

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/8915225>

# Structural Basis of Oligomannose Recognition by the *Pterocarpus angolensis* Seed Lectin

ARTICLE in JOURNAL OF MOLECULAR BIOLOGY · FEBRUARY 2004

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2003.11.043 · Source: PubMed

---

CITATIONS

37

READS

48

## 7 AUTHORS, INCLUDING:



[Henri De Greve](#)

Vrije Universiteit Brussel

115 PUBLICATIONS 4,098 CITATIONS

[SEE PROFILE](#)



[Francine Deboeck](#)

Vrije Universiteit Brussel

20 PUBLICATIONS 1,116 CITATIONS

[SEE PROFILE](#)



[Lode Wyns](#)

Vrije Universiteit Brussel

269 PUBLICATIONS 9,679 CITATIONS

[SEE PROFILE](#)



[Julie Bouckaert](#)

Université des Sciences et Technologies de Li...

86 PUBLICATIONS 2,272 CITATIONS

[SEE PROFILE](#)



## Structural Basis of Oligomannose Recognition by the *Pterocarpus angolensis* Seed Lectin

**Remy Loris<sup>1\*</sup>, Ivo Van Walle<sup>1</sup>, Henri De Greve<sup>1</sup>, Sonia Beeckmans<sup>2</sup>  
Francine Deboeck<sup>3</sup>, Lode Wyns<sup>1</sup> and Julie Bouckaert<sup>1</sup>**

<sup>1</sup>Laboratorium voor Ultrastructuur, Instituut voor Moleculaire Biologie Building E, Vrije Universiteit Brussel and Vlaams Instituut voor Biotechnologie Pleinlaan 2, B-1050 Brussel Belgium

<sup>2</sup>Laboratorium voor Scheikunde der Proteïnen, Instituut voor Moleculaire Biologie Vrije Universiteit Brussel Pleinlaan 2, B-1050 Brussel Belgium

<sup>3</sup>Laboratorium voor Genetische Virologie, Instituut voor Moleculaire Biologie Vrije Universiteit Brussel Pleinlaan 2, B-1050 Brussel Belgium

\*Corresponding author

The crystal structure of a Man/Glc-specific lectin from the seeds of the bloodwood tree (*Pterocarpus angolensis*), a leguminous plant from central Africa, has been determined in complex with mannose and five manno-oligosaccharides. The lectin contains a classical mannose-specificity loop, but its metal-binding loop resembles that of lectins of unrelated specificity from *Ulex europaeus* and *Maackia amurensis*. As a consequence, the interactions with mannose in the primary binding site are conserved, but details of carbohydrate-binding outside the primary binding site differ from those seen in the equivalent carbohydrate complexes of concanavalin A. These observations explain the differences in their respective fine specificity profiles for oligomannoses. While Man(α1-3)Man and Man(α1-3)[Man(α1-6)]Man bind to PAL in low-energy conformations identical with that of ConA, Man(α1-6)Man is required to adopt a different conformation. Man(α1-2)Man can bind only in a single binding mode, in sharp contrast to ConA, which creates a higher affinity for this disaccharide by allowing two binding modes.

© 2003 Elsevier Ltd. All rights reserved.

**Keywords:** lectin; carbohydrate recognition; legume lectin; mannose

### Introduction

Protein-carbohydrate recognition is a major form of inter-cellular communication and plays a role in many biologically important processes such as viral, bacterial, mycoplasmal and parasitic infections, targeting of cells and soluble components, fertilisation, cancer metastasis and growth and differentiation.<sup>1</sup>

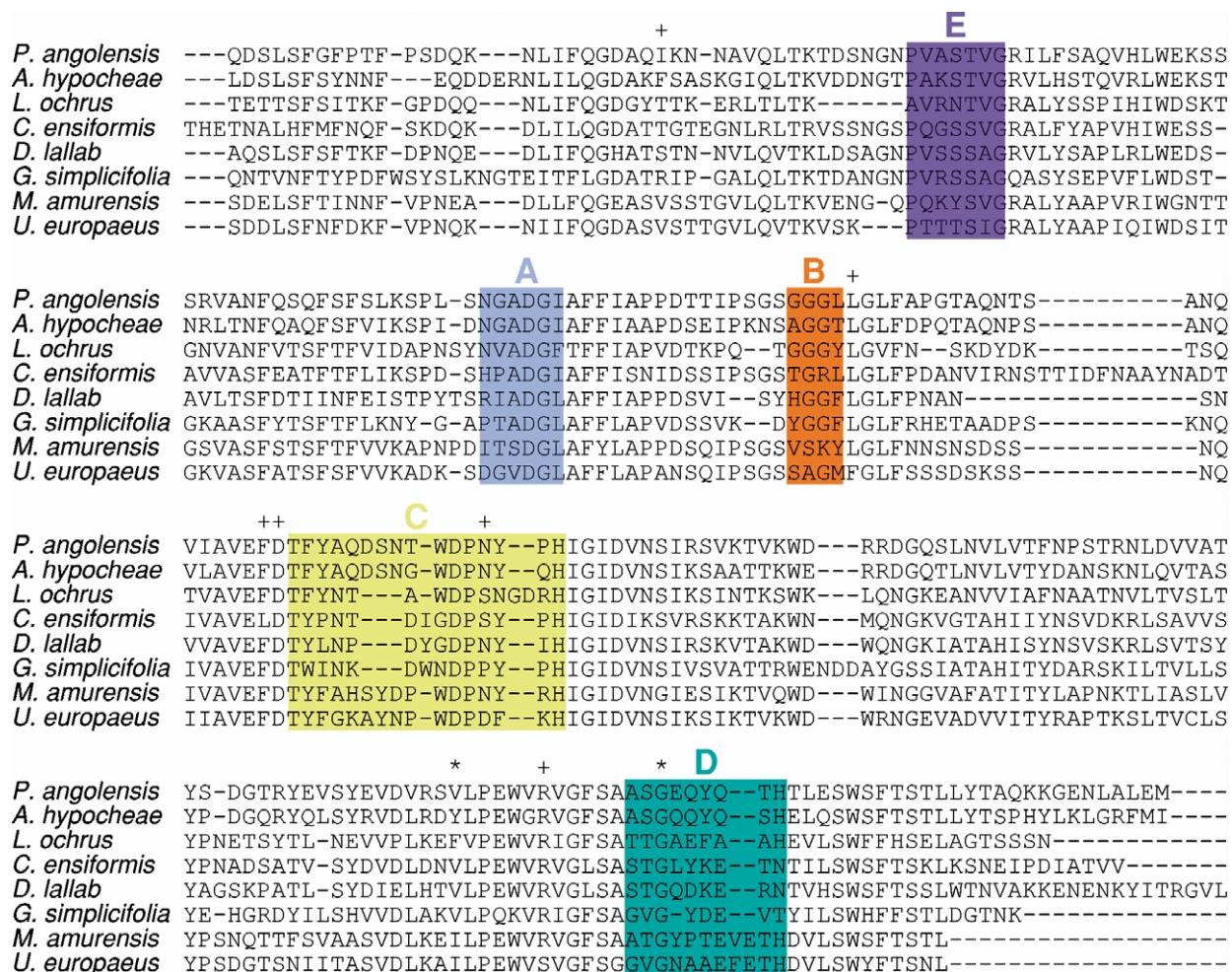
The specific recognition of an (oligo)saccharide by a protein is a much more complex problem

than other biologically relevant recognition processes such as protein-protein or protein-DNA interactions. The monosaccharide building blocks of a glycan are difficult to distinguish from each other due to the limited repertoire of functional groups involved. Apart from the occasional N-acetyl group or rarely a carboxylate group, one finds invariably a large abundance of hydroxyl groups interspersed with small aliphatic patches. In addition, the glycosidic bonds between two monosaccharides are rather flexible, especially the 1-6 linkage. As a consequence, a high entropic cost limits the binding affinities that can be obtained. The combination of flexibility and the difficulty in distinguishing monomeric building blocks allows oligosaccharides to mimic each other structurally, making the task of specific recognition a truly difficult one.

The lectins from leguminous plants have been considered as a model system for studying the molecular basis of protein-carbohydrate interactions for several decades. Among all known

Abbreviations used: ConA, concanavalin A from *Canavalia ensiformis*; PAL, *Pterocarpus angolensis* Man/Glc-specific seed lectin; MAL, *Maackia amurensis* leukoagglutinin; GS-IV, *Griffonia simplicifolia* lectin IV; UEA-II, *Ulex europaeus* lectin II; FRIL, Flt3 receptor interacting lectin from *Dolichos lablab*; LOL, *Lathyrus ochrus* lectin.

E-mail address of the corresponding author:  
reloris@vub.ac.be



**Figure 1.** Amino acid sequence of the *P. angolensis* lectin. The amino acid sequence of PAL as used in the crystal structure determination is aligned with those of other legume lectins discussed in the text. Residues where the crystal structure conflicts with the cDNA-derived sequences are indicated by an asterisk (\*). Positions where sequence variation was observed among the different cDNA clones sequenced are indicated by a +. The five stretches A-E that constitute the carbohydrate-binding site are indicated and shaded in different colours.

lectin families, animal, plant or microbial, they cover the widest range of carbohydrate specificities. Their carbohydrate recognition site consists of several loops, with differing degrees of variability.<sup>2,3</sup> The conformations of these loops are determined by the presence of a structural calcium ion and a transition metal ion,<sup>4-6</sup> the absence of which results in local unfolding and loss of carbohydrate-binding capacity. The vast amount of structural and biochemical data already available allows us to rationalise the structural determinants of carbohydrate specificity. Indeed, the crystal structures of more than 25 legume lectins have been determined<sup>3,7-20</sup> and are present in the Lectin Database†. Here, we present the crystal structure of the seed lectin (PAL) from the bloodwood tree (*Pterocarpus angolensis*; a leguminous plant from central Africa) in complex with a series of oligomannose ligands. This lectin belongs to the

Man/Glc specificity group and has a preference for Man(α1-2)Man and Man(α1-3)[Man(α1-6)]Man (L. Buts & S.B., unpublished results). Its fine specificity differs from that of other known Man/Glc-specific lectins, while its amino acid sequence suggests some peculiarities in the carbohydrate-binding sites not seen in other legume lectins with related specificities.

## Results and Discussion

### Amino acid sequence and overall structure of PAL

The amino acid sequence was deduced from 14 cDNA clones (data not shown). The lectin gene starts with an ATG start codon followed by a signal sequence corresponding to 34 amino acid residues. This signal peptide ends with a putative signal peptide cleavage site between residues Ser and Gln.<sup>21</sup> This glutamine is the N-terminal residue

† <http://www.cermav.cnrs.fr/lectines/>

observed in all X-ray structures and is present in the form of a cyclic glutamine residue. This is in agreement with the fact that the N terminus of the affinity-purified protein was proven to be blocked (data not shown).

At several places, heterogeneity was observed in the cDNA sequences: Ile/Thr26, Leu/Pro108, Phe/Leu129, Asp/His130, Asn/Asp143 and Gly/Arg212. In the electron density maps of our crystal structures, these residues were identified as Ile26, Leu108, Phe129, Asp130, Asn143 and Arg212. Since the sequences were determined starting from mRNA, it can be assumed that all the corresponding protein variants are indeed synthesised. With the exception of Ile/Thr26, all these residues are located at or near the carbohydrate-binding site. Especially Phe/Leu129 and Asp/His130 are known to be critical for carbohydrate and metal-binding and are conserved in other legume lectins. It is possible, therefore, that some of the cDNA clones correspond to proteins that are closely related to PAL but are inactive or have altered carbohydrate specificity.

**Figure 1** aligns the sequence of mature PAL as used in the crystal structure determination with those of other legume lectins that are discussed here. At two positions, the electron density was in disagreement with the cDNA-derived sequences: position 206 was interpreted as Val rather than Ala, since clear side-chain density was present. Val was chosen over Thr, as this side-chain is point-

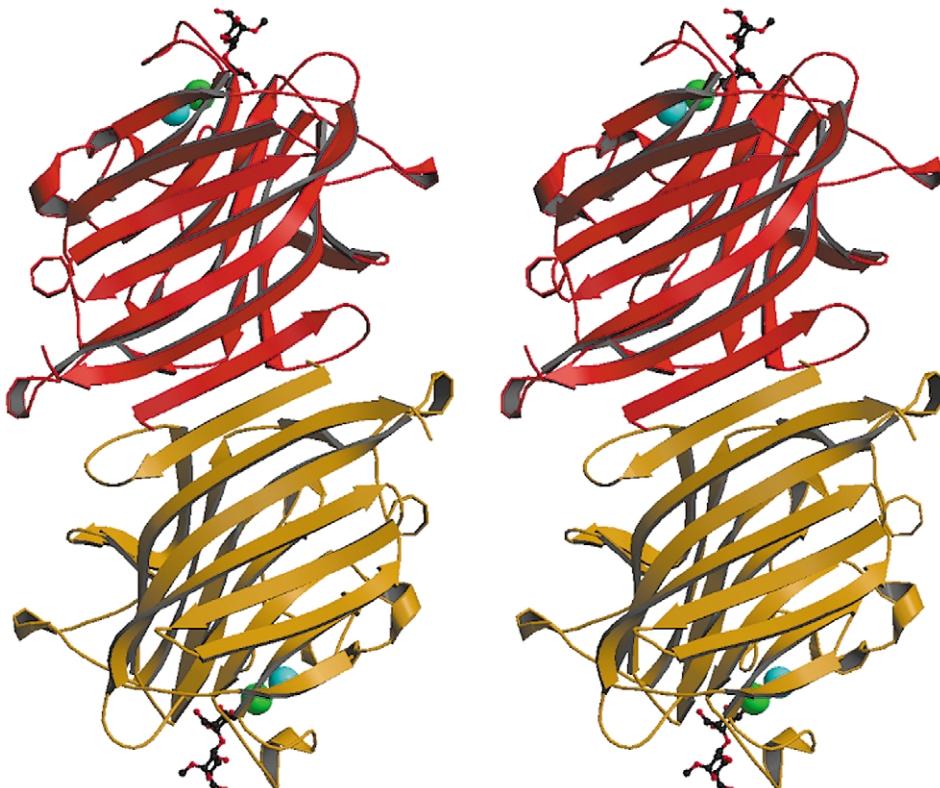
ing towards a hydrophobic area without any potential hydrogen bond donors or acceptors. At position 220, Gly was modelled instead of Arg. There is no electron density for an Arg side-chain (including its C<sup>B</sup> atom) in this otherwise well-defined region of the electron density map, and the presence of even a small side-chain at this position would clash with any sugar bound in the carbohydrate recognition site.

PAL has the highest level of sequence identity (64%) with the Man/Glc-specific lectin from *Arachis hypogae* (peanut), for which no crystal structure is available. Among those lectins for which the crystal structure is determined, the sialyllectose-specific MAL from *Maackia amurensis* (the Amur Maackia tree from Asia) is most closely related (43% sequence identity), while other Man/Glc-specific lectins such as ConA, FRIL and LOL show only 40–42% identity.

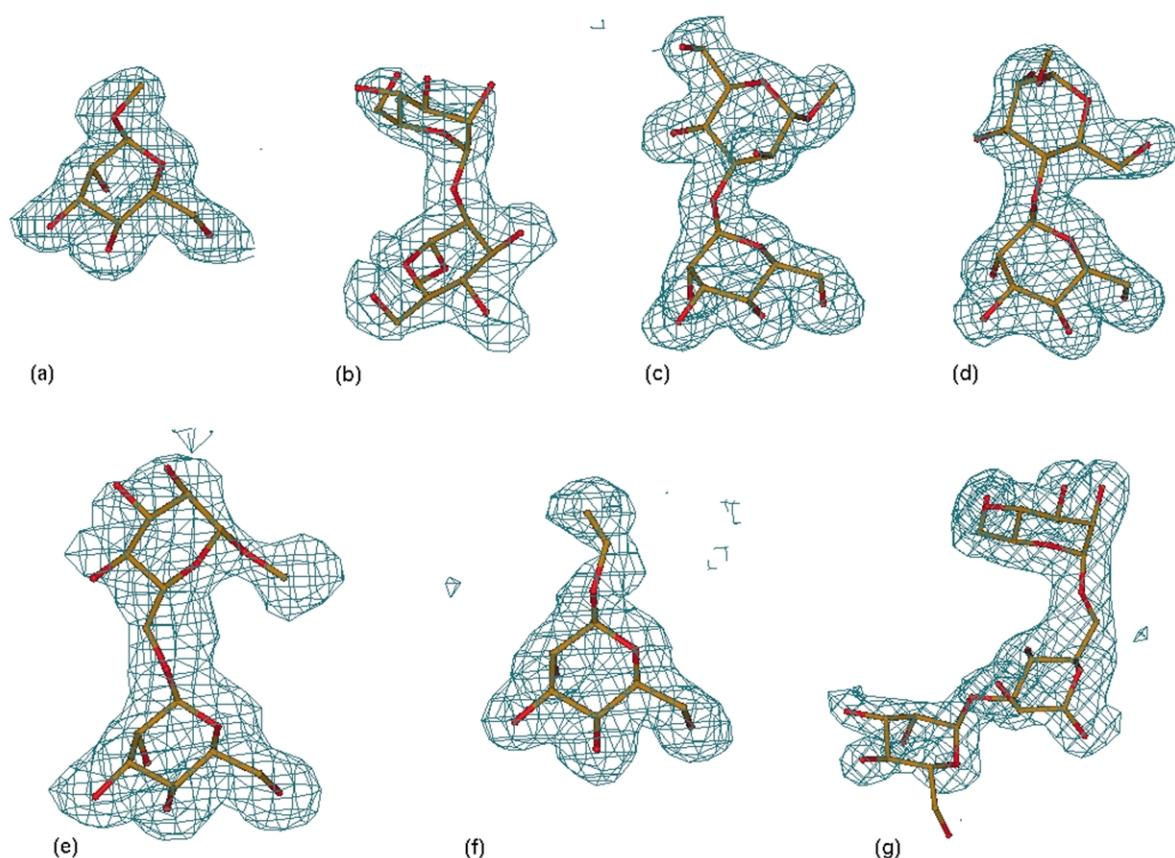
The overall structure of PAL has been described and is shown in **Figure 2**.<sup>22</sup> It consists of the typical legume lectin  $\beta$ -sandwich, the details of which are well known.<sup>23</sup>

All structures presented show good electron density for residues 1–238. Tyr239 was fit into the density, but the temperature factors for this residue remain high. Weak density is seen for residues Thr240 and Ala241. The bound carbohydrate molecules display clear electron densities in each of the complexes (**Figure 3**).

There are two legume lectin monomers in the



**Figure 2.** Overall structure of the *P. angolensis* lectin. Stereo cartoon representation of the PAL dimer (Man(α1-3)Man complex). One monomer is coloured orange, the other one yellow. Manganese ions are shown as light blue spheres and calcium ions as green spheres. Two bound molecules of Man(α1-3)Man are shown in ball-and-stick representation.



**Figure 3.** Experimental electron densities of the bound carbohydrates. (a)  $\alpha$ -Me-Mannose, (b)  $\text{Man}(\alpha 1\text{-}2)\text{Man}$ , (c)  $\text{Man}(\alpha 1\text{-}3)\text{Man}$ , (d)  $\text{Man}(\alpha 1\text{-}4)\text{Man}$ , (e)  $\text{Man}(\alpha 1\text{-}6)\text{Man}$ , (f)  $\text{Man}(\alpha 1\text{-}6)[\text{Man}(\alpha 1\text{-}3)]\text{Man}$  in monomer A and (g)  $\text{Man}(\alpha 1\text{-}6)[\text{Man}(\alpha 1\text{-}3)]\text{Man}$  in monomer B. In each case, except (g), density in the binding site of the A monomer is shown. In all cases, the electron density map is an  $F_{\text{obs}} - F_{\text{calc}}$  map contoured at  $3.0\sigma$  and obtained by taking the final refined model and deleting the carbohydrates in both binding sites.

asymmetric unit, together forming the canonical dimer (Figure 2), as has been observed in many other dimeric and tetrameric legume lectins.<sup>24,25</sup> As a consequence of the crystal packing, the carbohydrate-binding sites of subunits A and B in the asymmetric unit of the crystals are not equivalent. The binding site of subunit A is involved in crystal packing. In all carbohydrate complexes, the ligand bound to the binding site of subunit A makes contacts with protein atoms from a symmetry mate. The binding site of subunit B on the other hand is not involved in crystal packing. In all but one of the carbohydrate complexes, binding seems to be identical with both subunits. The only exception is the  $\text{Man}(\alpha 1\text{-}3)[\text{Man}(\alpha 1\text{-}6)]\text{Man}$  complex, where the conformation adopted in binding site B is prevented in site A due to a steric hindrance from a symmetry-related protein molecule.

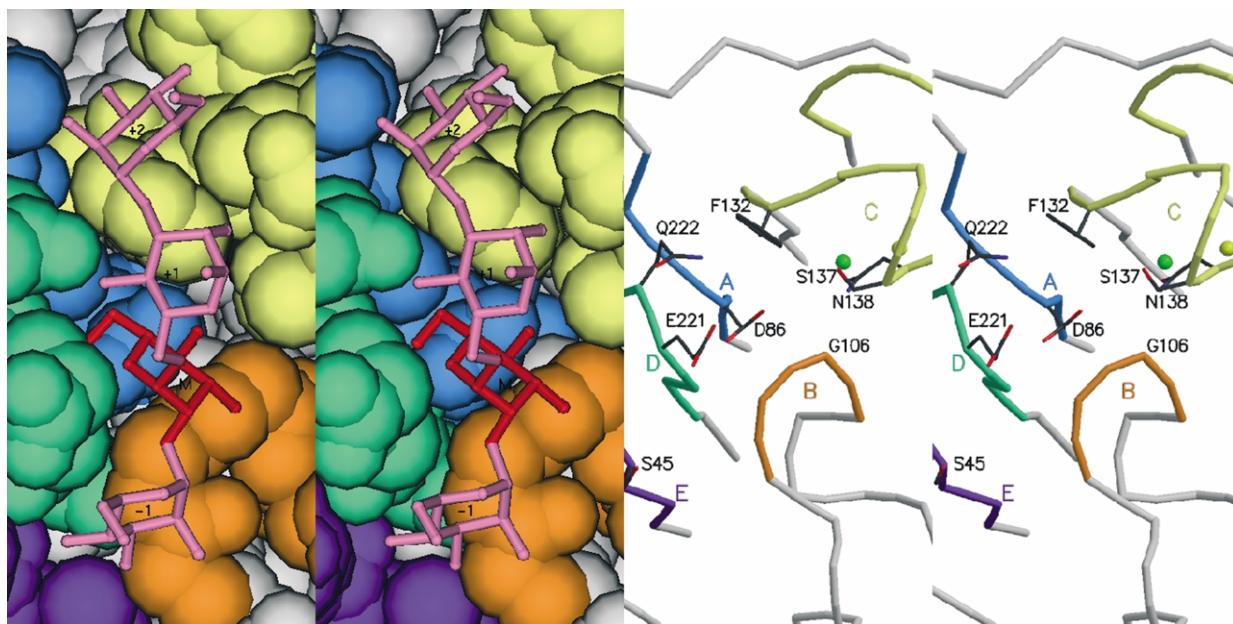
#### Carbohydrate-binding site

The carbohydrate-binding site of PAL can be described as one with a primary binding site that recognises glucose and mannose. This primary binding site lies in the centre of a shallow groove on the surface of the protein (Figure 4(a)). Exten-

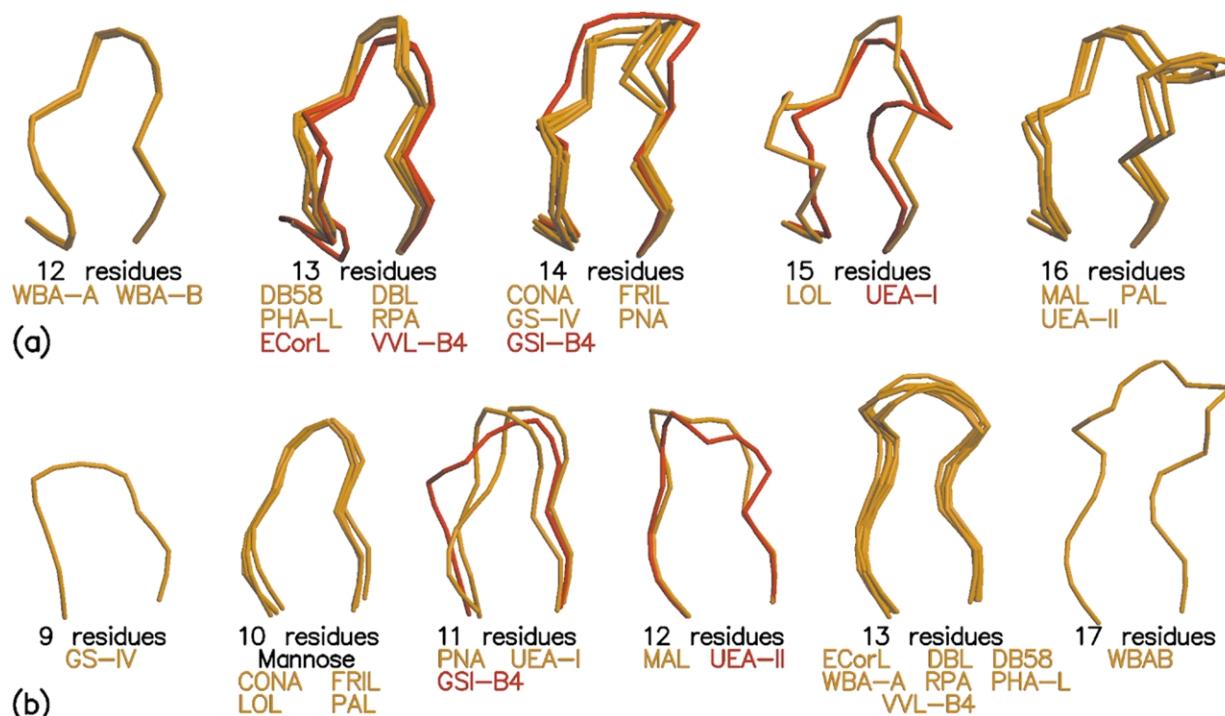
sions to the sugar bound in the primary binding site (M) can be made in two directions, following O1 or O2. Independent of the exact linkage that is present, these additional sugar residues occupy two particular regions of the lectin, which we call the  $-1$  (attached to O2) and  $+1$  (attached to O1) subsites. In the case of longer oligosaccharides, we speak of additional  $+2, +3; \dots$  or  $-2, -3, \dots$  subsites depending on their position relative to the primary binding site.

The carbohydrate-binding site of all legume lectins consists of residues belonging to five polypeptide stretches (termed A to E according to Sharma & Surolia<sup>2</sup>) (Figure 4(b)), which vary to different degrees between lectins with different specificities. Stretches A and B contain an essential aspartate residue (invariably preceded by a *cis*-peptide bond) and backbone NH group (usually from a glycine residue; Gly106 in PAL), respectively. The conformations of these two stretches do not vary much among different lectins. In the current structures, this picture is confirmed.

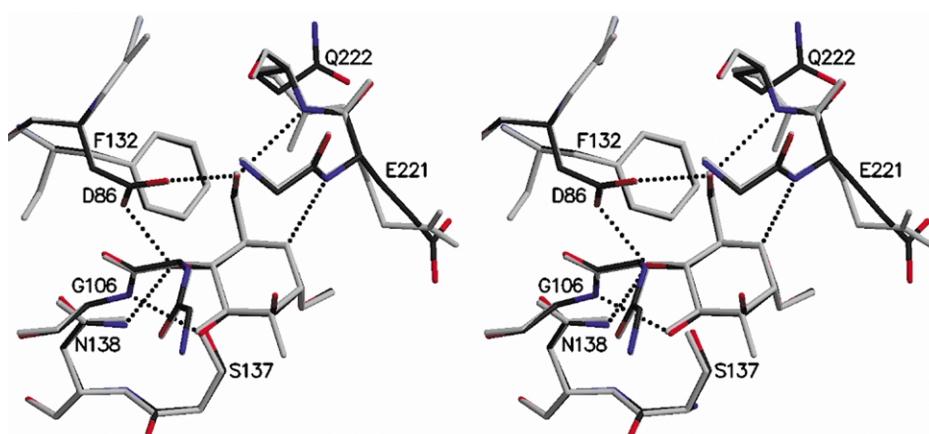
Stretch C is the metal-binding loop and wraps around the structurally important calcium and manganese ions. In the known crystal structures, five different loop sizes (12–16 residues) adopting



**Figure 4.** Architecture of the carbohydrate-binding site. (left) Stereo view of a CPK model of PAL with the four stretches that constitute the carbohydrate-binding site coloured: blue for stretch A containing *cis*-Asp86, orange for stretch B containing Gly106, yellow for metal-binding loop C and green for specificity loop D. These colours are similar to those used in Figure 1 and are maintained in all the other Figures. Superimposed is the ball-and-stick model of a virtual tetrasaccharide  $\text{Man}(\alpha 1\text{-}2)\text{Man}(\alpha 1\text{-}6)[\text{Man}(\alpha 1\text{-}3)]\text{Man}$  indicating the different subsites (M, primary binding site and  $-1$ ,  $+1$  and  $+2$  downstream and upstream subsites. (right) Stereo MOLSCRIPT representation showing the five stretches A–E that make up the carbohydrate-binding site. Side-chains of residues important for carbohydrate-binding are shown in ball-and stick and are labelled. The calcium and manganese ions are shown as large green and yellow spheres, respectively.



**Figure 5.** Loop conformations in the carbohydrate-binding sites of legume lectins. (a) conformations observed for loop C and (b) conformations observed in loop D suggesting the use of canonical loop conformations to modulate carbohydrate specificity in a way similar to that seen in the CDR loops of antibodies.



**Figure 6.** Interaction of mannose in the primary binding site. Stereo view of the superposition of the monosaccharide-binding sites of PAL complexes with Me $\alpha$ Man coloured and Me $\alpha$ Glc light grey. Selected residues are labelled. The complexes are structurally identical with the exception of the orientation of the O2 atom of the sugars. The difference in conformation for Glu221 is most likely due to the less well defined electron density of this residue in the Me $\alpha$ G structure and is probably not of biological relevance.

a total of eight different conformations are observed (Figure 5(a)). There is no simple relationship between monosaccharide specificity and the length and conformation of loop C. In PAL, this loop has a length of 16 residues and is identical with that seen in the two sialyllactose-specific lectins from *M. amurensis*<sup>18</sup> and in the chitobiose-specific lectin II from common gorse (*Ulex europaeus*).<sup>19</sup> The other Man/Glc-specific lectins for which crystal structures are available have loop lengths of either 14 (FRIL, ConA and related lectins) or 15 (LOL and related lectins) residues. While the backbone conformation of loop C is not a determinant for monosaccharide specificity, specific side-chains on this loop nevertheless do influence the nature of the sugar that can be accommodated in the binding site (see below).

In contrast to stretches A–C, stretch D does not interact directly with the structural calcium ion. It is highly variable in length, conformation and sequence and is often referred to as the monosaccharide specificity loop.<sup>3</sup> It is thought to be the prime determinant for monosaccharide as well as oligosaccharide specificity. In the PAL structure, the conformation adopted by this loop of ten residues is identical with that found in all other known crystal structures of Man/Glc-specific lectins<sup>15–17,26–28</sup> (Figure 5(b)), supporting this notion.

Finally, stretch E is found to interact with a bound carbohydrate in only a few cases, where it is part of the –1 subsite: *M. amurensis* leukoagglutinin (MAL) in complex with sialyllactose,<sup>19</sup> *Griffonia simplicifolia* lectin IV (GS-IV) in complex with the Le<sup>b</sup> tetrasaccharide<sup>29</sup> and ConA in complex with Man(α1-2)Man.<sup>30</sup> In most complexes of PAL, it is not involved in carbohydrate-binding, the exception being the Man(α1-2)Man complex.

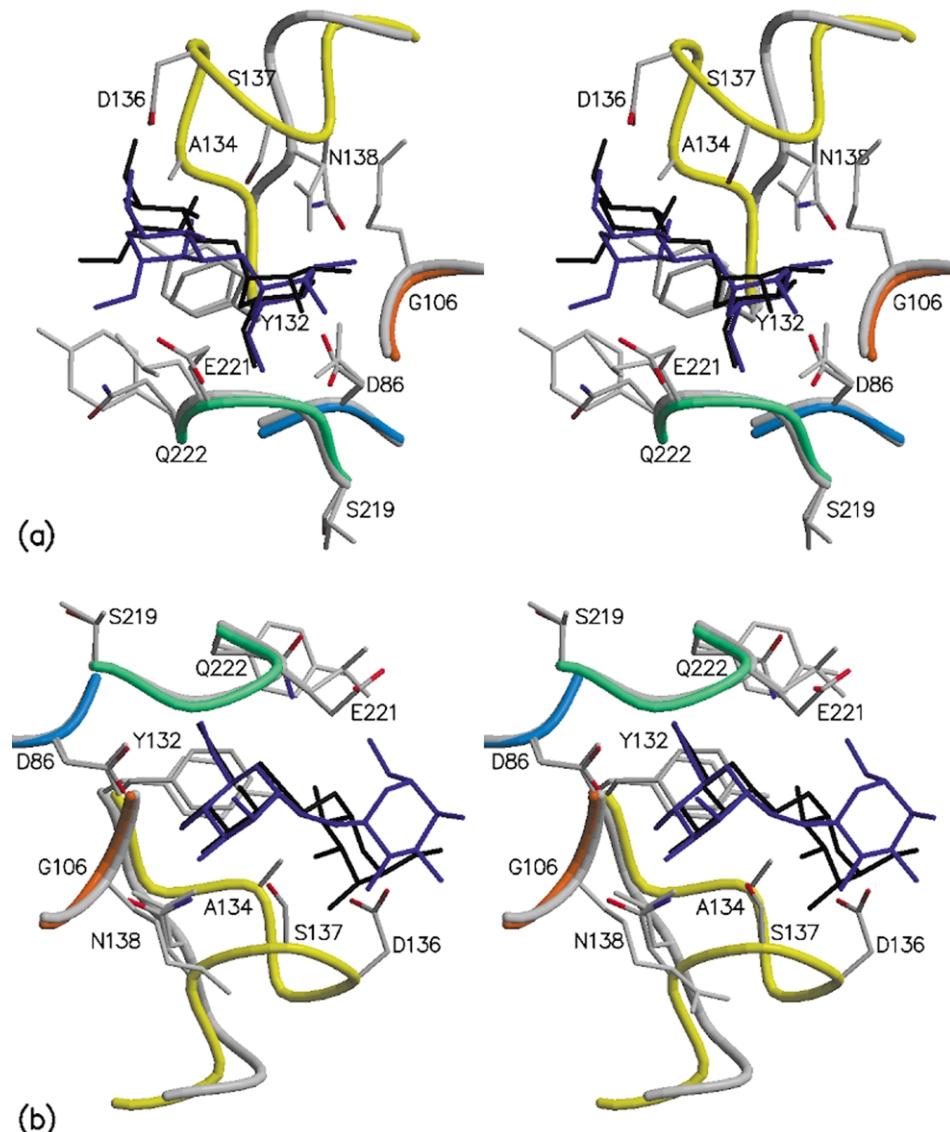
### Interactions in the primary binding site

Methyl- $\alpha$ -D-mannopyranoside (Me $\alpha$ Man) binds to the primary binding site in the same way as has been observed repeatedly for other Man/Glc-specific lectins (Figure 6).<sup>15,20,26–28</sup> The sugar interacts with the protein via a series of hydrogen bonds with the conserved Asp/Gly/Asn triad as well as with the backbone of specificity loop D (hydrogen bond between Glu221 and O5). In addition, the side-chain of Phe132 from the metal-binding loop C stacks favourably upon the sugar ring, while Gly220 and the side-chain of Glu221 also make favourable van der Waals contacts with the sugar.

As observed for most other Glc/Man-specific legume lectins, PAL has twice the affinity for Me $\alpha$ Man (our unpublished results) than for Me $\alpha$ Glc. In our crystal structure, the axial O2 of mannose makes van der Waals contacts with the C $\alpha$  atoms of Gly106 and Gly220, possibly forming CH···O hydrogen bonds (C–O distances 3.7 Å and 3.5 Å, respectively). This interaction is absent from the Me $\alpha$ Glc complex<sup>22</sup> due to the equatorial orientation of O2 in this sugar. As a consequence, a small void is present that is not filled with an ordered water molecule.

### Non-specific subsite interactions: Man(α1-3)-Man, Man(α1-4)Man and Man(α1-6)Man

The affinities of PAL for Man(α1-3)Man, Man(α1-4)Man and Man(α1-6)Man are essentially identical with that for Me $\alpha$ Man (our unpublished results). Clear density for the disaccharides has been observed in both subunits. The conformation adopted by the disaccharide is in each case identical for both subunits, despite the involvement of the binding site of subunit A in crystal packing. All bind with their non-reducing mannose in the



**Figure 7.** Recognition of Man(α1-3)Man and Man(α1-6)Man by PAL. (a) Stereo view of the PAL-binding site in complex with Man(α1-3)Man. The disaccharide is shown in blue and loops of PAL are coloured as in Figure 2. The equivalent loops and side-chains of the ConA-Man(α1-3)Man complex are shown in light grey. The ConA-bound disaccharide is shown in black. (b) A similar stereo view of the PAL-Man(α1-6)Man complex superimposed on the ConA-Man(α1-6)Man complex.

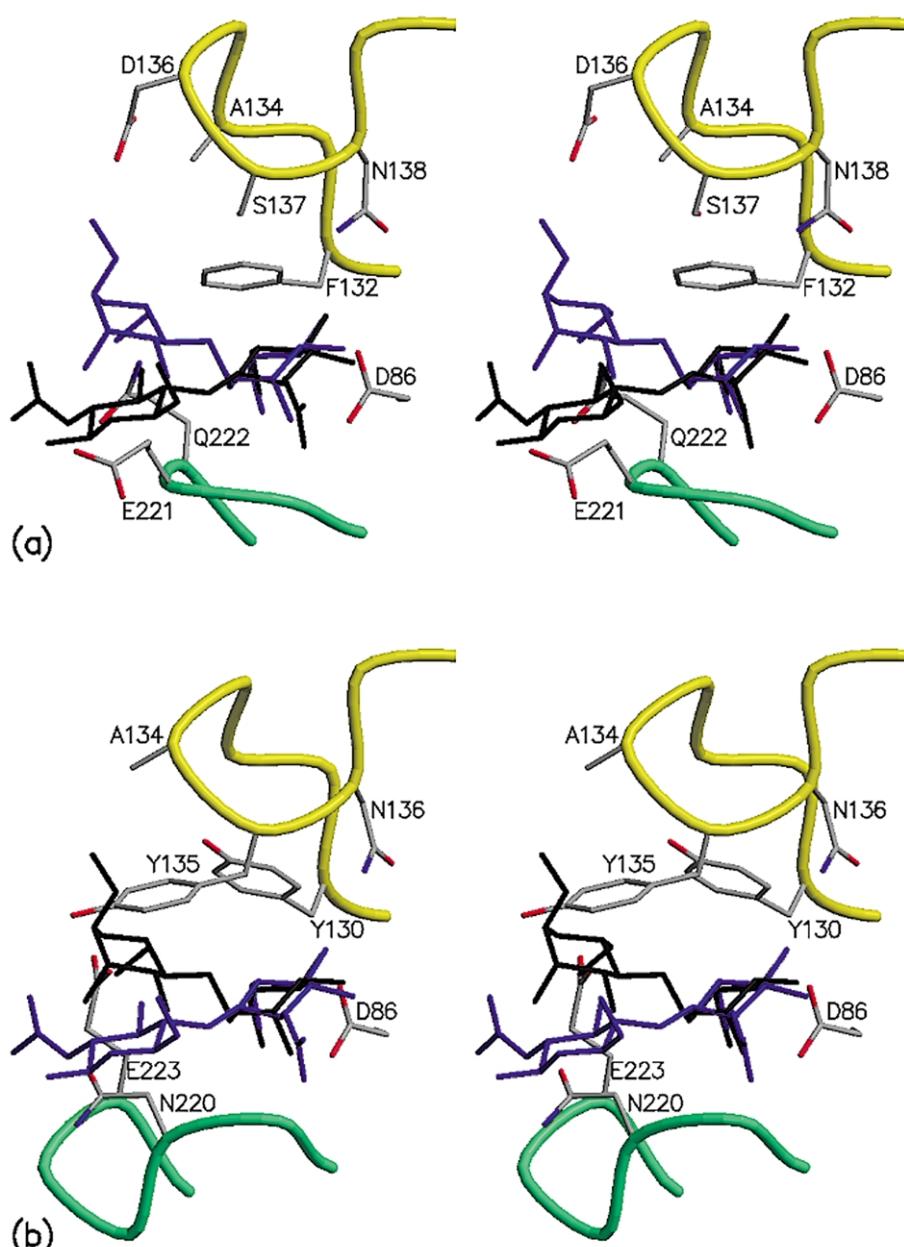
primary binding site, thus occupying subsite +1. The interactions with this subsite are sufficiently favourable to allow each of the disaccharides to adopt a unique conformation, but insufficient to result in an enhanced affinity. All three disaccharides make at least one additional hydrogen bond with amino acid residues in the +1 site. This situation is similar to what has been observed for ConA,<sup>31</sup> where Man(α1-6)Man binds with the same affinity as MeoMan, despite an additional hydrogen bond from the protein to the reducing mannose.

Figure 7 compares the binding modes of Man(α1-3)Man and Man(α1-6)Man with those observed previously in their complexes with ConA.<sup>31</sup> Man(α1-3)Man binds in an identical conformation and orientation to ConA and PAL,

although the residues that make up subsite +1 differ in both proteins (Figure 7(a)). Man(α1-6)Man recognition on the other hand is different in both proteins, and the disaccharides adopt different low-energy conformations (Figure 7(b)). The conformation found in ConA cannot be adopted in the PAL structure due to clashes with the metal-binding loop C, notably the side-chains of Ala134 and Asp136. The conformation found in PAL could, on the other hand, fit in the binding site of ConA, provided that the α-methyl group on O1 would be reoriented, but no hydrogen bond would be formed in the +1 site.

#### Specificity for $\alpha$ versus $\beta$ linkages

Similar to other Man/Glc-specific legume lectins,

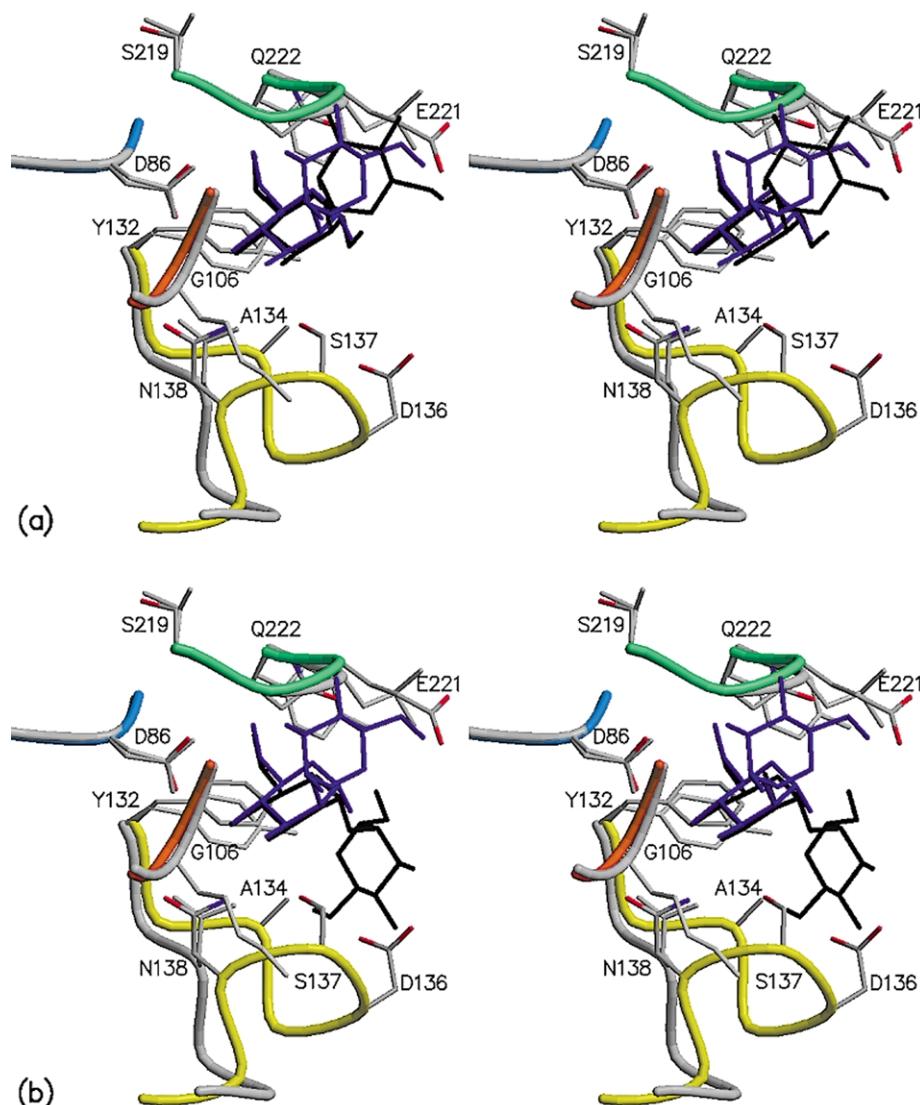


**Figure 8.** Binding of  $\text{Man}(\alpha 1\text{-}4)\text{Man}$  by PAL. (a) A stereo view of the PAL-binding site in complex with  $\text{Man}(\alpha 1\text{-}4)\text{Man}$  (blue). Loops are coloured as in Figure 2. In black, the co-ordinates of  $\text{GlcNAc}(\beta 1\text{-}4)\text{GlcNAc}$  as seen in the binding site of UEA-II are superimposed. In PAL, the  $\beta$  anomeric configuration is made impossible due to residues Glu221 and Gln222 in the specificity loop of PAL. (b) A stereo view of the UEA-II-binding site in complex with  $\text{GlcNAc}(\beta 1\text{-}4)\text{GlcNAc}$  (blue). In black, the co-ordinates of  $\text{Man}(\alpha 1\text{-}4)\text{Man}$  as seen in the binding site of PAL are superimposed. In UEA-II, the  $\alpha$  anomeric configuration is made impossible due to the bulky side-chain of Tyr135, which is replaced by Ser137 in PAL.

the *P. angolensis* lectin has a strict requirement for  $\alpha$ -linkages. The chitobiose-specific lectin II from *U. europaeus* (UEA-II), on the other hand, can accommodate only a  $\beta$ -linkage. From the different crystal structures of UEA-II-carbohydrate complexes,<sup>19</sup> it has been learned that the *N*-acetyl groups of chitobiose are not crucial for binding and that its primary binding site can accommodate mannose and glucose as efficiently.

Figure 8 compares the binding of  $\text{Man}(\alpha 1\text{-}4)\text{Man}$  on PAL to that of  $\text{GlcNAc}(\beta 1\text{-}4)\text{GlcNAc}$  (chito-

biose) to UEA-II. The bulky Tyr135 in metal-binding loop C prevents formation of an  $\alpha$ -linkage in the binding site of UEA-II and is replaced by the small Ser137 in PAL. Otherwise, the backbone conformations of this loop are very similar in PAL and UEA-II. On the other hand, Glu221 in specificity loop D of PAL sterically prevents a  $\beta$ -linkage from being accommodated in the binding site of this protein. In order to allow for a  $\beta$ -linkage, Glu221 needs to be at least truncated to Gly. In UEA-II, the specificity loop is two residues longer and



**Figure 9.** Recognition of  $\text{Man}(\alpha 1\text{-}2)\text{Man}$  by PAL. (a) A stereo view of the PAL- $\text{Man}(\alpha 1\text{-}2)\text{Man}$  complex superimposed on the equivalent ConA complex (non-reducing mannose bound in subsite  $-1$ ). Colouring is as in Figure 3. (b) An identical view of PAL, but superimposed on the ConA- $\text{Man}(\alpha 1\text{-}2)\text{Man}$  complex with the second mannose in subsite  $+1$ .

locally adopts a different conformation, allowing for a  $\beta$ -linked disaccharide to be bound. Thus, the selection for  $\alpha$  or  $\beta$ -glycosidic linkages appears to be controlled by a few specific amino acid residues and does not depend critically on the type of metal-binding loop, as was suggested earlier.<sup>19</sup>

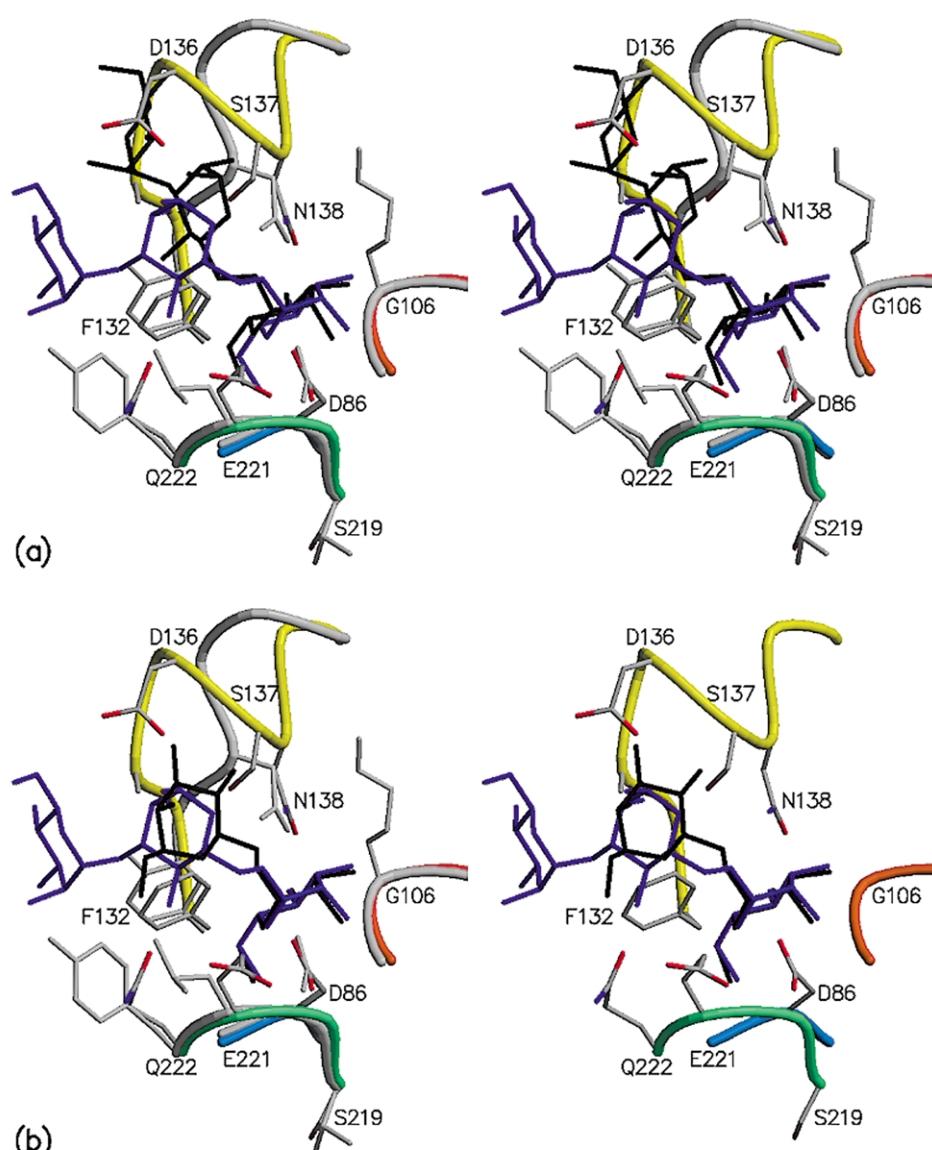
#### Man( $\alpha 1\text{-}2$ )Man binds with its reducing mannose in the primary binding site

In a typical Man/Glc-specific legume lectin, O2 of a bound Me $\alpha$ Man points towards the solvent. This means that, at least in theory, Man( $\alpha 1\text{-}2$ )Man would be able to bind with both its reducing or its non-reducing mannose in the primary binding site. Such a situation is observed in the complex of concanavalin A (ConA) with Man( $\alpha 1\text{-}2$ )Man,<sup>30</sup> and has been argued to contribute significantly to the

enhanced affinity of this lectin for Man( $\alpha 1\text{-}2$ )Man compared to Me $\alpha$ Man.<sup>32</sup>

PAL has a 2.5-fold higher affinity for Man( $\alpha 1\text{-}2$ )Man compared to Me $\alpha$ Man (our unpublished results). Nevertheless, in the current Man( $\alpha 1\text{-}2$ )Man complex only a single binding mode is observed, with the reducing mannose molecule in the primary binding site and the second mannose molecule in subsite  $-1$ . The second binding mode in the  $+1$  subsite is not possible in PAL, as it would lead to a steric clash with the two residues longer metal-binding loop C, in particular with the side-chains of Ala134, Asp136 and Ser137 (Figure 9).

In the observed binding mode to PAL, the conformation of the disaccharide roughly resembles the equivalent one found in the ConA complex, but is not identical. The difference in conformation



**Figure 10.** Recognition of the trisaccharide  $\text{Man}(\alpha 1\text{-}6)[\text{Man}(\alpha 1\text{-}3)]\text{Man}$  by PAL. Colouring is as in Figure 3. (a)  $\text{Man}(\alpha 1\text{-}6)[\text{Man}(\alpha 1\text{-}3)]\text{Man}$  bound by PAL. In light grey, the ConA- $\text{Man}(\alpha 1\text{-}6)[\text{Man}(\alpha 1\text{-}3)]\text{Man}$  complex is superimposed (the sugar itself in black). (b) An identical view of the  $\text{Man}(\alpha 1\text{-}6)[\text{Man}(\alpha 1\text{-}3)]\text{Man}$ -PAL complex, but with  $\text{Man}(\alpha 1\text{-}6)\text{Man}$  superimposed in black.

is due to a single amino acid substitution in loop B: Gly104 in PAL is replaced by Thr226 in ConA, forcing a  $30^\circ$  rotation around Phi (defined as the torsion angle O5-C1-O1-C'2) to prevent a van der Waals clash and to establish a hydrogen bond between O2 of  $\text{Man}(-1)$  and Thr226(OG1) (Figure 9).

#### Context-dependent protein–carbohydrate interactions in the $\text{Man}(\alpha 1\text{-}3)[\text{Man}(\alpha 1\text{-}6)]\text{Man}$ complex

$\text{Man}(\alpha 1\text{-}3)[\text{Man}(\alpha 1\text{-}6)]\text{Man}$  is the only carbohydrate for which there is a significant difference in binding in the two crystallographically independent lectin monomers. In the binding site of

monomer B, which is not involved in crystal packing, clear density is seen for the whole trisaccharide. Outside the primary binding site, five direct and one water-mediated hydrogen bonds are made between the protein and the carbohydrate.

In the binding site of monomer A, on the other hand, electron density is observed only for the mannose in the primary binding site. As there is no trace of density corresponding to the original  $\text{Man}(\alpha 1\text{-}3)\text{Man}$  (that was present before the soak with the trimannose), it has to be assumed that the trimannose is indeed bound in the binding site. The obviously favourable conformation of the trisaccharide as observed bound to monomer B is not accessible in monomer A because of severe steric conflicts with a symmetry-related protein

**Table 1.** X-ray data collection and refinement statistics

Carbohydrate	MeαMan	Man(α1-2)Man	Man(α1-3)Man	Man(α1-4)Man	Man(α1-6)Man	Man(α1-3)[Man(α1-6)]Man
<i>Unit cell parameters</i>						
<i>a</i> (Å)	56.82	56.32	56.88	57.28	56.88	56.85
<i>b</i> (Å)	83.67	82.98	83.03	83.37	83.38	83.61
<i>c</i> (Å)	123.00	122.00	122.94	122.96	122.57	122.96
Beamlne	BW7A	X31	BW7B	X31	X11	BW7A
Resolution(Å)	1.70	2.20	1.75	2.05	2.05	1.85
<i>N</i> <sub>meas</sub>	210,120	101,201	395,205	209,488	160,081	232,267
<i>N</i> <sub>unique</sub>	64,245	25,266	57,268	36,449	37,147	50,664
Completeness (%)	97.6	84.9	98.6	99.9	97.3	99.7
<i>R</i> <sub>merge</sub>	0.051	0.102	0.177	0.119	0.111	0.081
< <i>I</i> / <i>σ(I)</i> >	19.20	8.88	13.05	11.67	9.79	17.00
<i>R</i>	0.1889	0.1776	0.1837	0.1841	0.1824	0.1805
<i>R</i> <sub>free</sub>	0.2154	0.2178	0.2064	0.2179	0.2161	0.2052
PDB code	1ukg	1q8o	1q8p	1q8q	1q8s	1q8v

molecule. Apparently, when specific subsite interactions cannot be made, the trisaccharide is bound in a disordered way, similar to that observed in the complex of pea lectin with the same trisaccharide.<sup>28</sup>

Concanavalin A has a 40 times higher affinity for the trisaccharide than for MeαMan.<sup>33,34</sup> PAL favours the trisaccharide by only a factor of four (our unpublished results). Figure 10 compares trimannose-binding by ConA<sup>35,36</sup> and monomer B of PAL. Although there are many differences for the amino acid residues outside the primary binding site, some of the global features of the binding of the trisaccharide are remarkably similar for both lectins. In both cases, the non-reducing mannose that is linked to O6 is found in the primary binding site. The "middle" (reducing) mannose makes extensive interactions with the protein. These interactions require, however, the presence of the third mannose molecule (non-reducing O3-linked). The details of the binding, such as specific hydrogen bonds formed outside the primary binding site, nevertheless differ markedly for ConA and PAL.

## Materials and Methods

Purified *P. angolensis* seed lectin (PAL) was available from a previous study.<sup>22</sup> Internal amino acid sequences were obtained from four peptides after digestion of the purified *P. angolensis* lectin with trypsin. The amino acid sequences of these peptides (SPLSNGADGIAFFIA, and WDPDYPH) were used to design degenerate primers. Total RNA was extracted from ripening *P. angolensis* seeds (collected in Zimbabwe and stored at -80 °C) using the RNeasy Plant Mini Kit (Qiagen) and converted into double-stranded cDNA using the 5'RACE System for rapid amplification of cDNA ends (GibcoBRL) and the Q<sub>T</sub> primer.<sup>37</sup> This cDNA was amplified with the degenerate forward primer Muk-2:

5'-GGIGCIGAYGGIATHGCITYTT-3'

and the reverse primer Muk-5:

3'-ACCCTRGGICTRATRGGIGT-5'

After amplification, the PCR fragments were cloned

into the pUC18 using the Sure-Clone-Ligation kit (Amersham Pharmacia Biotech), and sequenced.

Specific primers Muk-6:

5'-CATCGCACCGCCGGATACTAC-3'

Muk-8:

3'-GAGTTTGTGCAGGCGCTTGG-5'

and Muk-9:

3'-GATCCCCAGAACGTGGACC-5'

were designed on the basis of the sequence information gathered from the PCR fragment obtained with the degenerate primers Muk-2 and Muk-5, to amplify and clone the nucleotide sequence corresponding to the 3'-end and the 5'-end of the *P. angolensis* mRNA.

The 3'-end of the cDNA was amplified with the Muk-6 and the Q<sub>0</sub> primer.<sup>37</sup> The 5'-end of the *P. angolensis* mRNA was amplified after adding a poly(C) tail to the 5'-end of the cDNAs (GibcoBRL). An initial PCR amplification was carried out with Muk-8 and RAAP (GibcoBRL), followed by a second PCR amplification using the nested primers Muk-9 and AUAP (GibcoBRL). The PCR fragments were cloned into the pUC18 using the Sure-Clone-Ligation kit (Amersham Pharmacia Biotech) and sequenced.

Full length cDNAs were subsequently amplified using primers Muk-32:

5'-CCCAAATATAATAAAAAGCGCTACCCA-TCTC-3'

(located in the 5'-untranslated sequence) Muk-14:

5'-CCTCCCTCTCCTTGCCTTCA-3':

(located in the signal peptide) or Muk-15:

5'-ATGCTACTGAACAAAGCATACT-3'

(located in the signal peptide) in combination with the Q<sub>0</sub> primer.<sup>37</sup> The PCR fragments were cloned in pUC18 and sequenced.

Sequencing was performed following the dideoxynucleotide termination method of Sanger *et al.*<sup>38</sup> using the Thermo Sequenase Radiolabeled Termination Cycle Sequencing kit (Amersham Pharmacia Biotech). Separation of the resulting DNA fragments was performed

on a 5% (w/v) polyacrylamide gel containing 8 M urea, and Tris-borate buffer (100 mM Tris, 100 mM boric acid). Detection of the sequencing ladders was achieved by autoradiography.

### Crystallisation and data collection

All carbohydrates were purchased from Dextra Laboratories (Reading, UK) and are methylated on their reducing O1, even when not specifically mentioned otherwise. Crystals of PAL in complex with Man(α1-3)Man were prepared as described for the methyl-α-D-glucose complex,<sup>22</sup> with the exception that 10 mM Man(α1-3)Man was used as a ligand instead of the glucoside. The complexes of the lectin with the other carbohydrates given in Table 1 were obtained by transferring the crystals of the Man(α1-3)Man complex to artificial mother liquor (200 mM calcium acetate, 100 mM sodium cacodylate (pH 6.5), 20% (w/v) PEG8000) containing increasing concentrations of the desired ligand (100 mM final concentration, reached in four steps).

All X-ray data were collected at room temperature on the EMBL beamlines of the DESY synchrotron (Hamburg, Germany). The data were processed with DENZO and SCALEPACK.<sup>39</sup> The statistics of the data collections are given in Table 1.

### Structure determination

The structure of the Man(α1-3)Man complex was determined by molecular replacement using the co-ordinates of lentil lectin (pdb code 1len<sup>40</sup>) as a search model. Two clear solutions were found with AMoRe<sup>41</sup> that together constructed the lectin dimer. Refinement was carried out using the mlf target of CNS 1.0.<sup>42</sup> Cross-validation, bulk solvent correction and anisotropic temperature factor scaling were used throughout. Rounds of slow-cool simulated annealing and restrained *B*-factor refinement using all available data were alternated with manual fitting in electron density maps using TURBO.<sup>43</sup> At the beginning of the study, the amino acid sequence was still unknown and a sequence was derived directly from the electron density maps. At the end of the refinement, the cDNA sequence became available, and this information was used to finalise the structure. At this stage, simulated annealing was abandoned in favour of conventional positional refinement and water molecules were fit into the electron density.

The structures of all other isomorphous lectin–carbohydrate complexes were determined using the refined co-ordinates of the Man(α1-3)Man complex (stripped from its carbohydrate ligands and water molecules) as the starting model. After rigid body refinement, a slow-cool stage was used to uncouple *R* and *R*<sub>free</sub>. From then on, restrained positional and *B*-factor refinement were alternated with manual fitting in electron density maps. The refinement statistics for all complexes are given in Table 1.

Superpositions of crystal structures were done using TURBO.<sup>43</sup> Figures 2–10 were produced using MOLSCRIPT<sup>44</sup> and Raster3D.<sup>45</sup>

### Data Bank accession numbers

The nucleotide sequences of partial and full-length cDNAs have been deposited at GenBank with accession numbers from AJ426054 to AJ426062.

Co-ordinates and structure factors were deposited at

the RCSB Protein Data Bank with as entries 1ukg, 1q80, 1q8p, 1q8q, 1q8s and 1q8v.

### Acknowledgements

This work was supported by the Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), the Onderzoeksraad of the V.U.B. (grants OZR846, OZR727 and OZR728), the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO) and the Directorate General of International Co-operation (DGIC). The authors acknowledge the use of synchrotron beamtime at the EMBL beamlines at the DORIS storage ring (Hamburg, Germany), and the Zimbabwe Forestry Commission (Seed Research Centre, Harare) for providing the seeds. W. Versées and S. De Vos collected the X-ray data on the Man(α1-6)Man complex. R.L. & J.B. are postdoctoral fellows of the FWO. I.V.W. received financial support from the IWT.

### References

- Taylor, M. E. & Drickamer, K. (2003). *Introduction to Glycobiology*, Oxford University Press, Oxford, UK.
- Sharma, V. & Surolia, A. (1997). Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. *J. Mol. Biol.* **267**, 433–445.
- Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). Legume lectin structure. *Biochim. Biophys. Acta*, **1383**, 9–36.
- Bouckaert, J., Loris, R., Poortmans, F. & Wyns, L. (1995). The crystallographic structure of metal-free concanavalin A at 2.5 Å resolution. *Proteins: Struct. Funct. Genet.* **23**, 510–540.
- Bouckaert, J., Dewallef, Y., Poortmans, F., Wyns, L. & Loris, R. (2000). Structural dissection of the conformational pathways of de- and re-metallization of concanavalin A. *J. Biol. Chem.* **275**, 19778–19787.
- Lescar, J., Loris, R., Mitchell, E., Gautier, C., Chazalet, V., Cox, V. et al. (2002). Isolectins I-A and I-B of *Griffonia (Bandeiria) simplicifolia*: crystal structure of metal-free GSII-B4 and molecular bases for metal binding and monosaccharide specificity. *J. Biol. Chem.* **277**, 6608–6614.
- Manoj, N., Srinivas, V. R., Surolia, A., Vijayan, M. & Suguna, K. (2000). Carbohydrate specificity and salt-bridge mediated conformational change in acidic winged bean agglutinin. *J. Mol. Biol.* **302**, 1129–1137.
- Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. & Suguna, K. (1998). Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 Å resolution. *J. Mol. Biol.* **276**, 787–796.
- Audette, G. F., Vandonselaar, M. & Delbaere, L. T. (2000). The 2.2 Å resolution structure of the O(H) blood-group-specific lectin I from *Ulex europaeus*. *J. Mol. Biol.* **304**, 423–433.
- Rabijns, A., Verboven, C., Rouge, P., Barre, A., Van Damme, E. J., Peumans, W. J. & De Ranter, C. J. (2001). Structure of a legume lectin from the bark of

- Robinia pseudoacacia* and its complex with N-acetyl-galactosamine. *Proteins: Struct. Funct. Genet.* **44**, 470–478.
11. Tempel, W., Tschampel, S. & Woods, R. J. (2002). The xenograft antigen bound to *Griffonia simplicifolia* lectin 1-B(4). X-ray crystal structure of the complex and molecular dynamics characterization of the binding site. *J. Biol. Chem.* **277**, 6615–6621.
  12. Svensson, C., Teneberg, S., Nilsson, C. L., Kjellberg, A., Schwarz, F. P., Sharon, N., Krengel, U. (2002). High-resolution crystal structures of *Erythrina cristagalli* lectin in complex with lactose and 2'-alpha-L-fucosyllactose and correlation with thermodynamic binding data. *J. Mol. Biol.* **321**, 69–83.
  13. Hamelryck, T. W., Loris, R., Bouckaert, J., Dao-Thi, M. H., Strecker, G., Imbert, A. et al. (1999). Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. *J. Mol. Biol.* **286**, 1161–1177.
  14. Buts, L., Dao-Thi, M. H., Loris, R., Wyns, L., Eztler, M. & Hamelryck, T. (2001). Weak protein–protein interactions in lectins: the crystal structure of a vegetative lectin from the legume *Dolichos biflorus*. *J. Mol. Biol.* **309**, 193–201.
  15. Hamelryck, T. W., Moore, J. G., Chrispeels, M. J., Loris, R. & Wyns, L. (2000). The role of weak protein–protein interactions in multivalent lectin–carbohydrate binding: crystal structure of cross-linked FRIL. *J. Mol. Biol.* **299**, 875–883.
  16. Sanz-Aparicio, J., Hermoso, J., Grangeiro, T. B., Calvete, J. J. & Cavada, B. S. (1997). The crystal structure of *Canavalia brasiliensis* lectin suggests a correlation between its quaternary conformation and its distinct biological properties from concanavalin A. *FEBS Letters*, **405**, 114–118.
  17. Wah, D. A., Romero, A., Gallego del Sol, F., Cavada, B. S., Ramos, M. V., Grangeiro, T. B. et al. (2001). Crystal structure of native and Cd/Cd-substituted *Dioclea guianensis* seed lectin. A novel manganese-binding site and structural basis of dimer-tetramer association. *J. Mol. Biol.* **310**, 885–894.
  18. Imbert, A., Gautier, C., Lescar, J., Perez, S., Wyns, L. & Loris, R. (2000). An unusual carbohydrate binding site revealed by the structures of two *Maackia amurensis* lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* **275**, 17541–17548.
  19. Loris, R., De Greve, H., Dao-Thi, M.-H., Messens, J., Imbert, A. & Wyns, L. (2000). Structural basis of carbohydrate recognition by lectin II from *Ulex europaeus*, a protein with a promiscuous carbohydrate-binding site. *J. Mol. Biol.* **301**, 987–1002.
  20. Rozwarski, D. A., Swami, B. M., Brewer, C. F. & Sacchettini, J. C. (1998). Crystal structure of the lectin from *Dioclea grandiflora* complexed with core trimannoside of asparagine-linked carbohydrates. *J. Biol. Chem.* **273**, 32818–32825.
  21. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6.
  22. Loris, R., Imbert, A., Beeckmans, S., Van Driessche, E., Read, J. S., Bouckaert, J. et al. (2003). Crystal structure of *Pterocarpus angolensis* lectin in complex with glucose, sucrose and turanose. *J. Biol. Chem.* **278**, 16297–16303.
  23. Chandra, N. R., Prabu, M. M., Suguna, K. & Vijayan, M. (2001). Structural similarity and functional diversity in proteins containing the legume lectin fold. *Protein Eng.* **14**, 857–866.
  24. Manoj, N. & Suguna, K. (2001). Signature of quaternary structure in the sequences of legume lectins. *Protein Eng.* **14**, 735–745.
  25. Srinivas, V. R., Reddy, G. B., Ahmad, N., Swaminathan, C. P., Mitra, N. & Surolia, A. (2001). Legume lectin family, the “natural mutants of the quaternary state”, provide insights into the relationship between protein stability and oligomerization. *Biochim. Biophys. Acta*, **1527**, 102–111.
  26. Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb, A. J., Dodson, E. J., Papiz, M. Z. et al. (1989). The structure of the saccharide-binding site of concanavalin A. *EMBO J.* **8**, 2189–2193.
  27. Bourne, Y., Roussel, A., Frey, M., Rouge, P., Fontecilla-Camps, J. C. & Cambillau, C. (1990). Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: fine specificity of the monosaccharide-binding site. *Proteins: Struct. Funct. Genet.* **8**, 365–376.
  28. Rini, J. M., Hardman, K. D., Einspahr, H., Suddath, F. L. & Carver, J. P. (1993). X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 Å resolution. *J. Biol. Chem.* **268**, 10126–10132.
  29. Delbaere, L. T., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S. & Dauter, Z. (1993). Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution. *J. Mol. Biol.* **230**, 950–965.
  30. Moothoo, D. N., Canan, B., Field, R. A. & Naismith, J. H. (1999). Man(α1-2)Man-OMe-concanavalin A complex reveals a balance of forces involved in carbohydrate recognition. *Glycobiology*, **9**, 539–545.
  31. Bouckaert, J., Loris, R., Hamelryck, T. & Wyns, L. (1999). The crystal structures of Man(α1-3)Man and Man(α1-6)Man complexed to concanavalin A. *J. Biol. Chem.* **274**, 29188–29195.
  32. Brewer, C. F. & Brown, R. D. (1979). Mechanism of binding of mono- and oligosaccharides to concanavalin A: a solvent proton magnetic relaxation dispersion study. *Biochemistry*, **18**, 2555–2562.
  33. Brewer, C. F. & Bhattacharyya, L. (1986). Specificity of concanavalin A binding to asparagine-linked glycopeptides. A nuclear magnetic relaxation dispersion study. *J. Biol. Chem.* **261**, 7306–7310.
  34. Mandal, D. K., Kishore, N. & Brewer, C. F. (1994). Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. *Biochemistry*, **33**, 1149–1156.
  35. Naismith, J. H. & Field, R. A. (1996). Structural basis of trimannoside recognition by concanavalin A. *J. Biol. Chem.* **271**, 972–976.
  36. Loris, R., Maes, D., Poortmans, F., Wyns, L. & Bouckaert, J. (1996). A structure of the complex between concanavalin A and methyl-3,6-di-O-(alpha-D-mannopyranosyl)-alpha-D-mannopyranoside reveals two binding modes. *J. Biol. Chem.* **271**, 30614–30618.
  37. Frohman, M. A. (1993). Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol.* **218**, 340–356.
  38. Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
  39. Otwinowski, Z. & Minor, W. (1997). Processing of

- X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
40. Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. & Wyns, L. (1993). Crystal structure determination and refinement at 2.3-Å resolution of the lentil lectin. *Biochemistry*, **32**, 8772–8781.
41. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. sect. A*, **50**, 157–163.
42. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. sect. D*, **54**, 905–921.
43. Roussel, A. & Cambillau, C. (1989). TURBO-FRODO. *Silicon Graphic Geometry Partner Directory*, pp. 71–78, Silicon Graphics, Mountain View, CA.
44. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
45. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photo-realistic molecular graphics. *Methods Enzymol.* **277**, 505–524.

Edited by R. Huber

(Received 28 August 2003; received in revised form 18 November 2003; accepted 21 November 2003)