Crystal Structure Determination and Refinement at 2.3-Å Resolution of the Lentil Lectin†

Remy Loris,*,‡ Jan Steyaert,‡ Dominique Maes,‡ John Lisgarten,‡ Richard Pickersgill,§ and Lode Wyns‡

Laboratorium voor Ultrastructuur, Instituut voor Moleculaire Biologie, Vrije Universiteit Brussel, Paardenstraat 65, B-1640 Sint-Genesius-Rode, Belgium, and AFRC Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 2EF, U.K.

Received December 21, 1992; Revised Manuscript Received April 8, 1993

ABSTRACT: We report on the X-ray structure determination of the orthorhombic crystal form of lentil lectin by molecular replacement using the pea lectin coordinates as a starting model. The structure was refined at 2.3-Å resolution with a combination of molecular dynamics refinement and classical restrained least-squares refinement. The final R value for all data $F_o > 1\sigma(F_o)$ between 7.0- and 2.3-Å resolution is 0.164%, and deviations from ideal bond distances are 0.014 Å. The C-terminus of the β-chain proved to be 23 amino acids longer than found in previous studies. This together with several inconsistencies between the previously determined amino acid sequence and the observed electron density forced a redetermination of the amino acid sequence of the protein. The overall structure is very similar to that of pea lectin and isolectin I of *Lathyrus ochrus*, the most prominent deviations being confined to loop regions and the regions of intermolecular contact. The largest difference between the pea and lentil lectin monomers is situated in the loop region of amino acids 73–79 of the β chain. There are no significant differences between the two crystallographic independent lentil lectin monomers in the asymmetric unit. The model includes 104 well-defined water molecules, of which a significant number have a counterpart in the pea lectin structure. As for the other legume lectins, each lentil lectin monomer contains one calcium ion in a highly conserved environment. On the contrary, the manganese binding sites are distorted with respect to the pea lectin and concanavalin A structures. The Aspβ121 side chain apparently does not ligate the Mn^{2+} ion. This difference is consistent in both lentil lectin monomers and agrees with earlier solution studies. Possible implications for oligosaccharide binding are discussed.

The legume lectins are a broad class of homologous proteins capable of recognizing specific polysaccharides and glycopeptides (Goldstein et al., 1980). Although some of them have been studied extensively, their physiological role still remains obscure. Several possible roles have been suggested including their function in seed maturation, cell wall assembly, defense mechanism, and involvement in rhizobial nodulation of legume roots. Nevertheless, plant lectins have found important applications as immunological and biochemical tools. Moreover they can be regarded as model systems for studying the molecular basis of protein–carbohydrate interactions as they occur in cell–cell recognition events [for a review, see Lis and Sharon (1986)]. The structures of several lectins, including some lectin carbohydrate complexes, are known at high resolution (Shaanan et al., 1991; Bourne et al., 1990; Einspahr et al., 1986; Reeke & Becker, 1986; Hardman et al., 1984; Becker et al., 1975). However, the specific mechanisms by which lectins can discriminate between different but closely related oligosaccharides are not yet fully understood. In order to gain detailed information on the structure–function relationships of these proteins, further knowledge on their high-resolution three-dimensional conformations is required.

† This work has been supported by a grant from the VLAB project of the Flemish government. R.L. received a grant from the IWONL during part of this work and is at present a Research Assistant of the Nationaal Fonds voor Wetenschappelijk Onderzoek Belgium. D.M. and J.S. are research associate and senior research assistant of the National Fonds voor Wetenschappelijk Onderzoek Belgium, respectively. The authors also acknowledge the receipt of NATO Grant No. 900270.

* Address correspondence to this author at Vrije Universiteit Brussel. (Telephone: 32-2-3590209; Fax: 32-2-3590390; e-mail: reloris@vub.ac.be).

‡ Vrije Universiteit Brussel.

§ Reading University.

All legume lectins bind one transition metal ion (usually manganese) and one calcium ion per monomer. The presence of these two metal ions is essential for their sugar binding activity. The geometry of the metal binding sites has proven to be a very conserved feature in the different legume lectin crystal structures reported to date. In all these structures both the transition metal and the calcium are hexacoordinated. Solution studies also point toward a conserved geometry around the metal ions. However, the experimental data on the transition metal binding site in solution favor a pentacoordinated transition metal ion if the calcium ion is also present (Kalb et al., 1979). This inconsistency between solution and single-crystal studies remains to be explained.

The lentil lectin is a homodimeric protein belonging to the class of two-chain lectins. The active protein consists of a dimer with each monomer containing an α and a β chain. It highly resembles the pea lathyrus and fava bean lectins in its primary structure, except for a presumed deletion at the C-terminus of the β chain. As the other *viciae* lectins, it shows a strong specificity for glycopeptides of the *N*-acetyllactosamine type derived from human lactalbumin (Debray et al., 1981). In this paper we present the X-ray structure determination of the orthorhombic crystal form of the lentil lectin at 2.3-Å resolution together with the gene structure and deduced amino acid sequence. The structure is compared in detail with those of pea and lathyrus lectin. The implications of the structural details for carbohydrate binding and posttranslational processing are discussed.

MATERIALS AND METHODS

Crystallization and Data Collection. Lentil lectin was purified using affinity chromatography on Sephadex G75 and crystallized as described previously (Loris et al., 1992). The

0006-2960/93/0432-8772$04.00/0 © 1993 American Chemical Society

Table I: Crystallization, Data Collection, and Refinement Statistics

crystallization conditions	20 mM phosphate, 20 mM acetate, 0.1 mM $CaCl_2$, 0.1 mM $MnCl_2$, pH 5.5, 3 mg/mL protein, and 40% MPD
space group	$P2_12_12_1$; a = 56.4 (1) Å, b = 7.64 (1) Å, c = 124.9 (1) Å
total number of observations	55 775 between 7 and 2.3 Å
number of unique reflections used	15 235 between 7 and 2.3 Å
final R value	0.164
$R_1 = \sum_i\sum_j\|<I_i> - I_{ij}\|/\sum_i\sum_j I_{ij}$	9.3% for all data
rms deviations from ideality (distance restraints; target values in parentheses)	bond lengths: 0.014 Å (0.020) bond angles: 0.032 Å (0.050) planar groups: 0.009 Å (0.015)
number of solvent molecules	104 water molecules 2 Ca^{2+} ions 2 Mn^{2+} ions 2 acetate ions

Table II: Completeness and Qulaity of the Data as a Function of the Resolution[a]

resolution shell (Å)	number of unique reflections	number theoretically possible	% completeness	R_{sym}
7.00–4.58	1828	1852	98.7	0.047
4.58–3.75	2272	2540	89.4	0.054
3.75–3.25	2418	3006	80.4	0.066
3.25–2.91	2476	3365	73.6	0.071
2.91–2.65	2267	3702	61.2	0.121
2.65–2.46	2012	3988	50.5	0.161
2.46–2.30	1962	4289	45.7	0.193
total	15 235	22 742	67.0	0.093

[a] Data were considered observed if $I > 1\sigma$. $R_{sym} = \sum_i\sum_j\|<I_i> - I_{ij}\|/\sum_i\sum_j I_{ij}$.

protein crystallizes in the orthorhombic space group $P2_12_12_1$ with one functional entity, a dimer, in the asymmetric unit. Details about the unit cell and the data collection statistics are summarized in Tables I and II. Diffraction intensities were collected up to a resolution of 2.3 Å on two crystals on a Xentronics/Siemens area detector using Cu Kα radiation generated from a rotating anode source operated at 50 kV and 70 mA. The crystals were mounted in an arbitrary orientation with respect to the camera axes and the detector distance was set to 160 mm to enable a cell edge larger than 100 Å to be resolved. The exposure time was 120 s/frame with each frame consisting of a 0.25° rotation in omega at 2θ = 25°. Data reduction was carried out with the XENGEN software (Howard et al., 1987).

The data were collected to a resolution of 2.3 Å. As can be seen in Table II, the higher resolution shells suffer from a low completeness. Based on the completeness, the effective resolution of these data was calculated to be 2.55 Å. The criterion used for including the highest resolution shell into the refinement was the resulting improvement of the quality of the electron density maps as compared to maps calculated at 2.4 or 2.5 Å.

Molecular Replacement. Molecular replacement was performed with the MERLOT package (Fitzgerald, 1988) on a MicroVAX 3300 computer using as a search model the 1.7-Å pea lectin dimer, present as entry 2LTN in the Brookhaven database (Bernstein et al., 1977). All nonidentical amino acid side chains were removed along with the 23 C-terminal residues and residues 73–74 and 107–108 of both β chains. Structure factors were calculated from this truncated model placed in a triclinic cell with $a = b = c = 100$ Å and $\alpha = \beta = \gamma = 90°$. The largest peak (5.83$\sigma$ above background) in the cross-rotation function was found around the three Eulerian angles $\alpha = 108.5°$, $\beta = 65.8°$, and $\gamma = 339.0°$. This solution was consistent in the different resolution ranges tested using data between 12.0 and 3.0 Å and was in agreement with the orientation of the noncrystallographic 2-fold axis almost parallel to the b axis as determined from a self rotation function.

The dimer, oriented according to the results of the rotation function, was subsequently submitted to a translation function search in the 10.0–4.0-Å resolution range. A single consistent set of peaks was found corresponding to a solution showing reasonably good contacts between symmetry-related units.

Refinement Strategy. Refinement was carried out using both restrained least-squares refinement with RESTRAIN (Haneef et al., 1985; Driessen et al., 1989) and simulated annealing refinement with X-PLOR (Brünger, 1989) implemented on a CRAY-YMP supercomputer. The course of the refinement is summarized in Table III. No noncrystallographic symmetry restraints were applied. Model building was carried out using the interactive graphics package FRODO (Jones, 1978) after inspection of difference electron density maps of the type $2F_o$–F_c and F_o–F_c. The final coordinate set contains 3510 protein atoms, 2 manganese, and 2 calcium atoms together with 104 symmetry independent water molecules distributed over the entire asymmetric unit. Structural comparison of the refined lentil lectin structure with other legume lectins available in the Brookhaven database were carried out with X-PLOR (Brünger, 1989) and programs available in the CCP4 package (Evans, 1991). Hydrogen bonds were identified using the criteria of Baker and Hubbard (1984). Molecular drawings were made by directly creating a PostScript output from FRODO. The final coordinates have been submitted to the Brookhaven Database (Bernstein et al., 1977).

Sequencing of the Lentil Lectin Gene. Genomic DNA from the leaves of lentil plants was prepared using an established

Table III: Refinement Strategy

step	resolution	no. of reflections	R value	description	observations
1	6.0–3.0	6599	0.43	rigid body refinement	
2	6.0–3.0	6599	0.30	restrained least-squares refinement	very poor density map for residues 1–10 of both β chains
3	6.0–2.5	10 117	0.29	restrained least-squares refinement	a limited number of substituted amino acid side chains could be located, mainly in the α chains
4	6.0–2.3	14 581	0.31	restrained least-squares refinement	density map still very noisy; few new features
5	6.0–2.3	14 581	0.29	molecular dynamics refinement; slow cooling protocol (2500–300 K) followed by 120 cycles of Jack–Levitt X-ray restrained energy minimalization	residues 158–180 inserted in both monomers
6	6.0–2.3	14 581	0.25	restrained least-squares refinement	remaining missing residues and side chains inserted; Tyrβ77 and Lysβ107 loops inserted
7	6.0–2.3	14 581	0.20	restrained least-squares refinement with individual B values	104 water molecules identified; Mn^{2+} binding site rebuilt; two acetate molecules identified
8	7.0–2.3	15 235	0.16	restrained least-squares refinement	

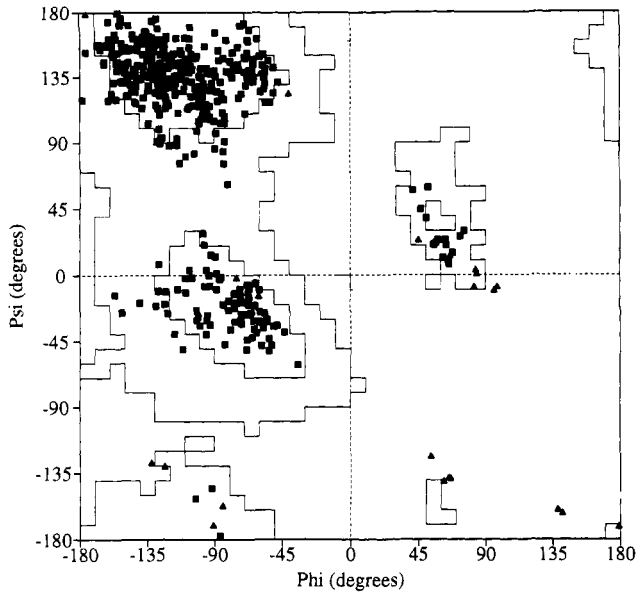

FIGURE 1: Ramachandran plot of the lentil lectin dimer, giving an indication of the stereochemical quality of the model. Glycine residues are represented by triangles while all other amino acids are represented by squares.

protocol (Dellaporta et al., 1983). Based on the previously published amino acid sequence of the lentil lectin (Foriers et al., 1981) and the cDNA sequence of pea lectin (Higgins et al., 1983), two optimal primers were designed (5′-ACCT AGATCTTCCAAGGAGATGG-3′ and 5′-TCAC-GGTACCCACTCAGGAACAAC-3) containing as few as possible amino acids with a codon usage larger than or equal to four and still covering the parts with unknown or unreliable sequence. *Bgl*II and *Kpn*I restriction sites (underlined) were built in in order to facilitate subcloning into pBluescript (Stratagene). These primers were used for PCR amplification of the lentil lectin gene fragment. A single amplification product of the expected length (592 bp) was obtained under standard conditions (Scharf, 1990) and sequenced using the Sanger dideoxy method (Sanger et al., 1977) after subcloning into pBluescript using the above mentioned built-in restriction sites.

RESULTS

Progress of the Refinement. The molecular replacement calculations using the pea lectin model yielded a solution that was consistent over all resolution ranges tested and that showed reasonably good contacts between symmetry-related units. The initial R value was 0.49 but dropped to 0.43 when the positions of both monomers were refined separately by rigid body refinement. After convergence of the restrained least-squares refinement, the computed electron density was still difficult to interpret with several breaks in the main chain density. Only very few of the substituted amino acid side chains could be placed in the electron density with certainty. Because no significant improvement to this situation was observed, even after several alternating rounds of refinement and map inspection, simulated annealing refinement was carried out using the program X-PLOR. Despite a relatively mediocre drop in R value, the resulting electron density was much clearer. Density for almost the complete backbone and some aromatic side chains of a 23 residues long stretch of amino acids extending from the C-termini of the β chains became visible in the F_o–F_c difference maps, albeit at the very low level of 0.12 $e/Å^3$. Probably because of the incompleteness of the current model, several parts of the model were artificially forced into wrong parts of the density. This phenomenon has also been observed by other authors (Bourne et al., 1990).

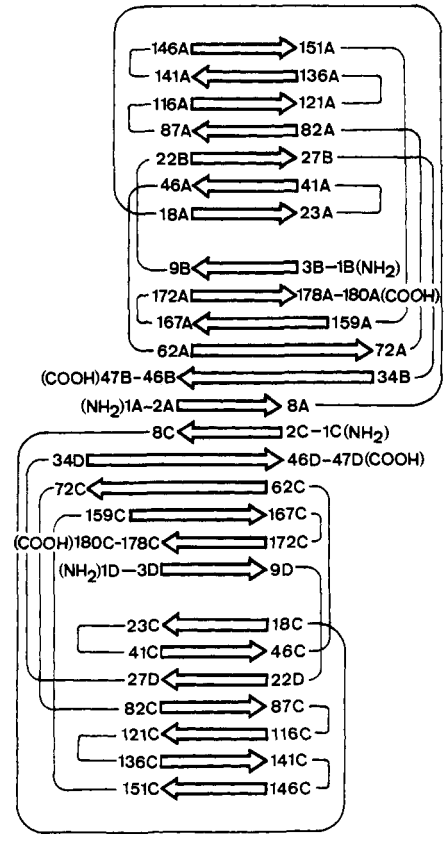

FIGURE 2: Topology of the secondary structure of the lentil lectin dimer. β strands are represented as arrows. The greek key β-barrel motifs are clearly visible. The numbering scheme adopted is the same as the one that was used for pea lectin by Einspahr et al. (1986): A and C correspond to the β chains while B and D correspond to the α chains of both monomers.

After building in this stretch of 23 residues and manually readjusting the parts of the structure that were forced into the wrong part of the density, the R-factor dropped to 28.4 at a 1σ level, and most of the missing side chains could be built in. Two regions, however, remained obscure: Serβ66–Tyrβ77 and Pheβ104–Lysβ111. The backbone in these regions could only be traced with confidence after we first removed a relatively large part of the structure around these areas. From that moment on, the electron density improved markedly, but it also became more obvious that the published amino acid sequence was in error in those two stretches. This together with the β chain terminal extension prompted us to redetermine the lentil lectin sequence. This newly determined sequence was in good agreement with the difference density seen at this stage. All missing side chains, together with 104 water molecules, could be fitted in during the next few cycles of refinement and model building. The final R value of the model is 0.164 for 15 235 independent reflections between 7.0 and 2.3 Å with a deviation from the ideal bond lengths of 0.014 Å, indicating good stereochemistry of the model. The Ramachandran plot (Ramachandran et al., 1963) shown in Figure 1 is characteristic for an all-β protein. No combination of Φ–Ψ angles is found outside the allowed regions of the Ramachandran plot.

Structure Description. As with the other legume lectins, the main structural features are two large antiparallel β-sheets stretched out over the whole lectin dimer. Figure 2 shows the diagram of the secondary structure. A stereo representation of the backbone tracing of the lentil lectin dimer and a ribbon drawing emphasizing the location of the secondary structure elements is given in Figure 3. The overall fold is the jellyroll Greek key β-barrel motif. Each monomer contains a calcium

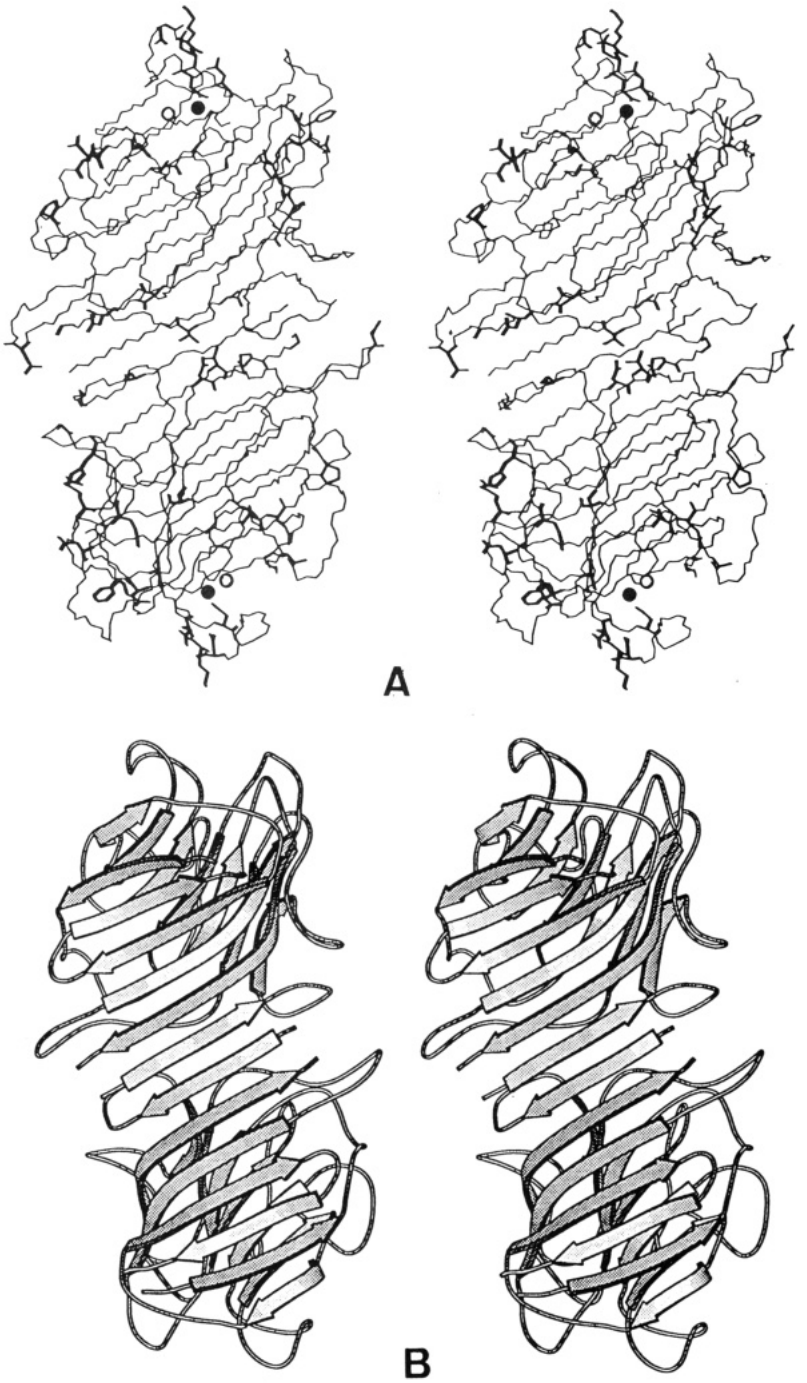

FIGURE 3: Overall view of the lentil lectin molecule. (A) Stereodrawing of the backbone tracing (N, C, Cα) of the lentil lectin dimer. The side chains that are substituted between pea and lentil lectin are shown in bold. The manganese and calcium ions are shown as solid and open spheres, respectively. (B) Ribbon diagram emphasizing the position of the secondary structure elements.

and a manganese atom in the vicinity of the putative carbohydrate recognition site.

At the resolution of 2.3 Å, no extended solvent networks could be identified. However, distributed over the entire surface of the lectin dimer, 104 symmetry-independent water molecules belonging to the first hydration shell were found.

The mean coordinate error of this structure, as determined by a Luzzati plot (Luzzati, 1952) is 0.23 Å. These errors will of course be larger in some of the more poorly defined loop regions. No discretely disordered side chains were observed. In the LOL I[1] structure, the last amino acid on the C-terminus of the β chain that can be seen clearly is Asnβ181, the side chain carboxylate group of which is involved in a salt bridge with the amino group of Valα1. The last amino acid of the β chain for which electron density can be observed is Proβ180. In monomer A of the lentil lectin structure this could be due to the fact that Valα1 is involved in intermolecular contacts, and as a result no salt bridge similar to the one observed in the lathyrus structure can be formed, thus leaving Asn181 unresolved.

A number of large side chains situated on the molecular surface have poorly defined densities and high *B* values. The following side chains have no detectable density beyond the CB atom: Thrβ27, Tyrβ77 and Asnβ161 of both monomers, Valβ41, Serβ106, and Valβ163 of monomer A, and Serβ75 of monomer B. All are located in the β chains.

[1] Abbreviations: MPD, 2-methyl-2,4-pentanediol; LOL I, isolectin I from *Lathyrus ochrus*; PsL, Pea (*Pisum sativum*) lectin; LcL, lentil (*Lens culinaris*) lectin; Con A, concanavalin A.

β-chain

```
          1        10        20        30        40        50        60        70
(1)  TETTSFSITKFSPDQQNLIFQGDGYTTKGKLTLTKAVKSTVGRALYSTPIHIWDRDTGNVANFVTSFTFV
(2)  TETTSFLITKFSPDQQNLIFQGDGYTTKEKLTLTKAVKNTVGRALYSSPIHIWDRETGNVANFVTSFTFV
(3)  TETTSFSITKFGPDQQNLIFQGDGYTTKERLTLTKAVRNTVGRALYSSPIHIWDSKTGNVANFVTSFTFV

                  80        90       100       110       120       130       140
(1)  IDASNSYNVADGFTFFIAPVDTKPQTGGGYLGVFNKEDYDKTSQTVAVEFDTFYATAWDPSNKERHIGID
(2)  INAPNSYNVADGFTFFIAPVDTKPQTGGGYLGVFNAEDYDKTSQTVAVEFDTFYATAWDPSRGDRHIGID
(3)  IDAPNSYNVADGFTFFIAPVDTKPQTGGGYLGVFNSKDYDKTSQTVAVEFDTFYNTAWDPSNGDRHIGID

                  150       160       170       180
(1)  VNSIKSVNTKSWNLQNGERANVVIAFNAATNVLTVTLTYPN
(2)  VNSIKSVNTKSWKLQNGEEANVVIAFNAATNVLTVSLTYPN
(3)  VNSIKSINTKSWKLQNGKEANVVIAFNAATNVLTVSLTYPN
```

α-chain

```
          1        10        20        30        40        50
(1)  VTSYTLNEVVPLKDVVPEWVRIGFSATTGAEFAAQEVHSWSFNSQLGHTSKS
(2)  VTSYTLSDVVSLKDVVPEWVRIGFSATTGAEYAAHEVLSWSFHSELSGTSSNKQ
(3)  ETSYTLNEVVPLKEFVPEWVRIGFSATTGAEFAAHEVLSWFFHSELAGTSSN
```

FIGURE 4: Comparison of the amino acid sequence of (1) the lentil lectin, (2) pea lectin, and (3) LOL I. Amino acid differences between the three lectins are shown in bold. The part of the lentil lectin sequence that was not covered by the PCR is shown in italic. No inconsistencies were found in these regions between the published amino acid sequence and the observed electron density.

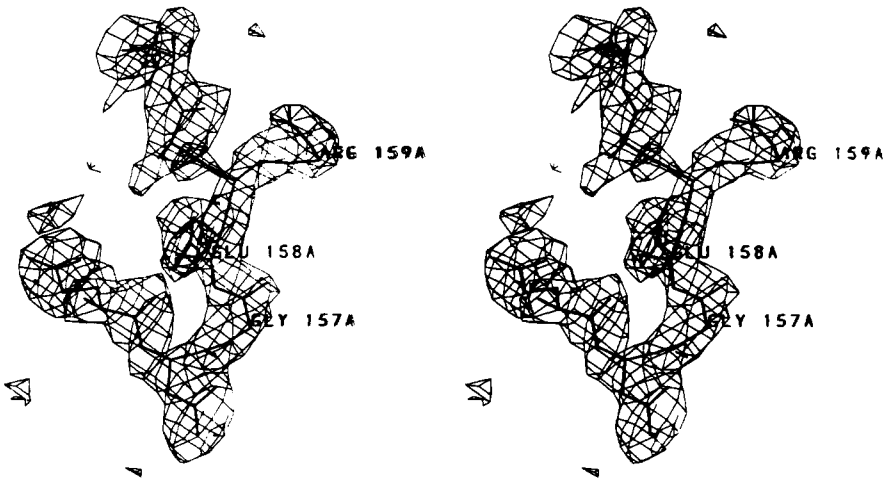

FIGURE 5: Electron density map around Argβ159 in monomer A. The arginine side chain is well defined as well as the covalent link between Glyβ157 and Lysβ158. The map was calculated using phases derived from the final structure after removal of residues Leuβ155A to Asnβ161A.

Gene Structure. It is generally accepted that the α and β subunits of the two chain lectins from the *Viciae* tribe are generated by a proteolytic cleavage at the Asn–X bond of a single chain precursor. Unexpectedly, the β chain of the lentil lectin was found to contain only 157 amino acids instead of 181 observed in the other *Viciae* lectins (Foriers et al., 1981). Because the missing part of the sequence corresponded to two strands of the highly conserved β-sheet structure, it was assumed that in the lentil lectin a second proteolytic cleavage takes place after Glyβ157.

During the course of the refinement, an extra stretch of 23 amino acids was identified on the C-terminus of both of the β chains. This stretch was clearly covalently linked to Glyβ157. Very poor agreement was also observed between the published amino acid sequence and the side chain densities of amino acids β70–β77 and β105–β108. Therefore, the lentil lectin gene was amplified by PCR and its DNA sequence determined. The deduced amino acid sequence is given in Figure 4 together with the amino acid sequences from pea lectin (Higgins et al., 1983) and isolectin I of *Lathyrus ochrus* (Richardson et al., 1984; Yarwood et al., 1985). The results from the sequencing of the lentil lectin gene indicate a much closer homology to the pea lectin than previously assumed. No heterogeneity was detected in the sequence derived from five different clones except for Tyrβ77, which was Phe in one clone. It cannot be concluded whether this is due to real sequence heterogeneity or to an error from the PCR amplification.

DISCUSSION

Posttranslational Processing. In contrast to what previously had been found (Foriers et al., 1981), the β chain of the lentil lectin has the same length as the β chain of the other two chain lectins. In the extra 23 amino acids that we identified for the C-terminus of the β chain, only one amino acid is substituted as compared to pea lectin: Gluβ159 is substituted for an arginine, the electron density of which showed up clearly in the difference map. The electron density of the region around Argβ159 is shown in Figure 5 and shows clearly the covalent link between Glyβ157 and Lysβ158.

We may thus conclude that the lentil lectin gene structure is similar to that of the other legume lectins and posttranslational processing occurs in the same way as has been demonstrated for pea lectin: the lentil lectin is encoded as a single polypeptide chain that after removal of a 20 amino acid leader sequence is proteolytically cleaved into an α chain and

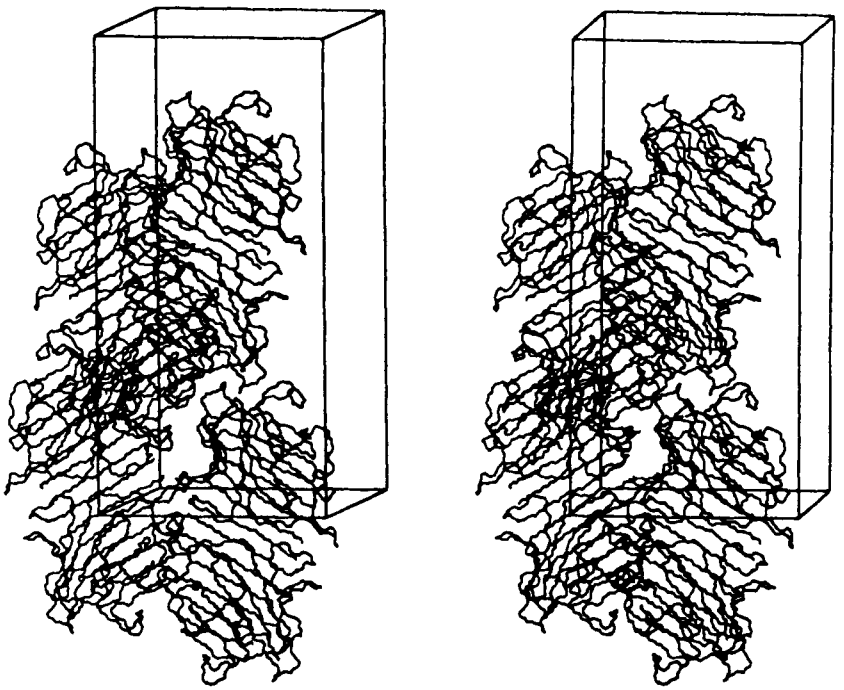

FIGURE 6: Stereodrawing of the packing of the orthorhombic lentil lectin dimers in their unit cell viewed along the *c* axis.

a β chain. No evidence was found for a second cleavage of the β chain after Glyβ157 as was suggested by the work of Foriers et al. (1981).

Packing Analysis. A stereodrawing of the packing of the lectin dimer in the unit cell is given in Figure 6. The lectin dimers are packed in a very loose structure with large solvent channels. Interestingly, the presumed carbohydrate recognition site faces one of these channels, suggesting the possibility of soaking in oligosaccharides. Indeed preliminary experiments involving α-D-mannopyranoside and *N*-acetylmuramyl-D-Ala-D-isoGln showed that our lentil crystals withstand soaking with small haptenic ligands for several days. Such soaking studies were not possible for the other legume lectin crystals reported to date because of unfavorable packing contacts.

The number of crystal packing interactions in the lentil lectin crystals is significantly lower than those observed in the crystals of the pea and *L. ochrus* lectin (Table IVa). Most of the crystal packing contacts are mediated by at least one water molecule. Indeed, only seven symmetry-independent direct intermolecular hydrogen bond contacts are observed (Table IVb), five of which invole a main chain–side chain interaction. The other two consist of a strong hydrogen bond between the amino terminus of the α chain of the first monomer with the carbonyl of Proβ94 of the same monomer in a symmetry-related position and a side chain–side chain interaction between Lysβ145A (NZ) and Glnβ155B (OE1). The other hydrogen bonding contacts are between the two crystallographically different monomers, and four of them are reciprocal, e.g., similar contacts are made by both monomers.

Overall Structure of the Lentil Lectin. The high sequence homology of the lentil lectin with the lectins from pea and *L. ochrus* is also reflected in its overall three-dimensional structure. A plot of the per residue rms distances of the main chain atoms after superposition of the different lentil lectin and pea lectin monomers is given in Figure 7. Both monomers are quite similar, the rms coordinate difference being only 0.33 Å for the backbone atoms and 0.89 Å for the side chain atoms. Similar values were obtained for the comparison of pea lectin and isolectin I of *L. ochrus* (Bourne et al., 1990). These values together with the rms coordinate differences between the different lectins are listed in Table V. As was also observed for the *L. ochrus* structure, both monomers are not completely identical, monomer I giving a better overall fit to both pea lectin monomers. Interestingly, monomer I also displays the largest absolute difference with respect to the two pea lectin monomers, situated around Tyrβ77. Other significant differences between pea lectin and lentil lectin can be found in the backbone around Thrβ27, the side chain packing around Thrβ48 and to a lesser extent around Serβ7.

As can be seen in Figure 7, there are no significant differences in the backbone coordinates of both lentil lectin monomers. The largest deviations between both monomers

Table IV: Crystal Packing Contacts

(a) Comparison of the Intermolecular Contacts in the Lentil, Pea, and *Lathyrus ochrus* Lectin Structures[a]

	lentil			pea			lathyrus		
	A	B	T	A	B	T	A	B	T
total number of contacts	66	54	120	222	124	346	89	117	226
hydrogen bonds	8	6	14	18	12	30	7	5	12
salt bridges	0	0	0	3	3	6	4	4	8
van der Waals contacts	58	48	106	201	109	310	78	108	186
no. of amino acids involved in crystal contacts	15	13	28	44	30	74	23	25	58

(b) Intermolecular Hydrogen Bond Contacts in the Lentil Lectin Crystals

molecule 1	molecule 2	distance (Å)
Pro74A O	Asn148B ND2	2.47
Ser76A N	Glu188B OE1	3.37
Asp91A O	Tyr109B OD1	2.91
Pro94A O	Val181A N	2.67
Lys145A NZ	Gln155B OE1	3.40
Asn148A ND2	Pro74B O	2.89
Glu188A OE2	Ser76B N	3.04

[a] Listed here are the total number of atoms for each molecule to be involved in crystal packing interactions. A, B, T correspond to the A and B monomers and the complete dimer, respectively. The number of symmetry-independent interactions thus corresponds to half of the number listed under T. Packing contacts were determined with the program X-PLOR using the criteria $2.4 < d < 3.5$ for hydrogen bonds, $1.75 < d < 3.5$ for salt bridges, and $d < 4.0$ for van der Waals contacts. The values for the lathyrus lectin are taken from Bourne *et al.* (1990a).

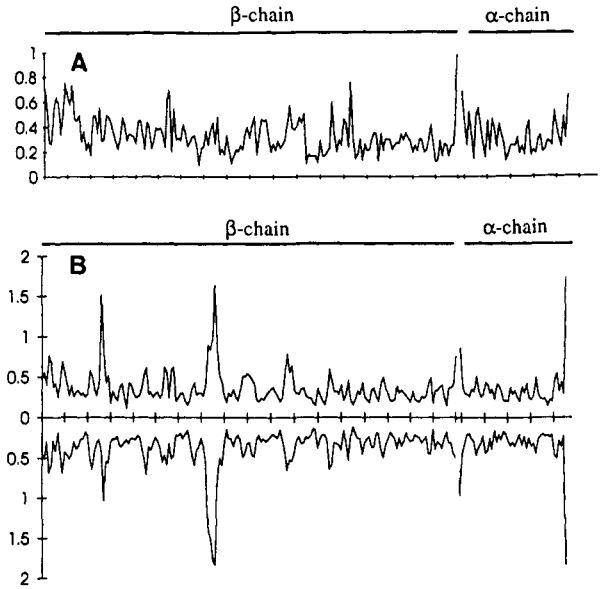

FIGURE 7: (A) Per residue backbone rms deviations between the two lentil lectin monomers. (B) Per residue backbone rms deviations between lentil lectin and pea lectin. (Upper) lentil lectin monomer A and pea lectin monomer A and (lower) lentil lectin monomer B and pea lectin monomer B.

Table V: Rms Deviations (Å) between the Backbone Atoms for the Different Legume Lectin Structures

	LcL 2	PsL 1	PsL 2
LcL 1	0.33	0.36	0.40
LcL 2		0.38	0.42
PsL 1			0.23

are found in the N-terminal and C-terminal residues of the α and β chains and more localized around Thrβ56 and Aspβ134. By far the largest difference between pea and lentil lectin is found around Tyrβ77. This part of the structure forms a flexible loop. Here the two lentil lectin conformations are clearly more similar to each other than to pea lectin. This region is characterized by relatively large temperature factors and was manually rebuilt. The side chain density could not be observed beyond the CB atom for Tyrβ77 in either of the two lentil lectin monomers. In the pea lectin structure, Tyrβ77 is involved in crystal packing interactions either by hydrogen bonding of the tyrosine OH atom (A monomer) or by extensive van der Waals contacts with side chains of a neighboring molecule (B monomer).

A second region with relative large backbone deviations from the pea lectin model can be found around Thrβ27. In the pea lectin structure the side chain of this residue in monomer A points roughly in the opposite direction as compared to the one in monomer B. The backbone conformation of both lentil lectin monomers resembles more closely the one observed in pea lectin monomer B. The Thrβ27 region was reported to give the largest deviations between pea lectin and *L. ochrus* lectin. Deviations in the Tyrβ77 region however are less pronounced in the pea–lathyrus comparison.

In general, amino acid side chains substituted between pea and lentil lectin occupy the same volume in space and require only local adjustments of the three-dimensional structure. The distribution of the amino acid substitutions between pea and lentil lectin are shown in Figure 3. As can be seen, almost all amino acid substitutions are located on the surface of the protein. Most of them are located in loop regions. Those that occur in β-sheets have their side chains systematically facing the solvent.

Table VI: Metal Binding Geometry

	Calcium Binding Site					
ligand	LcL 1	LcL 2	PsL 1	PsL 2	LOL 1	Con A
Asp 121 OD1	2.98	2.57	2.47	2.40	2.22	2.46
Asp 121 OD2	2.24	2.45	2.66	2.65	2.52	2.50
Phe 123 O	2.37	2.39	2.30	2.24	2.08	2.08
Asn 125 OD1	2.24	2.17	2.23	2.35	1.89	2.60
Asp 129 OD2	2.77	2.83	2.42	2.39	2.60	2.45
water	2.15	2.10	2.34	2.34	2.29	2.27
water	2.09	1.99	2.39	2.39	2.27	2.45
	Manganese Binding Site					
ligand	LcL 1	LcL 2	PsL 1	PsL 2	LOL	Con A
Glu 119 OE2	2.65	2.62	2.22	2.21	2.29	2.31
Asp 121 OD1	3.57	3.29	2.07	1.93	2.21	2.28
Asp 129 OD1	2.64	2.89	2.19	2.23	2.28	2.28
His 136 NE2	2.56	2.46	2.24	2.19	2.44	2.2
water	ND[a]	ND[a]	2.25	2.19	2.38	2.20
water	ND[a]	ND[a]	2.15	2.03	2.32	2.32

[a] Not determined. The water ligands of the manganese atoms did not show as separate peaks in the electron density maps.

Two amino acid substitutions are of particular interest as they constitute part of the monomer–monomer interface. The first one is Serβ7, which in pea lectin is a leucine. This substitution may account for the small but consistent rms difference that is observed in the backbone of the N-terminal residues as compared to pea lectin. The second one is more interesting and involves Serβ48 in pea lectin to be replaced by Thrβ48. This side chain is completely buried in the interface, and some local changes in the monomer–monomer interactions are necessary in order for the structure to accommodate for the presence of an extra CG2 atom. In the lentil lectin structure, OG1 of this residue is hydrogen bonded to the OG1 of the same residue of the second monomer. In the pea lectin structure, the OG atom of Serβ48 occupies roughly the same location as the lentil lectin CG2 atom of Thrβ48 and makes contact via a water bridge with the ND2 atom of Asnβ17 and with the OH atom of Tyrβ46, both of the opposite monomer. In the lentil lectin structure the latter interaction is substituted by a water-bridged hydrogen bond between OH of Tyrβ46 and OG of Serβ4, again of opposite monomers.

Detailed Analysis of the Metal Binding Sites. In the other legume lectin structures determined to date, each monomer contains one manganese and one calcium ion in a highly conserved environment. The presence of a *cis* peptide bond between Alaβ80 and Aspβ81 is essential for providing a correct geometry of the calcium binding sites. All of the protein ligands that form the manganese and calcium sites in the pea and lathyrus lectins have been conserved in the lentil lectin and therefore no significant differences are expected between these three structures. As can be seen in Table VI, the calcium binding site in the lentil lectin structure is indeed very similar to that in the other two chain lectins and Con A. The largest difference between the lentil lectin and the other two chain lectins concerns the position of the carboxylate group of Aspβ121 which normally bridges both metal ions.

In the case of the manganese binding sites, the situation is somewhat different. It was noted during the refinement that the ligand distances to the manganese ions remained unusually long. Moreover, the difference in the refined *B* values of the manganese ions in both monomers, as compared to the calcium ions in the lentil lectin structure, were remarkably large. Because this was considered problematic, the manganese binding regions of both lentil lectin monomers have been heavily modeled using omit maps and different conditions for the refinement. We are confident now that the side chain

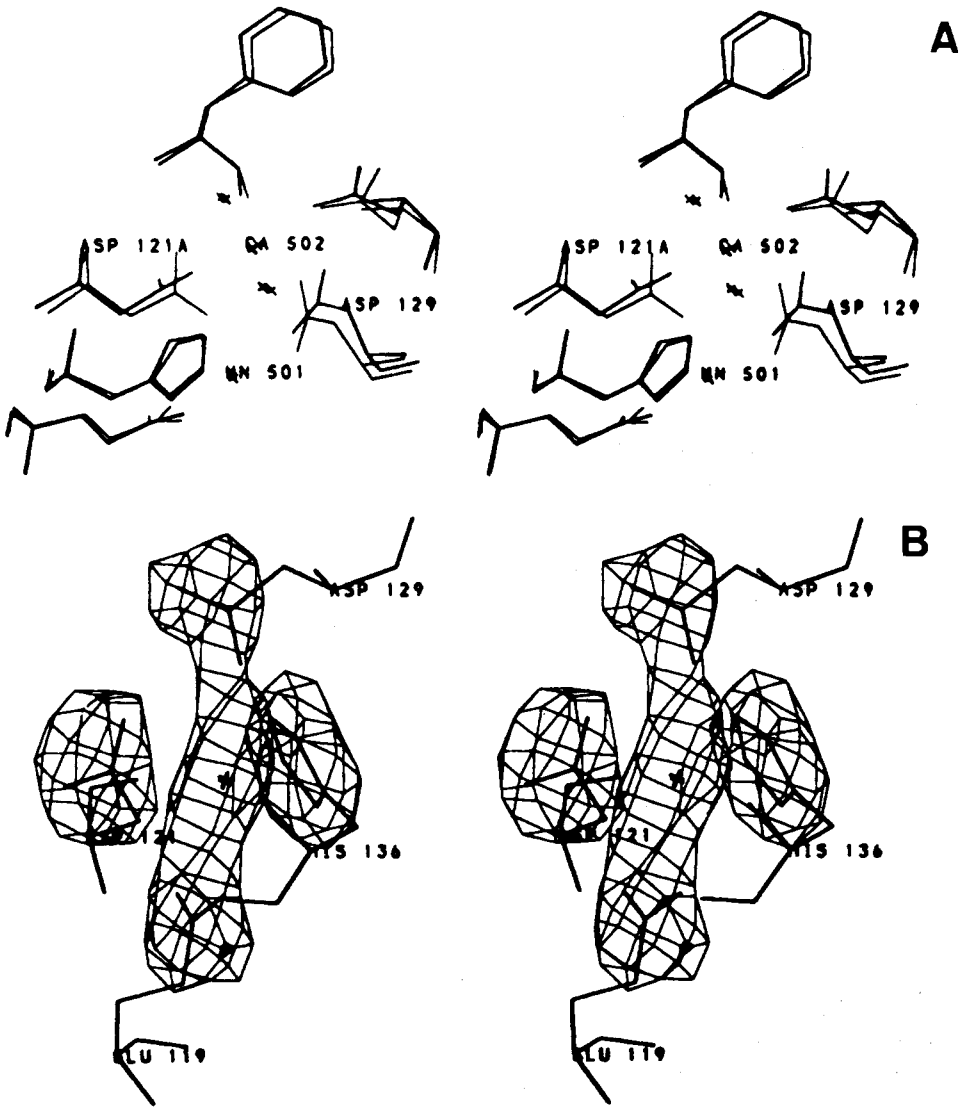

FIGURE 8: Stereodrawing of (A) the superposition of the metal binding region of the lentil lectin monomer A (thick lines) and pea lectin monomer A (thin lines) showing the Mn^{2+} and Ca^{2+} ions, their protein ligands, and the two conserved water ligands coordinating the Ca^{2+} ion and (B) F_o–F_c omit map of the manganese environment. The manganese ion was removed from the final coordinate file and the side chains of Gluβ119, Aspβ129 and Hisβ136 truncated to CB. The side chain of Aspβ121 was truncated to CA. The map was calculated after refinement of this truncated structure reached convergence and is shown at an arbitrary level.

orientations in this region are correct within the limits imposed by the data at the resolution obtained and that the differences between the lentil and pea lectin structures are significant. In the structures of ConA, pea lectin, and lathyrus lectin, the *B* values of metal ions are comparable and of the same order of magnitude as the calcium *B* value in our structure. The manganese *B* values in the lentil lectin structure, on the contrary, are approximately twice as large as the corresponding calcium *B* values. This difference in *B* values is similar in both monomers of the lentil lectin which were refined independently. The main structural difference in the manganese binding sites between pea and lentil lectin is found in the side chain orientation of Aspβ121. As can be seen in Figure 8, the carboxyl group of this residue is turned 90° around the CB–CG bond with respect to its orientation in the pea lectin structure. Therefore none of the carboxyl oxygen atoms ligate the manganese ion as their distance is far too large for ligand binding. As can be seen in Table VIb and Figure 8B, the other protein ligands are moved somewhat away from the manganese ligands, resulting in a considerable anisotropic movement of the manganese ion in both monomers.

In the pea and lathyrus lectin structures as well as in Con A, both the calcium and the manganese ion have two water ligands. Despite their short distances to the metal, similar water ligands could be identified in the calcium binding sites of the lentil lectin. The positions of these waters are virtually identical to those of the corresponding water molecules in the pea lectin structure. This is in full agreement with proton and deuteron nuclear magnetic relaxation dispersion and solid-state NMR spin-echo investigation studies on the lentil and pea lectins (Bhattacharyya et al., 1985; Marchetti et al., 1988) that also indicate that water ligation and environment of the metals in pea and lentil lectin are similar if not identical. The two water ligands which were expected around each of the manganese ions could not be identified in the lentil lectin structure. Since there is no obvious reason for the absence of these molecules, their identification is most likely hampered by the presumed anisotropic motion of the Mn^{2+} ion they ligate.

Interestingly, a pentavalent coordination of the transition metal (as observed in the lentil lectin structure) is in complete agreement with the early EXAFS studies on Con A by Kalb (Gilboa) and co-workers (Kalb et al., 1979). These authors observed a decrease of the transition metal coordination number from six to five upon binding of calcium in the S2 site. In our structure, the Aspβ121 ligand is turned away from the manganese in favor of the calcium. More recent XAFS studies in the crystal and in solution (Lin et al., 1990, 1991) also

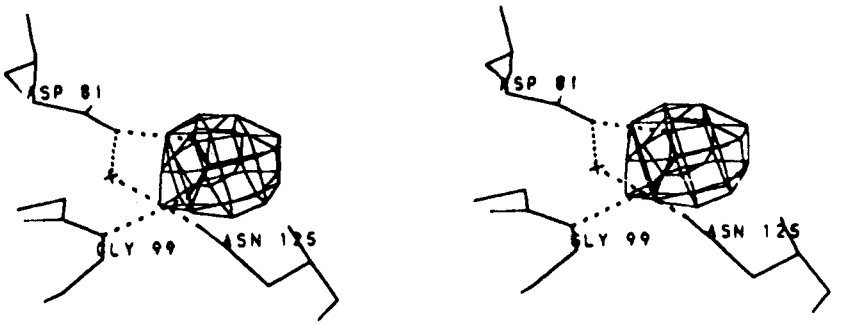

FIGURE 9: F_o–F_c omit map of the acetate environment. The acetate molecule was removed from the final coordinate file, and the map was calculated after eight cycles of restrained refinement. The final coordinates are superimposed on this map. Hydrogen bonds are indicated by dashed lines.

suggest significant differences of the transition metal coordination between the crystal and solution structure. Therefore we consider it a possibility that the lentil lectin crystal structure is more closely related to the solution structure than the densely packed crystal structures of Con A and pea lectin. However, further studies will be necessary to confirm or contradict this hypothesis.

Solvent Structure. Approximately half of the 104 solvent sites identified in the lentil lectin crystal seem to have a counterpart in the pea lectin structure despite side chain substitutions and differences in crystal packing. These conserved water sites do not only involve the metal binding sites (the water molecules ligating the Ca^{2+} cations) and the monomer–monomer interface (three water molecules close to or on the noncrystallographic 2-fold axis relating both monomers) but also a number of water molecules that are located at the surface of the molecule and have a large solvent-exposed area. One of the conserved waters on the monomer-monomer interface is located on the noncrystallographic 2-fold axis and bridges the hydroxyl groups of the Tyrβ46 residue of both protomers. Moreover side chain substitutions seem to be such that the substitution has a minimal effect on the surrounding water structure. For example, pea lectin Hisβ215 interacts with Thrβ122 via a water bridge. In the lentil lectin structure, Hisβ215 is replaced by a glutamine but the NE2 atom is located in the same position as the histidine NE2 atom allowing the same water bridge to be formed, illustrating both the importance of structural water in protein structures and the equivalence of His and Gln that is observed in certain situations (Lowe et al., 1985). It is also interesting to note that mutation of Trpβ128 in pea lectin, which is one of the amino acid residues that stabilizes a conserved water site, leads to increased temperature sensitivity or even complete denaturation (Stubbs et al., 1991).

It is expected that if more complete and higher resolution data would become available, the number of apparently conserved water sites will be increased since in the pea lectin structure several of these waters are part of more extended water networks on the surface of the protein.

Implications for Oligosaccharide Recognition. Recently, the structure of a complex between isolectin I from *L. ochrus* and an octasaccharide derived from human lactotransferrin ([β-Gal-(1–4)-β-GlcNac-(1–2)-α-Man-(1–3,6)]$_2$-β-Man-(1–4)β-GlcNac) was described at 2.3-Å resolution (Bourne et al., 1992). This structure will be referred to as LOLI·OCTA. In this structure, an significant number of the polar interactions between the lectin and the octasaccharide are mediated via water bridges. This finding is not surprising since thermodynamic studies on other legume lectins strongly emphasize the importance of water as a mediator in protein–carbohydrate recognition (Lemieux, 1989; Lemieux et al. 1990). Since the lentil and pea lectins have a very similar (if not identical) specificity as the lathyrus lectin, it was expected that at least some of the water sites that are conserved between pea and lentil lectin might have a role in oligosaccharide binding. Only two of the nine residues that are implicated in the water-mediated hydrogen bonding scheme between the octasaccharide and the lectin in the LOLI·OCTA complex have a conserved water ligand. These are Glyβ98 and Thrβ122. However, neither of those two water molecules can be correlated with a water bridge in the LOLI·OCTA structure. Thus the conserved water sites in the lentil lectin structure can be classified as structural and not as functional.

In the pea lectin and LOL I structures, the monosaccharide binding site contained four well-ordered water molecules that mimic some of the key polar hydrogen bonds that occur upon carbohydrate recognition. This situation is not reflected in the lentil lectin structure. Difference maps of the monosaccharide binding site systematically showed a blob of electron density, the center of which was 4 Å away from the nearest protein atom (Figure 9). After carefully inspecting this region, this density was finally identified as a possible acetate ion forming a hydrogen bond with Glyβ99 (N), a key polar grouping in the different LOL I–monosaccharide complexs. This identification is mainly based on the content of the crystallization mixture and the shape of the density. This density is consistently present for both monomers, and its presence is independent of the resolution range that is used for refinement and map calculation. For a final identification, however, a more complete and higher resolution set of data will be necessary.

Of all the LOL I residues that interact directly with the octasaccharide, all but one are conserved in the lentil lectin sequence. The sole residue substituted is Serβ39, which is Asn in both lathyrus and pea lectin. In the LOLI·OCTA structure, OD1 of Asnβ39 forms two hydrogen bonds with GlcNac5, one of the key residues of the octasaccharide that interact strongly with the lectin: a direct hydrogen bond to O5 and a water-mediated hydrogen bond to N2. It is most likely that, in the lentil lectin, OG of Serβ39 takes over this role. The coordinate rms difference around residue β39 between the pea and lentil lectins is very low, and therefore it is more likely that the largest conformational adaptation to this mutation may be found in the oligosaccharide binding to this region. Indeed, saccharide binding outside the monosaccharide binding site in the different LOL I–saccharide structures has proven to be flexible. For example the trisaccharide α-Man-(1–3)-β-Man-(1–4)β-GlcNac adopts a different conformation as part of the LOLI-OCTA structure as compared to LOL complexes with this trisaccharide only (Bourne et al., 1992).

The displacement of the loop Asnβ75–Asnβ78, which in the LOLI·OCTA complex together with loop Asnβ132–Aspβ134 forms a cleft that binds saccharide residues GlcNac5′

and Gal6′ in a partly hydrophobic environment, is likely to be due to an inherent flexibility of this loop as indicated by the high *B* values. The β132–β134 loop on the other hand does not change its backbone conformation when pea and lentil lectin are compared, despite the variability in residue β133 (Gly in LOLI, Lys in lentil lectin and Arg in pea lectin). The presence of the large Lys and Arg side chains at this position in the lentil and pea structures respectively should not strongly affect oligosaccharide binding since in both structures this side chain is facing the solvent away from any bound oligosaccharide.

Except for the *L. ochrus* lectin, the only other mannose/glucose specific lectin for which the protein–sugar interactions in the monosaccharide binding site have been described in detail is Con A (Derewenda et al., 1989). This structure has been compared to the complexes of LOL with mannose and glucose (Bourne et al., 1990). For both proteins, the monosaccharides were found to interact with the protein in a similar way. Differences in fine specificity between LOL and Con A for saccharides with small substituents on O3 were attributed to the presences of two aromatic side chains (Tyrβ100 and Trpβ128) that are expected to make favorable van der Waals contacts with these substituents. The equivalent residues in Con A are two glycines, thus preventing such an hydrophobic interaction. As there are no differences between lentil lectin and LOL in the residues that make up the monosaccharide binding site, these conclusions remain valid for lentil lectin as well.

Conclusions. We described the refined crystal structure of the lentil lectin at 2.3-Å resolution. During our investigations, it was necessary to revise the amino acid sequence of the β chain, which had to be extended to contain 180 amino acids, similar to the other two chain *Viciae* lectins. This structure is the first high-resolution legume lectin structure determined by crystallography that exhibits significant structural differences in the transition metal binding site. A detailed description of the metal binding region is given. A number of conserved water sites was identified. Further comparative studies of the lentil lectin are now possible: beside the orthorhombic crystal form described in this paper, a monoclinic form was also grown under the same conditions. Moreover, preliminary experiments on both crystal forms indicate they are not damaged if soaked with small oligosaccharides, enabling a detailed study of the molecular basis of the carbohydrate specificity of this lectin.

ACKNOWLEDGMENT

We thank Maria Vanderveken for excellent technical assistance. Thanks to Dr. J. Naismith and Dr. J. Helliwell for making available the ConA–Cd coordinates prior to publication. We also thank Dr. G. Reeke for providing the partially refined favin coordinates, Dr. L. Delbaere for providing the GS4 coordinates, and Dr. Y. Bourne for providing the LOL I coordinates. Crystallographic coordinates have been submitted to the Brookhaven Protein Data Bank (1LAL).

REFERENCES

Baker, E. N., & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol. 44*, 97–179.

Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A., & Edelman, G. M. (1975) *J. Biol. Chem. 250*, 1513–1514.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 111*, 535–542.

Bhattacharyya, L., & Brewer, C. F. (1985) *Biochemistry 24*, 4985–4990.

Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rougé, P., & Fontecilla-Camps, J. C. (1990a) *J. Mol. Biol. 214*, 571–584.

Bourne, Y., Roussel, A., Frey, M., Rougé, P., Fontecilla-Camps, J. C., & Cambillau, C. (1990b) *Proteins: Struct., Funct., Genet. 8*, 365–376.

Bourne, Y., Rougé, P., & Cambillau, C. (1992) *J. Biol. Chem. 267*, 197–203.

Brünger, A. T. (1990) *X-PLOR Version 2.1: A System for Crystallography and NMR*, Yale University, New Haven, CT.

Dellaporta, S. L., Wood, J., & Hick, J. B. (1983) *Plant Mol. Biol. Rep. 1*, 19–21.

Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb (Gilboa), A. J., Dodson, Dodson, E. J., Papiz, M. Z., Wan, T., & Campbell, J. (1989) *EMBO J. 8*, 2189–2193.

Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G., & Moss, D. S. (1989) *J. Appl. Crystallogr. 22*, 510–516.

Einspahr, H., Parks, H. E., Suguna, K., Subramanian, E., & Suddath, F. L. (1986) *J. Biol. Chem. 261*, 16518–16527.

Evans, P. R. (1991) in *Crystallographic Computing 5: From Chemistry to Biology* (Moras, D., Podjarny, A. D., & Thierry, J. C., Eds.) pp 136–144, Oxford University Press, Oxford.

Fitzgerald, P. M. D. (1988) *J. Appl. Crystallogr. 21*, 273–278.

Foriers, A., Lebrun, E., Van Rapenbusch, R., De Neve, R., & Strosberg, A. D. (1981) *J. Biol. Chem. 256*, 5550–5560.

Goldstein, I. J., Hughes, R. C., Monsigny, M., Osawa, T., & Sharon, N. (1980) *Nature (London) 285*, 66–66.

Haneef, I., Moss, D. S., Stanford, M. J., & Borkakoti, N. (1985) *Acta Crystallogr. A41*, 426–433.

Hardman, K. D., Agarwal, R. C., & Feisner, M. J. (1984) *J. Mol. Biol. 157*, 69–89.

Howard, A., Gililand, G., Finzel, B., Poulos, T., Ohlendorf, D., & Salemme, F. (1987) *J. Appl. Crystallogr. 20*, 383–387.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Kalb (Gilboa), A. J., Stern, A., & Heald, S. M. (1979) *J. Mol. Biol. 135*, 501–506.

Lemieux, R. U. (1989) *Chem. Soc. Rev. 18*, 347–374.

Lemieux, R. U., Delbaere, L. T. J., Beierbeck, H., & Spohr, U. (1990) Lecture presented at CIBA Foundation Symposium No. 104. Host–guest molecular interactions: from chemistry to biology, July 3–5 1990, London, U.K.

Lin, S.-L., Stern, A., Kalb (Gilboa), A. J., & Zhang, Y. (1990) *Biochemistry 29*, 3599–3603.

Lin, S.-L., Stern, A., Kalb (Gilboa), A. J., & Zhang, Y. (1991) *Biochemistry 30*, 2323–2332.

Lis, H., & Sharon, N. (1986) *Annu. Rev. Biochem. 53*, 35–67.

Loris, R., Lisgarten, J., Maes, D., Pickersgill, R., Korber, F., Reynolds, C., & Wyns, L. (1992) *J. Mol. Biol. 223*, 579–581.

Lowe, D. M., Fersht, A. R., Wilkinson, A. J., Carter, P., & Winter, G. (1985) *Biochemistry 24*, 5106–5109.

Luzzati, V. (1952) *Acta Crystallogr. 5*, 802–810.

Marchetti, P. S., Bhattacharyya, L., Ellis, P. D., & Brewer, C. F. (1988) *J. Magn. Reson. 80*, 417–426.

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963) *J. Mol. Biol. 7*, 95–99.

Reeke, N. R., & Becker, J. W. (1986) *Science 234*, 1108–1111.

Richardson, M., Rougé, P., Sousa-Cavada, B., & Yarwood, A. (1984) *FEBS Lett. 175*, 76–81.

Sanger, F., Nicklen, S., & Coulson, A. R. (1975) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Scharf, S. J. (1990) in *PCR Protocols: A Guide to Methods and Applications* (Innis, D., Gelfand, J., Sninsky, & White, T., Eds.) pp 13–20, Academic Press, New York.

Shaanan, B., Lis, H., & Sharon, N. (1991) *Science 254*, 862–866.

Stubbs, M. E., Pang, H., Dunn, R. J., & Carver, J. P. (1991) Abstract No. 12.30 of the 11th International Symposium on Glycoconjugates. *Glycoconjugate J. 8*, 235–236.

Yarwood, A., Richardson, M., Sousa-Cavada, B., & Rougé, P. (1985) *FEBS Lett. 184*, 104–109.