar atoms. As is expected these waters exhibit the highest B-factors as they are more mobile due to having less stabilising interactions.

Number of bonds per water oxygen to protein atoms	Number of waters	Average B-factor ($Å^2$)
0	70	59.4
1	131	55.2
2	53	48.8
3	10	38.0
4	2	31.7

Table 71.   Table to show the number of bonds made from each water molecule to protein atoms and their average B-factors.

The 7 water molecule sites that are conserved in all legume lectin crystal structures[140] have all been identified within both subunits of the cubic concanavalin A model, with the exception of the water in the saccharide binding site. This has been displaced by the methyl-glucoside molecule on binding, with the sugar O4 occupying the position formerly occupied by the water oxygen. Below are the distances from key saccharide binding site residues to the sugar O4 atom for comparison against the distances from the same key saccharide binding site residues to the water that is expelled on saccharide binding.

Structure\ Residue	Asp 208 OD1	Arg 228 N	Asn 14 ND2
1.65 Å X-ray saccharide-free concanavalin A water oxygen (W4)	2.64 Å	3.62 Å	3.00 Å
Saccharide-bound D_2O cubic concanavalin A sugar O4 atom	2.57 Å	3.34 Å	2.85 Å

Table 72.   Comparison of the position of the sugar O4 atom with one of the conserved waters that is expelled on binding.

When we look at the positions occupied by the conserved waters, general trends can be seen. First of all we see that most waters are found at the surface of the protein as opposed to being buried within the protein. This is to be expected from hydrophilic/hydrophobic considerations, *i.e.* most hydrophilic residues are found at the exterior of the protein where they have favourable interactions with polar water molecules, whilst most hydrophobic residues are found within the interior of the protein so as to avoid unfavourable interactions with charged or polar species. Of the waters at the surface, most are seen on the opposite side to the 12-stranded β-sheet side, where the molecule is made up of more loop regions and some β-sheet. Of the waters found on the β-sheet side, they are mainly situated once again near to the loop regions that interconnect the β-strands, particularly at 3 main loop regions: loops 56-59, 68-71, 118-122.
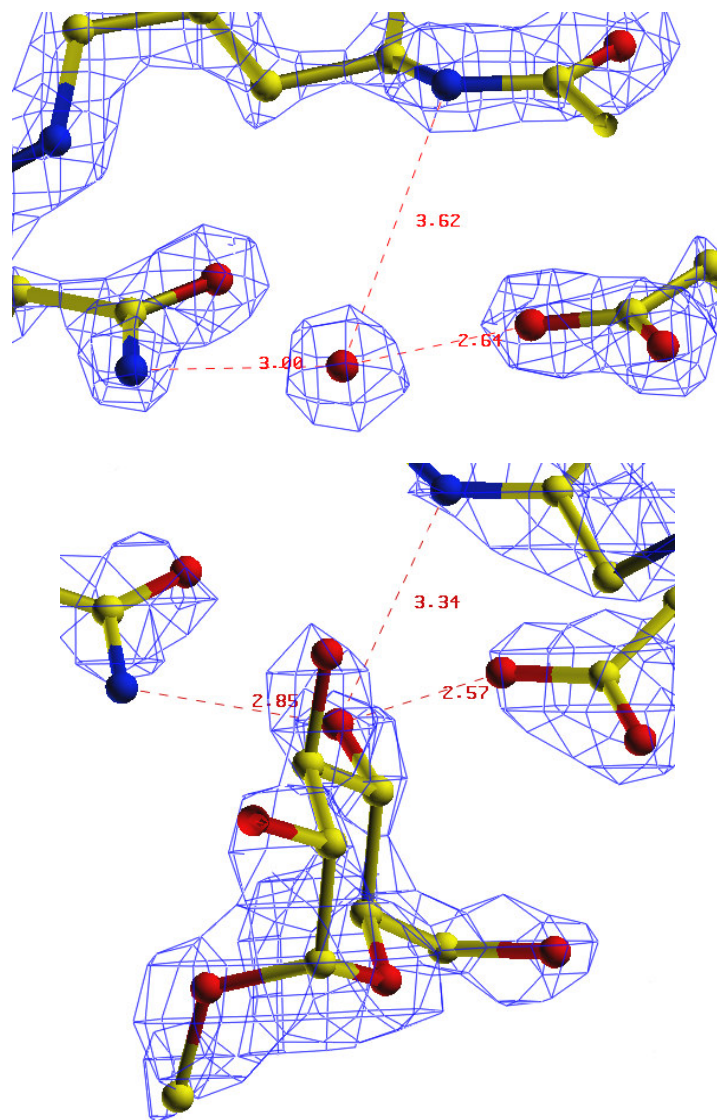
Figure 163.   *Top,* a water oxygen atom in the 1.65 Å saccharide-free X-ray model, identified in the saccharide binding site, at a very similar position to the sugar O4 oxygen atom of the bound glucoside in the 2.1 Å saccharide bound complex.   Water oxygen to key protein residue distances are given.   *Bottom,* the glucoside bound at the binding site with the sugar O4 to key protein residue distances given for comparison.

The highest density of water molecules is seen at 2 main areas, these being the dimer interface and the apex of each sub-unit. The inter sub-unit interaction is produced by many hydrogen bonds and van der Waals interactions between sub-units, but also *via* water-mediated interactions. These water molecules at the dimer interface help to tether the 2 monomers by forming networks of hydrogen bonds.

When an analysis of the closest (up to a cut-off of 0.2A) conserved waters was done, more than a third of these were seen to be at the dimer interface. As expected, the other area with a high degree of conservation is the area at the apex of each sub-unit, containing the saccharide binding site, the Ca^{2+} ion site and the Mn^{2+} ion site. Below in **Table 73** the first 36 (up to a cut-off of 0.2 Å) conserved water positions between the data presented here and the model of Bradbrook *et al.,* (1998)[8] are given.

Dimer Interface	13/36
Apex	8/36
Side opposite to the 12-stranded β-sheet side	7/36
12-stranded β-sheet side	3/36
β-bulge motif	2/36
β-hairpin	1/36
Others	2/36

Table 73.   The positions of the closest 36 conserved water sites between the D_2O and H_2O cubic concanavalin A models.

6.10 Concluding remarks

The three-dimensional structure of the complex between methyl-α-D-glucopyranoside and D_2O cubic concanavalin A has been determined to 2.1 Å resolution with an improvement in the structural precision being achieved (more data) compared to previously[8]. The D_2O cubic structure was then compared against the previous H_2O cubic structure in order to identify if there are any visible differences between the two *i.e.* does the D_2O soaking affect the structure? If so, this would have important consequences for neutron protein crystallography, where the crystals are preferentially soaked in D_2O, to reduce the background incoherent scattering from hydrogen atoms. The two structures were found to be similar. Metal to ligand distances and sugar-lectin hydrogen bond interactions were compared and found to be the same within the precision of the structures. However, the B-factors of the D_2O structure are higher than for the H_2O structure and this warrants further study.

The water structure was analysed and it was shown that many of the waters that are found in conserved positions between the H_2O and D_2O structures function as to stabilise the overall quaternary structure. For example many of the conserved waters are found at the dimer interface where they are involved in water mediated interactions with both subunits and hence, assist in the tethering of the two monomers together. The other main area at which many conserved water sites are found is the apex of the subunit, which includes the Mn^{2+} and Ca^{2+} binding sites and the saccharide binding site. All of the waters at the metal sites are conserved as well as some in the saccharide binding site which are involved in water mediated hydrogen bonds between the protein and sugar.

Chapter 7
Future work

7.1   Future work

The practicality of measuring useful neutron diffraction data on biological molecules of a large unit cell is of considerable interest to the crystallographic community studying biochemical mechanisms involving hydrogen atoms. The determination of the hydrogen atom positions for the 50 kDa dimer complex of concanavalin A with glucoside represents a real challenge for both SR X-ray and neutron techniques. At present the largest published structure determined on LADI is the 33 kDa aspartic protease endothiapepsin[224]. As described earlier, (**Chapter 6**) with large unit cells the crowding of spots on a fixed radius detector becomes limiting. On the current LADI instrument this can be overcome in part by using narrower band-pass filters (*e.g.* $\delta\lambda/\lambda$ = 10%, however, this would then increase data collection times) or by tuning to longer wavelengths. However, in order to keep the data collection times down to a minimum while still attempting to overcome the problem of spatial overlap, we have begun testing a new detector configuration at the LADI position. The new device provides a very large active area (800 x 800 mm) on a flat plate mount which allows the crystal to detector distance to be varied and optimized to meet the requirements of the particular system. With an optimized detector distance with this new detector configuration, we plan to collect a full neutron data set at medium resolution (~2 Å) and thus elucidate the positions of the sugar recognition hydrogen atoms (as deuterium) for the cubic concanavalin A as an exemplar of the larger unit cell/molecular weight challenge.

Secondly, after successfully cooling a large protein-crystal to cryotemperatures, which can be seen as a real breakthrough in terms of the size of crystal being frozen, I aim to collect neutron diffraction data on other such protein crystals at cryotemperatures. The improved density definition that can be achieved by reducing the dynamic disorder this way can be further enhanced by the expression of fully deuterated proteins. Thirdly, it has also been shown again here that the 'X then N' refinement strategy can work, but also suggests that a comparison with an 'X with N' approach would be useful, in which one would refine a test protein structure using both types of data combined, as was attempted by A. Wlodawer *et al.,* (1982)[306] for ribonuclease-A; the software used then has apparently been lost!

References

1. L. Stryer, 1995. *Biochemsitry,* (4th edition)

2. D. Horton, **1990**. *'Development of carbohydrate nomenclature'* in 'The terminology of biotechnology: a multidisciplinary problem' K. L. Loening (Ed.)

3. E. Fischer, **1890**. *Ber.,* **23** p2114

4. T. W. Rademacher, R. B. Parekh, R. A. Dwek, **1988**. *Annu. Rev. Biochem.,* **57** p785-838

5. T. Feizi, **1991**. *Trends Biochem. Sci.,* **16** p84-86

6. N. K. Vyas, **1991**. *Curr. Opin. Struct. Biol.,* **1** p732-740

7. N. Sharon & H. Lis, **1990**. *Chem. Brit.,* p679-682

8. G. M. Bradbrook, T. Gleichmann, S. J. Harrop, J. Habash, J. Raftery, A. J. Kalb (Gilboa), J. Yariv, I. H. Hillier, J. R. Helliwell, **1998**. *J. Chem. Soc., Faraday Trans.,* **94** p1603-1611.

9. D. N. Moothoo, **1988**. *PhD Thesis,* University of St. Andrews.

10. F. A. Quiocho, **1991**. *Trans. Am. Crystallogr. Assoc.,* **25** p23-35

11. L. N. Johnson, J. Cheetham, P. J. McLaughlin, K. R. Acharya, D. Barford, D. C. Phillips, **1988**. *Curr. Topics Microbiol. Immunol.,* **139** p81-134

12. L. N. Johnson, K. R. Acharya, M. D. Jordan & P. J. McLaughlin, **1990**. *J. Mol. Biol.,* **211** p645-661

13. M. Cygler, D. R. Rose, & D. R. Bundle, **1991**. *Science,* **253** p442-445

14. F. A. Quiocho, **1990**. *Phil. Trans. R. Soc. Lond.,* **326** p341-351

15. F. A. Quiocho & N. K. Vyas, **1984**. *Nature,* **310** p381-386

16. N. K. Vyas, M. N. Vyas, & F. A. Quiocho, **1988**. *Science,* **242** p1290-1295

17. F. A. Quiocho, **1986**. *Annu. Rev. Biochem.,* **55** p287-315

18. Y. Bourne, A. Roussel, M. Frey, P. Rouge, J. C. Fonntecilla-Camps, C. Cambillau, **1990**. *Proteins: Struct., Funct. Genet.,* **8** p365-376

19. M. I. Page & W. P. Jencks, **1971**. *Proc. Natl. Acad. Sci. USA,* **68**(8) p1678-83

20. F. A. Quiocho, D. K. Wilson & N. K. Vyas, **1989**. *Nature,* **340** p404-427

21. N. K. Vyas, M. N. Vyas & F. A. Quiocho, **1991**. *J. Biol. Chem.,* **266** p5226-5237

22.   N. C. J. Strynadka & M. N. G. James, **1991**. *J. Mol. Biol.,* **220** p401-424

23.   F. A. Quiocho, **1989**. *Pure Appl. Chem.,* **61** p1293-1306

24.   N. Sharon, **1993**. *Trends Biochem. Sci.,* **18** p221-226

25.   W. C. Boyd & E. Shapleigh, **1954**. *Science,* **119** p419

26.   I. J. Goldstein, R. C. Hughes, M. Monsigny, T. Osawa & N. Sharon, **1980**. *Nature,* **285** p66

27.   J. Kocourek, V. Horejsi, **1983**. *'Lectins. Biology, Biochemistry, Clinical Biochemistry'* **3** p3-6 T. C. Bog-Hansen, G. A. Spengler (Eds.), Walter de Gruyter, Berlin.

28.   H. Franz, P. Ziska, J. Mohr, **1982**. *Acta Histochem.,* **71** p19-21

29.   I. E. Liener, N. Sharon, I. J. Goldstein, **1986**. *'The Lectins: Properties, Functions and Applications in Biology and Medicine'* Academic Press, Orlando.

30.   N. Sharon & H. Lis, **1989**. *Science,* **246** p227-234

31.   N. Sharon & H. Lis, **2002**. *J. Agri. Food Chem.,* **50** p6586-6591

32.   K. Drickamer & M. E. Taylor, **1993**. *Annu. Rev. Cell Biol.,* **9** p237-264

33.   D. C. Kilpatrick, **2002**. *Biochim. Biophys. Acta,* **1572** p187-197

34.   J. M. Charcot & C. Robin, **1853**. *C. R. Mem. Soc. Biol.,* **5** p44-50

35.   E. Leyden, **1872**. *Arch. Pathol. Anat.,* **54** p324-344

36.   S. J. Ackerman, S. E. Corrette, H. F. Rosenberg, J. C. Bennett, D. M. Mastrianni, A. Nicholson-Weller, P. F. Weller, D. T. Chin, D. G. Tenen, **1993**. *J. Immunol.,* **150** p456-468

37.   G. J. Swaminathan, D. D. Leonidas, M. P. Savage, S. J. Ackerman, K. R. Acharya, **1999**. *Biochem.,* **38** p13837-13843

38.   S. Flexner & H. Noguchi, **1902**. *J. Exp. Med.,* **6** p277-301

39.   K. Drickamer, **1998**. *J. Biol. Chem.,* **263** p9557-9560

40.   H. J. Gabius, S. Andre, H. Kaltner, H-C. Siebert, **2002**. *Biochim. Biophys. Acta,* **1572** p165-177

41.   R. Loris, **2002**. *Biochim. Biophys. Acta,* **1572** p198-208

42.   T. Nagai, S. Kawabata, F. Shishikura, H. Sugita, **1999**. *J. Biol. Chem.,* **274** p37673-37678

43.   M. J. Chrispeels & N. V. Raikhel, **1991**. *The Plant Cell,* **3** p1-9

44.   W. J. Peumans & E. J. M. Van Damme, **1996**. *Histochem. J.,* **27** p253-271

45.   V. A. Hilder, **1995**. *Transgenic Res.,* **4** p18-25

46.   H. Lis & N. Sharon, **1986**. *Annu. Rev. Biochem.,* **55** p35-67

47.   J. B. Sumner, **1919**. *J. Biol. Chem.,* **37** p137-142

48.   K. D. Hardman & C. F. Ainsworth, **1972**. *Biochem.,* **11** p4910

49.   Z. Derewenda, J. Yariv, J. R. Helliwell, A. J. Kalb (Gilboa), E. J. Dodson, M. Z. Papiz, T. Wan, J. Campbell, **1989**. *EMBO J.* **8** p2189-2193.

50.   A. Deacon, T. Gleichmann, A. J. Kalb (Gilboa), H. Price, J. Raftery, G. Bradbrook, J. Yariv & J. R. Helliwell, **1997**. *J. Chem. Soc., Faraday Trans.,* **93** p4305-4312

51.   J. Habash, J. Raftery, R. Nuttall, H. J. Price, C. Wilkinson, A. J. Kalb (Gilboa), J. R. Helliwell, **2000**. *Acta Cryst.,* **D56** p541-550

52.   J. Habash, J. Raftery, S. Weisgerber, A. Cassetta, M. S. Lehmann, P. Hoghoj, C. Wilkinson, J. W. Cambell & J. R. Helliwell, **1997**. *J. Chem. Soc., Faraday Trans.,* **93** p4313-4317

53.   N. Sharon & H. Lis, **1990**. *FASEB J.,* **4** p3198-3208

54.   H. Lis & N. Sharon, **1998**. *Chem. Rev.,* **98** p637-674

55.   P. Argos, M. G. Rossmann, **1980**. *J. Mol. Evol.,* **15** p169-179

56.   G. Chelvanayagam, J. Heringa & P. Argos, **1992**. *J. Mol. Biol.,* **228** p220-242

57.   W. Tempel, S. Tschampel, & R. J. Woods, **2002**. *J. Biol. Chem.,* **277**(8) p6615-6621

58.   R. Banerjee, K. Das, R. Ravishankar, K. Suguna, A. Surolia, M. Vijayan, **1996**. *J. Mol. Biol.,* **259** p281-296

59.   R. Loris, T. Hamelryck, J. Bouckaert, L. Wyns, **1998**. *Biochim. Biophys. Acta.,* **1383** p9-36

60.   A. Imberty, C. Gautier, J. Lescar, S. Perez, L. Wyns & R. Loris, **2000**. *J. Biol. Chem.,* **275** p17541-17548.

61.   K. D. Hardman, R. C. Agarwal, M. J. Frieser, **1982**. *J. Mol. Biol.,* **157** p69-89

62.   J. Bouckaert, R. Loris, F. Poortmans, L. Wyns, **1995**. *Proteins: Struct. Func. Genet.,* **23** p510-524

63.   J. Bouckaert, R. Loris, F. Poortmans, L. Wyns, **1996**. *J. Biol. Chem.,* **271** p16144-16150

64. N. Manoj, V. R. Srinivas, & K. Suguna, **1999**. *Acta Cryst.,* **D55** p794-800

65. B. Shaanan, H. Lis & N. Sharon, **1991**. *Science,* **254** p862-866.

66. A. Surolia, N. Sharon & F. P. Schwarz, **1996**. *J. Biol. Chem.,* **271** p17697-17703

67. R. Adar & N. Sharon, **1996**. *Eur. J. Biochem.,* **239** p668-674.

68. S. Elgavish & B. Shaanan, **1998**. *J. Mol. Biol.,* **1998 277** p917-932.

69. R. Arango, E. Rodriguez-Arango, R. Adar, D. Belenky, F. L. Loontiens, S. Rozenblatt & N. Sharon, **1993**. *FEBS Lett.,* **330** p133-136.

70. G. M. Bradbrook, J. R. Forshaw & S. Perez, **2000**. *Eur. J. Biochem.,* **267** p4545-4555

71. R. Banerjee, S. C. Mande, V. Ganesh, K. Das, V. Dhanaraj, S. K. Mahanta, K. Suguna, A. Surolia, M. Vijiyan, **1994**. *Proc. Natl. Acad. Sci. USA,* **91** p227-231

72. A. Dessen, D. Gupta, S. Sabesan, C. F. Brewer, J. C. Sacchettini, **1995**. *Biochemistry,* **34** p4933-4942.

73. M. M. Prabu, R. Sankaranarayanan, K. D. Puri, V. Sharma, A. Surolia, M. Vijayan, & K. Suguna, **1998**. *J. Mol. Biol.,* **276** p787-796

74. N. Manoj, V. R. Srinivas, A. Surolia, M. Vijayan & K. Suguna, **2000**. *J. Mol. Biol.,* **302** p1129-1137

75. V. R. Srinivas, G. B. Reddy & A. Surolia, **1999**. *FEBS letts.,* **450** p181-185

76. N. Mitra, V. R. Srinivas, T. N. C. Ramya, N. Ahmad, G. B. Reddy & A. Surolia, **2002**. *Biochem.,* **41** p9256-9263

77. S. Acharya, S. R. Patanjali, S. U. Sajjan, B. Gopalakrishnan & A Surolia, **1990**. *J. Biol. Chem.,* **265** p11586-11594

78. E. Osinaga, D. Tello, C. Batthyany, M. Bianchet, G. Tavares, R. Duran, C. Cervenansky, L. Camoin, A. Roseto & P. M. Alzari, **1997**. *FEBS Letts.,* **412** p190-196

79. A. Babino, D. Tello, A. Rojas, S. Bay, E. Osinaga, P. M. Alzari, **2003**. *FEBS Letts.,* **536** p106-110

80. J. Lescar, R. Loris, E. Mitchell, C. Gautier, V. Chazalet, V. Cox, L. Wyns, S. Perez, C. Breton, A. Imberty, **2002**. *J. Biol. Chem.,* **277**(8) p6608-6614

81. C. Svensson, S. Teneberg, C. L. Nilsson, A. Kjellberg, F. P. Schwarz, N. Sharon, U. C. Krengel, **2002**. *J. Mol. Biol.,* **321** p69-83

82.   A. Rabijns, C. Verboven, P. Rouge, A. Barre, E. J. Van Damme, W. J. Peumans, C. J. De Ranter**, 2001**. *Proteins: Struct. Func. Genet.,* **44** p470-478

83.   A. Rabijns, C. Verboven, E. J. Van Damme, W. J. Peumans, C. J. De Ranter, **2000**. *Acta Cryst.,* **D56** p1638-1640

84.   M. Dao-Thi, T. W. Hamelryck, J. Bouckaert, F. Korber, V. Burkow, F. Poortmans, M. Etzler, G. Strecker, L. Wyns & R. Loris, **1998**. *Acta Cryst*., **D54** p1446-1449

85.   T. W. Hamelryck, R. Loris, J. Bouckaert, M. Dao-Thi, G. Strecker, A. Imberty, E. Fernandez, L. Wyns, M. E. Etzler, **1999**. *J. Mol. Biol.,* **286** p1161-1177.

86.   L. Buts, M. H. Dao-Thi, R. Loris, L. Wyns, M. E. Etzler, T. W. Hamelryck, **2001**. *J. Mol. Biol.,* **309** p193-201

87.   H. Einspahr, E. H. Parks, K. Suguna, E. Subramanian & F. L. Suddath, **1986**. *J. Biol. Chem.,* **261** p16518-16527

88.   J. M. Rini, K. D. Hardman, H. Einspahr, F. L. Suddath & J. P. Carver, **1993**. *J. Biol. Chem.,* **268** p10126-10132

89.   Y. Bourne, C. Abergel, C. Cambillau, M. Frey, P. Rouge & J. C. Fontecilla-Camps, **1990**. *J. Mol. Biol.,* **214** p571-584

90.   Y. Bourne, P. Rouge & C. Cambillau, **1990**. *J. Biol. Chem.,* **265** p18161-18165

91.   Y. Bourne, P. Rouge & C. Cambillau, **1992**. *J. Biol. Chem.,* **267** p197-203

92.   Y. Bourne, A. Ayouba, P. Rouge & C. Cambillau, **1994**. *J. Biol. Chem.,* **269** p9429-9435

93.   Y. Bourne, J. Mazurier, D. Legrand, P. Rouge, J. Montreuil, G. Spik & C. Cambillau, **1994**. *Structure,* **2** p209-219

94.   T. W. Hamelryck, J. G. Moore, M. J. Chrispeels, R. Loris, L. Wyns, **2000**. *J. Mol. Biol.,* **299** p875-883

95.   R. Loris, J. Steyaert, D. Maes, J. Lisgarten, R. Pickersgill, L. Wyns, **1993**. *Biochem.,* **32** p8772-8781

96.   F. Casset, T. W. Hamelryck, R. Loris, J. Brisson, C. Tellier, M. Dao-Thi, L. Wyns, F. Poortmans, S. Perez & A. Imberty, **1995**. *J. Biol. Chem.,* **270** p25619-25628

97.   R. Loris, D. Van Overberge, M. Dao-Thi, F. Poortmans, N. Maene, L. Wyns, **1994**. *Proteins: Struct. Func. Genet.,* **20** p330-346

98.   D. A. Wah, A. Romero, F. Gallego del Sol, B. S. Cavada, M. V. Ramos, T. B. Grangeiro, A. H. Sampaio, J. J. Calvete, **2001**. *J. Mol. Biol.,* **310** p885-894

99.   J. Sanz-Aparicio, J. Hermoso, T. B. Grangeiro, J. J. Calvete, B. S. Cavada, **1997**. *FEBS Letts*., **405** p114-118

100.   D. A. Rozwarski, B. M. Swami, C. F. Brewer, J. C. Sacchettini, **1998**. *J. Biol. Chem.,* **273** p32818-32825

101.   T. K. Dam, R. Roy, D. Page, C. F. Brewer, **2002**. *Biochem,* **41** p1351-1358

102.   T. K. Dam, S. Oscarson, C. F. Brewer, **1998**. *J. Biol. Chem.,* **273** p32812-32817

103.   T. K. Dam, S. Oscarson, J. C. Sacchettini, C. F. Brewer, **1998**. *J. Biol. Chem.,* **273** p32826-32832

104.   T. K. Dam, R. Roy, S. K. Das, S. Oscarson, C. F. Brewer, **2000**. *J. Biol. Chem.,* **275** p14223-14230

105.   D. Gupta, S. Oscarson, T. S. Raju, P. Stanley, E. J. Toone, C. F. Brewer, **1996**. *Eur. J. Biochem.,* **242** p320-326

106.   G. N. Reeke & J. W. Becker, **1986**. *Science,* **234** p1108-1111

107.   R. Loris, A. Imberty, S. Beeckmans, E. Van Driessche, J. S. Read, J. Bouckaert, H. De Greve, L. Buts, L. Wyns, **2003**. *J. Biol. Chem.,* **278** (18) p16297-16303

108.   L. T. J. Delbaere, M. Vandonselaar, L. Prasad, J. W. Quail, K. S. Wilson, Z. Dauter, **1993**. *J. Mol. Biol.,* **230** p950-965

109.   T. W. Hamelryck, M. H. Dao-Thi, F. Poortmans, M. J. Chrispeels, L. Wyns, R. Loris, **1996**. *J. Biol. Chem.,* **271** p20479-20485

110.   R. N. Knibbs, I. J. Goldstein, R. M. Ratcliffe, N. Shibuya, **1991**. *J. Biol. Chem.,* **266** p83-88

111.   W. Wang & R. D. Cummings, **1986**. *J. Biol. Chem.,* **263** p4576-4585

112.   M. H. Dao-Thi, P. Rizkallah, L. Wyns, F. Poortmans, R. Loris, **1998**. *Acta Cryst.,* **D54**, p844-847

113.   R. Loris, H. De Greve, M. H. Dao-Thi, J. Messens, A. Imberty, L. Wyns, **2000**. *J. Mol. Biol.,* **301**(4) p987-1002

114.   G. F. Audette, M. Vandonselaar, L. T. J. Delbaere, **2000**. *J. Mol. Biol.,* **304** p423-433

115. T. K. Dam, B. S. Cavada, T. B. Grangeiro, C. F. Santos, F. A. de Sousa, S. Oscarson, C. F. Brewer, **1998**. *J. Biol. Chem.,* **273** p12082-12088

116. H. Debray, D. Decout, G. Strecker, G. Spik, J. Montreuil, **1981**. *Eur. J. Biochem.,* **117** p41-55

117. K. Zhu, R. A. Bressan, P. M. Hasegawa, L. L. Murdock, **1996**. *FEBS Lett.,* **390** p271-274.

118. H. G. Beisel, S. I. Kawabata, S. Iwanaga, R. Huber, W. Bode, **1999**. *EMBO J.* **18** p2313– 2322.

119. T. E. Mirkov, M. J. Chrispeels, **1993**. *Glycobiology* **3** p381-387.

120. R. R. Van Eijsden, B. S. De Pater, J. W. Kijne, **1994**. *Glycoconjugate J.* **11** p375-380.

121. R. Loris. F. Casset, J. Bouckaert, J. Pletinckx, M. H. Dao-Thi, F. Poortmans, A. Imberty, S. Perez, L. Wyns, **1994**. *Glycoconjugate J.* **11** p507-517.

122. V. Sharma & A. Surolia, **1997**. *J. Mol. Biol.,* **267** p433-445

123. P. A. Karplus & C. Faerman **1994**. *Curr. Opin. Struct. Biol.,* **4** p770-776

124. G. Otting, E. Liepinsh, K. Wuthrich, **1991**. *Science,* **254** p974-980

125. X. D. Cheng & B. P. Schoenborn, **1991**. *J. Mol. Biol.,* **220** p381-399

126. S. H. Koenig, R. D. Brown 3rd, R. Ugolini, **1992**. *Mag. Res. in Med.,* **193** p77-83

127. W. Gu, A. E. Garcia, B. P. Schoenborn, **1996**. *Neutrons in biology,* **64** p289-298

128. B. W. Matthews, **1968**. *J. Mol. Biol.,* **33** p491-497

129. H. Savage & A. Wlodawer, **1986**. *Methods Enzymol.,* **127** p162-183

130. W. C. Hamilton & J. A. Ibers, **1968**. *'Hydrogen bonding in solids'* Benjamin, New York

131. P. Schuster, G. Zundel, G. Sandorfy (Eds.), **1975**. *'The hydrogen bond'* **2** North Holland Publ., Amsterdam

132. B. Pederson, **1974**. *Acta Cryst.,* **B30** p289

133. G. Ferraris & M. Franchini-Angela, **1972**. *Acta Cryst.,* **B28** p3572

134. G. Chiari & G. Ferraris, **1982**. *Acta Cryst.,* **B38** p2331

135. T. Steiner & W. Saenger, **1993**. *Acta. Cryst.,* **D49** p592-593

136. T. Steiner & W. Saenger, **1993**. *J. Am Chem. Soc.,* **115** p4540-4547

137. Z. Derewenda & U. Derewenda, **1995**. *J. Mol. Biol.,* **252** p248-262

138. H. Savage & J. L. Finney, **1986**. *Nature,* **322** p717-720

139. H. Savage, **1986**. *Biophys. J.,* **50** p967-980

140. R. Loris, P. P. G. Stas & L. Wyns, **1994**. *J. Biol. Chem.* **269** p26722-26733.

141. E. J. Toone, **1994**. *Curr. Opin. Struct. Biol.,* **4** p719-728

142. J. H. Naismith & R. A. Field, **1996**. *J. Biol. Chem.,* **271** p972-976

143. R. Ravishankar, M. Ravindran, K. Suguna, A. Surolia & M. Vijayan, **1997**. *Curr. Science,* **72** p855-861

144. J. H. Naismith, C. Emmerich, J. Habash, S. J. Harrop, J. R. Helliwell, W. N. Hunter, J. Raftery, A. J. Kalb (Gilboa), J. Yariv, **1994**. *Acta Cryst.,* **D50** p847-858

145. C. Emmerich, J. R. Helliwell, M. Redshaw, J. H. Naismith, S. J. Harrop, J. Raftery, A. J. Kalb (Gilboa), J. Yariv, Z. Dauter, K. S. Wilson, **1994**. *Acta Cryst.,* **D50** p749-756

146. I. Tsyba & R. Bau, **2002**. *Chemtracts-Inorg. Chem.,* **15** p233-257

147. B. P. Schoenborn, **1996**. *'Neutrons in Biology, Basic Life Sciences'* **64** p1-16 B. P. Schoenborn & R. B. Knott (Eds.), Plenum Press, New York

148. J. R. Helliwell, **1997**. *Nature Struct. Biol.,* **4** p874-876

149. D. A. A. Myles, C. Bon, P. Langan, F. Cipriani, J. C. Castagna, M. S. Lehmann, C. Wilkinson, **1998**. *Physica B,* **241–243** p1122-1130

150. N. Niimura, Y. Minezaki, T. Nonaka, J. C. Castagna, F. Cipriani, P. Hoghoj, M. S. Lehmann, C. Wilkinson, **1997**. *Nature Struct. Biol.,* **11** p909-914

151. B. P. Schoenborn, **1992**. *SPIE*, **1738** p192-198

152. C. Wilkinson & M. S. Lehmann, **1991**. *Nucl Instrum. Meths.,* **A310** p411-415

153. F. Cipriani, J. C. Castagna, M. S. Lehmann, C. Wilkinson, **1995**. *Physica,* **B213** p975-977

154. C. Rausch, T. Bücherl, R. Gähler, H. Seggern, A. Winnacker, **1992**. *SPIE*, **1737** p255-263

155. C. Wilkinson, A. Gabriel, M. S. Lehmann, T. Zemb, F. Né, **1992**. *SPIE*, **1737** p324-329

156. N. Niimura, Y. Karasawa, I. Tanaka, J. Miyahara, K. Kahashi, H. Saito, S. Koizumi & M. Hidaka, **1994**. *Nucl. Instrum. Meths.,* **A349** p521-525

157. D. W. Cruickshank, J. R. Helliwell & K. Moffat, **1987**. *Acta Cryst.,* **A43** p656-674

158. J. R. Helliwell, **1985**. *J. Mol. Struct.,* **130** p63-91

159. A. J. Kalb (Gilboa), D. A. A. Myles, J. Habash, J. Raftery & J. R. Helliwell, **2001**. *J. Appl. Cryst.,* **34** p454-457

160. Z. R. Korszun, **1997**. *Methods Enzymol.,* **276** p218-232

161. D. L. Worcester, **1988**. *J. Appl. Cryst.,* **21** p669-674

162. S. J. Cooper, D. Brockwell, J. Raftery, D. Attwood, J. Barber & J. R. Helliwell, **1998**. *Chem. Commun.,* p1063-1064

163. T. R. Gamble, K. R. Clauser & A. A. Kossiakoff, **1994.** *Biophys. Chem.,* **53** p15-26

164. K. B. Schowen, R. L. Schowen, **1982**. *Methods Enzymol.,* **87** p551

165. D. S. Berns, **1963**. *Biochem.,* **6** p1377

166. A. Hattori, H. L. Crespi, J. J. Katz, **1965**. *Biochem*., **4** p1213

167. D. J. Brockwell, **1996**. *PhD Thesis,* University of Manchester.

168. J. P. Derrick, L. Y. Lian, G. C. Roberts, W. V. Shaw, **1992**. *Biochem.,* **31** p8191

169. R. J. Brennan, A. Awan, J. Barber, E. Hunt, K. L. Kennedy & S. Sadegholnejat, **1994**. *J. Chem. Soc., Chem. Commun.,* p1615

170. J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips & V. C. Shore, **1960**. *Nature,* **185** p422-427

171. S. E. V. Phillips, **1980**. *J. Mol. Biol.,* **142** p531-554

172. S. E. V. Phillips & B. P. Schoenborn, **1981**. *Nature,* **292** p81-82

173. A. Jack & M. Levitt, **1978**. *Acta Cryst.,* **A34** p931-935

174. J. C. Hanson & B. P. Schoenborn, **1981**. *J. Mol. Biol.,* **153** p117-146

175. J. C. Norvell, A. C. Nunes & B. P. Schoenborn, **1975**. *Science,* **190** p568-70

176. R. Diamond, **1971**. *Acta Cryst.,* **A27** p436-452

177. X. D. Cheng & B. P. Schoenborn, **1990**. *Acta Cryst.,* **B46** p195-208

178. W. A. Hendrickson & J. A. Konnert, **1980**. *Biomolec. Struct. Func. Conf. and Evolution* **1** p43-57

179. W. A. Hendrickson, **1985**. *Methods Enzymol.,* **115** p252-270

180. B. P. Schoenborn, **1988**. *J. Mol. Biol.,* **201** p741-749

181. N. V. Raghavan & B. P. Schoenborn, **1984**. *Neutrons in biology,* **27** p247-259

182. M. L. Bender & F. J. Kezdy, **1964**. *J. Am. Chem. Soc.,* **86** p3704-3714

183. P. W. Inward & W. P. Jencks, **1965**. *J. Biol. Chem.,* **240** p1986-1996

184. D. M. Blow, **1976**. *Acc. Chem. Res.,* **9** p145-152

185. J. Kraut, **1977**. *Ann. Rev. Biochem.,* **46** p331-358

186. R. M. Stroud, L. M. Kay, R. E. Dickerson, **1974**. *J. Mol. Biol.,* **83** p185

187. A. A. Kossiakoff & S. A. Spencer, **1980**. *Nature,* **288** p414-416

188. J. E. Cain, J. C. Norvell & B. P. Schoenborn, **1976**. *Brookhaven Symp. Biol.,* **27** p43-50

189. J. L. Chambers & R. M. Stroud, **1977**. *Protein Data Bank (Brookhaven National Lab.)*

190. S. T. Freer, R. A. Alden, C. W. Carter, J. Kraut, **1975**. *J. Biol. Chem.,* **250** p46

191. J. L. Chambers & R. M. Stroud, **1977**. *Acta Cryst.,* **B33** p1824-1837

192. A. A. Kossiakoff & S. A. Spencer, **1981**. *Biochem.,* **20** p6462-6474

193. M. M. Teeter, S. M. Roe, N. H. Heo, **1993**. *J. Mol. Biol.,* **230** (1) p292-311

194. A. Yamano & M. M. Teeter, **1994**. *J. Biol. Chem.,* **269** (19) p13956-13965

195. C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing, C. Lecomte, **2000**. *Proc. Natl. Acad. Sci. USA,* **97** (7) p3171-3176

196. M. M. Teeter, A. Yamano, B. Stec, U. Mohanty, **2001**. *Proc. Natl. Acad. Sci. USA,* **98** (20) p11242-11247

197. M. Llinas, A. DeMarco, J. T. Lecomte, **1980**. *Biochemistry* **19** p1140-1145

198. W. A. Hendrickson & M. M. Teeter, **1981**. *Nature,* **290** p107-113

199. M. M. Teeter, A. A. Kossiakoff, **1984**. *'Neutrons in Biology, Basic Life Sciences'* **27** p335-348. B. P. Schoenborn (Ed.), Plenum Press, New York

200. A. A. Kossiakoff & S. Shetyn, **1984**. *Nature,* **311** p582-583

201. A. A. Kossiakoff, M. Ultsch, S. White, C. Eigenbrot, **1991**. *Biochem.,* **30** p1211-1221

202. R. Bott, M. Ultsch, A. A. Kossiakoff, T. Graycar, B. Katz, S. Power, **1988**. *J. Biol. Chem.,* **263** p7895-7906

203. J. Hermans & J. E. McQueen, **1974**. *Acta Cryst.,* **A30** p730-739

204. H. P. Avey, M. O. Boles, C. H. Carlisle, S. A. Evans, S. J. Morris, R. A. Palmer, B. A. Woolhouse, S. Shall, **1967**. *Nature,* **213** p557

205. G. Kartha, J. Bello, D. Harker, **1967**. *Nature,* **213** p862-865

206. D. D. Leonidas, R. Shapiro, L. I. Irons, N. Russo, K. R. Acharya, **1997**. *Biochemistry,* **36** p5578-5588

207. A. Wlodawer & L. Sjolin, **1984**. *Neutrons in biology,* **27** p349-363

208. A. Wlodawer, R. Bott & L. Sjolin, **1982**. *J. Biol. Chem.,* **257** p1325

209. A. Wlodawer & L. Sjolin, **1982**. *Proc. Natl. Acad. Sci. USA,* **79** p1418-1422

210. S. A. Mason, G. A. Bentley, G. J. McIntyre, **1984**. *'Neutrons in Biology, Basic Life Sciences'* **27** p323-334. B. P. Schoenborn (Ed.), Plenum Press, New York

211. J. J. Abel, **1926**. *Proc. Natl. Acad. Sci. USA,* **122** p132-136

212. D. Crowfoot, **1935**. *Nature,* **135** p591-592

213. A. Wlodawer & H. Savage, **1989**. *Acta Cryst.,* **B45** p99-107

214. E. J. Dodson, G. G. Dodson, D. C. Hodgkin, C. D. Reynolds, **1979**. *Can. J. Biochem.,* **57** p469-479

215. P. G. Lenhert & D. C.Hodgkin, **1961**. *Nature,* **192** p937-938

216. H. Savage, P. F. Lindley, J. L. Finney & P. A. Timmins, **1987**. *Acta Cryst.,* **B43** p280-295

217. H. Savage, **1986**. *Biophys. J.,* **50** p947-965

218. F. Cipriani, J. C. Castagna, C. Wilkinson, P. Oleinek, & M. S. Lehmann, **1996**. *J. Neutron Res.,* **4** p79-85

219. T. Imoto, L. N. Johnson, A. C. T. North, D. C. Phillips, J. A. Rupley, **1972**. *'The Enzymes'* **7** p665-868. P. D. Boyer (Ed.), Academic Press, New York

220. M. C. Vaney, S. Maignan, M. Ries-Kautt & A. Ducruix, **1996**. *Acta Cryst.,* **D52** p505-517

221. D. C. Phillips, **1966**. *Sci. Am.,* **215**(5) p78-90

222. C. Bon, M. S. Lehmann & C. Wilkinson, **1999**. *Acta. Cryst.,* **D55** p978-987

223. M. A. Walsh, T. Schneider, L. C. Sieker, Z. Dauter, V. Lamzin & K. Wilson, **1998**. *Acta. Cryst.,* **D54** p522-547

224. L. Coates, P. T. Erskine, S. P. Wood, D. A. A. Myles, J. B. Cooper, **2001**. *Biochemistry,* **40** p13149-13157

225. F. Shu, V. Ramakrishnan, B. P. Schoenborn, **2000**. *Proc. Natl. Acad. Sci., USA,* **97** p3872-3877

226. B. P. Schoenborn, **1984**. *'Neutrons in Biology, Basic Life Sciences'* **27** p261-

281. B. P. Schoenborn (Ed.), Plenum Press, New York

227. S. J. Rinderle, I. J. Goldstein, K. L. Matta, R. M. Ratcliffe, **1989**. *J. Biol. Chem.,* **264** p16123-16131

228. M. J. Swamy, D. Gupta, S. K. Mahanta & A. Surolia, **1991**. *Carbohydr. Res.,* **213** p59-67

229. A. M. Wu, **1984**. *Mol. Cell Biochem.,* **61** p131-141

230. H. Clansen *et al.,* **1988**. *Mol. Immunol.,* **25** p199-204

231. D. O'Keefe & L. Ashman, **1982**. *Clin. Exp. Immunol.,* **48** p329-338

232. C. K. Cling & J. K. Rhodes, **1989**. *Br. J. Cancer.,* **59** p949-953

233. P. L. Zabel, A. A. Noujaim, A. Shysh, J. Bray, **1983**. *Eur. J. Nucl. Med.,* **8** p250-254

234. N. Zebda, M. Bailly, S. Brown, J. F. Dore, O. Berthier-Vergnes, **1994**. *J. Cell Biochem.,* **54** p161-173

235. Y. Reisner, M. Biniaminov, E. Rosenthal, N. Sharon, B. Ramot, **1979**. *Proc. Natl. Acad. Sci. USA,* **76** p447-451

236. M. M. Prabu, K. Saguna & M. Vijayan, **1999**. *Proteins: Struct. Func. Genet.,* **35** p58-69

237. S. Jones & J. M. Thornton, **1995**. *Prog. Biophys. Mol. Biol.,* **63** p31-65

238. W. W. Fish, L. M. Hamlin & R. L. Miller, **1978**. *Arch. Biochem. Biophys.,* **190** p693-698

239. H. Decastel, H. de Boeck, Y. Goussault, C. K. de Bruyne, F. G. Loontiens & J. P. Frenoy, **1985**. *Arch. Biochim. Biophys.,* **240** p811-819

240. D. M. Salunke, M. Islam Khan, A. Surolia & M. Vijayan, **1982**. *J. Mol. Biol.,* **154** p177-178

241. D. M. Salunke, M. Islam Khan, A. Surolia & M. Vijayan, **1983**. *FEBS Letts.,* **156** p127-129

242. M. E. Pereira, E. A. Kabat, R. Lotan, N. Sharon, **1976**. *Carbohydr. Res.,* **51** p107-118

243. R. Lotan, E. Skutelsky, D. Danon & N. Sharon, **1975**. *J. Biol. Chem.,* **250** p8518-8523

244. G. F. Springer, **1984**. *Science,* **224** p1198-1206

245. T. Irimura, T. Kawaguchi, T. Terao, T. Osawa, **1975**. *Carbohydr. Res.,* **39**

p317-327

246. D. M. Salunke, J. M. Swamy, M. Islam Khan, S. C. Mande, A. Surolia & M. Vijayan, **1985**. *J. Biol. Chem.,* **250** p13576-13579

247. R. Ravishankar *et al.*, **1996**. *Prog. Biophys. Mol. Biol.,* **65** Suppl 1 p33

248. N. M. Young & R. P. Oomen, **1992**. *J. Mol. Biol.,* **228** p924-934

249. W. Weis & K. Drickamer, **1996**. *Annu. Rev. Biochem.,* **65** p441-473.

250. T. Majumdar & A. Surolia, **1978**. *Prep. Biochem.,* **8** p119-131

251. Collaborative Computational Project, Number 4, **1994**. *Acta Cryst.,* **D50** p760-763

252. A. T. Brunger, **1992**. *Nature,* **335** p472-475

253. A. L. U. Roberts & A. T. Brunger, **1995**. *Acta Cryst.,* **D51** p990-1002

254. F. C. Bernstein, T. F. Koetzel, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi & M. Tasumi, **1977**. *J. Mol. Biol.,* **112** p535-542

255. T. A. Jones, J. Y. Zou, S. W. Cowan & M. Kjeldgaard, **1991**. *Acta Cryst.,* **A47** p110-119

256. C. Ramakrishnan & G. N. Ramachandran, **1965**. *J. Biophys.,* **5** p909-933

257. R. A. Laskowski, M. W. MacArthur, D. S. Moss & J. M. Thornton, **1993**. *J. Appl. Cryst.,* **26** p283-291

258. P. Manikandan, R. Carmieli, T. Shane, A. J. Kalb (Gilboa), D. Goldfarb, **2000**. *J. Am. Chem. Soc.,* **122**, p3488-3494

259. J. W. Becker, G. N. Reeke, J. L. Wang, B. A. Cunningham & G. M. Edelman, **1975**. *J. Biol. Chem.,* **250** p1513-1524

260. A. J. Kalb (Gilboa) & A. Levitzki, **1968**. *Biochem. J.,* **109** p669-672

261. M. Shoham, A. J. Kalb (Gilboa), I. Pecht, **1973**. *Biochem.,* **12** p1914-1917

262. S. J. Harrop, J. R. Helliwell T. Wan, A. J. Kalb (Gilboa), L. Tong & J. Yariv, **1996**. *Acta Cryst.,* **D52** p143-155

263. E. Meirovitch, Z. Luz, A. J. Kalb (Gilboa), **1974**. *J. Am. Chem. Soc.,* **96** p7538

264. J. Greer, H. Kaufman & A. J. Kalb (Gilboa), **1970**. *J. Mol. Biol.,* **48** p365-366

265. J. R. Helliwell, J. Habash, D. W. Cruickshank, M. M. Harding, T. J. Greenhough, J. W. Campbell, I. J. Clifton, M. Elser, P. A. Machin, M. Z. Papiz & S. Zurch, **1989**. *J. Appl. Cryst.,* **22** p483

266. J. W. Campbell, **1995**. *J. Appl. Cryst.,* **28** p228

267. J. W. Campbell, Q. Hao, M. M. Harding, N. D. Nguti & C. Wilkinson, **1998**. *J. Appl. Cryst.,* **31** p496

268. S. Arzt, J. W. Campbell, M. M. Harding, Q. Hao & J. R. Helliwell, **1999**. *J. Appl. Cryst.,* **32** p554

269. Z. Otwinowski & W. Minor, **1997**. *Methods Enzymol.,* **276** p307-326

270. V. S. Lamzin & K. S. Wilson, **1993**. *Acta Cryst.,* **D49** p129-143

271. J, Wenzierl & A. J. Kalb (Gilboa), **1971**. *Febs Letts.,* **18** p268-270

272. G. N. Reeke, J. W. Becker & G. M. Edelman, **1975**. *J. Biol. Chem.,* **250** p1525-1547

273. J. H. Naismith, J. Habash, S. Harrop, J.R. Helliwell, W. N. Hunter, T. C. M. Wan, S. Weisgerber, A.J. Kalb Gilboa , J. Yariv, **1993**. *Acta Cryst.,* **D49** p561-571

274. A. T. Brunger, P. D. Adams, G. M. Clore, W. L. Delano, P. Gros & R. W. Grosse-Kunstleve, **1998**. *Acta. Cryst.,* **D54** p905-921

275. J. Yariv, A. J. Kalb (Gilboa), M. Z. Papiz, J. R. Helliwell, S. J. Andrews & J. Habash, **1987**. *J. Mol. Biol.,* **195** p759-760

276. J. Yariv, A. J. Kalb (Gilboa), J. R. Helliwell, M. Z. Papiz, E. R. Bauminger & I. Nowik, **1988**. *J. Biol. Chem.,* **263** p13508-13510

277. D. K. Mandal, N. Kishore & C. F. Brewer, **1994**. *Biochem.,* **33** p1149-1156

278. M. Dani, F. Manca, G. Rialdi, **1981**. *Biochim. Biophys. Acta,* **667** p108

279. M. C. Chervenak & E. J. Toone, **1995**. *Biochem.,* **34** p5685-5695

280. A. Vladimir, B. Makarov, K. Andrews, P. E. Smith & B. Montgomery Pettitt, **2000**. *Biophysical J.,* **79** p2966–2974

281. X. D. Cheng & B. P. Schoenborn, **1991**. *Acta Cryst.,* **A47** p314-317

282. C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing & C. Lecomte, **2000**. *Proc. Natl. Acad. Sci. USA,* **97** p3171–3176

283. J. P. Bouqiere, J. L. Finney, H. F. Savage, **1994**. *Acta Cryst.,* **B50** p566-578

284. I. Rayment, **2002**. *Structure,* **10** p147-151

285. N. Chayen & R. Hilgenfeld, **2002**. *Trends in biotech.,* **20** p320-321

286. N. Chayen, **2002**. *Trends in biotech.,* **20** p98

287. E. Garman & C. Nave, **2002**. *J. Sync. Rad.,* **9** p327-328

288. P. O'Neill, D. L. Stevens, E. Garman, **2002**. *J. Sync. Rad.,* **9** p329-332

289. T. M. Kuzay, M. Kazmierczak, B. J. Hsich, **2001**. *Acta Cryst.,* **D57** p69-81

290. E. H. Snell, R. A. Judge, M. Larson, M. J. van der Woerd, **2002**. *J. Sync. Rad.,* **9** p361-367

291. R. Ravelli, P. Theveneau, S. McSweeney, M. Caffrey, **2002**. *J. Sync. Rad.,* **9** p355-360

292. T. Teng & K. Moffat, **2002**. *J. Sync. Rad.,* **9** p198-201

293. P. Sliz, **2003**. *Structure,* **11** p13-19

294. T. Teng & K. Moffat, **2002**. *J. Sync. Rad.,* **7** p313-317

295. V. Cherezov, K. M. Riedl, M. Caffrey, **2002**. *J. Sync. Rad.,* **9** p333-341

296. M. Weik, R. Ravelli, G. Kryger, S. McSweeney, M. L. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, J. L. Sussman, **2000**. *Proc. Natl. Acad. Sci. USA,* **97** p623-628

297. R. Ravelli & S. McSweeney, **2000**. *Structure,* **8** p315-328

298. M. Weik, J. Berges, M. L. Raves, P. Gros, S. McSweeney, I. Silman, J. L. Sussman, C. Houee-Levin, R. Ravelli, **2002**. *J. Sync. Rad.,* **9** p342-346

299. W. Burmeister, **2000**. *Acta Cryst.,* **D56** p328-341

300. K. Moffat & R. Henderson, **1995**. *Curr. Opin. Struct. Biol.,* **5** p656-663

301. K. Moffat, **1995**. *Biotechnology,* **13** p133

302. C. Nave, **1998**. *Acta Cryst.,* **D54** p848-853

303. J. R. Helliwell, **1992**. Macromolecular Crystallography with Synchrotron Radiation. Cambridge University Press.

304. H. D. Bellamy, E. H. Snell, J. Lovelace, M. Pokross, G. E. O. Borgstahl, **2000**. *Acta Cryst.,* **D56** p986-995

305. C. Nave, **1999**. *Acta Cryst.,* **D55** p1663-1668

306. A. Wlodawer & W. A. Hendrickson, **1982**. *Acta Cryst.,* **A38** p239-247

307. R. Carmieli, P. Manikandan, B. Epel, A. J. Kalb (Gilboa), A. Schnegg, A. Savitsky, K. Möbius, & D. Goldfarb, **2003**. *Biochemistry* **42** p7863-7870

308. R. Carmieli, P. Manikandan, A. J. Kalb (Gilboa), & D. Goldfarb, **2001**. *J. Am. Chem. Soc.,* **123(34)** p8378-8386