Conformation, Protein-Carbohydrate Interactions and a Novel Subunit Association in the Refined Structure of Peanut Lectin-Lactose Complex

R. Banerjee, K. Das, R. Ravishankar, K. Suguna, A. Surolia and M. Vijayan*

*Molecular Biophysics Unit
Indian Institute of Science
Bangalore-560 012, India*

The structure of the complex of the tetrameric peanut lectin with lactose has been refined to an R-value of 16.4% using 2.25 Å resolution X-ray diffraction data. The subunit conformation in the structure is similar to that in other legume lectins except in the loops. It has been shown that in the tertiary structure of legume lectins, the short five-stranded sheet plays a major role in connecting the larger flat six-stranded and curved seven-stranded sheets. Furthermore, the loops that connect the strands at the two ends of the seven-stranded sheet curve toward and interact with each other to produce a second hydrophobic core in addition to the one between the two large sheets. The protein-lactose interactions involve the invariant features observed in other legume lectins in addition to those characteristic of peanut lectin. The "open" quaternary association in peanut lectin is stabilised by hydrophobic, hydrogen-bonded and water-mediated interactions. Contrary to the earlier belief, the structure of peanut lectin demonstrates that the variability in quaternary association in legume lectins, despite all of them having nearly the same tertiary structure, is not necessarily caused by covalently bound carbohydrate. An attempt has been made to provide a structural rationale for this variability, on the basis of buried surface areas during dimerisation. A total of 45 water molecules remain invariant when the hydration shells of the four subunits are compared. A majority of them appear to be involved in stabilising loops.

*Corresponding author

Introduction

Lectins are a group of multivalent carbohydrate-binding proteins which have received considerable attention in recent years on account of their ability to specifically bind different cell-surface carbohydrates (Sharon & Lis, 1989a,b). The first lectin to be X-ray analysed was concanavalin A (con A), isolated from the legume *Canavalia ensiformis* (Hardman & Ainsworth, 1972; Becker *et al.*, 1975). The other lectins subsequently analysed include those from cereal (Wright, 1989), viral (Wilson *et al.*, 1981) and animal (Weis *et al.*, 1991, 1992; Liao *et al.*, 1994; Lobsanov *et al.*, 1993) sources. However, those from legume lectins remain the most extensively studied family of lectins. The legume lectins whose three-dimensional structures have been determined include, apart from con A, pea lectin (Einspahr *et al.*, 1986), favin (Reeke & Becker, 1986), isolectin I from *Lathyrus ochrus* (LOLI; Bourne *et al.*, 1990a), lectin IV from *Griffonia simplicifolia* (GS4; Delbaere *et al.*, 1993), *Erythrina corallodendron* lectin (EcorL; Shaanan *et al.*, 1991), lentil lectin (Loris *et al.*, 1993) and soybean agglutinin (Dessen *et al.*, 1995). They differ in glycosylation and the number of polypeptide chains in each subunit. However, the subunits in all of them have essentially the same tertiary structure, the now well-known legume lectin fold, as is to be expected from the sequence homology (Young & Oomen, 1992) among lectins from

Abbreviations used: PNA, peanut agglutinin; con A, concanavalin A; GS4, lectin IV from *Griffonia simplicifolia*; EcorL, *Erythrina corallodendron* lectin; LOLI, isolectin I from *Lathyrus ochrus*; r.m.s., root-mean-square.

different leguminous sources. All of them are dimers except con A and soybean agglutinin, each being a dimer of a dimer. Con A, pea lectin, favin, LOLI, lentil lectin and soybean agglutinin dimerise in a similar manner. Interactions involving covalently bound sugar have been suggested to be responsible for the other modes of dimerisation found in GS4 and EcorL. The X-ray analysis of several carbohydrate complexes of legume lectins have provided a detailed picture of lectin-carbohydrate interactions (Derewenda *et al.*, 1989; Reeke & Becker, 1986; Rini *et al.*, 1993; Bourne *et al.*, 1990b,c, 1992, 1994; Delbaere *et al.*, 1993; Young & Oomen, 1992; Sharon, 1993; Naismith *et al.*, 1994). Yet another interesting development in recent years is the discovery of the legume lectin fold in animal lectins (Liao *et al.*, 1994; Lobsanov *et al.*, 1993) and other unrelated proteins (Crennell *et al.*, 1994; Emsley *et al.*, 1994).

Peanut (*Arachis hypogaea*) agglutinin (PNA) is a homotetrameric non-glycosylated protein, M_r 110,000, with specificity for the tumour-associated T-antigenic disaccharide Galβ1,3GalNAc. The single chain in each subunit is 236 amino acid residues long and is homologous to the subunits in other legume lectins (Young & Oomen, 1992). The protein in association with lactose was first crystallised at neutral pH in an orthorhombic form (Salunke *et al.*, 1982) and then at acidic pH in two monoclinic forms and one triclinic form (Salunke *et al.*, 1983). All of them contained a tetramer in the asymmetric unit. The structure of the orthorhombic form was solved by the multiple isomorphous replacement method. A preliminary account of the

structure at 2.95 Å resolution based on α-carbon positions and with almost exclusive emphasis on quaternary association has already been reported (Banerjee *et al.*, 1994). PNA has an unusual, open quaternary structure, which, among other things, explains the failure of attempts at structure solution using molecular replacement. The structure was subsequently refined at 2.25 Å resolution. The refined structure, presented here, provides detailed information on the tertiary structure of the protein with emphasis on those features of the legume lectin fold which have not been sufficiently well appreciated, PNA-lactose interactions, subunit association and the water structure associated with the protein. A possible rationalisation of the unanticipated variability in subunit association in legume lectins is also attempted here.

Results and Discussion

Accuracy, conformation angles, displacement parameters and hydrogen bonds

The Luzzati plot (Luzzati, 1952) calculated at the end of the refinement indicated a coordinate error of 0.12 Å in atomic positions. The Ramachandran plot (Ramachandran & Sasisekharan, 1968) corresponding to the structure is given in Figure 1. The φ, ψ values of all the residues are in the allowed region. The side-chain conformational angles by and large follow the expected distribution (Bhat *et al.*, 1979; McGregor *et al.*, 1987). The maximum, the minimum and the average B-values for main-chain atoms are 63.2, 5.2 and 20.5 Å^2 respectively. The
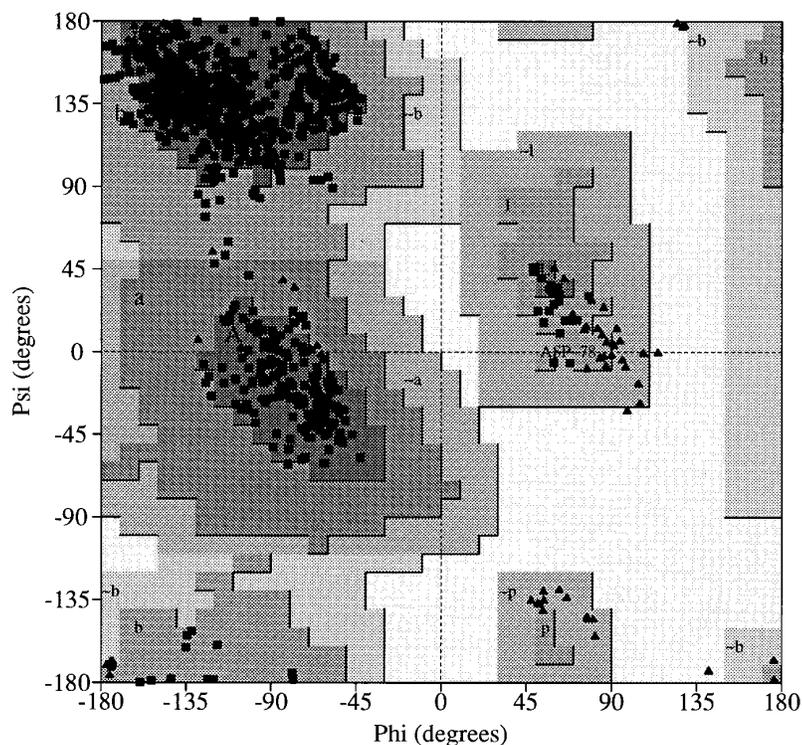
Figure 1. Ramachandran plot produced with PROCHECK (Laskowski *et al.*, 1993). The squares and triangles represent non-glycine and glycine residues respectively.
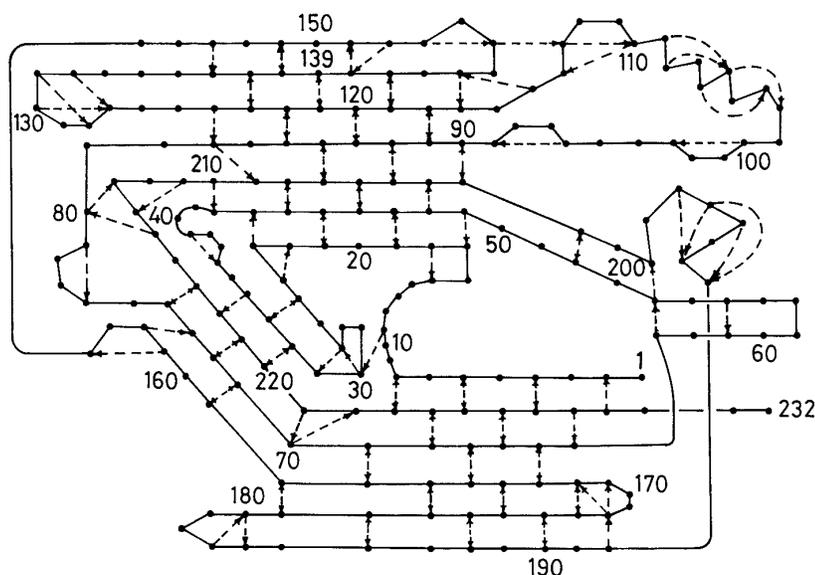
Figure 2. Schematic representation of the main-chain to main-chain hydrogen bonds present in all subunits. A hydrogen bond which connects Gly103 N and Gly210 O is not shown in the Figure.

corresponding values for side-chain atoms, atoms in the lactose molecules and water molecule oxygen atoms are 68.1, 2.0 and 21.8; 57.9, 17.4 and 36.5; and 65.5, 5.8 and 36.8 Å^2 respectively. Hydrogen bonds have been analysed using the program CONTACT in the CCP4 package (SERC Daresbury Laboratory, 1994). Interactions between the protein atoms are considered as hydrogen bonds if the distance between the donor and the acceptor is less than 3.5 Å and if the following angle criterion is satisfied; in the case of hydrogen bonds where the nitrogen atom is the donor, the cut-off for the N-H...O angle has been taken as 120°. When an oxygen atom is the donor, an angle cut-off of 90° has been taken for the C-O...O angle. When water molecule oxygen atoms are involved, only a distance cut-off of 3.6 Å has been used.

Secondary and tertiary structure

The four subunits in the molecule have essentially the same tertiary structure. The root-mean-square (r.m.s.) deviation in α-carbon positions between pairs of subunits when they are superposed (Rossmann & Argos, 1976) range from 0.24 to 0.41 Å. The corresponding values for all main-chain atoms are 0.30 and 0.48 Å respectively. Figure 2 schematically illustrates the main-chain to main-chain hydrogen bonds present in all the four subunits in addition to indicating the topological relationship among the different segments of the polypeptide chain. As expected, the tertiary fold in each subunit is very similar to that in other legume lectins (Table 1; Figure 3), except in relation to the loops. This jelly roll lectin fold, in its essentials, has been observed in several other proteins (Liao *et al.*, 1994; Crennell *et al.*, 1994; Lobsanov *et al.*, 1993; Emsley *et al.*, 1994) and has therefore assumed added importance. Yet some of the features of this fold are not sufficiently appreciated. The emphasis here is on these features.

As in other legume lectins, the framework of the molecule (Figures 2 and 3) consists of three sheets, a six-stranded flat sheet (residues 2 to 8, 64 to 70, 162 to 168, 173 to 179, 186 to 192 and 222 to 229) at the back of the subunit (sheet 1), a seven-stranded curved sheet (18 to 23, 42 to 49, 84 to 90, 117 to 124, 136 to 143, 149 to 153 and 203 to 210) towards the front of the subunit (sheet 2) and a small five-stranded sheet (25 to 27, 31 to 34, 71 to 74, 160 to 161 and 217 to 220), sheet 3, which has a major role in holding the two larger sheets together. This role of the third sheet has not so far been clearly enunciated even in the rare instances where its existence has been acknowledged.

Loops make up 54% of structure. Loops 91 to 116, 125 to 135 and 144 to 148 connect adjacent strands in sheet 2, the first and the third at the bottom and the second at the top. Loop 91 to 116 and to some extent loop 144 to 148 curve up and loop 125 to 135 curves down in such a way that they interact among themselves and with the curved sheet 2 (Figure 4). These interactions involve main chain as well as side-chain atoms. The arrangement of the loops is such that it gives rise to a second hydrophobic core between sheet 2 and the loops in addition to that between sheets 1 and 2 (Figure 5).

Table 1. Superposition of the monomers of a few typical legume lectins

	GS4 (Å)	Con A (Å)	Pea (Å)	EcorL (Å)
PNA	1.47 (0.83)	1.91 (1.02)	1.64 (0.64)	1.77 (0.68)
GS4		1.64 (1.04)	1.71 (0.66)	2.04 (0.85)
Con A			1.81 (0.89)	1.85 (0.94)
Pea				1.42 (0.49)

Table shows the r.m.s. deviations (Å) in C$^\alpha$ positions. The values given in parentheses are the r.m.s. deviations in C$^\alpha$ positions when only the β-sheet regions are superimposed.
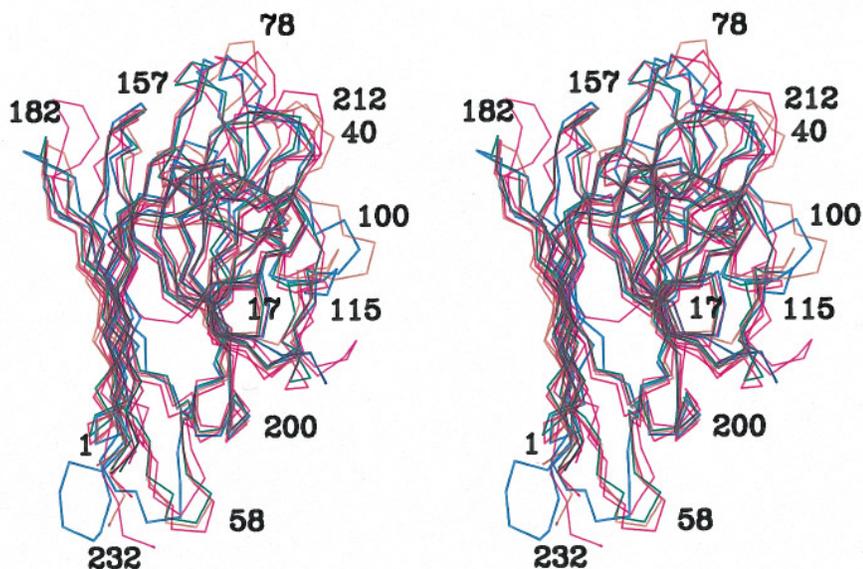
Figure 3. Stereo view of the superposition of the subunits of con A (blue), pea lectin (green), EcorL (red) and GS4 (pink) on to PNA subunit 1 (brown). The residue numbering corresponds to the PNA sequence.

Metal binding

As in the case of other legume lectins, each monomer of PNA contains one calcium ion and one manganese ion. The details of metal coordination are almost identical in the four subunits, which in turn are nearly the same as in other legume lectins. This is hardly surprising as the metal-binding region is totally conserved in these lectins. Asp123 OD1, Asp123 OD2, Tyr125 O, Asn127 OD1, Asp132 OD2 and two water molecules coordinate to the calcium ion while Glu121 OE2, Asp123 OD2, Asp132 OD1, His137 NE2 and two other water molecules coordinate to the manganese ion. The calcium-manganese distances vary between 4.13 and 4.39 Å in the four subunits.

Lectin-carbohydrate interactions

It is well known (Young & Oomen, 1992) that residues in four loops, 91 to 106, 125 to 135, 75 to 83 and 211 to 216, make up the carbohydrate binding pocket towards the top right of the subunits in legume lectins. The electron density corresponding to the lactose molecule is well defined in the present structure (Figure 6) and it permits a detailed analysis of the protein-sugar interactions illustrated in Figure 7(a). The hydrogen bonds involved and the residues which are within a distance of 4 Å from any sugar atom are listed in Table 2. The most important invariant loci for interactions with carbohydrate in almost all legume lectins are (in the PNA numbering scheme) Asp83 OD1, Asp83 OD2, Gly104 N and Asn127 ND2. The mutual disposition of these atoms are also the same in all of them. For example, the constellation of these four atoms superpose with an r.m.s. deviation of 0.23 Å for pea lectin and PNA and 0.31 Å for EcorL and PNA. Another invariant feature of sugar-binding sites in legume lectins (Sharon, 1993; Vyas, 1991) is the presence of an aromatic residue (Phe or Tyr) at the site corresponding to Tyr125 in PNA, which stacks against the sugar ring. Ala82 is yet another invariant residue. These invariant features belong to
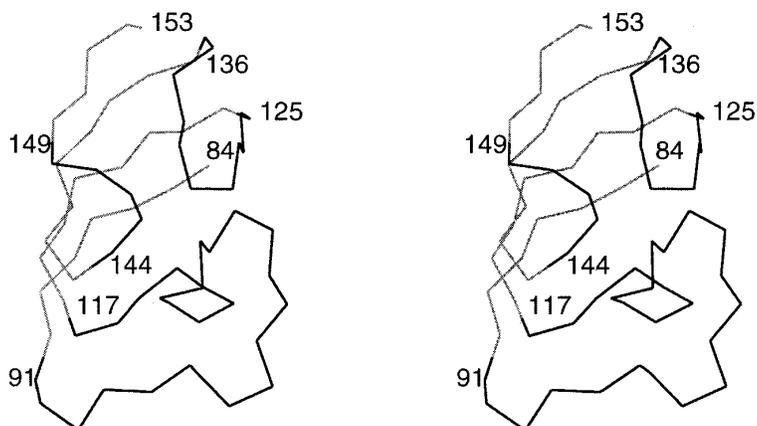
Figure 4. Stereo view of the front sheet (represented by grey lines) and the loops (black lines) that connect the strands in the sheet.
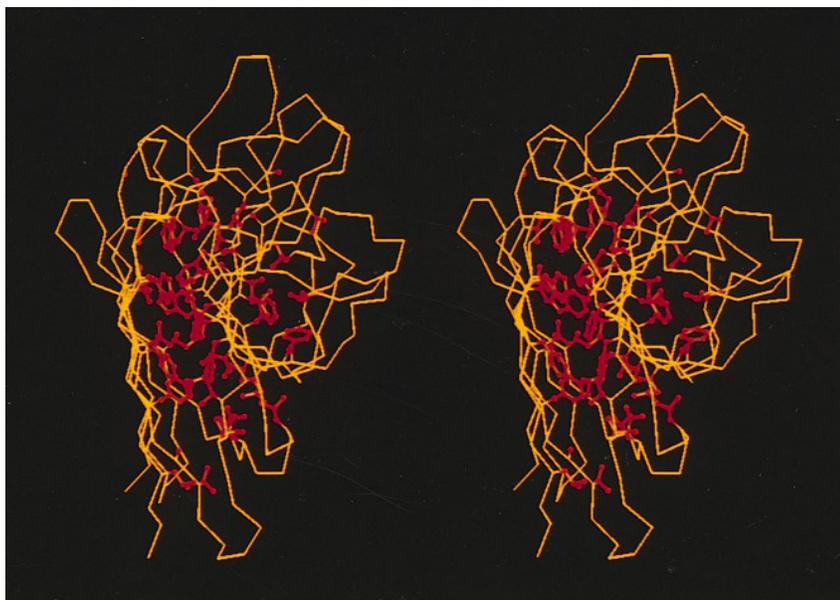
Figure 5. Stereo view of the buried hydrophobic residues (red) in the subunit with the main-chain shown in yellow.

the first three loops mentioned earlier while interactions involving loop 211 to 216, which exhibit considerable variability, appear to be primarily responsible for the sugar specificity of each legume lectin. Indeed, this loop is highly variable in terms of length, sequence and conformation.

In addition to the invariant hydrogen bonds, the galactose ring in lactose is involved in two more hydrogen bonds with PNA, O4 with Ser211 OG and O6 with Asp80 OD2. Ser211 OG interacts with O3 in the glucose ring also. Yet another interaction with this ring involves O3 and Gly213 N. Two water molecules are involved in bridging the carbohydrate with the protein molecule connecting O2 of galactose to Gly104 N as well as Glu129 OE2 (Figure 7(a); Table 2).

It is instructive to compare PNA-lactose interactions with those in the lactose complex of the galactose-binding protein EcorL. A superposition of the corresponding regions in the two complexes is shown in Figure 7(b). The sugar molecules have indeed nearly the same position and orientation in the two complexes. In addition to the invariant hydrogen bonds, O4 of the galactose ring is

hydrogen bonded to Ala218 N and O6 to Gln219 NE2 in the EcorL complex. Gln219 is also hydrogen bonded to O3 in the glucose ring. The other residues in EcorL which are less than 4 Å from the sugar atoms are Tyr106, Phe131 and Ala88. Phe131 in EcorL corresponds to Tyr125 in PNA. The same is true for Ala88 in EcorL and Ala82 in PNA. Tyr106 in EcorL and Gly103 in PNA occur at equivalent positions. Ala218 in EcorL corresponds to Ser211 in PNA. While the peptide nitrogen atom of the former hydrogen bonds to O4 in the galactose ring in the EcorL complex, the side-chain OG of the latter hydrogen bonds to the same sugar atom in the PNA complex. The residues which hydrogen bond to O6 in the galactose ring (Asp80 in PNA and Gln219 in EcorL) are not equivalent. However, in the three-dimensional structure of PNA, Asp80 and Gly214, which correspond in sequence to Gln219 in EcorL, are close to each other. It may also be noted that Gly214 is within 4 Å of the sugar molecule. Thus, the geometry of the lectin-sugar interaction is very similar in PNA and EcorL although differences in detail exist. The difference in the conformation of the fourth sugar-binding loop (Figure 7(b)) is
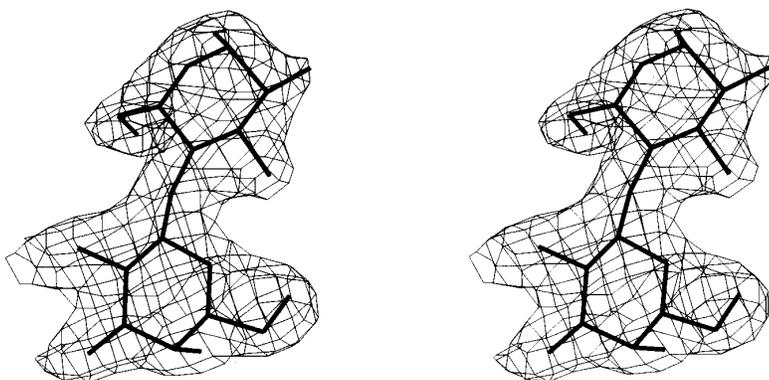
Figure 6. Stereo view of the electron density corresponding to the lactose molecule in subunit 1.
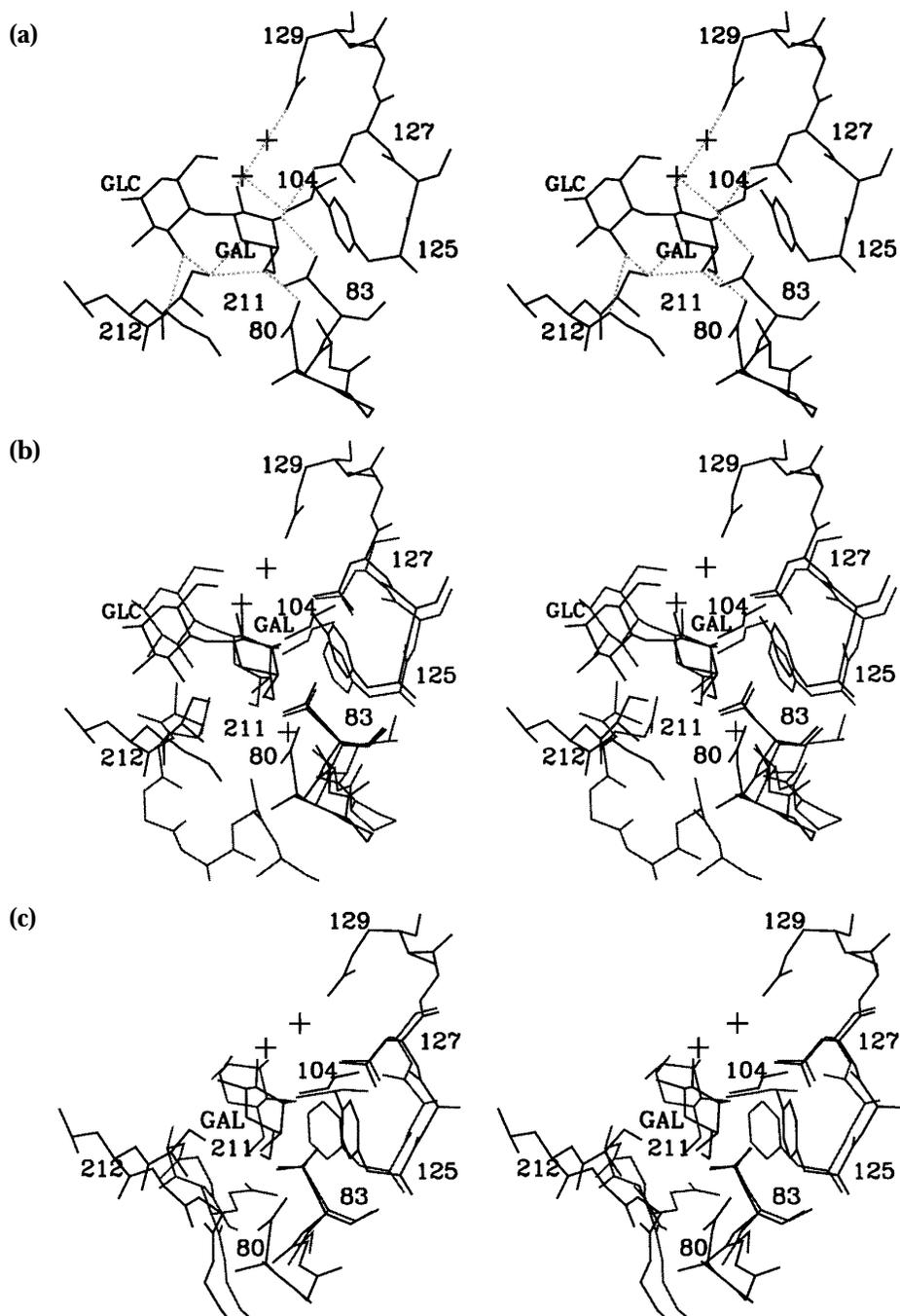
Figure 7. (a) Stereo view of the protein-lactose interactions in PNA. The crosses represent water molecules and the broken lines hydrogen bonds. Superposition of the carbohydrate binding region in PNA (dark lines) and that in (b) EcorL and (c) pea lectin.

particularly noteworthy. Also the water bridge in EcorL between the sugar and the protein molecules does not bear any relation to those in PNA.

As illustrated in Figure 7(c), which shows the superposition of the sugar binding regions of PNA and pea lectin along with the respective sugars, the similarity between the geometry of lectin-carbohydrate interaction in PNA and that in mannose-binding lectins is much less striking. The invariant features in the carbohydrate-binding region referred to earlier constitute the only element of similarity.

The position of the sugar ring itself is substantially different. The relation between the sugar rings in the two structures involves a rotation of about 22° and translation of about 2.25 Å with respect to an internal coordinate system involving 83 OD1, 83 OD2, 104 N and 127 ND2.

Asymmetry in quaternary association

The symmetry principles underlying the association of subunits in multimeric proteins were first

Table 2. Protein-sugar interactions

Protein atom	Sugar atom	Distance (Å) in			
		Subunit 1	Subunit 2	Subunit 3	Subunit 4
A. *Hydrogen bonds*					
Asp83 OD1	Galactose O3	2.67	2.43	2.58	2.49
Gly104 N	Galactose O3	3.08	2.92	3.07	2.81
Asn127 ND2	Galactose O3	2.86	3.00	3.12	3.02
Asp83 OD2	Galactose O4	2.59	2.68	2.64	2.55
Ser211 OG	Galactose O4	2.62	2.82	2.66	2.76
Ser211 OG	Galactose O5	3.12	3.34	3.09	3.16
Asp80 OD2	Galactose O6	3.33	3.39	3.36	2.98
Ser211 OG	Glucose O3	3.34	3.31	3.58	2.98
Gly213 N	Glucose O3	2.92	2.98	3.29	3.28

B. *Water bridges*
Galactose O2 is connected to N Gly104 and OE2 Glu129 through water bridges

C. *Residues within 4 Å from the sugar*
Asp80, Ala82, Asp83, Gly103, Gly104, Tyr125, Asn127, Ser211, Leu212 and Gly214

clearly enunciated by Monod *et al.* (1965). A closed structure is possible only if the subunit assembly has a point group symmetry. Monod *et al.* had originally suggested that only those point groups which lead to isologous association are favoured. Subsequent work by Cornish-Bowden & Koshland (1971) indicated that heterologous association is also possible. Thus the only possible symmetries for a tetrameric protein with identical subunits are 222 with a maximum of three or a minimum of two types of isologous interfaces, and a 4-fold axis with a single type of heterologous interface repeated four times (Klotz *et al.*, 1975). The crystal structures of a large number of multimeric proteins are now available (Bernstein *et al.*, 1977) and they conform to the symmetry principles to a surprisingly high degree.

The most interesting aspect of the quaternary structure of peanut lectin, illustrated in Figure 8, is

Figure 8. Quaternary structure of PNA. The lactose molecules bound to the subunits are also shown. Prepared with RIBBONS (Carson, 1992).
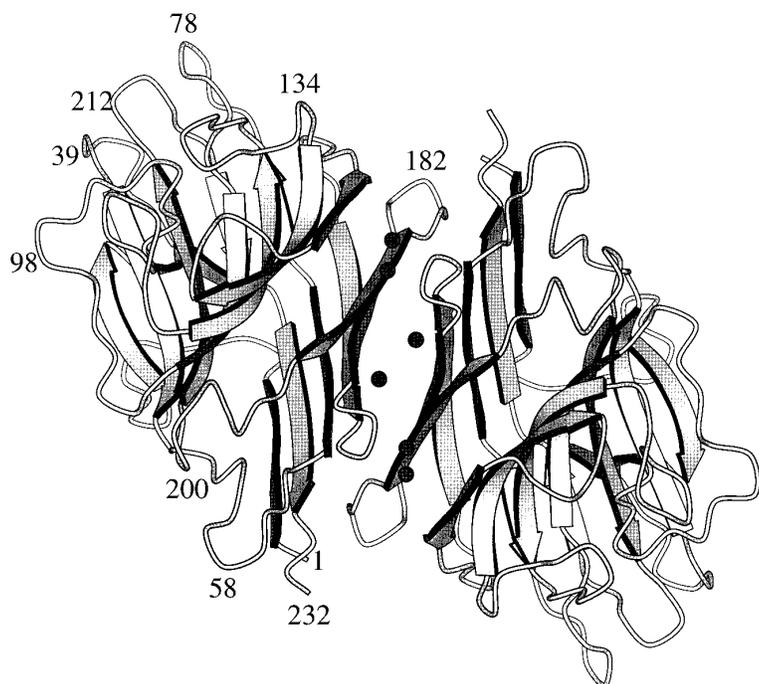
Figure 9. The PNA dimer (subunits 1 and 4). The water molecules (represented by small filled circles in this and the subsequent diagrams) at the interface are also shown. Prepared with MOLSCRIPT (Kraulis, 1991).

that it has neither 222 nor 4-fold symmetry (Banerjee *et al.*, 1994). One half of the molecule, consisting of subunits 1 and 4, is related to the other half, subunits 2 and 3, by a molecular dyad P. Subunits 1 and 4 are related by a local 2-fold axis R1 while 2 and 3 are related by another local 2-fold axis R2. R1 and R2 are inclined with respect to P by 73° and −73° and are skewed by 12 Å on either side of P. The combination of P, R1 and R2 gives rise to an irrational screw axis perpendicular to and passing through all three. It relates subunits 3 and 1 (146° and 24 Å) and 4 and 2 (−146° and −24 Å). The tetramer has three subunit interfaces: that between 1 and 2, that between 1 and 4 which is related to that between 2 and 3 by the molecular dyad, and that between 3 and 4. The binding region corresponding to the 3-4 interface is vacant in subunits 1 and 2 while that corresponding to 1-2 is vacant in subunits 3 and 4. Thus the tetramer has an open structure.

The best characterised example of a relatively open structure, and perhaps the only one observed so far, is that of a dimeric form of hexokinase (Steitz *et al.*, 1976) in which the two identical monomers are related by a rotation of 156° and a translation of 13.8 Å, thus leaving part of the binding region exposed. Further association is prevented as the exposed binding regions in the dimer are not sterically accessible for interactions with other monomers. The possibility of steric factors being responsible, as in the case of hexokinase, for the prevention of polymerisation of the open structure of PNA was explored. One subunit was attached through model building to each subunit such that all four vacant binding sites are appropriately occupied. The addition of these four subunits to the tetramer did not lead to any serious steric problems.

Thus, the absence of polymerisation of PNA molecules, despite their open structure, cannot be readily explained in terms of simple steric considerations.

Dimerisation

It has been conclusively established that the PNA molecule is a very stable tetramer at physiological pH even at as low a concentration as 2 μg/ml, with a stability higher than that of other well-known tetrameric proteins such as con A and haemoglobin (Lotan *et al.*, 1975; Fish *et al.*, 1978; Decastel *et al.*, 1985; Senear & Teller, 1981a,b; Ip & Ackers, 1977; Banerjee *et al.*, 1994). The tetramer, however, dissociates into dimers at low pH (Fish *et al.*, 1978; Decastel *et al.*, 1985). Chemical cross-linking studies have shown that even at neutral pH the molecule exists as a dimer of a dimer (Salunke *et al.*, 1985). Based on structural arguments, it was concluded that subunits 1 and 4, and subunits 2 and 3 constitute the natural dimers in the tetramer (Banerjee *et al.*, 1994). An analysis of intersubunit interactions, outlined below, based on high-resolution refinement confirmed this conclusion. Therefore, 1-4 and 2-3 will be referred to as PNA dimers where appropriate in further discussion.

As illustrated in Figure 9, the formation of the PNA dimer involves the back-to-back association of monomers through the flat β-sheet. The β-sheets from the two monomers are inclined with respect to each other by 86° (Figure 10) and have extensive contacts between them. These contacts involve as many as 22 residues (1, 3, 64 to 66, 149, 164, 166, 168, 171, 173, 175, 177, 179, 183 to 185, 187 to 189, 226 and 228) distributed over all six strands of the sheets. Residues with small hydrophilic side-chains (eight
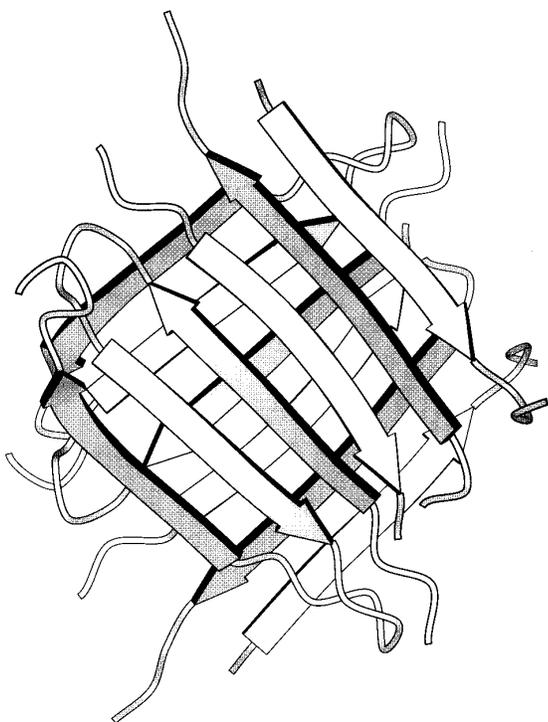
Figure 10. The disposition of the two flat β sheets in the inter-subunit interface in the dimer.

threonine residues, two serine residues and two aspartic acid residues) account for more than half of them; of the remaining ten, one is a glycine, three are alanine residues, three are isoleucine residues, and one each of leucine, phenylalanine and lysine. Thus, residues with comparatively short side-chains dominate the monomer-monomer interface. There are about 90 contacts with interatomic distances of less than 4 Å (89 between subunits 1 and 4 and 91 between 2 and 3) between the monomers. Of these, 11 may be described as hydrogen bonds (Thr3 OG1...OD1 Asp184, Ser64 OG...OG1 Thr187, Thr187 OG1...OD2 Asp168, Ile185 N...OG1 Thr228 and Ile188 N...OD1 Asp168 and their 2-fold equivalents and Ser175 OG...OG Ser175). Furthermore, there are bridges involving six water molecules in the 1-4 dimer and nine in the 2-3 dimer. Calculation of the accessible surface area using the algorithm of Lee & Richards (1971) showed that dimerisation results in the burial of a 1920 Å^2 surface area of which 71% is non-polar. Thus, dimerisation results from substantial hydrophobic interactions, hydrogen bonds and water bridges.

Association of dimers in the tetramer

The association between the two dimers in the tetrameric assembly primarily involves two interfaces, namely those between subunits 1 and 2, and between subunits 3 and 4. The 1-2 interface brings about, as illustrated in Figure 11(a), the side-by-side alignment of the two flat β-sheets. The 65

interatomic contacts involve residues 2 and 3 from the first strand in the flat β- sheet, 231 from the second strand of the same sheet, 48 from the second strand of the curved sheet, and 12 to 17, 48, 51, 53 and 201 from the loops that connect the two sheets. They include side-chain to side-chain hydrogen bonds between Ser12 and Glu2, and between Arg53 and Glu13. The residues at the interface have a mixed composition with one each of glycine, alanine, threonine, asparagine and tyrosine residues, and two each of proline, serine, glutamic acid and arginine residues. The surface area buried in the 1-2 interface is rather low at 1385 Å^2 with a non-polar component of 50%. The most remarkable feature of this interface is the large number of water bridges between the monomers. As many as 19 water molecules participate in these bridges. Quite remarkably, as illustrated in Figure 11(b), all the peptide CO and NH groups of residues 3, 5 and 7 in the first strand of the flat sheet are involved in these bridges. In addition to these residues, many in the 9 to 17 loop and residues 51, 53, 201 and 231 take part in the water bridges.

The second dimer-dimer interface (Figure 11(c)) brings together the top regions of subunits 3 and 4, and is made up of several residues in sheet 3 (Leu27, Asn31, Gln33, Glu72, Lys74, Val160, Ile217 and Leu219), in a turn that connects two adjacent strands in this sheet (Ser28 and Asn29) and in the loops at the top of the subunit (Asn9, Ser10, Leu37, Lys77 and Gly158) and Arg221. The 58 interatomic contacts between the subunits include six hydrogen bonds (Lys74 Nz...O Asn9, Gln33 NE2...OG Ser28, Lys74 Nz...O Asn29, Arg221 NH2...OE2 Glu72, Arg221 NH1...OE1 Glu72 and Arg221 NH2...O Gly158) and their 2-fold equivalents. Five of these six involve the side-chains of Lys74 or Arg221. In addition, nine water molecules bridge the two subunits. The surface area buried in the association between subunits 3 and 4 is also low at 1246 Å^2 of which 39% is non-polar.

Variability in the quaternary association of legume lectins

All legume lectins studied so far are dimers or tetramers made up of dimers. Thus the first or the only step in subunit association in all of them is dimerisation. The dimerisation in con A, the most thoroughly studied of the proteins in this family, involves the anti-parallel side-by-side alignment of the two flat β-sheets leading to the formation of a contiguous 12-stranded β-sheet, six strands coming from each subunit (Hardman & Ainsworth, 1972; Reeke *et al.*, 1975). The same mode of dimerisation exists in pea lectin (Einspahr *et al.*, 1986), favin (Reeke & Becker, 1986), LOLI (Bourne *et al.*, 1990a), lentil lectin (Loris *et al.*, 1993) and soybean agglutinin (Dessen *et al.*, 1995). Different modes of dimerisation were observed in GS4 (Delbaere *et al.*, 1990) and EcorL (Shaanan *et al.*, 1991), although both involve the flat β-sheet. Interactions involving
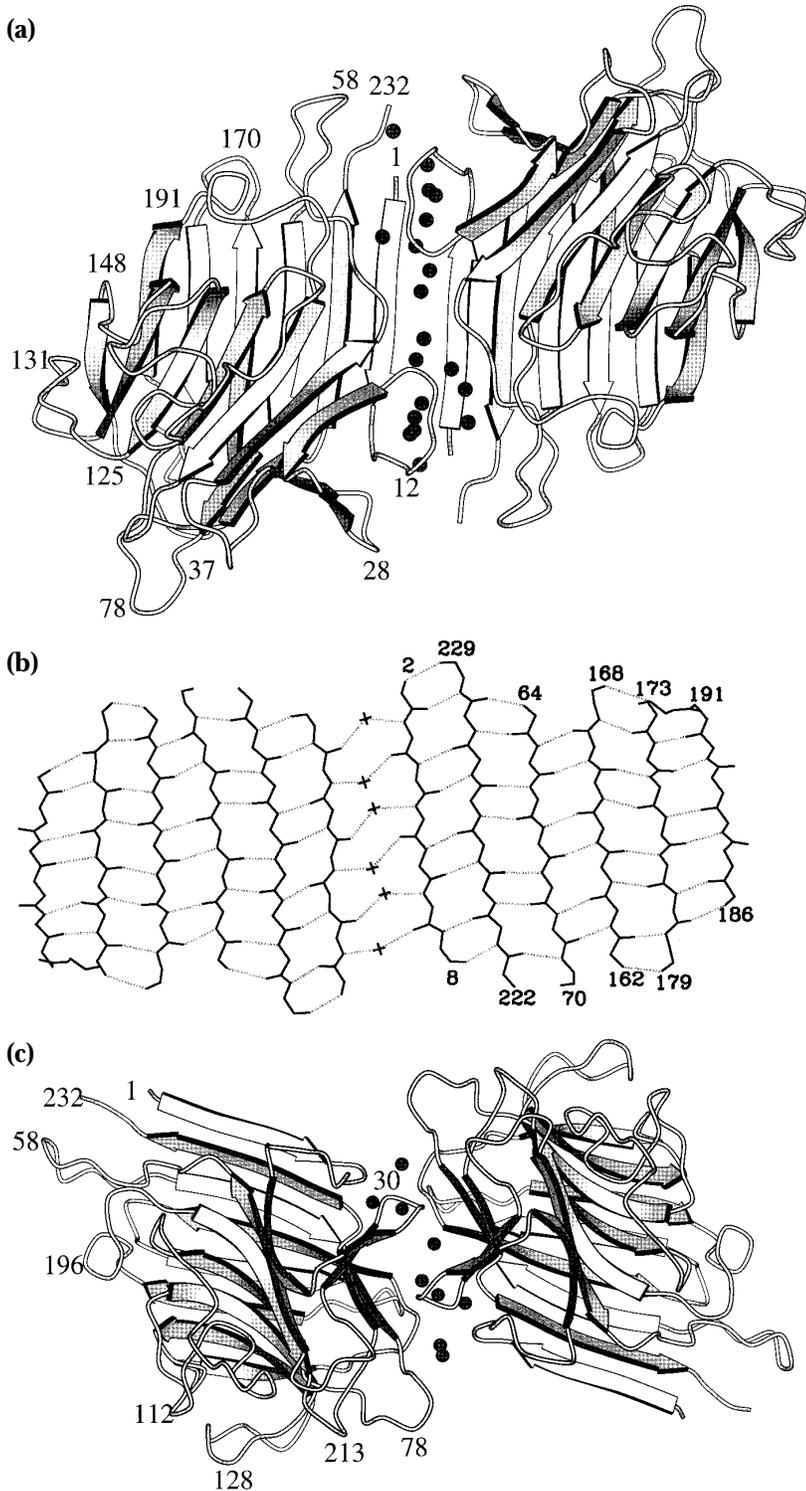
Figure 11. (a) 1-2 interface in the interaction between two dimers in the PNA tetramer and (b) the water bridges that connect the two flat sheets in this interface. (c) The 3-4 interface.

covalently bound sugar were suggested to be responsible for the non-occurrence of the canonical mode observed in con A in these two lectins. Thus, according to the accepted wisdom, legume lectins have an intrinsic propensity to form canonical dimers and interactions involving covalently linked sugar are responsible for the occurrence of other modes of dimerisation.

PNA is not a glycoprotein and obviously interactions involving covalently linked carbohydrate cannot play any role in subunit association in it. Yet dimerisation in PNA is very similar to that in GS4 with a back-to-back association of the two subunits. Furthermore, the 1-2 interface in PNA has a close resemblance to the monomer-monomer interface in the canonical con A dimer. In both cases,

the flat β-sheets align in an antiparallel side-by-side fashion. In con A, they join to form a 12-stranded contiguous antiparallel sheet while in the 1-2 interface in PNA, main-chain to main-chain hydrogen bonds between the two six-stranded sheets are replaced by water bridges. The 3-4 interface in PNA has so far not been observed in any other legume lectin. However, the occurrence in PNA of a GS4-type dimerisation and the presence of an interface of the type that occurs in the canonical legume lectin dimer, demonstrate that the variability in quaternary association is not necessarily caused by interactions involving covalently bound sugar; the origin of the variability lies in the protein itself.

The reason for the variability in the quaternary structure of legume lectins despite all of them having essentially the same tertiary structure, is not immediately obvious. It appeared worthwhile to explore this problem by comparing subunit association in PNA and con A in some detail, for the following reasons. Both are tetrameric lectins of known three-dimensional structure. Both have single-chain subunits although the chain termini in the two are related by circular permutation. Neither contains covalently linked sugar.

The first step in the comparison consisted of the construction, by appropriate superposition of a con A-type tetramer using PNA subunits and a PNA-type tetramer using con A subunits. The resulting models did not contain prohibitively large steric interactions. Therefore, the differences in the subunit arrangements in the two lectins did not appear to be explainable in terms of simple large-scale steric interactions alone. Subsequent efforts were concentrated at the dimeric level. Specifically, four different models were carefully examined: PNA dimer, con A dimer, PNA dimer constructed with con A subunits (con A in PNA dimer) and con A dimer constructed with PNA subunits (PNA in con A dimer). All four models were optimised through energy minimisation using the DISCOVER package interfaced with INSIGHT (Biosym Technologies, V, 2.3.0 and 2.9.5). The surface area buried on dimerisation in each of the four energy-minimized models was evaluated. The results, given in Table 3, indicate that the non-polar surface area buried on dimerisation is substantially greater in the PNA dimer than in the PNA in con A dimer and in the con A dimer than

Table 3. Surface area buried on dimerisation

Model	Buried area (Å²)		
	Total	Polar	Non-polar
PNA dimer	2246	772	1475
Con A dimer	2875	1125	1741
PNA in con A dimer	2453	1386	1067
Con A in PNA dimer	2720	1469	1241

Table shows the surface area (Å²) buried on dimerisation in the optimised models of PNA dimer, con A dimer, PNA in con A dimer and con A in PNA dimer.

in the con A in PNA dimer. Thus it would appear that the distribution of surface residues in the PNA subunit leads to a greater hydrophobic stabilisation of the natural PNA dimer than the con A-type dimer, while the reverse is true in the case of the con A subunit.

The effect of the different modes of dimerisation in PNA and con A on tetramerisation was then investigated. When the two dimers in PNA (subunits 1 and 4, and subunits 2 and 3) were independently moved so as to make the 1-2 interface the same as in the con A dimer, several hundred contacts of less than 2 Å developed between subunits 3 and 4. On the other hand, when two canonical con A dimers were arranged such that one PNA 1-4 type interface existed between them, the buried surface area on association of the two dimers decreased by nearly half from what it is in the actual con A tetramer. Thus, it appears probable that the differences between the tetrameric arrangements in con A and PNA necessarily follow from their different modes of dimerisation.

Hydration

The final refined model contains 563 water molecule oxygen atoms. A water oxygen atom was considered to interact with another oxygen atom or a nitrogen atom if the distance between the two is less than 3.6 Å. When this criterion is used, 21 water molecules do not interact with any other molecule while 45 interact only with other water molecules. Of the remaining 497, 476 interact with one tetramer while 21 interact with two symmetry-related tetramers. Consequently, each tetramer is surrounded by and interacts with 518 water sites which is considered to constitute its hydration shell. The number of water molecules in the hydration shell of each subunit varies from 120 to 156.
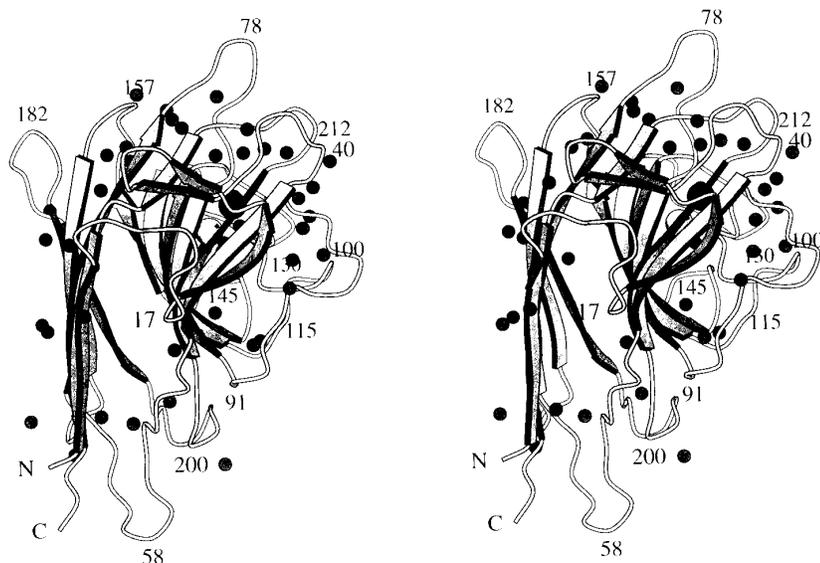
It is of interest to compare the hydration shells of the four chemically equivalent, but crystallographically independent, subunits in the tetramer in the asymmetric unit. In particular, it is instructive to examine the extent to which the hydration shells of individual subunits reflect the molecular and local dyads that relate the subunits themselves. The variation in the number of water molecules in the four hydration shells itself indicates their substantial non-equivalence, but a more rigorous analysis is necessary to quantify the extent of the variation.

As in the case of our earlier studies on protein hydration (Kodandapani *et al.*, 1990; Madhusudan & Vijayan, 1991; Radha Kishan *et al.*, 1995), a water molecule in the hydration shell of one subunit and one in the hydration shell of another were considered invariant if they interact with at least one common protein atom and if the distance between the two is less than 1.8 Å when the two subunits along with their hydration shells are superimposed on each other. When this criterion is used, the number of invariant water molecules between pairs of subunits is found to vary between 65 and 74 except in the pair consisting of subunits 1 and 2, in

Figure 12. Stereo view of the distribution of invariant water molecules around the PNA subunit.

which the number of invariant water molecules is higher (at 88), presumably on account of the large number of water bridges that connect the two subunits. Thus, on a rough and ready liberal estimate, only about half the water molecules in any pair of subunit hydration shells reflect the non-crystallographic equivalence of the subunits themselves.

The number of water molecules which are invariant with respect to all of the four subunits is still lower at 45. These include the six which were found by Loris *et al.* (1994) to exist in the structures of all legume lectins and their complexes. Four of them are involved in metal coordination, two of which also interact with protein atoms that hydrogen bond to the carbohydrate ligand. Of the remaining two, one connects the 144 to 148 loop and the C terminus of the 136 to 143 strand while the other stabilises the β-bulge contiguous to sheet 2 between loops 50 to 63 and 193 to 202.

The distribution of invariant water molecules in the subunit, including the six referred to above, is illustrated in Figure 12. As many as nine of them are associated exclusively with sheet 1. Of these, three occur in the monomer-monomer interface in the dimer and one in the 1-2 interface. Four interconnect different strands in the sheet while the remaining one is close to the amino terminus of the 186 to 192 strand. Only two water molecules are associated exclusively with sheet 2; they are attached to separate strands. Another invariant water molecule occurs in the junction between sheets 1 and 3, and interacts with Asn9 O, Asn31 ND2 and Arg221 O, in all of the four subunits. It also serves as a bridge between subunits 3 and 4. Five invariant water molecules are associated with loops 50 to 63 and 193 to 202 which connect sheets 1 and 2 at the bottom of the subunit. One of the five is also involved in the interface between subunits 1 and 2.

As many as 15 invariant water molecules, including the four that coordinate to the metal ions,

are found in the front loops 91 to 116, 125 to 135, 144 to 148 and the adjoining strands. Seven of them, like the one illustrated in Figure 13(a), interact primarily with the 91 to 116 loop, lending additional stability to the peptide fold. One of these seven water molecules also interacts with O2 and O3 of galactose. Each of a majority of the remaining invariant water molecules in this region interacts with two of the three loops. Another distinct yet contiguous region containing a number (13) of invariant water molecules is primarily made up of sheet 3 and the loops that connect sheets 2 and 3 (35 to 41, 211 to 216, 75 to 83 and 154 to 159) at the top of the molecule. Nearly half of them connect two loops each. For example, three water molecules which connect two loops are shown in Figure 13(b). There are interactions through protein atoms or invariant water molecules between the two regions. It is perhaps significant that a majority of the invariant water molecules occur in the substantially contiguous region made up of seven loops at the top and the front of the subunit. The observed hydrogen bonds and other direct interactions involving protein atoms do not appear to be enough to account for the stability and integrity of this region. It is probable that a majority of the invariant water molecules are involved in stabilising the loops and in ensuring the correct structural relations among them and with the rest of the molecule.

Materials and Methods

Crystallisation
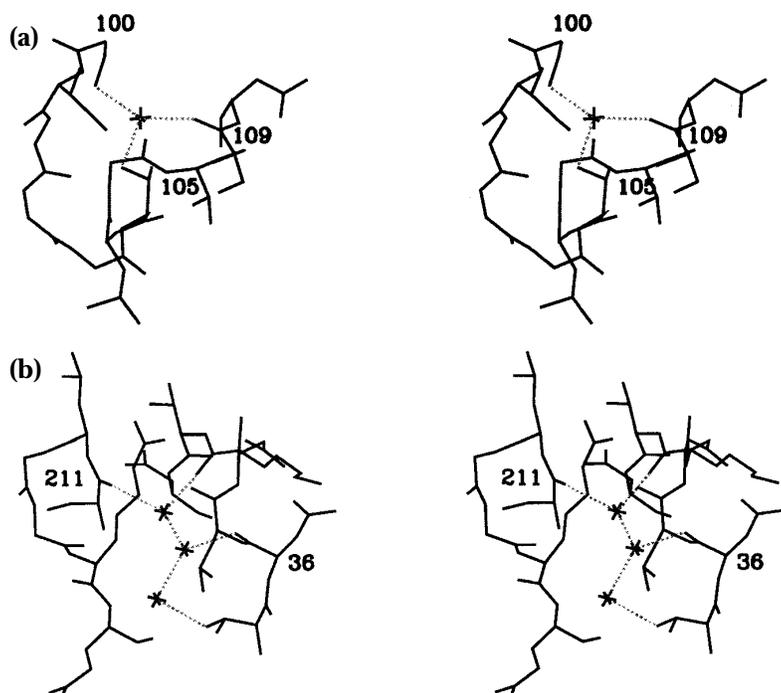
Figure 13. Stereo diagrams illustrating the role of (a) an invariant water molecule (represented by star) in stabilising the 91 to 117 loop and (b) three invariant water molecules involved in interconnecting loops 35 to 41 and 211 to 216.

Data collection

Intensity data were collected on a Siemens area detector system mounted on a GX 20 rotating anode generator and processed using the XENGEN suite of programs (Howard *et al.*, 1987). Two data sets were collected from two different crystals and the data were merged. The details of data collection are given in Table 4.

Refinement

The structure of the PNA tetramer that resulted from the work at 2.95 Å resolution (Banerjee *et al.*, 1994) was used as the starting model for the high-resolution refinement. This model consisted of 6919 protein atoms, which accounted for 97% of the total number of atoms in the tetramer, and one calcium ion and one manganese ion in each subunit. To start with, several cycles of refinement of positional parameters were carried out using the restrained least squares program PROLSQ (Hendrickson & Konnert, 1980) in the CCP4 package (SERC Daresbury Laboratory, 1994). Data in the 2.95 to 2.25 Å shell were introduced in six steps, each step involving six cycles of refinement. Restraints for β-sheets and non-crystallographic symmetry were used in these refinement cycles, but were discarded in further work. Simulated annealing using X-PLOR (Brunger, 1992) was performed at this stage, followed by several cycles of PROLSQ refinement of positional parameters. A subsequent $2F_o - F_c$ map was

Table 4. Details of data collection

	Crystal 1	Crystal 2
Resolution (Å)	2.90	2.25
No. of observations	76,203	97,006
No. of unique reflections	24,145	48,309
R_{merge} (%)	6.8	6.7

		Combined data		
Resolution range (Å)	Number of observations	Number of unique reflections	Completion (%)	R_{merge} (%)
∞ to 3.99	57,970	10,874	96.9	5.28
3.99 to 3.17	48,800	10,275	94.8	7.25
3.17 to 2.77	27,178	9252	85.7	11.09
2.77 to 2.52	17,935	8388	78.2	12.66
2.52 to 2.34	14,219	7467	69.8	15.08
2.34 to 2.25	7107	4508	42.3	15.12
Total	173,209	50,564	88.0	7.90

$R_{merge} = \Sigma |I - \langle I \rangle| / \Sigma \langle I \rangle$

Table 5. Refinement parameters

Resolution limits (Å)	10 to 2.25
Final R-factor (%)	16.4
No. of reflections used ($F > 3\sigma(F)$)	44,510
No. of protein atoms	6976
No. of solvent molecules	563
No. of carbohydrate atoms	92
Other ions	4 Mg^{2+}, 4 Ca^{2+}
Average B-value (protein atoms) ($Å^2$)	21.1
Average B-value (all atoms) ($Å^2$)	22.5
	R.m.s. deviations from standard values
Bond length (Å)	0.010
Angle distance (Å)	0.038
Planarity (Å)	0.010
Chiral volume ($Å^3$)	0.124
Non-bonded contacts (Å)	0.175

used for adjustments in the model using FRODO (Jones, 1978). A further eight cycles of refinement of positional parameters followed by one cycle of B-refinement resulted in an R-factor of 0.249. The subsequent $2F_o − F_c$ map had clear density for both rings of the lactose molecule in each subunit. The R-factor reduced to 0.228 after a few cycles of refinement of the positional and thermal parameters of the protein and sugar atoms. Successive $2F_o − F_c$ and $F_o − F_c$ maps, interspersed between refinement cycles, were used to locate water molecules and adjust the protein model. An omit-type map (Vijayan, 1980; Bhat & Cohen, 1984) was also used to check the model. The refinement converged at 16.4% for 44,510 reflections with $F > 3\sigma(F)$ in the resolution range between 10 and 2.25 Å (Table 5). The final atomic coordinates of PNA were deposited (2PEL) with the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973, USA.

References

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A. & Vijayan, M. (1994). Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proc. Natl Acad. Sci. USA*, **91**, 227–231.

Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975). The covalent and three-dimensional structure of concanavalin A. *J. Biol. Chem.* **250**, 1513–1524.

Bernstein, F. C., Koetzel, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Bhat, T. N. & Cohen, G. H. (1984). OMITMAP: an electron density map suitable for the examination of errors in a macromolecular model. *J. Appl. Crystallog.* **17**, 244–248.

Bhat, T. N., Sasisekharan, V. & Vijayan, M. (1979). An analysis of side-chain conformation in proteins. *Int. J. Peptide Protein Res.* **13**, 170–184.

Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rouge, P. & Fontecilla-Camps, J.-C. (1990a). X-ray crystal structure determination and refinement at 1.9 Å resolution of isolectin I from the seeds of *Lathyrus ochrus*. *J. Mol. Biol.* **214**, 571–584.

Bourne, Y., Roussel, A., Frey, M., Rouge, P., Fontecilla-Camps, J.-C. & Cambillau, C. (1990b). Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose. *Proteins: Struct. Funct. Genet.* **8**, 365–376.

Bourne, Y., Rouge, P. & Cambillau, C. (1990c). X-ray structure of a (α-Man(1-3)β-Man(1-4)GlcNAc)-lectin complex at 2.1 Å resolution. *J. Biol. Chem.* **265**, 18161–18165.

Bourne, Y., Rouge, P. & Cambillau, C. (1992). X-ray structure of biantennary octasaccharide-lectin complex refined at 2.3-Å resolution. *J. Biol. Chem.* **267**, 197–203.

Bourne, Y., Mazurier, J., Legrand, D., Rouge, P., Montreuil, J., Spik, G. & Cambillau, C. (1994). Structures of a legume lectin complexed with the human lactotransferrin NZ fragment, and with an isolated biantennary glycopeptide: role of the fucose moiety. *Structure*, **2**, 209–219.

Brunger, A. T. (1992). *X-PLOR Manual, Version 3.1*, Yale University, New Haven, CT.

Carson, M. (1992). *RIBBONS Manual, Version 2.0*, University of Alabama at Birmingham, Alabama.

Cornish-Bowden, A. J. & Koshland, D. E., Jr (1971). The quaternary structure of proteins composed of identical subunits. *J. Biol. Chem.* **246**, 3092–3102.

Crennell, S., Garman, E., Laver, G., Vimr, E. & Taylor, G. (1994). Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure*, **2**, 535–544.

Decastel, H., de Boeck, H., Goussault, Y., de Bruyne, C. K., Loontiens, F. G. & Frenoy, Jean-Pierre (1985). Effect of pH on oligomeric equilibrium and saccharide-binding properties of peanut agglutinin. *Arch. Biochem. Biophys.* **240**, 811–819.

Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Pearlstone, J. W., Carpenter, M. R., Smillie, L. B., Nikrad, P. V., Spohr, U. & Lemieux, R. U. (1990). Molecular recognition of a human blood group determinant by a plant lectin. *Can. J. Chem.* **68**, 1116–1121.

Delbaere, L. T. J., Vandonselaar, M., Prasad, L. Quail, J. W., Wilson, K. S. & Dauter, Z. (1993). Structure of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution. *J. Mol. Biol.* **230**, 950–965.

Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb (Gilboa), A. J., Dodson, E. J., Papiz, M. Z., Wan, T. &

Campbell, J. (1989). The structure of the saccharide-binding site of concanavalin A. *EMBO J.* **8**, 2189–2193.

Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F. & Sacchettini, J. C. (1995). X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. *Biochemistry*, **34**, 4933–4942.

Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E. & Suddath, F. L. (1986). The crystal structure of pea lectin at 3.0-Å resolution. *J. Biol. Chem.* **261**, 16518–16527.

Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B. & Wood, S. P. (1994). Structure of pentameric human serum amyloid P component. *Nature*, **367**, 338–345.

Fish, W. W., Hamlin, L. M. & Miller, R. L. (1978). The macromolecular properties of peanut agglutinin. *Arch. Biochem. Biophys.* **190**, 693–698.

Hardman, K. D. & Ainsworth, C. F. (1972). Structure of concanavalin A at 2.4 Å resolution. *Biochemistry*, **11**, 4910–4919.

Hendrickson, W. A. & Konnert, J. H. (1980). Incorporation of stereochemical information into crystallographic refinement. In *Computing in Crystallography* (Diamond, R., Ramaseshan, S. & Venkatesan, K., eds), pp. 13.01–13.23, Indian Academy of Sciences, Bangalore.

Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. & Salemme, F. R. (1987). The use of an imaging proportional counter in macromolecular crystallography. *J. Appl. Crystallog.* **20**, 383–387.

Ip, S. H. C. & Ackers, G. K. (1977). Thermodynamic studies on subunit assembly in human hemoglobin. *J. Biol. Chem.* **252**, 82–87.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallog.* **11**, 268–272.

Klotz, I. M., Darnall, D. W. & Langerman, N. R. (1975). Quaternary structure of proteins. In *The Proteins*, vol. 3, (Neurath, H. & Hill, R. L., eds), pp. 293–411, Academic Press, New York and London.

Kodandapani, R., Suresh, C. G. & Vijayan, M. (1990). Crystal structure of low humidity tetragonal lysozyme at 2.1-Å resolution. *J. Biol. Chem.* **265**, 16126–16131.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

Liao, D.-I., Kapadia, G., Ahmed, H., Vasta, G. R. & Herzberg, O. (1994). Structure of S-lectin, a developmentally-regulated vertebrate β-galactoside binding protein. *Proc. Natl Acad. Sci. USA*, **91**, 1428–1432.

Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H. & Rini, J. M. (1993). X-ray crystal structure of the human dimeric S-lac lectin, L-14-II in complex with lactose at 2.9-Å resolution. *J. Biol. Chem.* **268**, 27034–27038.

Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. & Wyns, L. (1993). Crystal structure determination and refinement at 2.3-Å resolution of the lentil lectin. *Biochemistry*, **32**, 8772–8781.

Loris, R., Stas, P. G. P. & Wyns, L. (1994). Conserved waters in legume lectin crystal structures. *J. Biol. Chem.* **269**, 26722–26733.

Lotan, R., Skutelsky, E., Danon, D. & Sharon, N. (1975). The purification, composition, and specificity of the anti-T lectin from peanut (*Arachis hypogaea*). *J. Biol. Chem.* **250**, 8518–8523.

Luzzati, V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallog.* **5**, 802–810.

Madhusudan & Vijayan, M. (1991). Rigid and flexible regions in lysozyme and the invariant features in its hydration shell. *Curr. Science*, **60**, 165–170.

Majumdar, T. & Surolia, A. (1978). A large scale preparation of peanut agglutinin on a new affinity matrix. *Prep. Biochem.* **8**, 119–131.

McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987). Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**, 295–310.

Monod, J., Wyman, J. & Changeux, J.-P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118.

Naismith, J. H., Emmerich, C., Habash, J., Harrop, S., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb, Gilboa, A. J. & Yariv, J. (1994). Refined structure of concanavalin A complexed with methyl α-D-mannopyranoside at 2.0 Å resolution and comparison with the saccharide-free state. *Acta Crystallog. sect. D*, **50**, 847–858.

Radha Kishan, K. V., Chandra, N. R., Sudarsanakumar, C., Suguna, K. & Vijayan, M. (1995). Water-dependent domain motion and flexibility in ribonuclease A and the invariant features in the hydration shell. An X-ray study of two low humidity crystal forms of the enzyme. *Acta Crystallog. sect. D*, **51**, 703–710.

Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–438.

Reeke, G. N., Jr & Becker, J. W. (1986). Three-dimensional structure of favin: saccharide binding-cyclic permutation in leguminous lectins. *Science*, **234**, 1108–1111.

Reeke, G. N., Jr, Becker, J. W. & Edelman, G. M. (1975). The covalent and three-dimensional structure of concanavalin A. *J. Biol. Chem.* **250**, 1525–1547.

Rini, J. M., Hardman, K. D., Einspahr, H., Suddath, F. L. & Carver, J. P. (1993). X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 Å resolution. *J. Biol. Chem.* **268**, 10126–10132.

Rossmann, M. G. & Argos, P. (1976). The taxonomy of protein structure. *J. Mol. Biol.* **105**, 75–95.

Salunke, D. M., Islam Khan, M., Surolia, A. & Vijayan, M. (1982). Crystallization and preliminary X-ray studies of the anti-T lectin from peanut (*Arachis hypogaea*). *J. Mol. Biol.* **154**, 177–178.

Salunke, D. M., Islam Khan, M., Surolia, A. & Vijayan, M. (1983). Preparation and preliminary X-ray studies of three acidic pH crystal forms of the anti-T lectin from peanut (*Arachis hypogaea*). *FEBS Letters*, **156**, 127–129.

Salunke, D. M., Swamy, J. M., Islam Khan, M., Mande, S. C., Surolia, A. & Vijayan, M. (1985). Arrangement of subunits in peanut lectin. Rotation function and chemical cross-linking studies. *J. Biol. Chem.* **260**, 13576–13579.

Senear, D. F. & Teller, D. C. (1981a). Thermodynamics of concanavalin A dimer-tetramer self-association: sedi-

mentation equilibrium studies. *Biochemistry*, **20**, 3076–3083.

Senear, D. F. & Teller, D. C. (1981b). Effects of saccharide and salt binding on dimer-tetramer equilibrium of concanavalin A. *Biochemistry*, **20**, 3083–3091.

SERC Daresbury Laboratory, Warrington, WA4 4AD, UK (1994). The *CCP4 Suite: Programs for Protein Crystallography. Acta Crystallog. sect. D*, **50**, 760–763.

Shaanan, B., Lis, H. & Sharon, N. (1991). Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*, **254**, 862–866.

Sharon, N. (1993). Lectin-carbohydrate complexes of plants and animals: an atomic view. *Trends Biochem. Sci.* **18**, 221–226.

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

Sharon, N. & Lis, H. (1989b). Lectins as cell recognition molecules. *Science*, **246**, 227–246.

Steitz, T. A., Fletterick, R. J., Anderson, W. F. & Anderson, C. M. (1976). High resolution X-ray structure of yeast hexokinase, an allosteric protein exhibiting a non-symmetric arrangement of subunits. *J. Mol. Biol.* **104**, 197–222.

Vijayan, M. (1980). Phase evaluation and some aspects of the Fourier refinement of macromolecules. In *Computing in Crystallography* (Diamond, R., Ramaseshan, S. & Venkatesan, K., eds), pp. 19.01–19.26, Indian Academy of Sciences, Bangalore.

Vyas, N. K. (1991). Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**, 732–740.

Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. (1991). Structure of the calcium dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science*, **254**, 1608–1615.

Weis, W. I., Drickamer, K. & Hendrickson, W. A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature*, **360**, 127–134.

Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature*, **289**, 366–378.

Wright, C. S. (1989). Comparison of the crystal structures of two wheat germ isolectins. *J. Mol. Biol.* **209**, 475–487.

Young, N. M. & Oomen, R. P. (1992). Analysis of sequence variation among legume lectins. *J. Mol. Biol.* **228**, 924–934.