THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Structural plasticity of peanut lectin: an X-ray analysis involving variation in pH, ligand binding and crystal structure

OPEN ACCESS

Structural plasticity of peanut lectin: an X-ray analysis involving variation in pH, ligand binding and crystal structure

**S. Kundhavai Natchiar,
A. Arockia Jeyaprakash,
T. N. C. Ramya, C. J. Thomas,
K. Suguna, A. Surolia and
M. Vijayan***

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

Correspondence e-mail: mv@mbu.iisc.ernet.in

Until recently, it has only been possible to grow crystals of peanut lectin when complexed with sugar ligands. It is now shown that it is possible to grow peanut lectin crystals at acidic pH in the presence of oligopeptides corresponding to a loop in the lectin molecule. Crystals have also been prepared in the presence of these peptides as well as lactose. Low-pH crystal forms of the lectin–lactose complex similar to those obtained at neutral pH have also been grown. Thus, crystals of peanut lectin grown under different environmental conditions, at two pH values with and without sugar bound to the lectin, are now available. They have been used to explore the plasticity and hydration of the molecule. A detailed comparison between different structures shows that the lectin molecule is sturdy and that the effect of changes in pH, ligand binding and environment on it is small. The region involving the curved front β-sheet and the loops around the second hydrophobic core is comparatively rigid. The back β-sheet involved in quaternary association, which exhibits considerable variability, is substantially flexible, as is the sugar-binding region. The numbers of invariant water molecules in the hydration shell are small and they are mainly involved in metal coordination or in stabilizing unusual structural features. Small consistent movements occur in the combining site upon sugar binding, although the site is essentially preformed.

1. Introduction

Lectins, found in plants, animals and microorganisms, are carbohydrate-binding proteins that mediate a variety of biological processes by specifically recognizing diverse sugar structures (Lis & Sharon, 1998; Loris *et al.*, 1998; Vijayan & Chandra, 1999). Those from leguminous sources constitute the most thoroughly studied family of lectins. Tetrameric peanut (*Arachis hypogaea*) agglutinin (PNA), $M_r = 110\,000$, belongs to this family. Each subunit is 236 residues long and is non-glycosylated. It is galactose-specific at the monosaccharide level and binds to the tumour-associated disaccharide Galβ1-3GalNAc, commonly known as T-antigen, with high specificity.

A remarkable feature of the peanut lectin molecule is its unusual 'open' quaternary structure without point-group symmetry (Banerjee *et al.*, 1994). The structure also demonstrates that the variability in quaternary association in legume lectins is not necessarily caused by interactions involving covalently bound sugar. Subsequent detailed investigation involving peanut lectin and winged-bean lectins (Prabu *et al.*, 1998; Manoj *et al.*, 2000) demonstrated that legume lectins constitute a family of proteins in which small changes in essentially the same tertiary structure leads to large variations in quaternary association (Prabu *et al.*, 1999). The origin of

these variations could be traced to those in the sequences (Manoj & Suguna, 2001). The structures of a number of PNA–carbohydrate complexes determined in this laboratory (Banerjee *et al.*, 1996; Ravishankar *et al.*, 1997, 1998, 1999) led to elucidation of the geometrical features of lectin–carbohydrate interactions, with special reference to PNA. In particular, interactions involving water bridges were shown to be a strategy for generating ligand specificity. The studies also led to the structure-based redesign of the carbohydrate-binding site of the lectin (Sharma *et al.*, 1996, 1998; Adhikari *et al.*, 2001). In another effort, further insights into PNA–carbohydrate interactions were obtained from molecular-dynamics studies (Pratap *et al.*, 2001).

The early structural studies on PNA were carried out on an orthorhombic form grown at neutral pH. The lectin is known to dissociate into dimers at very low pH. It was therefore interesting to examine the quaternary structure close to the pH of dissociation, especially in view of its unusual nature. The variation of the nature of lectin–carbohydrate interactions as a function of pH also deserved examination. Therefore, two crystal forms of the lectin grown in the presence of lactose at pH 4.6, the lowest pH at which crystals grew, were subsequently analysed (Ravishankar *et al.*, 2001). One form was monoclinic, while the other was triclinic. The structures demonstrated the invariance of the quaternary arrangement with respect to the variation in pH. The molecules at low pH contained lactose-bound carbohydrate-binding sites as well as empty ones. More interestingly, in some instances a loop from a neighbouring molecule interacted with the binding site. Solution studies at low pH indicated association of the lectin with peptides having sequences corresponding to the sequence of amino acids in this loop. This prompted the study of PNA crystals grown at acidic pH in the presence of these peptides, which is reported here. Crystals were also grown in the presence of a peptide and sugar (lactose). In none of the studies could a specific PNA–peptide interaction be demonstrated. However, it is clear that the peptides influence the crystallization of PNA. Prior to these results, crystals of PNA could be grown only in the presence of a sugar ligand. In the present study, PNA could be crystallized even in the absence of sugar if an appropriate peptide were present. The structures reported here and those analysed earlier provide a number of views of the PNA molecule in different environmental conditions and with or without sugar bound at the combining sites. These results permit, among other things, a comprehensive analysis of the rigid and flexible regions of the molecule, the invariant features in its hydration shell and the plasticity of the combining site. Also permitted is an assessment of the effect of environment and ligands on crystallization and the three-dimensional structure of the lectin.

2. Materials and methods

Peanut lectin was prepared by affinity chromatography on cross-linked arabinogalactan (Majumdar & Surolia, 1978). Two decapeptides IWSSATGNVA (10IA) and PVRIWSSATG (10PG) with overlapping sequences that form parts of the loop that interacts with a carbohydrate-binding site in a neighbouring molecule, were custom synthesized by Xcyton, Bangalore and Mimotopes Pvt Ltd, Australia, respectively. Two nonapeptides, one with Thr removed from the first decapeptide (9IA) and the other with Arg removed from the second decapeptide (9PG) were custom synthesized by Xcyton. The sequences of the peptides were confirmed using sequencing (Shimadzu PSO-1, Japan).

2.1. Crystallization

Crystals did not grow in the presence of peptides when the conditions used for crystallizing the original low-pH form were employed. Therefore, a full search of the Hampton conditions (Jancarik & Kim, 1991; Cudney *et al.*, 1994) was carried out to crystallize PNA at low pH in the presence of the peptides, in the presence of lactose, in the presence of each peptide and lactose and in the absence of both peptide and lactose. No crystals grew in the last set of experiments when neither peptide nor lactose was present. The solutions used in the experiments that yielded crystals were the following: (i) 4.2 mg ml^{-1} protein solution in 0.05 M sodium acetate buffer containing 0.2 M NaCl and 0.05% sodium azide, (ii) the same protein solution but containing 2.1 mM lactose, (iii) two precipitant solutions, both in 0.1 M sodium cacodylate buffer pH 6.5, but one containing 30% PEG 8K and 0.2 M ammonium sulfate and the other containing 18% PEG 8K and 0.2 M calcium acetate and (iv) solutions containing 4.4–5.5 mM peptide in the buffer used to make the protein solution (20% DMSO was added in the case of the decapeptides to make them soluble). In the hanging-drop method used, each drop contained 4–5 µl of the protein solution with or without lactose, 1.5–2.5 µl of a peptide solution and 1.5–2.5 µl of the precipitant solution. The drops were equilibrated against the appropriate precipitant solution. Examination using calibrated strips indicated that the pH of the drops varied between 4.6 and 5.0. The original low-pH forms of the PNA–lactose complex were grown from the acetate buffer with PEG 8K as the precipitant, using the bulk method. Crystals under similar conditions were grown employing the hanging-drop method using a 4.2 mg ml^{-1} protein solution in 0.05 M sodium acetate buffer containing 0.2 M NaCl, 0.05% sodium azide and 2.1 mM lactose and a precipitant solution of 12% PEG 8K in the same buffer. The drop was made up of 6 µl protein solution and 3 µl precipitant solution.

2.2. Data collection and processing

Data were collected from all crystals at 293 K using MAR Research imaging plates mounted on Rigaku rotating-anode generators. The crystal-to-detector distance varied between 120 and 150 mm. Data were processed using *DENZO* and scaled using *SCALEPACK* from the *HKL* suite of programs (Otwinowski, 1993). Details of the data-collection statistics are given in Table 1.

Table 1
Crystal data and data-collection statistics.

Values in parentheses refer to the last resolution shell. See text (§3.1) for the nomenclature used in this table and Table 2.

	PNALacA	PNALacB	PNALac9IA	PNALac9PG	PNA9IA	PNA10PG	PNA9PG
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters							
a (Å)	128.61	128.73	128.03	128.28	127.85	128.13	127.15
b (Å)	125.75	125.53	124.03	124.64	126.14	125.80	125.80
c (Å)	76.05	76.05	75.71	75.75	84.83	84.74	84.73
β (°)					116.2	116.2	116.2
Resolution (Å)	2.15	2.90	2.40	2.50	2.70	3.00	3.50
Last shell resolution range (Å)	2.17–2.15	2.92–2.90	2.42–2.40	2.52–2.50	2.72–2.70	3.03–3.00	3.59–3.50
No. observations	391685 (18544)	191532 (8910)	190413 (9336)	120907 (5785)	241010 (12407)	110293 (2737)	58800 (1908)
No. unique reflections	67601 (3325)	26610 (1304)	46230 (2267)	41998 (2143)	64462 (3220)	43166 (1185)	26281 (775)
Reflections with $I = 0$	3451 (387)	1068 (128)	2662 (304)	2352 (236)	2941 (399)	2407 (100)	1427 (73)
Completeness (%)	99.9 (100)	97.6 (96.5)	96.4 (95.3)	96.5 (98.5)	97.4 (96.9)	95.5 (94.0)	94.7 (97.4)
R_{merge} (%)	13.1 (57.5)	15.1 (52.8)	11.9 (58.4)	12.7 (46.1)	16.0 (56.3)	16.7 (35.7)	18.2 (30.3)
Multiplicity	5.8 (5.6)	7.2 (6.8)	4.1 (4.1)	2.9 (2.7)	3.7 (3.9)	2.6 (2.3)	2.2 (2.5)

Table 2
Refinement parameters.

	PNALacA	PNALacB	PNALac9IA	PNALac9PG	PNA9IA	PNA10PG	PNA9PG
No. protein atoms	6976	6976	6976	6976	13952	13952	13952
No. sugar atoms	92	92	92	92	—	—	—
No. solvent atoms	423	189	399	362	360	260	—
No. ions	8	8	8	8	16	16	16
Resolution limits (Å)	20–2.15	20–2.90	20–2.40	20–2.50	20–2.70	20–3.00	20–3.50
R factor (%)	18.4	16.3	17.7	17.3	21.0	17.4	18.7
R_{free} (%)	21.8	22.2	21.9	22.4	26.4	24.1	29.2
No. reflections	64350	25507	42158	38448	61717	42215	26046
R.m.s. deviations from ideal values							
Bond length (Å)	0.005	0.011	0.009	0.014	0.009	0.010	0.010
Bond angle (Å)	1.4	1.8	1.9	2.5	2.1	2.3	2.3
Dihedral angle (°)	26.4	27.0	26.6	26.7	26.7	26.5	27.3
Improper angle (°)	0.85	1.23	0.93	1.49	1.02	1.26	2.0
Residues in Ramachandran plot (%)							
Most favoured region	90.8	89.5	89.1	89.8	86.2	83.7	75.6
Additionally allowed region	9.1	10.4	10.9	10.0	13.4	15.9	23.8
Generously allowed region	0.1	0.1	0.0	0.2	0.4	0.4	0.5
Disallowed region	0.0	0.0	0.0	0.0	0.0	0.0	0.2
Estimated coordinate error from Luzzati plot	0.23	0.26	0.25	0.25	0.32	0.29	0.42

2.3. Refinement and structure analyses

All crystals grown in the present study are isomorphous to the original PNA–lactose (Banerjee *et al.*, 1996) crystals grown at pH 7 or to the PNA–lactose (Ravishankar *et al.*, 2001) crystals grown at pH 4.6 (Table 1). Hence, the coordinates of PNA in these crystals (PDB codes 2pel and 1cr7) were used in refinement using *CNS* (Brünger *et al.*, 1998). In all cases 40–50 cycles of rigid-body refinement were carried out first. Positional refinement with a maximum-likelihood target was performed next. Sugar molecules were fitted using *FRODO* (Jones, 1978) where appropriate density in $2F_o - F_c$ and $F_o - F_c$ maps existed in the binding sites. Water O atoms were added initially using density with peak heights greater than 3.0σ in $F_o - F_c$ maps and 1.0σ in $2F_o - F_c$ maps. In the subsequent cycles of refinement and fitting, the cutoff values were reduced to 2.5σ in $F_o - F_c$ maps and 0.8σ in $2F_o - F_c$ maps. Final positional and temperature-factor refinements were carried out until R and R_{free} converged. The isotropic temperature factors of atoms were refined individually in structures with resolution of 2.5 Å or better and group B factors were used in those with lower resolution. Bulk-solvent

corrections and overall anisotropic B-factor corrections were used throughout the refinement. The quality of the refined models was checked using *PROCHECK* (Laskowski *et al.*, 1993). The final refinement statistics are given in Table 2. The structures of the original orthorhombic crystals grown at neutral pH (Banerjee *et al.*, 1996) and monoclinic crystals grown at acidic pH (Ravishankar *et al.*, 2001) were further refined using *CNS* in the same way as the structures reported were refined in order to facilitate proper comparison. The rigid and flexible regions of the protein were delineated using the *Error-inclusive Structure Comparison and Evaluation Tool* (*ESCET*; Schneider, 2002). Molecular superpositions were performed using the program *ALIGN* (Cohen, 1997).

3. Results and discussion

### 3.1. Effect of lactose and peptides on crystallization

The space groups and the unit-cell parameters of the crystals obtained in the experiments described in the previous section are given in Table 1. As indicated earlier, the lectin

Table 3
PNALacO, PNALacA and PNALacB: r.m.s. deviations (Å) in C^{α} positions on superposition of subunits.

	PNALacO				PNALacA				PNALacB			
	A	B	C	D	A	B	C	D	A	B	C	D
PNALacO												
A		0.17	0.21	0.23	0.14	0.19	0.21	0.23	0.19	0.22	0.25	0.26
B			0.23	0.26	0.20	0.15	0.25	0.29	0.24	0.21	0.29	0.30
C				0.16	0.20	0.23	0.16	0.20	0.23	0.24	0.20	0.24
D					0.21	0.25	0.19	0.16	0.24	0.28	0.25	0.20
PNALacA												
A						0.13	0.16	0.17	0.12	0.17	0.20	0.20
B							0.20	0.21	0.17	0.13	0.23	0.24
C								0.15	0.18	0.22	0.13	0.19
D									0.21	0.25	0.20	0.13
PNALacB												
A										0.17	0.19	0.20
B											0.22	0.22
C												0.19

does not crystallize in the absence of sugar or peptide. However, interestingly, crystals of PNA could be grown, for the first time in the absence of sugar, in the presence of the peptides 9PG, 9IA and 10PG. They are isomorphous to the monoclinic crystals of the PNA–lactose complex (PNALacM) originally obtained at low pH from acetate buffer. Crystals also grew in the presence of lactose on the one hand and 9PG (PNALac9PG) or 9IA (PNALac9IA) on the other. Both are isomorphous to the orthorhombic form of the PNA–lactose complex (PNALacO) grown at neutral pH (Banerjee *et al.*, 1994). Crystals grown in the presence of only lactose from a mixture of acetate and cacodylate buffers (PNALacA) and those grown from acetate buffer using the hanging-drop method (PNALacB) are isomorphous to the orthorhombic form obtained at neutral pH.

Only the orthorhombic form is obtained at neutral pH irrespective of the sugar used (Banerjee *et al.*, 1996; Ravishankar *et al.*, 1997, 1998, 1999). At lower pH, both orthorhombic and monoclinic forms could be obtained in the presence of lactose. Two conditions at low pH involving the
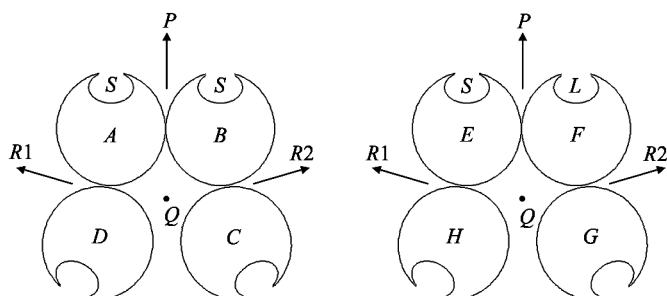
Figure 1
Schematic representation of the two tetramers in the monoclinic crystals. *S* and *L* indicate the occupation of the combining site by lactose and a loop from a neighbouring molecule, respectively. Four combining sites are unoccupied. *P* represents the molecular dyad. *R*1 and *R*2 represent the twofold axes that relate *A* and *D*, and *B* and *C*, respectively. *P*, *R*1 and *R*2 do not intersect, but pass through a line represented by *Q*. In the orthorhombic crystals, which grow only in the presence of sugar, the crystallographic asymmetric unit contains one tetramer in which all the subunits are sugar-bound.

presence of lactose and a peptide yielded crystals. Both of them are orthorhombic. When only the peptides are present in the low-pH crystallization solution, only monoclinic crystals grew. It was shown earlier (Ravishankar *et al.*, 2001) that the orthorhombic form is related to the monoclinic form through a shear in the former along the *a* axis by 37.5 Å and rotation of the tetramer by 35° about *a*. The change in crystal packing is apparently caused by a strong interaction of the 49–63 loop in a subunit of the molecule with the carbohydrate-binding site in a subunit of a neighbouring molecule. The peptides used in the present work correspond to stretches in this loop. The results indicate that the interaction of the peptide with the lectin molecule is not strong enough to replace that involving the loop of the intact molecule, perhaps owing to the flexibility of the peptide molecules. In contrast, the lactose molecules are able to compete, but not very strongly at low pH, with the loop for the binding site. Therefore, at low pH in the presence of lactose, orthorhombic crystals grew in some conditions, while in some others monoclinic crystals grew. At neutral pH, the lectin–carbohydrate interactions are strong enough to exclude the possibility of monoclinic crystals.

3.2. Effect of pH, lactose, the peptides and packing on the overall structure

Contrary to expectations, none of the carbohydrate-binding sites are occupied by peptide molecules, even in the absence of lactose. No ordered peptide molecule could be identified in any of the structures. Nevertheless, the peptides in the medium do induce the ordering of the protein molecules, even in the absence of sugar.

Orthorhombic PNALacO was grown at neutral pH in the presence of lactose. PNALacA and PNALacB are isomorphous to PNALacO and were obtained in the pH range 4.5–5.0, again in the presence of lactose. The three structures have bound lactose in all the subunits. A comparison between them would serve as a measure of the effect of pH on molecular structure. The r.m.s. deviations in C^{α} positions between corresponding subunits (subunit *A* in one and subunit *A* in another *etc.*) in pairs of structures have values lying between 0.12 and 0.21 Å (Table 3). In fact, none of these deviations have values greater than 0.5 Å. For perspective, when r.m.s. deviations in C^{α} positions among the four subunits in the same tetramer in the same structure were calculated, they were found to vary between 0.13 and 0.26 Å. Thus, the variations in individual subunits caused by the change in pH are comparable to those among the subunits in the tetramer. The r.m.s. deviation when the whole tetramer of PNALacA is superposed on that of PNALacB is as low as 0.14 Å (Table 4). Both the crystals were grown at acidic pH. The two tetramers superpose on that in PNALacO, grown at neutral pH, with r.m.s. deviations of 0.26 and 0.29 Å, respectively. Thus, change of pH has probably some effect on the structure, but it is small. Thus, the tertiary and the quaternary structure remain essentially unchanged in the pH range 4.5–7.

PNALacA, PNALac9IA and PNALac9PG were grown in the same manner except for the presence of a peptide in the

Table 4
R.m.s. deviations (Å) in C^α positions on superposition of tetramers.

	PNALacA	PNALacB	PNALac9IA	PNALac9PG	PNALacM1	PNALacM2	PNA9IA1	PNA9IA2	PNA10PG1	PNA10PG2
PNALacO	0.26	0.29	0.42	0.43	0.50	0.54	0.53	0.54	0.54	0.55
PNALacA		0.14	0.20	0.23	0.50	0.54	0.43	0.44	0.47	0.50
PNALacB			0.23	0.23	0.53	0.56	0.45	0.47	0.48	0.52
PNALac9IA				0.13	0.60	0.64	0.47	0.47	0.54	0.57
PNALac9PG					0.63	0.66	0.49	0.49	0.56	0.58
PNALacM1						0.20	0.30	0.32	0.24	0.27
PNALacM2							0.31	0.30	0.27	0.24
PNA9IA1								0.17	0.24	0.26
PNA9IA2									0.27	0.22
PNA10PG1										0.52

Table 5
R.m.s. deviations (Å) in C^α positions on superposition of subunits in PNALacM and PNA9IA.

	PNALacM								PNA9IA							
	A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H
PNALacM																
A		0.21	0.29	0.28	0.17	0.25	0.28	0.29	0.20	0.26	0.31	0.33	0.19	0.26	0.31	0.32
B			0.26	0.29	0.22	0.19	0.26	0.27	0.23	0.21	0.29	0.31	0.24	0.21	0.28	0.28
C				0.27	0.30	0.29	0.17	0.26	0.29	0.29	0.22	0.30	0.28	0.30	0.22	0.28
D					0.30	0.31	0.28	0.20	0.29	0.29	0.30	0.24	0.29	0.31	0.30	0.23
E						0.24	0.29	0.29	0.19	0.26	0.32	0.33	0.18	0.27	0.32	0.31
F							0.29	0.29	0.25	0.22	0.33	0.32	0.23	0.20	0.32	0.32
G								0.26	0.29	0.28	0.23	0.29	0.27	0.28	0.21	0.29
H									0.27	0.28	0.32	0.24	0.28	0.29	0.30	0.22
PNA9IA																
A										0.20	0.27	0.29	0.14	0.22	0.28	0.28
B											0.26	0.28	0.21	0.17	0.28	0.27
C												0.27	0.27	0.28	0.16	0.26
D													0.29	0.30	0.27	0.18
E														0.21	0.29	0.29
F															0.28	0.28
G																0.24

crystallization medium in the case of the latter two. Here again the r.m.s. deviations between corresponding subunits remain less than 0.15 Å and those between tetramers less than 0.25 Å. PNA9PG, PNA9IA and PNA10PG grew in the presence of peptides. All of them are monoclinic and isomorphous to PNALacM grown in the presence of lactose, but without any peptide in the medium. All four were crystallized at nearly the same pH. Of these, PNA9PG and PNA10PG have resolutions of 3 Å or worse. PNA9IA and PNALacM, however, have comparable resolutions. Both contain two tetramers in the asymmetric unit. Following Ravishankar *et al.* (2001), the first tetramer is made up of subunits *ABCD* and the second subunit *EFGH* (Fig. 1). In terms of their location in the tetramer, *A* is equivalent to *E*, *B* to *F*, *C* to *G* and *D* to *H*. A comparison between the two should provide a measure of the combined effect of lactose and the peptides. Even in this comparison the r.m.s. deviations in C^α positions between corresponding subunits are less than 0.25 Å (Table 5). That between corresponding tetramers is 0.3 Å.

The two monoclinic crystals provide good systems for exploring the effect of ligand binding on overall structure. In PNALacM, the binding sites in *A*, *B* and *E* are occupied by lactose, while that in *F* interacts with a loop from a neighbouring molecule. The binding sites in the remaining four subunits are empty. In PNA9IA, the interaction with the loop

exists in subunit *F*. The binding sites in the other seven subunits are empty. The r.m.s. deviations in C^α positions where the corresponding subunits are superposed (*A* and *E etc.*) are in the range 0.17–0.20 Å in PNALacM. The range is 0.14–0.18 Å in PNA9IA. The superposition of lactose subunits *A*, *B* and *E* in PNALacM with the remaining five subunits, which have either an empty binding site or one interacting with a loop, yields r.m.s. deviations in the range 0.24–0.31 Å when the superposition of corresponding subunits are excluded. The same kind of superpositions yield r.m.s. values in a similar range in PNA9IA, in spite of the fact that subunits *A*, *B* and *E* also have empty subunits in the structure. In the next step, subunits in one structure were superposed on those in the other. When the superposition is between corresponding subunits (*A* on *A* or *E etc.*), the r.m.s. deviation in C^α positions ranged between 0.18 and 0.24 Å. It is interesting to compare this with superpositions involving subunits with empty binding sites on the one hand and sugar-bound subunits on the other. These superpositions led to r.m.s. deviations ranging from 0.23 to 0.33 Å. Thus, the structures of corresponding subunits are closer to each other than those of subunits with the same kind of occupancy (empty or sugar-bound) of the binding site. In other words, the effect of ligand binding is not greater than that caused by differences in the position of the subunits in the quaternary structure.

It is also interesting to examine the effect of crystal packing on the overall structure of the molecule. PNALacM, PNALacA and PNALacB were grown at acidic pH in the presence of lactose. The first is monoclinic with two tetramers in the asymmetric unit, while the other two are orthorhombic with one molecule in the asymmetric unit. The two crystallographically independent molecules in the monoclinic crystals have very similar structures, with an r.m.s. deviation on superposition of the two tetramers of 0.20 Å (Table 4). The C^α positions in the tetramers in PNALacA and PNALacB superpose with an r.m.s. deviation of 0.14 Å. However, the tetramers in the orthorhombic crystals on one hand and those in the monoclinic crystals on the other superpose with r.m.s. deviations ranging from 0.50 to 0.56 Å, values larger than those caused by change in pH, ligand binding and the presence or absence of lactose or the peptides.

Admittedly, the r.m.s. deviations used in the above discussion have small values and one often relies on small differences. However, each r.m.s. deviation represents an average over about 230 values in the case of subunits or more than 900 values in the case of tetramers. Furthermore, the results are entirely internally consistent. For instance, in monoclinic PNALacM each of the four r.m.s. deviations obtained by superposing corresponding subunits is less than 0.20 Å, while all the r.m.s. deviations (14 in all) obtained when the sugar-bound subunits A, B, E are superposed on the other five subunits (excluding corresponding ones) are greater than 0.24 Å. Thus, the overall conclusions arrived at on the basis of r.m.s. deviations appear to be reliable. The main conclusion undoubtedly is that the lectin molecule is very sturdy and the effect of variation in pH, ligand binding and the presence of the peptides around the molecule on the overall tertiary and quaternary structure are less than that of normal differences in crystal packing or position of the subunit in the quaternary assembly.

3.3. Rigid and flexible regions of the molecule

Although the effects of environment or ligand binding on the structure of the peanut lectin molecule are small, they nevertheless exist. The effects vary from one region of the molecule to another. In fact, the delineation of the rigid and flexible regions of the molecule by comparing structures determined under different conditions has been found to be useful and informative in earlier studies in this laboratory on lysozyme and ribonuclease A (Madhusudan & Vijayan, 1991; Radha Kishan *et al.*, 1995; Sadasivan *et al.*, 1998; Biswal *et al.*, 2000). The number of available peanut lectin structures determined in crystals grown under different conditions is large enough to attempt a similar exercise in the case of the lectin. The number of subunits in structures determined to a resolution of 3 Å or better involving crystals grown in the presence of lactose or the peptides or both works out to be 52. These structures form a good database for such an attempt.

When using the *Error-inclusive Structure Comparison and Evaluation Tool* (*ESCET*; Schneider, 2002), the parameter σ is employed to divide a subunit into rigid and flexible regions.

This parameter is calculated from the error estimate in the coordinates of the structures being compared. At the 0.7σ level as the lower limit, nearly half (120) of a total of 232 defined residues in the subunit fall into the rigid regions and the remainder into the flexible regions. At the 0.8σ level, only 35 residues fall into the flexible regions. These 35 may be designated as belonging to the most flexible regions. The rigid, flexible and most flexible regions defined in this manner are illustrated in Fig. 2.

A subunit of the peanut lectin molecule and indeed of any legume lectin molecule is made up of a six-stranded nearly flat back β-sheet (residues 2–8, 64–70, 162–168, 173–179, 186–192 and 222–229), a curved, seven-stranded front β-sheet (18–23, 42–49, 84–90, 117–124, 136–143, 149–153 and 203–210) and a small five-stranded sheet (25–27, 31–34, 71–74, 160, 161 and 217–220), which has a major role in holding together the two larger sheets and loops (Banerjee *et al.*, 1996). The main hydrophobic core of the subunit is bound by the three sheets. Loop 91–116 and to some extent loop 144–148 curve up and loop 125–135 curves down in such a way that they interact among themselves and with the curved sheet. The arrange-
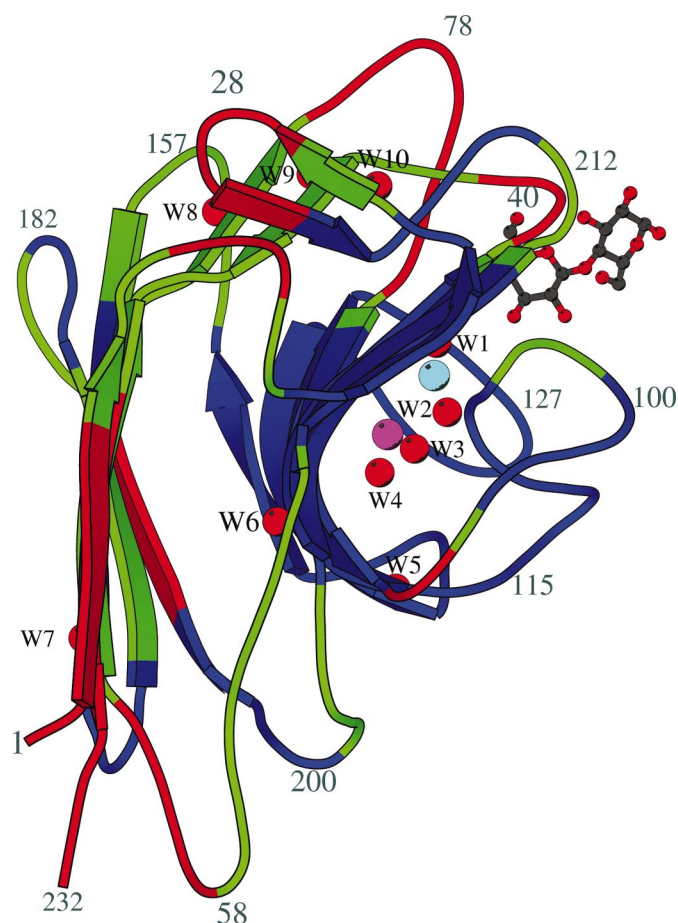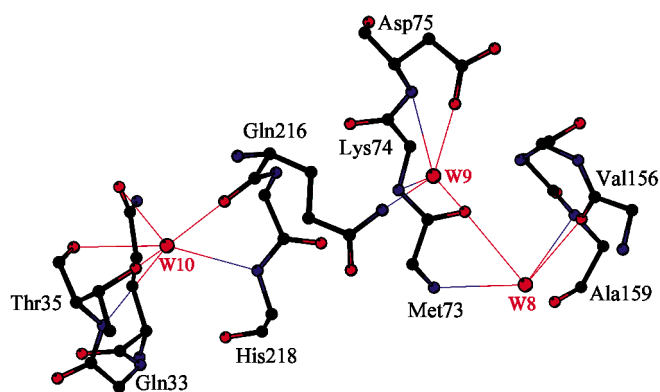

Figure 2
Rigid (blue), flexible (green) and most flexible (red) regions in a PNA subunit. Pink and cyan balls represent metal ions. Red balls represent invariant water molecules obtained through the superposition of 16 subunits along with their hydration shells. The lactose molecule is in ball-and-stick representation. This and the subsequent figures were prepared using *MOLSCRIPT* (Kraulis, 1991).

ment of loops is such that it gives rise to a second hydrophobic core between the curved sheet and the loops (Banerjee *et al.*, 1996). The carbohydrate-binding site, made up of loops 91–106, 125–135, 75–83 and 211–216, is situated approximately on the top of the second hydrophobic core. The site is also close to the two metal ions in the structure.

Interestingly, much of the curved sheet and the three loops that along with this sheet give rise to the second hydrophobic core are in the rigid region. The residues (121, 123, 125, 127, 132 and 137) that interact with the metal ions belong to this region. Metal binding could be a factor that confers rigidity to this region. The small β-sheet is substantially flexible. A vast majority of the residues in the flat back β-sheet are flexible. Some even belong to the most flexible region. This sheet is invariably involved in quaternary interactions and usually occurs in intersubunit interfaces (Prabu *et al.*, 1999). Legume lectins are known to exhibit considerable variability in quaternary association (Vijayan & Chandra, 1999) and the flexibility of this sheet is perhaps necessary to accommodate this variability. The above results are also consistent with the observation of a monomeric unfolding intermediate of peanut lectin, which retains much of the carbohydrate-binding specificity and activity of the native lectin (Reddy, Srinivas *et al.*, 1999; Reddy, Bharadwaj *et al.*, 1999).

Of the four loops that constitute the carbohydrate-binding site, those made up of residues 75–83, 91–106 and 125–135 are relatively invariant in legume lectins, while the D loop (211–216), which is primarily involved in conferring specificity, is variable in length and composition (Sharma & Surolia, 1997). Not surprisingly, the D-loop is in a flexibile region. Interestingly, loop 75–83 is also almost entirely in a most flexible region. The central stretch of loop 91–106 forms part of the major rigid region around the second hydrophobic core, but the region of the loop close to the sugar is flexible but not highly flexible. Loop 125–135 is wholly in the major rigid region. Therefore, there is no direct correlation between conservation of size and sequence and flexibility. There are ten residues (80, 82, 83, 103, 104, 125, 127, 211, 212 and 214) that are in contact with the sugar. Of these, only two (125 and 127) are rigid on the basis of the criterion adopted in the present analysis for distinguishing between rigid and flexible regions.

Of the rest, four can be considered as highly flexible. This does not mean that the sugar-binding site is not sufficiently preformed. As shown earlier, the structure of the molecule as a whole is well conserved with respect to environmental effects and ligand binding. The sugar-binding site is a relatively flexible region of the somewhat rigid molecule.

3.4. Invariant water molecules

Structures of peanut lectin and its complexes determined to a resolution of 2.5 Å or better were used to examine the invariant features in the hydration shell of the protein. As in

Figure 3
The interactions of the invariant water molecules in a locally convoluted feature involving the small β-sheet.
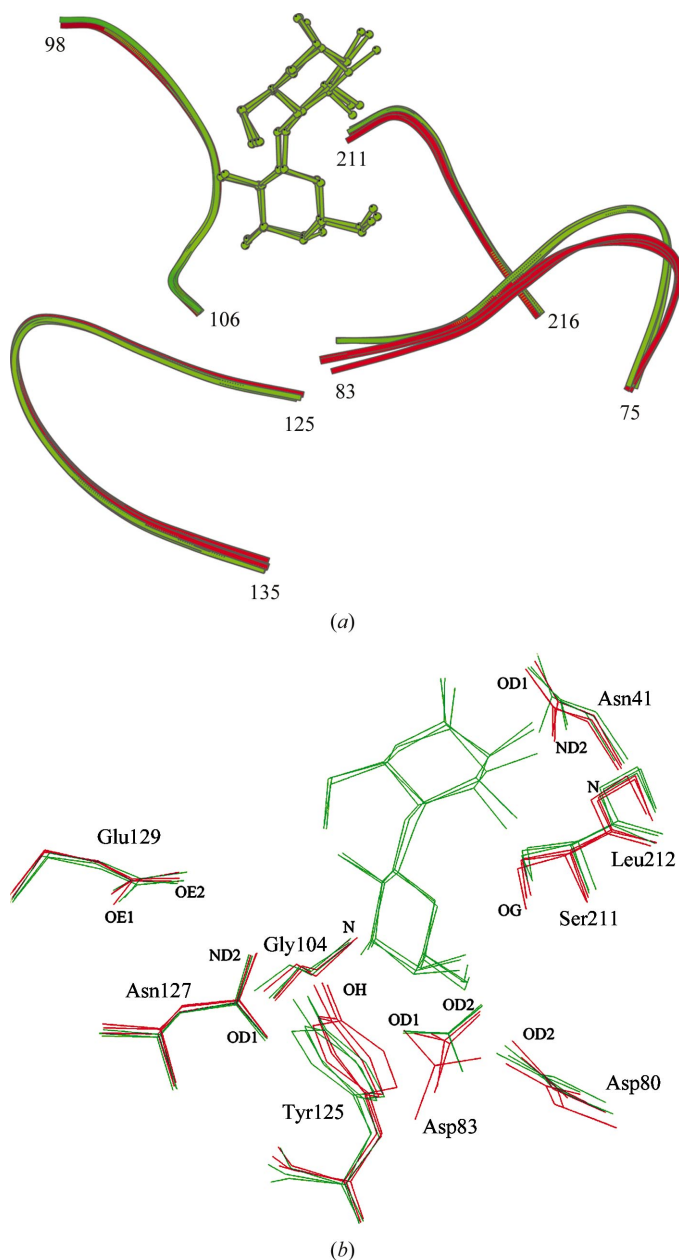
Figure 4
(*a*) Superpositions of the carbohydrate-binding loops in sugar-bound subunits *A*, *B* and *E* in PNALacM (green) and that of the same loops in the empty subunits *A*, *B* and *E* in PNA9IA (red). (*b*) Superpositions of the side chains close to the position of the sugar in the loop illustrated in (*a*). The same colour code is used.

our earlier studies on hydration and its consequences (Sadasivan *et al.*, 1998; Biswal *et al.*, 2000), a water molecule at a distance of 3.6 Å or less from a protein N or O atom of a subunit was considered to be part of its hydration shell. When two subunits are superposed along with their hydration shell, a water molecule in the hydration shell of one and a water molecule in that of the other are considered equivalent if the two water molecules have at least one common interaction with the same protein atom and if the two are within 1.8 Å of each other on superposition. The water molecules that are equivalent in all pairwise superpositions are considered as invariant. The above criterion, which is rather stringent especially when one is dealing with 16 subunits as in the present case, led to the identification of ten invariant water molecules (Fig. 2). Four of them are involved in metal binding embedded in the major rigid region (W1 to W4). They were found in an earlier study to be invariant in all legume lectin structures available at that time (Loris *et al.*, 1994). The same is true of two more water molecules, one involved with a β-bulge (W5) and the other in a hairpin bend (W6). Of the remaining four invariant water molecules in peanut lectin, one (W7) interconnects two strands in the back β-sheet close to a hairpin bend. The remaining three (W8–W10) appear to stabilize a somewhat locally convoluted structural feature involving the small β-sheet (Fig. 3). One of them (W8) additionally stabilizes the 156–159 β-bend and connects the bend to residue 73 in the small β-sheet. Another water molecule (W9) primarily holds together residues 73 and 74 in the sheet and the residue (75) immediately following them. It also appears to interact with residue 216, which is just outside the sheet on the same side as 75. Yet another invariant water molecule (W10) connects the 216–218 stretch to the 33–35 stretch. 217 and 218 on one hand and 33 and 34 on the other belong to the small sheet. Residue 35, like 216, is just outside the sheet on the same side as 75.

PNA is an asymmetric tetramer. The r.m.s. deiviations on superposition among subunits from different structures indicated slight structural differences among subunits depending upon their location in the tetramer. Therefore, it was of interest to examine the hydration of the tetramer as a whole. Superposition of the four tetramers and their hydration shells in the four structures led to the identification of 170 invariant water molecules associated with the tetramer. The number of these associated solely with one subunit varied between 24 and 48. In addition, there were 16 water molecules associated with pairs of subunits. Interestingly, invariant water molecules common to all the four subunits turned out to be the same ten obtained by the superposition of the 16 subunits along with their separate hydration shells, lending additional credence to the method employed for identifying invariant water molecules. The 16 invariant water molecules associated with pairs of subunits are among those involved in intersubunit interactions through water bridges (Banerjee *et al.*, 1996). These interactions between subunits *A* and *B* are particularly interesting. These two subunits have an arrangement similar to that in a concanavalin dimer, except that the two flat β-sheets do not form a contiguous 12-stranded sheet. Instead, each

hydrogen bond that connects the two sheets is replaced by a water bridge in PNA (Banerjee *et al.*, 1996). There are six water molecules involved in these water bridges. Three of them remain invariant.

3.5. Lectin–ligand interactions and the binding site

In the structures grown in the presence of lactose, in all the subunits in which the ligand is bound the peanut lectin–sugar interactions originally described in the orthorhombic crystals (Banerjee *et al.*, 1996) are almost entirely preserved, in spite of the variation in pH and crystal packing and the presence of peptides. Likewise, the same interactions occur in cases where the binding site is occupied by a loop from a neighbouring molecule.

A comparison between the sugar-bound sites in subunits *A*, *B* and *E* in PNALacM and the corresponding empty sites in PNA9IA can be used to explore the geometrical differences, if any, between empty and sugar-bound combining sites. The superposition of the sugar-binding loops in the two sets of structures is shown in Fig. 4(*a*). Three of the loops do not undergo any significant conformational change. The 73–83 loop, however, exhibits conformational change in all the three subunits. It may be recalled that this loop falls almost entirely in the most flexible region. The conformational change of the loop is the same in the three subunits, indicating its significance. The conformational change is such that the C^α positions in residues 77, 78 and 83 move consistently by about 1 Å on ligand binding. 77 moves towards the other three loops, while the movement of 78 cannot be described in any simple manner. In any case, these residues are at a tip of the molecule far away from the location of the sugar. 83 moves away from the 125–135 loop towards the 211–216 loop. The distance of its C^α from the position of the centre of the mass of the galactose residue in the complexes, however, remains largely unchanged. The changes in side chains in the immediate neighbourhood of the molecule, illustrated in Fig. 4(*b*), caused by ligand binding are also small. The most conspicuous, though small, change is that in the side chain of Tyr125, which stacks against the galactose ring. The change is caused primarily by the rotation and the translation of the residue, the conformation of which as a whole remains nearly the same. There is considerable movement in the side chain of Asp83. Here again, change in the conformational angles is small. However, translations, which are greater than 1 Å in a few cases, occur. The above observations confirm that the combining site of the lectin is substantially preformed, with some flexibility in a couple of residues.

References

Adhikari, P., Bachhawat Sikder, K., Thomas, C. J., Ravishankar, R., Jeyaprakash, A. A., Sharma, V., Vijayan, M. & Surolia, A. (2001). *J. Biol. Chem.* **276**, 40734–40739.

Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1996). *J. Mol. Biol.* **259**, 281–296.

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A. & Vijayan, M. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 227–231.

Biswal, B. K., Sukumar, N. & Vijayan, M. (2000). *Acta Cryst.* D**56**, 1110–1119.

Brünger, A. T., Adams, P. D. & Rice, L. M. (1998). *Curr. Opin. Struct. Biol.* **8**, 606–611.

Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

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

Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.

Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.

Lis, H. & Sharon, N. (1998). *Chem. Rev.* **98**, 637–674.

Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). *Biochim. Biophys. Acta*, **1383**, 9–36.

Loris, R., Stas, P. P. & Wyns, L. (1994). *J. Biol. Chem.* **269**, 26722–26733.

Madhusudan, R. & Vijayan, M. (1991). *Curr. Sci.* **60**, 165–170.

Majumdar, T. & Surolia, A. (1978). *Prep. Biochem.* **8**, 119–131.

Manoj, N., Srinivas, V. R., Surolia, A., Vijayan, M. & Suguna, K. (2000). *J. Mol. Biol.* **302**, 1129–1137.

Manoj, N. & Suguna, K. (2001). *Protein Eng.* **14**, 735–745.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. & Suguna, K. (1998). *J. Mol. Biol.* **276**, 787–796.

Prabu, M. M., Suguna, K. & Vijayan, M. (1999). *Proteins Struct. Funct. Genet.* **35**, 58–69.

Pratap, J. V., Bradbrook, G. M., Reddy, G. B., Surolia, A., Raftery, J., Helliwell, J. R. & Vijayan, M. (2001). *Acta Cryst.* D**57**, 1584–1594.

Radha Kishan, K. V., Chandra, N. R., Sudarsanakumar, C., Suguna, K. & Vijayan, M. (1995). *Acta Cryst.* D**51**, 703–710.

Ravishankar, R., Ravindran, M., Suguna, K., Surolia, A. & Vijayan, M. (1997). *Curr. Sci.* **72**, 855–861.

Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1999). *Acta Cryst.* D**55**, 1375–1382.

Ravishankar, R., Surolia, A., Vijayan, M., Lim, S. & Kishi, Y. (1998). *J. Am. Chem. Soc.* **120**, 11297–11303.

Ravishankar, R., Thomas, C. J., Suguna, K., Surolia, A. & Vijayan, M. (2001). *Proteins Struct. Funct. Genet.* **43**, 260–270.

Reddy, G. B., Bharadwaj, S. & Surolia, A. (1999). *Biochemistry*, **38**, 4464–4470.

Reddy, G. B., Srinivas, V. R., Ahmad, N. & Surolia, A. (1999). *J. Biol. Chem.* **274**, 4500–4503.

Sadasivan, C., Nagendra, H. G. & Vijayan, M. (1998). *Acta Cryst.* D**54**, 1343–1352.

Schneider, T. R. (2002). *Acta Cryst.* D**58**, 195–208.

Sharma, V., Srinivas, V. R., Adhikari, P., Vijayan, M. & Surolia, A. (1998). *Glycobiology*, **8**, 1007–1012.

Sharma, V. & Surolia, A. (1997). *J. Mol. Biol.* **267**, 433–445.

Sharma, V., Vijayan, M. & Surolia, A. (1996). *J. Biol. Chem.* **271**, 21209–21213.

Vijayan, M. & Chandra, N. (1999). *Curr. Opin. Struct. Biol.* **9**, 707–714.