The Crystal Structure of Pea Lectin at 3.0-Å Resolution*

(Received for publication, March 31, 1986)

Howard Einspahr†
From Physical and Analytical Chemistry, The Upjohn Company, Kalamazoo, Michigan 49001

Elizabeth H. Parks, Kaza Suguna‡, and Eswara Subramanian††
From the Department of Biochemistry, Institute of Dental Research, and Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, Alabama 35294

F. L. Suddath
From the School of Chemistry, Georgia Institute of Technology, Atlanta, Georgia 30332

The structure of pea lectin has been determined to 3.0-Å resolution based on multiple isomorphous replacement phasing to 6.0-Å resolution and a combination of single isomorphous replacement, anomalous scattering, and density modification to 3.0-Å resolution. The pea lectin model has been optimized by restrained least squares refinement against the data between 7.0- and 3.0-Å resolution. The final model at 3.0 Å gives an R factor of 0.24 and a root mean square deviation from ideal bond distances of 0.02 Å. The two monomers in the asymmetric unit are related by non-crystallographic 2-fold symmetry to form a dimer. Monomers were treated independently in modeling and refinement, but are found to be virtually identical at this resolution. The molecular structure of the pea lectin monomer is very similar to that of concanavalin A, the lectin from the jack bean. Similarities extend from secondary and tertiary structures to the occurrence of a cis-peptide bond and the pattern of coordination of the Ca

Among the variety of carbohydrate-binding proteins that have been isolated from a wide range of sources is a group, including both proteins and glycoproteins, that is distinguished solely by in vitro properties. The proteins in this group, the lectins, share the ability to agglutinate cells or precipitate complex carbohydrates (polysaccharides, glycoproteins, glycolipids). Although the lectins are, by definition, not of immune origin, they are characterized by a high degree of specificity in binding to some saccharides in preference to others. Lectins have been isolated from bacteria and molds, lichens, sponges, fish, sera and eggs, animal cell membranes, and, very commonly, from plant seeds. The physiological role of lectins in plants is still debated; however, in the legumes, substantial evidence points to a specific involvement of lectins during infection of plant root hairs by their respective bacterial symbionts, the Rhizobia (1). Rhizobium colonization results in the formation of root nodules that are essential to the conversion of nitrogen into ammonia in the legumes (2). The nitrogen fixed by this symbiotic relationship is of considerable agricultural and economic importance.

Despite the lack of detailed knowledge of the physiological role of lectins in plants, this group of proteins has proved to be exceptionally useful as molecular tools for isolating carbohydrate-containing molecules, for typing blood and for studying the structure of blood group substances, the mitogenic stimulation of lymphocytes, and the agglutination of cells. Immobilized on a support, lectins have been used to isolate cell populations and to classify cells on the basis of their cell surface glycoproteins and glycolipids. A large number of purified plant lectins are commonly available for these purposes. The first lectin obtained in purified form, concanavalin A (ConA), was isolated from the seeds of a legume, the jack bean (Canavalia ensiformis) (3). Concentrated study of this lectin has resulted in determinations of its amino acid sequence (4, 5) and its three-dimensional crystal structure (6-8). In addition, a considerable volume of literature describes the metal ion requirements and carbohydrate-binding properties (9, 10, and references therein) of ConA from both spectroscopic and structural viewpoints.

With few exceptions, all legume seeds so far examined have been found to contain lectins. While these lectins may differ in carbohydrate specificity and quaternary structure, they are found to share a number of biochemical properties. All are composed of monomers of about 25,000 to 30,000 daltons. Some monomers are glycosylated, other are not. Some have subunit structure, others consist of a single chain. All require Ca

†† This work was supported in part by Grants GM-36610, DE-00106, DE-07026, and DE-02670 from United States Public Health Service and equipment Grant PCM-8212839 from the National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

††† The abbreviations used are: ConA, concanavalin A; pCMBS, para-chloromercuribenzenesulfonate; MIR, multiple isomorphous replacement; SIRAS, single isomorphous replacement and anomalous scattering.

16518
been grouped together because they show similar (but not identical) preferences for α-methyl-D-glucose and α-methyl-D-mannose monosaccharides. They differ, however, in subunit structure and quaternary structure. ConA exists at physiological pH as a tetramer of four identical continuous polypeptide chains of 237 amino acid residues each. At lower pH, the tetramers dissociate to form dimers. Pea lectin, lentil lectin, and fava bean monomers are each composed of two different polypeptide chains, an α-chain of about 6,000 daltons and a β-chain of about 18,000 daltons. The quaternary structure of these lectins is dimeric, consisting of two αβ monomers. In addition, favin, unlike the others, is glycosylated.

Recent amino acid sequence data from a number of legume lectins, including lentil (11), pea (12), broadbean (13), soybean (14), sainfoin (15), and French bean (16), confirm an unusual aspect of the relationship among the legume lectin sequences first described by Cunningham and co-workers (13) for the broadbean lectin favin. When the sequences are aligned to reflect similarities, they are seen to fall into two groups that differ strikingly in termini. While there is a relatively strong one-to-one correspondence in sequence throughout the chains, one group, ConA and mucana lectin (17), represents a permutation approximately 180° out of phase with that of all the others, a circular permutation of sequences with respect to amino and carboxyl termini. Fig. 1 illustrates the amino acid sequence permutation of pea lectin and ConA.

Fig. 1. A schematic representation of the circularly permuted amino acid sequences of pea lectin and ConA showing differences in chain termini. The two chains of pea lectin (α and β) are represented by a circular band. A concentric stippled band represents the single chain of ConA positioned to maximize sequence homology with pea lectin.

Sequential analysis of cDNAs transcribed from mRNAs of favin (18) and pea lectin (12) correspond directly to the peptide sequences of these proteins and show that synthesis is initiated at the amino terminus of the β-subunit, continues uninterrupted through the α- and β-subunits, and terminates at the carboxyl terminus of the α-subunit, cleavage of subunits occurring post-translationally. Unexpectedly, the recently determined sequence of the ConA mRNA, rather than corresponding directly to the ConA peptide sequence, shows instead the same permutation as the favin and pea lectin messages and protein sequences (19). Among the paucity of explanations for this peculiarity is the suggestion that the mutation corresponding to the ConA peptide sequence is the result of post-translational ligation of the initial amino and carboxyl termini as well as post-translational cleavage to create the new termini (19). This ligation would occur at a position originally thought to be a site of proteolytic cleavage in ConA that may now be interpreted as a partial failure of ligation.

The unusual biochemical complexities of these related lectins, differences in sequence permutation, differences in quaternary structure, and subtle differences in carbohydrate-binding specificity, can now be examined in terms of the molecular structures of pea lectin and ConA. The structure of pea lectin has been determined at 3.0-Å resolution, an extension of the results reported at 6.0-Å resolution (20). Preliminary accounts of this work have been reported (21, 22). A report of the structure at 5.0 Å has been published by a group at the USSR Academy of Sciences in Moscow (23).

**EXPERIMENTAL PROCEDURES**

**Protein Purification and Crystallization**—The isolation and crystallization of pea lectin have been described previously (20). Crystals of pea lectin used in this study have space group P212121, with unit cell dimensions of a = 50.73(2) Å, b = 61.16(2) Å, and c = 136.59(8) Å. These two pea lectin monomers per asymmetric unit.

**Heavy Atom Deriuatiues**—Conditions for preparation of heavy atom derivatives were screened as described previously (20). The original single-site uranyl nitrate derivative could not be reproduced. Instead, multisite uranyl nitrate derivative crystals with a variety of substitution sites and occupancies were prepared by the following method: native pea lectin crystals were soaked in solutions slowly increased to a uranyl nitrate concentration of 2 mM over a 36-h period, and then left at this final concentration for 12 h more before being mounted for data collection. A second derivative was prepared by soaking native pea lectin crystals for 45 h in 5 mM para-chloromercuribenzenesulfonate (pCMBS) solution in artificial mother liquor.

**Data Collection**—Three-dimensional diffraction data were collected from native and derivative crystals at room temperature on a Picker diffractometer. Data were collected by an ω-step scan procedure in shells of 100 reflections starting at high and going to low resolution. For the 3.0-Å resolution native, the 6.0-Å resolution pCMBS derivative, and the 4.4-Å resolution uranyl nitrate derivative data sets, each reflection was sampled by 1-s counts at each of 37 ω-steps. Steps were spaced 0.02° apart and were symmetrically disposed about calculated peak positions. Data were collected from six additional uranyl nitrate derivative crystals. One uranyl nitrate derivative data set extended to 6.6-Å resolution; the other five data sets spanned the range from 6.0- to 3.0-Å resolution. These data sets were collected by a peak top step-scan method (24). Each reflection was measured for 2 s at each of five to seven steps in ω. Steps were spaced 0.05° to 0.09° apart and were symmetrically disposed about predicted peak positions. Background counts for 1 s were taken at displacements of 0.4° in ω to both sides of predicted peak positions. Throughout the range of each data set, Friedel mates for all unique reflections were collected. Crystal alignment was monitored for each reflection. Seven to nine standard reflections were monitored for decay in intensity every 100 to 200 reflections. No data were included for which decay estimates were greater than 30%.

**Data Processing**—Integrated intensities and variances were determined for the 3.0-Å resolution native, 6.0-Å resolution pCMBS, and 4.4-Å resolution uranyl nitrate data as described previously (20). Integrated intensities and variances for the remaining data were determined by application of a Gaussian-fit method (24). The intensities and variances were corrected for Lorentz and polarization effects. Absorption corrections were applied by the method of North et al. (25). Linear decay corrections were applied to the intensities based on the average decay observed for a subset of the periodically collected standard reflections corresponding to the range of sin ω/λ of the data set. Anomalous differences for the derivatives were computed from the measured intensities of Friedel mates. The uranyl nitrate derivative data sets were not merged together due to differences in substitution sites and occupancies among the crystals from which data were collected. The derivative data were scaled against the native data first with a relative Wilson (26) scaling procedure and then in shells similar to those used for data collection.
Initial Phase Determination—A cross-difference electron density map was calculated based on differences between pCMBS derivative data and the native data together with centrodipole phase angles calculated from the isomorphous and anomalous contributions from the previously determined single-site uranyl nitrate derivative (20). Six substitution sites were found. These sites were consistent with the isomorphous and anomalous difference Patterson maps. Centric least squares refinement (27) yielded a model consisting of four major and two minor sites with an R value (R = \[\sum_{i=1}^{n} |F_i - |F_i|/\sum_{i=1}^{n} |F_i|\]) of 0.49.

A combination of three methods was used to determine the sites of substitution in the 5.6-Å resolution and 4.4-Å resolution derivative data sets. These methods were: 1) interpretations of isomorphous difference Patterson maps; 2) interpretations of cross-difference electron density maps based on centrodipole phase angles derived from the isomorphous and anomalous contributions of the pCMBS and corels; and 3) direct methods (MULTAN) based on isomorphous differences as described by Wilson (28). A total of eight binding sites was found.

The 5.6-Å resolution data set was modeled as three major and four minor sites. Centroid least squares refinement of this model gave an R value of 0.54. The 4.4-Å resolution data set was modeled as two major sites and a single minor site not found in the 5.6-Å set; centrodipole least squares refinement yielded an R value of 0.54. The most prominent site and a single minor site not found in the 5.6-Å set; centrodipole least squares refinement yielded an R value of 0.54. The most prominent site was interpreted as a polypeptide segment, and the minor site, a minor site. After additional cycles of refinement, the R value decreased to 0.54. A hybrid model was calculated based on MIR-SIRAS phases from 15.0 Å to 4.5 Å and newly calculated phases from 4.5 Å to 3.0 Å. The complete main chain for both monomers could be traced in this new map. At various places in the map the disposition of side chains was not clear. Amino acid residues were fit to the electron density by application of the MIR-SIRAS phases to the 6.0-Å data set. When the 6.0-Å data set was re-examined, a minor site was discovered. This site corresponded to a major site common to the 5.6- and 4.4-Å data sets.

The eight different uranyl sites described above were used to model the five partial derivative data sets that spanned the 6.0- to 3.0-Å resolution range. Not all of the sites were found in each data set. Exhaustive centroid least squares refinement yielded models consisting of six to eight sites for the individual data sets with R values ranging from 0.49 to 0.58.

In all cases, only the heavy atom position and occupancy parameters were refined; isotropic temperature factors were fixed at 15.0 Å². Centroid phase angles were calculated by the method of Blow and Crick (29). Anomalous scattering information was incorporated according to the procedure of Matthews (30). Phasing contributions were obtained from all the derivative data that are described above as well as from the 6.0-Å resolution uranyl derivative described previously (20).

The final combined phases of 8776 reflections had a mean figure of merit of 0.69, ranging from 0.90 in the lowest resolution shell to 0.58 at the highest resolution shell. The low figure of merit at higher resolution reflects reduced phasing information in the region of 5.6 to 3.0-Å resolution, where only SIRAS phases were available.

Electron Density Map Interpretation, Model Building, and Refinement—Electron density maps were calculated on a 1-Å grid. The maps were contoured at equal but arbitrary intervals at a scale of approximately 1 e/Å³ and were plotted on acetate film sections perpendicular to the nonecrystallographic 2-fold axis relating the two monomers of the dimer in the asymmetric unit.

An initial interpretation of the chain tracing was made from a map calculated based on the MIR-SIRAS phases. With this interpretation, approximately 60% of the residues of the known sequence of pea lectin (13) could be placed in the electron density. A second tentative chain tracing of approximately 70% of the total dimer was obtained with the GRINCH system of programs (31) located at the University of North Carolina. The GRINCH interpretation was in agreement with and extended the initial map interpretation. Amino acid residues were fit to the electron density by application of the FRODO system of programs (32) on an Evans and Sutherland Multipricture System linked to a VAX 11/750 computer. When regions of density at turns and loops were not clear, a model of these regions was not built. Approximately 80% of the total dimer was interpreted in terms of a detailed model with side chains. Another few percent were interpreted in terms of a polyglycine backbone without side chains. This interpretation resulted in 16 individual polyglycine segments for the dimer.

The initial model building exercise was aided significantly by knowledge of the locations of the 4 Ca²⁺ ions and Mn²⁺ ions in the asymmetric unit. The positions of these ions were determined from film data collected by tuning the wavelength (1.680 Å) close to the MnK edge at 1.896 Å in order to enhance the primary scattering coefficient f°. Anomalous difference Fourier maps were calculated using the partial derivative data between the MIR-SIRAS phases to 3.0 Å as described by Kraut (46). The positions of the Mn²⁺ ions were obvious in these maps and clearly distinguished from the adjacent Ca²⁺ peaks. The positions in fractional coordinates (x, y, z) for monomer A, Mn(1) = (0.38, 0.69, 0.57) and Ca(1) = (0.32, 0.67, 0.57), and for monomer B, Mn(2) = (0.65, 0.40, 0.78) and Ca(2) = (0.71, 0.46, 0.80), with uncertainties of about 0.01. These data were collected at the Daresbury Synchrotron Radiation Source and a comparison of the quality of the anomalous difference maps from film and diffractometer data has been reported (33).

Bricogne’s double-sort technique (34) was used to average the electron density map about the nonecrystallographic molecular 2-fold axis. The averaged map was of lower quality than the MIR-SIRAS map and, therefore, was not used. The reasons for this unsuccessful application of symmetry averaging are not fully understood, but an improvement may have been a contributing factor at the time.

The initial model was subjected to rigid body refinement with CORELS (35). The 16 segments of the partial model were treated as rigid domains in the refinement. In five cycles (20.0 to 6.0 Å data), the R value between Fobs and Fcalc for the partial model was reduced from 0.57 to 0.46 in the sensitive region from 7.0 to 3.0 Å. A hybrid model was calculated based on MIR-SIRAS phases from 15.0 to 4.5 Å and newly calculated phases from 4.5 to 3.0 Å. The complete main chain for both monomers could be traced in this new map. At various places in the map the disposition of side chains was not clear.

The third model was then subjected to restrained least-squares refinement with each amino acid residue taken as a rigid domain. After six cycles, the R value decreased from 0.345 to 0.240 (7.0 to 3.0 Å data). This was followed by two cycles of restrained least-squares refinement using the Hendrickson-Konrath program (36) to improve the geometry of the model. While the root mean square deviation from ideal bond distances decreased from 0.055 to 0.026 Å, the R value increased to 0.392. Phases were calculated from the refined model and were combined with MIR-SIRAS phases (Ref. 37 and references therein) by use of Bricogne’s modification of the Sim weighting scheme (34). The combined phases and observed structure factors were used as input into Wang’s (38) system of programs for density modification. The resultant map was compared to a map calculated with Fobs and combined phases and to a second density-modified map calculated with MIR-SIRAS phases and Fobs. The combined phases density-modified map appeared to be the best of the three maps and was used to rebuild the model. All residues were fitted to the electron density except for residues 225 to 239 in one monomer which were omitted due to the absence of interpretable electron density. (The numbering of pea lectin residues used here is that of Higgins et al. (12) Fig. 4.)

The third model was then subjected to restrained least-squares refinement with each amino acid residue taken as a rigid domain. After six cycles, the R value decreased from 0.40 to 0.27 in eight cycles (7.0 to 3.0 Å data with a 3σ cut off). Four additional cycles of refinement including individual isotropic temperature factors for all atoms resulted in a further decrease in the R factor to 0.22 (5.0 to 3.0 Å data with 3σ cut off). The root mean square deviation from ideal bond distances was 0.021 Å.

Structure factors and phases were calculated from the refined model and used to calculate a new electron density map in which an attempt was made to reduce bias due to the model. For this new map, the electron density was set to zero in a slab corresponding to approximately 10% of the volume of the asymmetric unit. The modified asymmetric unit was then backtransformed to obtain new calculated structure factors and phases. Calculated and observed structure factors were scaled together. Calculated phases were combined with MIR-SIRAS phases as before. The combined phases together with coefficients of w(\[w_{calc} = 1/F_{obs}^2\]) were used to calculate the electron density within the slab described above. This process was repeated with successive slabs of the calculated map until the entire asymmetric unit of the map had been constructed.

The final model was adjusted to fit the electron density in this new map with FRODO. Residues 235 to 239 were not included in the final model in either monomer as the electron density is not interpretable in these regions. The final model was subjected to restrained least-squares refinement in which refinement of thermal parameters was limited to 10% of the volume of the asymmetric unit. In 16 cycles, the R value was reduced from 0.345 to 0.240 (7.0 to 3.0 Å data at 3σ cut off). The root mean square deviation from ideal bond distances is 0.029 Å.
RESULTS AND DISCUSSION

Heavy Atom Substitution Sites—The isomorphous and anomalous contributions from a single-site uranyl nitrate derivative were used to phase the initial 6.0-Å native data set (20). While the molecular boundary of the pea lectin dimer and the β-sheet secondary structure were clearly visible in the 6.0-Å resolution electron density map, it was anticipated that another derivative would be required to determine the structure to 3.0-Å resolution. A pCMBS derivative was prepared and data were collected to 6.0-Å resolution. Six pCMBS sites were located by examining difference Patterson and difference electron density maps. These sites were clustered around the imidazole rings of the 2 His-230 residues in the dimer. The four major binding sites correspond to mercury binding to both ring nitrogens in each monomer. The two minor mercury binding sites appear to be close to alternative locations of the imidazole rings in both monomers.

Seven overlapping uranyl nitrate derivative data sets were collected. Difference Patterson and difference electron density maps as well as direct methods were used to determine the uranyl binding sites for the two complete lower resolution data sets (5.6 and 4.4 Å). A total of eight binding sites were located. Substitution sites and occupancies varied from data set to data set. The two sites in the 6.0-Å data set, which are found in all of the data sets, are related by the molecular 2-fold axis and are located in charged pockets consisting of residues Glu-205, His-51, Arg-55, and Asp-54. Another pair of 2-fold related sites is located near Glu-158 in both monomers. The other four sites are not paired: one is near Asp-134 and Lys-145 in monomer B, a second is near Thr-27 in monomer A, a third is near Lys-153 in monomer B, and the last is located between the two monomers, near Asp-54, Glu-56, and Glu-205 in monomer A and Ser-12 and Lys-10 in monomer B. The uranyl ion has been found to bind primarily to carboxylate groups and occasionally to hydroxyl side chains (39). The binding sites found in pea lectin are consistent with these observations, except for the site near Lys-153B. (The designations A and B are used arbitrarily but consistently to distinguish between the two monomers in the asymmetric unit.)

The Pea Lectin Monomer—The cDNA complementary to pea lectin mRNA has been cloned and sequenced (12) and indicates that pea lectin is synthesized in pre-pro form. During processing to the mature form, a leader sequence is removed co-translationally and the resultant pro-lectin is cleaved post-translationally to yield the β- and α-subunits. During the post-translational cleavage, several residues are apparently removed from the linker region connecting the β- and α-chains. Protein sequencing results indicate that the aminoterminal residue of the α-chain is Val-188 (46), while the carboxyl-terminal residue of the β-chain is Tyr-179. The electron density in this region, however, suggests the presence of Pro-180 and possibly also Asn-181 in both monomers. These residues have not been included in the model in an effort to avoid biasing our interpretation; however, density consistent with extra residues consistently returns in both the combined phase maps and the final map. Another possible post-translational modification is the removal of several amino acid residues from the carboxyl terminus of the α-chain. Protein sequencing results indicate that the α-chain terminates at Ser-239; in the electron density maps no residues are apparent past Ser-234 in either monomer.

The electron density is weak in several other areas of the model. These areas include residues 27-30, 54-58, and 75-77 in both monomers. These areas all consist of turns on the surface of the molecule. Although the density is weak, the path of the polypeptide is clearly outside normal values. Five of these residues are glycines. One residue is Asn-171, which is the fourth residue of a 4-residue turn. The geometry of this turn is similar to that of a class of 4-residue turns described by Sibanda and Thornton (40), in which the conformation at the fourth residue, usually glycine or asparagine, commonly places it in the +/+ quadrant of the Ramachandran plot. Asn-39, Leu-101, Arg-133, and Asn-142 all appear to be clearly defined by the density. The equivalent

Fig. 2. The Ramachandran plots for the two crystallographically independent monomers of pea lectin.
residues to Leu-101 (Leu-230) and Asn-142 (Lys-30) in ConA also have positive $\phi$ values. There are no equivalent residues for Asn-39 and Arg-133 in ConA due to insertions or deletions. The density for Glu-29, the remaining residue with unusual conformation, is very weak in both monomers, and confirmation of torsion angles in this turn region must await extension to higher resolution.

Another unusual conformational feature in pea lectin is the presence of a cis-peptide bond between residues 80 and 81. The cis-peptide bond is also present in ConA between residues 207 and 208 and is thought to be important for the Ca$^{2+}$ binding site as described later. This cis-peptide region was omitted from the early refinement runs, but was built into the model when the density of the combined phase maps indicated it was present. The cis-peptide was restrained in the model when the density of the combined phase maps indicated it was present. The cis-peptide was restrained in the final two rounds of refinement. The electron density in the framework of the molecule. Along the way, extensive random coil region is overlaid onto the front face of the molecule, and the chain of the $\alpha$-subunit winds its way toward termini and as expected these regions differ in the two monomers. The transformation in orthogonal coordinates is:

$$
R = \begin{bmatrix}
-0.9037 & -0.4144 & -0.1076 \\
-0.4019 & 0.7345 & 0.5489 \\
-0.1476 & 0.5374 & -0.8303 \\
\end{bmatrix} T = \begin{bmatrix}
71.877 \\
-24.4330 \\
131.663 \\
\end{bmatrix}
$$

An $\alpha$-carbon tracing of the dimer is shown in Fig. 5.

The two pea lectin monomers of the dimer are joined such that the back $\beta$-sheets of both monomers come together to form a 12-stranded $\beta$-sheet spanning the back of the dimer. Five main chain hydrogen bonds are formed between residues 1 and 9, 3 and 7, and 5 and 5 between monomers. Several other intermonomer contacts that may involve hydrogen bond formation, both side chain to side chain or side chain to main chain, are possible. These contacts include Thr-10 to Thr-9 Oy, Asp-54 O61 to Lys-10 N7, Glu-2 Oel to Gln-15 N2, Trp-206 N2 to Asn-17 O4i, Ser-48 Oy to Tyr-46 OH, Tyr-46 OH to Ser-48 Oy, and Asp-14 O to Try-206 Ne, where the first atom in each contact comes from monomer A and the second from monomer B. Apart from the interaction between residues 46 and 48, these latter contacts do not reflect the molecular 2-fold symmetry. The intermonomer interactions are illustrated in Fig. 6.

Comparison of Pea Lectin and ConA Tertiary and Quaternary Structure—As can be seen in Fig. 7, the structures of pea lectin and ConA are very similar, and both molecules have essentially the same overall $\beta$-sheet structure. However, the structures of the two molecules differ in several of the loops between $\beta$-strands and also in the regions of the carboxyl and amino termini. In the loops that differ between the two molecules, there are either insertions or deletions in pea lectin with respect to ConA. In addition, the amino acid sequence of ConA is circularly permuted with respect to pea lectin and the ConA monomer consists of a single polypeptide chain, while the pea lectin monomer consists of $\alpha$- and $\beta$-chains. Thus pea lectin and ConA have different carboxyl and amino termini and as expected these regions differ in the two molecules. The transformation that superimposes the ConA monomer on the A monomer of pea lectin was calculated by least squares based on 185 eq $\alpha$-carbon atoms plus the Mn$^{2+}$ and Ca$^{2+}$ ions; the root mean square difference between atom
FIG. 4. A diagram indicating the topology of pea lectin (in Richardson's notation: 1x, +1, +9, -7, +1, +1, +3, -1, -1, -5, +9). The NH₂ terminus of the β-chain is represented by the box at the left. The dotted line represents the peptide removed by processing of the pea lectin precursor.

FIG. 5. A stereo pair representing the α-carbon tracing of the pea lectin dimer. The noncrystallographic 2-fold axis is approximately perpendicular to the page. The ribbon portion of the drawing represent β-strands.

FIG. 6. A stereo pair representing preliminary interpretations of the monomer-monomer interactions at the dimer interface.
Fig. 7. Stereo pairs of a pea lectin monomer (top) and a ConA monomer (bottom). Chain termini are numbered, as are the positions on ConA where protein ligation is suggested to occur.

Fig. 8. A stereo pair of the superimposed α-carbon tracings of pea lectin (heavy lines) and ConA (light lines). The dimers are aligned for optimal overlap of corresponding α-carbon positions in the lower monomers.

positions is 0.87 Å. The transformation in orthogonal coordinates is:

\[
R = \begin{bmatrix}
-0.3389 & 0.1639 & 0.9284 \\
-0.3789 & -0.9251 & 0.0251 \\
0.8612 & -0.3425 & 0.3756
\end{bmatrix},
\]

\[
T = \begin{bmatrix}
0.8612 & -0.3425 & 0.3756 & 0.8612 & -0.3425 & 0.3756
\end{bmatrix}
\]

All equivalent α-carbon positions differing by more than 2.5 Å after the initial transformation were eliminated from the final calculation.

It should be noted that all sites suggested to be subjects of post-translational modification lie on the surface of the pea lectin and ConA molecules easily accessible to enzymatic action. In particular, the proposed site of ligation in ConA, in the loop at about position 120, is firmly embedded in the back β-sheet. In pea lectin, the corresponding site, at residues 1 and 234, is not connected by a peptide bond. The close correspondence between pea lectin and ConA structures in this region suggests that, were the ConA chains not ligated, the structure would nevertheless be very similar to that seen with termini lying in close proximity and in a not unfavorable orientation for ligation.

The four monomers in the ConA tetramer are paired (I–II and III–IV) by a special relationship between monomers such that the tetramer appears to be a dimer of ConA dimers (6, 7). The relationship between monomers in the ConA dimer involves a 2-fold symmetric assembly of monomer β-sheets to form a 12-stranded antiparallel dimer-wide sheet. This relationship is similar to that relating monomers in the dimer of pea lectin and, in fact, the structures of the pea lectin and ConA dimers are also very similar.

Based on the same atoms described above from both monomers, the root mean square difference between superimposed
pea lectin and ConA dimers was calculated to be 1.3 Å. The main regions of interaction between monomers in forming the dimer are the same in ConA and pea lectin. These interactions include hydrogen bonding between β-strands (such as residues 1-9: 9-1 in pea lectin, corresponding to 123-131: 131-123 in ConA), close contact of the loop 14-17 in one monomer to the side chain of Trp-206 in the other monomer (136-139 with 88 in ConA), and the close contact between loops 46-48 in the two monomers (176-177 in ConA). However, specific side chain interactions between the monomers are not conserved; in most of these cases, the amino acid residues involved in these interactions are different in pea lectin and ConA.

Pea lectin and ConA differ in that the ConA dimers associate to form tetramers at physiological pH, while pea lectin dimers do not. The 12-stranded β-sheets of the ConA dimers are twisted such that the two dimers are slightly wrapped around each other in the tetramer. This twist is different in the pea lectin dimer, as can be seen in Fig. 8 in which the lower monomers of the two dimers were optimally superimposed, the difference in twist being seen in the upper pair. It is possible that the change in twist is in part responsible for the lack of tetramer formation in pea lectin. It will be interesting to see if the twists of the 12-stranded β-sheets in lentil lectin and favin, neither of which form tetramers, are closer to that in pea lectin than to that in ConA. In addition to the twist of the β-sheets, the contacting residues between the sheets are also expected to be important in the formation of the tetramer. In ConA, there are 54 residues involved in the interactions between monomer I and monomers III and IV. Of these 54 residues, only 12 are conserved in pea lectin, but no interacting pairs are conserved. The same situation is observed for lentil lectin and favin.

Metal-binding Regions—Fig. 9 shows a comparison of the metal-binding loops in ConA and pea lectin. The protein ligands to the Ca²⁺ and Mn²⁺ ions are conserved in the two proteins and, as can be seen in the figure, the two binding
sites appear to be very similar. Glu-119, Asp-121, and Asp-129 are unidentate ligands of the Mn$^{2+}$ ion in pea lectin as are Glu-8, Asp-10, and Asp-19 in ConA. The fourth protein ligand to Mn$^{2+}$ is provided by His-136 in pea lectin and His-24 in ConA. Asp-121 appears to be a bidentate ligand to Ca$^{2+}$ in pea lectin as is Asp-10 in ConA. Phe-123 provides a carbonyl oxygen ligand to Ca$^{2+}$ in pea lectin where Tyr-12 provides this same function in ConA. The remaining two protein ligands to Ca$^{2+}$ in pea lectin are provided by Asn-125 and Asp-129, Asp-129 serving to bridge the metal ions. In ConA, these ligands are provided by Asn-14 and Asp-19. Asp-19 also bridging the metal ions. During the last two rounds of refinement, the metal-ligand distances were very lightly restrained. The overall ligand pattern did not change significantly during refinement. The electron density in the final map indicates that the ligand assignments listed above are correct and are not biased by this restraining of distances during refinement.

In ConA, there are two water ligands for Mn$^{2+}$ and two water ligands for Ca$^{2+}$. At 3.0-Å resolution, it is not possible to tell whether these ligands are also present in pea lectin. In ConA, one of the water ligands to Ca$^{2+}$ is also hydrogen-bonded to O33 and O of Asp-208. The proximity of these two protein atoms to the binding site is a result of the cis-peptide bond between residues 207 and 208. In pea lectin, an equivalent cis-peptide is located between residues 80 and 81 and the carbonyl and carboxylate oxygen atoms of Asp-81 are at the appropriate distances from the Ca$^{2+}$ ion for H bonding to a bridging water molecule.

Carbohydrate Binding Sites—The structure of an α-methyl-D-mannoside ConA complex has been determined to 6-Å resolution (42, 43). The use of iodinated sugars permitted the location of the α-methyl-D-mannose binding site in the vicinity of the side chains of residues Tyr-12, Tyr-100, Asp-16, and Asp-208, some 10–14 Å from the Mn$^{2+}$ position. This structure has been used along with the refined coordinates and binding specificity data to build a model with computer graphics to better define the orientation of the mannose in the binding site. The modeling studies suggest that residues Asn-14, Gly-165, and Asp-208 involve in binding α-methyl-D-mannoside (8, 10). Of these residues, only 3, Asn-125 (Asn-14 of ConA), Gly-216 (Gly-98 of ConA), and Asp-208 (Asp-208 of ConA), are conserved in pea lectin. Studies with α-methyl-D-mannoside-soaked pea lectin crystals have not so far been successful, although x-ray diffraction patterns show them to tolerate soaking in concentrations of sugar as high as 4 M for 24 h. A pea lectin complex with a trimannose has been crystallized and data were collected to 2.5-Å resolution. Preliminary results based on a rotation-translation solution with native pea lectin coordinates indicate that the carbohydrate binds in a similar location to that in ConA, but details of the interaction are not yet available.

Summary—The overall similarity of pea lectin and ConA is readily discernible at 3.0-Å resolution. However, at this resolution it is possible to see differences in the structures that are necessitated by both the two-chain nature of pea lectin and the insertions and deletions that are present with respect to ConA. Most of the insertions and deletions occur in surface loops that allow small changes without substantially altering the course of the polypeptide chain. The ligands surrounding the metal sites are identical to those found in ConA, even though there are minor alterations in the course of the backbone folding near the metal sites. The disposition of the two monomers in the pea lectin dimer suggest a subtle, but significant difference that may contribute to the preservation of the dimeric nature of pea lectin at neutral pH in contrast to ConA, which forms tetramers. The pea lectin structure is likely to be the prototype of all of the presently known two-chain lectins, as well as the single-chain peanut, soybean, and safflow lectins.

High resolution (1.83 Å) native data for pea lectin have been collected. A further understanding of the magnitude and the significance of the differences in the crystallographically independent monomers must await refinement at higher resolution, where details of the metal binding sites and disposition of the water molecules should also be discernible with confidence. The ability of pea lectin crystals to diffract to the unusually high resolution of at least 1.2 Å suggests that many ordered water molecules and strong intermolecular contacts are to be expected.

Acknowledgments—We acknowledge Drs. Charles Bugg, Steve Ealick, and Trevor Greenough at University of Alabama at Birmingham for helpful discussions and advice. The technical assistance of Reid Kerr and Mark Veve is also appreciated.

REFERENCES
1. Wolpert, J. S., and Albersheim, P. (1976) Biochem. Biophys. Res. Commun. 70, 729–737
2. Dilworth, M., and Glenn, A. (1984) Trends Biochem. Sci. 9, 519–522
3. Sumner, J. B. (1919) J. Biol. Chem. 37, 137–143
4. Wang, J. L., Cunningham, B. A., Waxdal, M. J., and Edelman, G. M. (1975) J. Biol. Chem. 250, 1490–1502
5. Cunningham, B. A., Wang, J. L., Waxdal, M. J., and Edelman, G. M. (1975) J. Biol. Chem. 250, 1503–1512
6. Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J., and Wang, J. L. (1977) Proc. Natl. Acad. Sci. U. S. A. 69, 2580–2584
7. Hardman, K. D., and Ainsworth, C. F. (1972) Biochemistry 11, 4910–4919
8. Hardman, K. D., Agarwal, R. C., and Freiser, M. J. (1982) J. Mol. Biol. 157, 89–86
9. Brewer, C. F., Brown, R. D., III, and King, S. H. (1983) J. Bioul. Struct. Dynamics 1, 961–977
10. Carver, J. P., Mackenzie, A. E., and Hardman, K. D. (1985) Biopolymers 24, 49–63
11. Foriers, A., Lebrun, E., Van Rapenbusch, R., deNeve, R., and Strosberg, A. D. (1985) J. Biol. Chem. 256, 5550–5560
12. Higgins, T. J. V., Chandler, P. M., Zurawski, G., Button, S. C., and Spencer, D. (1983) J. Biol. Chem. 258, 9544–9549
13. Hopp, T. P., Hempley, J. J., and Cunningham, B. A. (1982) J. Biol. Chem. 257, 4473–4483
14. Hempley, J. J., and Cunningham, B. A. (1983) Trends Biochem. Sci. 8, 100–102
15. Kouchalakos, R. N., Bates, O. J., Bradshaw, R. A., and Hapner, G. D. (1984) Biochemistry 23, 1024–1830
16. Hoffman, L. M., Ma, Y., and Barker, R. F. (1982) Nucleic Acid Res. 10, 7819–7828
17. Richardson, M., Campos, F. D. A. P., Moreira, R. A., Ainouz, I. L., Begbie, R., Watt, W. B., and Pusztai, A. (1984) Eur. J. Biochem. 144, 101–111
18. Hempley, J. J., Mostov, K. E., and Cunningham, B. A. (1982) J. Biol. Chem. 257, 7903–7909
19. Carrington, D. M., Auffret, A., and Hanke, D. E. (1985) Nature 313, 64–67
20. Meehan, E. J., Jr., McDuffie, J., Einspahr, H., Bugg, C. E., and Suddath, F. L. (1982) J. Biol. Chem. 257, 12275–12282
21. Einspahr, H., Sugana, K., Bugg, C. E., and Suddath, F. L. (1983) Program and Abstracts, American Crystallographic Association Meeting, August 1–5, Snowmass, CO, Abstract PD 15, American Crystallographic Assn., New York
22. Einspahr, H., Sugana, K., Bugg, C. E., and Suddath, F. L. (1984) Collected Abstracts: Thirteenth International Congress of Crystallography, August 9–18, Hamburg, Federal Republic of Ger-

S. Y. Guan, H. Einspahr, and F. L. Suddath, unpublished results.
J. Rini, K. D. Hardman, J. P. Carver, H. Einspahr, E. H. Parks, and F. L. Suddath, unpublished results.
Crystal Structure of Pea Lectin

23. Riskulov, R. R., Dobrokhotova, Z. D., Kuzev, S. V., Lobanov, Yu. D., Lubnin, M. Yu., Mokulskaya, T. D., Myshko, G. E., Proskudina, L. T., Rogacheva, M. M., Saprykina, L. F., Khrenov, A. A., and Mokulskii, M. A. (1984) FEBS Lett. 165, 97–101
24. Hanson, J. C., Watenpaugh, K. D., Sieker, L., and Jensen, L. H. (1979) Acta Cryst. A35, 616–621
25. North, A. C. T., Phillips, D. C., and Mathews, F. S. (1968) Acta Cryst. A24, 351–359
26. Wilson, A. J. C. (1942) Nature 150, 151–152
27. Rossmann, M. G. (1960) Acta Cryst. 13, 221–226
28. Wilson, K. S. (1978) Acta Cryst. B34, 1599–1608
29. Blow, D. M., and Crick, F. H. C. (1959) Acta Cryst. 12, 794–802
30. Matthews, B. W. (1960) Acta Cryst. 12, 794–802
31. Williams, T. V. (1982) Ph.D. thesis, University of North Carolina
32. Jones, T. A. (1978) J. Appl. Cryst. 11, 298–272
33. Einspahr, H., Suguna, K., Suddath, F. L., Ellis, G., Helliwell, J. R., and Papiz, M. Z. (1985) Acta Cryst. B41, 336–341
34. Bricogne, G. (1976) Acta Cryst. A32, 832–847
35. Sussman, J. L., Holbrook, S. R., Church, G. M., and Kim, S. H. (1977) Acta Cryst. A33, 809–804
36. Hendrickson, W. A., and Konnert, J. H. (1980) in Computing in Crystallography (Diamond, R., Ramaseshan, S., and Venkatesan, K., eds) pp. 13.01–13.23, Indian Academy of Sciences, Banjalore
37. Rice, D. W. (1961) Acta Cryst. A37, 491–500
38. Wang, B. C. (1983) Lecture Notes for School on Direct Methods and Macromolecular Crystallography, pp. 1–7, Medical Foundation of Buffalo, Buffalo, NY
39. Blundell, T. L., and Johnson, L. N. (1976) Protein Crystallography, pp. 204, Academic Press Ltd., London
40. Sibanda, B. L., and Thornton, J. M. (1985) Protein Crystallography, pp. 204, Academic Press Ltd., London
41. Richardson, J. S. (1977) Nature 268, 495–500
42. Hardman, K. D., and Ainsworth, C. F. (1976) Biochemistry 15, 1120–1128
43. Becker, J. W., Reeke, G. N., Jr., Cunningham, B. A., and Edelman, G. M. (1976) Nature 259, 406–409
44. Lesk, A. M., and Hardman, K. D. (1982) Science 216, 539–540
45. Kraut, J. (1968) J. Mol. Biol. 35, 511–512
46. Richardson, C., Behnke, W. D., Freisheim, J. H., and Blumenthal, K. M. (1978) Biochem. Biophys. Acta 537, 310–319