

Signature of quaternary structure in the sequences of legume lectins

N.Manoj and K.Suguna¹

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

¹To whom correspondence should be addressed.
E-mail: suguna@mbu.iisc.ernet.in

Legume lectins exhibit a wide variety of oligomerization and sugar specificity while retaining the characteristic jelly-roll tertiary fold. An attempt has been made here to find whether this diversity is reflected in their primary structures by constructing phylogenetic trees. Dendrograms based on sequence alignment showed clustering related to the oligomeric nature of legume lectins. Though the clustering primarily follows the oligomeric states, it also appears to correlate with different sugar specificities indicating an interdependence of these two properties. Analysis of the structure-based alignment and the alignment of the sequences of the carbohydrate-binding loops alone also revealed the same features. By a close examination of the interfaces of the various oligomers it was also possible, in some cases, to pinpoint a few key residues responsible for the stabilization of the interfaces.

Keywords: interfaces/legume lectin/phyletic tree from crystal structures/quaternary structure/sequence analysis

Introduction

Lectins are multivalent carbohydrate-binding proteins of non-immune origin with a high degree of specificity for cell-surface carbohydrates. They are present in a variety of organisms, performing diverse biological functions involving cellular recognition and interaction (Weis and Drickamer, 1996; Lis and Sharon, 1998). Legume lectins, the most abundant of all classes of lectins, are widely studied in terms of their structural and biological characterization (Loris *et al.*, 1998; Bouckaert *et al.*, 1999; Vijayan and Chandra, 1999). On the basis of their monosaccharide specificity, legume lectins can be classified into five groups: mannose/glucose (Man/Glc), galactose/*N*-acetylgalactosamine (Gal/GalNAc), *N*-acetylglucosamine (GlcNAc), fucose and *N*-acetylneuraminic acid. The mode of carbohydrate binding by legume lectins has been understood in terms of the sugar-combining site made up of four loops (Young and Oomen, 1992; Sharma and Surolia, 1997). The diverse sugar specificities of these lectins have been explained, at least at the monosaccharide level, in terms of the variations in the length, composition and interactions of one of the four loops. Crystal structures of a number of legume lectins representing all five classes of sugar specificity have been determined. Also, the sequences of a larger number of legume lectins are available in the sequence databases. These lectins exhibit a high degree of homology, with a sequence identity ranging from 28 to 99% among those with known three-dimensional structure. The tertiary structures of

all of them are the same except for some variations in the loops. However, they exhibit a wide range of carbohydrate specificity and oligomeric structure. Thus, they are good candidates for studies on evolutionary relationships of structure and function.

Legume lectins of known three-dimensional structure, their oligomeric state and carbohydrate specificity are listed in Table I. As illustrated in Figure 1, the protomer of each is made up of a six-stranded nearly flat 'back' β -sheet, a seven-stranded curved 'front' β -sheet, a short five-membered β -sheet at the 'top' of the molecule and several loops that connect the sheets. All of them are dimers or tetramers that can be considered as dimers of dimers. Each tetramer has three types of interfaces. These interfaces have varying degrees of similarity, ranging from very close to broad, with those found in the dimeric proteins. All these interfaces involve the six-stranded back β -sheet of the monomer in one way or the other and it is possible to describe each of them in terms of the mutual disposition of the sheets in the two participating subunits. The observed modes of quaternary association have been rationalized in terms of hydrophobic surface buried on oligomerization, interaction energy and shape complementarity (Prabu *et al.*, 1999). Dimerization in a majority of instances involves a side-by-side arrangement, resulting in a contiguous 12-stranded β -sheet with the dyad axis perpendicular to the β -sheet. This kind of association first observed in ConA (Hardman and Ainsworth, 1972) may be described as II-type (Jones and Thornton, 1995). All abbreviations of lectin names are listed in Table II. Dimerization in other instances involves different kinds of back-to-back association of the six-stranded β -sheets (named as X1, X2, X3 and X4 types).



Fig. 1. Subunit of peanut agglutinin.

Table I. Legume lectins with known three-dimensional structures available in the Protein Data Bank (PDB)

Lectin	Specificity	Abbreviation	Oligomeric state	PDB code	Nature of interface(s)
<i>Phaseolus vulgaris</i>	Complex	PHAL	4	1FAT	Two II-types and two X1-types
<i>Glycine max</i>	Gal/GalNAc	SBA	4	2SBA	
<i>Ulex europeaus</i>	GlcNAc	UEAII	4	1QOO	
<i>Dolichos biflorus</i>	GalNAc	DBL	4	1LU1	
<i>Maackia amurensis</i>	SialylLactose	MAL	4	1DBN	
<i>Canavalia ensiformis</i>	Man/Glc	ConA	4	5CNA	Two II-types and two X2-types
<i>Canavalia brasiliensis</i>	Man/Glc	AZD	4	1AZD	
<i>Dolichos lablab</i>	Man/Glc	DIAB	4	1QMO	
<i>Dioclea grandiflora</i>	Man/Glc	DGL	4	1DGL	
<i>Arachis hypogaea</i>	Gal	PNA	4	2PEL	One II-type, two X4-types and one unusual
<i>Psophocarpus tetragonolobus</i>	Gal/GalNAc	WBAI	2	1WBL	X3-type
		WBAIL	2	1F9K	
<i>Erythrina corallodendron</i>	Gal/GalNAc	EcorL	2	1LTE	X4-type
<i>Griffonia simplicifolia</i> IV	Complex	GS4	2	1LED	
<i>Lathyrus ochrus</i> I	Man/Glc	LOLI	2	1LOA	II-type
<i>Pisum sativum</i>	Man/Glc	PSL	2	1RIN	
<i>Lens culinaris</i>	Man/Glc	LENL	2	1LES	
<i>Ulex europeaus</i>	Fucose	UEAI	2	1FX5	
Dolichos biflorus (stem and leaf lectin)	GalNAc	DB58	2	1LUL	X1-type

Table II. Some legume lectins for which information on quaternary structure or/and specificity are available

Lectin	Source	Specificity	Sequence ID
WBAI	<i>Psophocarpus tetragonolobus</i>	GalNAc	LEC_PSOTE
WBAIL	<i>Psophocarpus tetragonolobus</i>	GalNAc	Q9SM56
EcorL	<i>Erythrina corallodendron</i>	GalNAc	LEC_ERYCO
EvarL	<i>Erythrina variegata</i>	GalNAc	JX0289
DBL	<i>Dolichos biflorus</i>	GalNAc	LEC1_DOLBI
DB58	<i>Dolichos biflorus</i>	GalNAc	LEC5_DOLBI
SJAL	<i>Sophora japonica</i>	GalNAc	LECS_SOPJA
VML	<i>Vatairea macrocarpa</i>	GalNAc	LECS_VATMA
CSII	<i>Cytisus scoparius</i>	GalNAc	LEC2_CYTSC
SBA	<i>Glycine max</i>	GalNAc	LEC_SOYBN
GS4	<i>Griffonia simplicifolia</i>	Complex	LEC4_GRISI
BPL	<i>Bauhinia purpurea</i>	GalNAc	LEC_BAUPU
PHAL	<i>Phaseolus vulgaris</i>	Complex	PHAL_PHAVU
PHAE	<i>Phaseolus vulgaris</i>	Complex	PHAE_PHAVU
PNA	<i>Arachis hypogea</i>	Gal	LECG_ARAHY
LAA	<i>Laburnum alpinum</i>	GlcNAc	LEC1_LABAL
UEAII	<i>Ulex europeaus</i>	GlcNAc	LEC2_ULEEU
GS2	<i>Griffonia simplicifolia</i>	GlcNAc	Q41263
UEAI	<i>Ulex europeaus</i>	L-Fuc	LEC1_ULEEU
LTA	<i>Lotus tetragonolobus</i>	L-Fuc	LEC_LOTTE
LOLI	<i>Lathyrus ochrus</i>	Man/Glc	LECB_LATOC
PSL	<i>Pisum sativum</i>	Man/Glc	LEC_PEA
LENL	<i>Lens culinaris</i>	Man/Glc	LEC_LENCU
Favin	<i>Vicia faba</i>	Man/Glc	LEC_VICFA
MTA	<i>Medicago truncatula</i>	Man/Glc	LEC_MEDTA
LSL	<i>Lathyrus sphaericus</i>	Man/Glc	LEC_LATSP
OVL	<i>Onobrychis vicifolia</i>	Man/Glc	LEC_ONOVI
ConA	<i>Canavalia ensiformis</i>	Man/Glc	CONA_CANEN
AZD	<i>Canavalia brasiliensis</i>	Man/Glc	G222590
DGL	<i>Dioclea grandiflora</i>	Man/Glc	LECA_DIOGR
DIAB	<i>Dolichos lab lab</i>	Man/Glc	LECA_DOLLA
BMA	<i>Bowringia mildibraedii</i>	Man/Glc	LEC_BOWMI
MAL	<i>Maackia amurensis</i>	SialylLactose	G257094

Various types of dimeric associations that are observed in the structures of legume lectins are schematically shown in Figure 2. Among the dimeric lectins, PSL (Einspahr *et al.*, 1986), Favin (Reeke and Becker, 1986), LOLI (Bourne

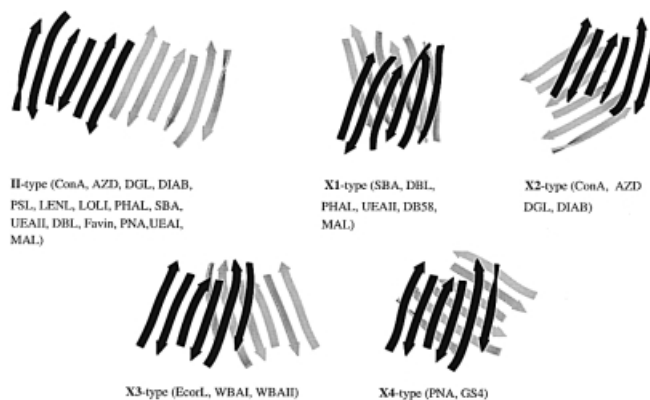


Fig. 2. Schematic representation of the back β -sheets involved in the different modes of dimerization in legume lectins.

et al., 1990), LENL (Loris *et al.*, 1993) and UEAI (Audette *et al.*, 2000) associate in II-type fashion (Figures 2 and 3a). In the 10 tetramers of known structure except PNA (ConA, AZD, DIAB, DGL, SBA, PHAL, UEAI, DBL and MAL), subunits 1 and 2, and 3 and 4 associate in a side-by-side fashion (II-type) (Figures 4 and 5). All these tetramers can be considered as resulting from II-type associations of X-type dimers. The 1–4 and the 2–3 interfaces in SBA (Dessen *et al.*, 1995), PHAL (Hamelryck *et al.*, 1996), UEAI (Dao-thi *et al.*, 1998), DBL (Hamelryck *et al.*, 1999) and MAL (Imberty *et al.*, 2000) are of one kind of back-to-back type (X1-type) while those of ConA, AZD (Sanz-Aparicio *et al.*, 1997), DIAB (Protein Data Bank code: 1QMO) and DGL (Rozwarski *et al.*, 1998) form another type (X2-type). DB58 (Hamelryck *et al.*, 1999), a lectin closely related to DBL but dimeric in nature exhibits the X1-type back-to-back interface (Figures 2 and 3b). PNA represents a unique case of a tetramer without 4-fold or 222 symmetry (Banerjee *et al.*, 1994, 1996). Consequently, the 1–2 and the 3–4 interfaces are not equivalent. It is

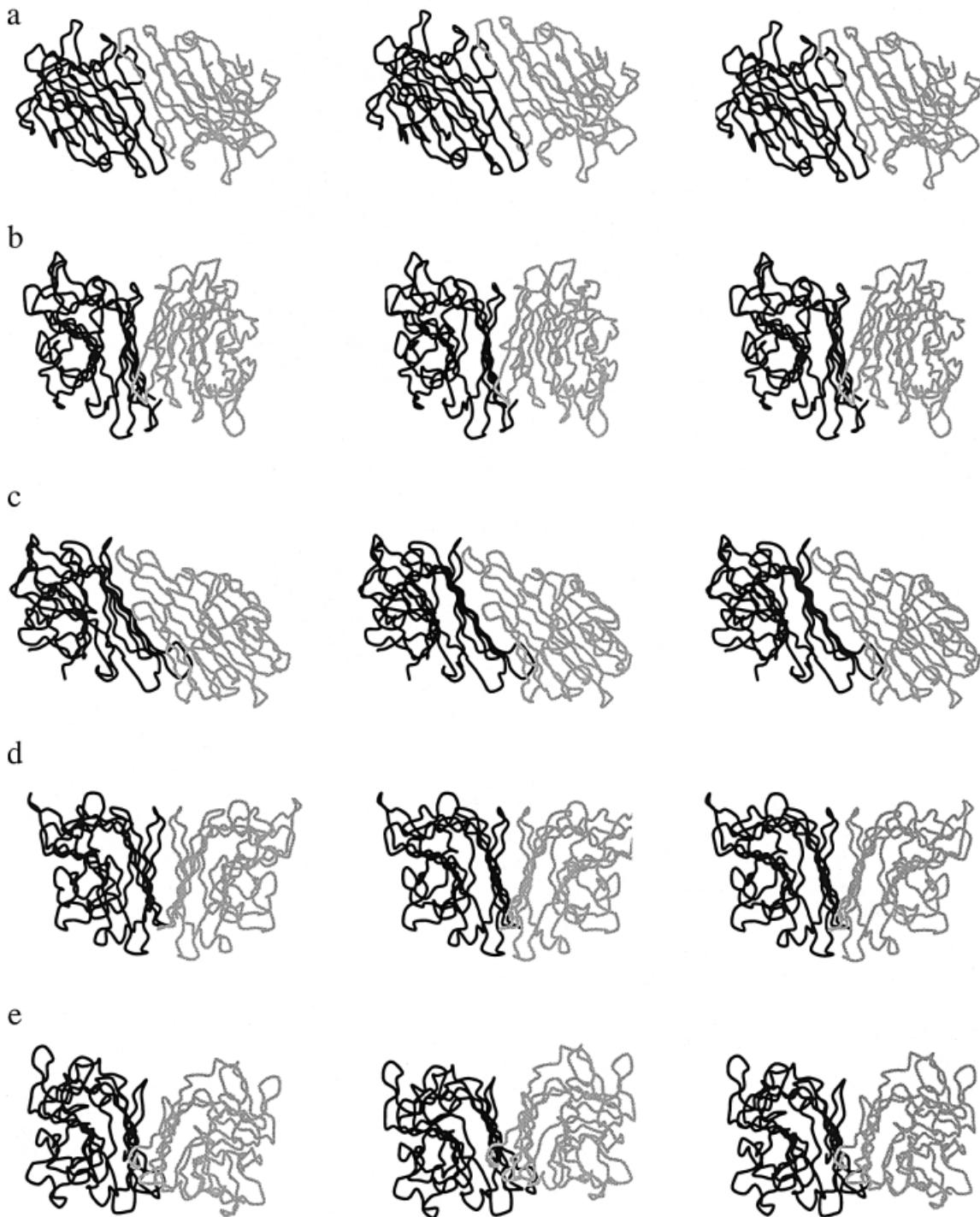


Fig. 3. Stereo views of the dimers of (a) PSL, (b) DB58, (c) ConA, (d) EcorL and (e) GS4.

believed that the 3–4 interface is an incidental consequence of the presence of two dimers with an X4-type interface (1–4 and 2–3) associating with one II-type interface (1–2). Although 1–2 is a side-by-side interface, the two six-membered sheets do not form a contiguous 12-stranded β -sheet, but are connected through a number of interfacial water molecules. The dimeric lectins, EcorL (Shaanan *et al.*, 1991), WBAl (Prabu *et al.*, 1998) and WBAlI (Manoj *et al.*, 2000) exhibit one kind of back-to-back association (X3-type) while GS4 (Delbaere *et al.*, 1993) exhibits an interface (X4-type) similar to that in PNA (Figure 2 and

3d, e). Thus, all the oligomerization modes observed so far in legume lectins can be explained in terms of the formation of two classes of dimers (II-type and X-type) and further association of two of these dimers into tetramers. It is interesting that although there are four different kinds of X-type interfaces, in all cases the majority of the inter-subunit contacts come from the same fourth, fifth and sixth strands of the back β -sheet and a few additional contacts in each type coming from residues elsewhere in the back β -sheet.

Swamy *et al.* analysed the sequences of four legume

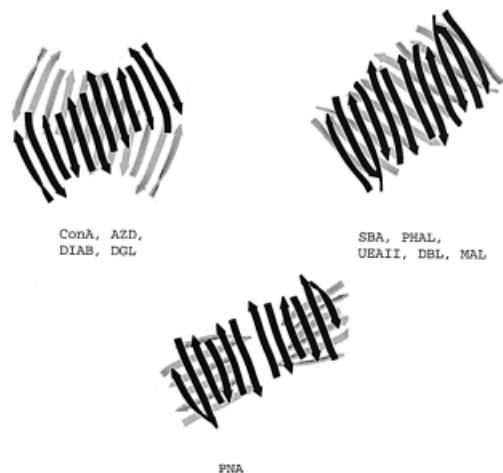


Fig. 4. Schematic representation of the back β -sheets in the tetrameric legume lectins. Subunits 1 and 2 which form a II-type interface are shown in black while subunits 3 and 4 are in grey. This figure and Figure 5 are reproduced from Prabu *et al.* (Prabu *et al.*, 1999).

lectins available at that time and concluded that all of them (ConA, LENL, SBA and Favin) would have the same secondary structural features but classified them into two pairs: ConA–SBA and LENL–Favin (Swamy *et al.*, 1985). However, indications that the sequences of legume lectins also hold the key to their quaternary association emerged from the structural studies that were carried out in our laboratory on peanut lectin (Banerjee *et al.*, 1994, 1996) and winged bean agglutinin (Prabu *et al.*, 1998). To further explore these indications, a detailed analysis of legume lectin sequences has been taken up. As a sufficient number of sequences and structures are now available, it appeared possible to arrive at meaningful and statistically significant conclusions from this type of study. In this paper we present an attempt to understand and identify relationships, if any, within various classes of carbohydrate specificities and modes of oligomerization.

Materials and methods

Sequence and structure sources

The list of all the legume lectin structures whose coordinates are available was obtained from the 3D Lectin Data Bank on World Wide Web URL: <http://www.cermap.cnrs.fr/databank/lectine>. The coordinates were obtained from the Protein Data Bank (Berman *et al.*, 2000) and the sequences from the SWISSPROT data bank (Bairoch and Apweiler, 1997).

Comparisons of legume lectins available in the Protein Data Bank

Multiple alignment of sequences. The multiple sequence alignment was performed using the program MULTALIGN from the AMPS suite of programs (Barton, 1990). This program uses the Needleman and Wuncsh algorithm (Needleman and Wuncsh, 1970; Barton and Sternberg, 1987) with a fixed gap penalty of 8 and the Dayhoff's mutation data matrix (Schwartz and Dayhoff, 1978). There were 19 sequences of legume lectins whose coordinates were available in the Protein Data Bank (Table I). The program ORDER was used to perform cluster analysis and ordering of the sequences by similarity and to construct a dendrogram from the output from a MULTALIGN pairwise run. The cluster

analysis uses the significance scores for the alignment calculated from the mean random score, the match score and standard deviation (SD) score of randomizations to generate a tree file. The lengths of the branches were adjusted to reflect the pairwise alignment scores.

Multiple alignment of sequence based on structures. The alignment of sequences based on three-dimensional structures was performed using the program STAMP (SStructural Alignment of Multiple Proteins) (Russel and Barton, 1992). STAMP makes use of the rigid body least squares superposition of C_{α} positions (Rossmann and Argos, 1975) for expressing the probability of equivalence of residue structural equivalence. A preliminary multiple sequence alignment is performed using sequence information, which then determines an initial superposition of the structures. A structure comparison algorithm is applied to all pairs of proteins in the superimposed set and a similarity tree calculated. Multiple sequence alignments are then generated by following the tree from the branches to the root. At each branch point of the tree, a structure-based sequence alignment and coordinate transformations are output, with the multiple alignment of all structures output at the root. The tree topologies and branch lengths for the phylogenetic tree were determined from the sequence distance matrices using the program KITSCH from the PHYLIP suite of programs (Felsenstein, 1985). This method accounts for unequal rates of change among the proteins by adjusting distances so that the branch lengths, from the root of the tree to the tip of each of its leaves, are equidistant.

Comparisons of legume lectins available in the sequence database

An analysis of the sequences of legume lectins for which some information about their carbohydrate specificity and/or quaternary structures are available in the sequence databases was also performed. The list of 33 lectins and their sources and specificities are given in Table II. These sequences were aligned and their phylogenetic tree constructed using the programs available in the Wisconsin Sequence Analysis package [Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, WI]. The program PILEUP was used for the multiple sequence alignment and the program PAUPSEARCH (Phylogenetic Analysis using Parsimony) was used for constructing a phylogenetic tree starting from an aligned sequence that is optimal according to parsimony criteria. The program constructs a neighbour-joining tree and the best tree is the one with the minimum sum of branch lengths based on a corrected distance matrix calculated from the aligned sequences. Confidence values were evaluated using bootstrapping replications and a consensus bootstrap tree was obtained. The program PAUPDISPLAY was used to plot the tree.

Results and discussion

The alignment scores as expressed in units of SD above the mean background obtained for comparison of unrelated sequences of identical length and amino acid composition calculated by MULTALIGN ranges from 15.1 to 26.6 SD for all the pairs of sequences. The total number of residues conserved in all sequences is 69, out of which 21 residues are identical. There are nine sites where significant gaps have been introduced. Figure 6 shows the unrooted dendrogram obtained on the basis of the sequence alignment. The

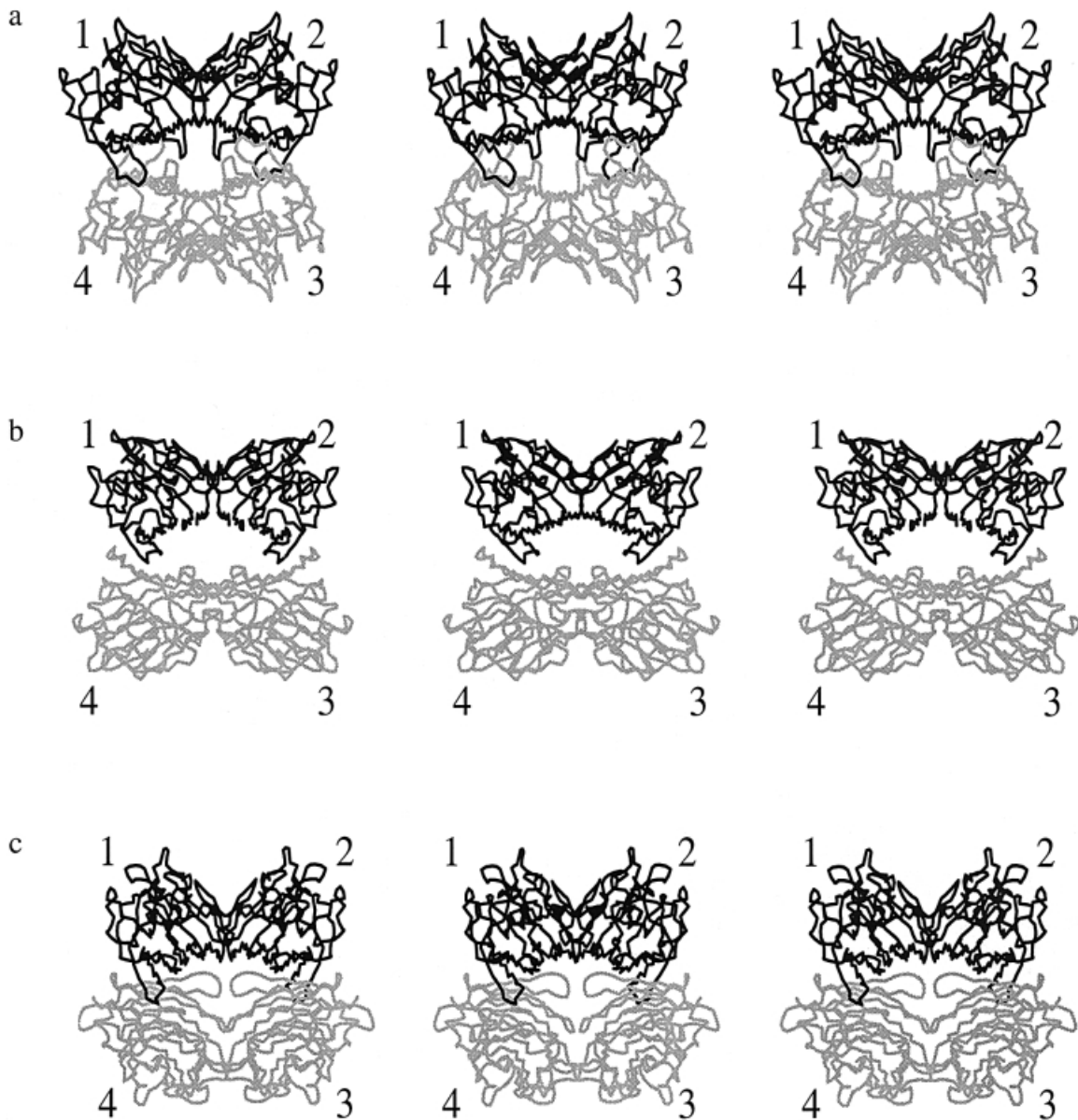


Fig. 5. Stereo view of the tetramers in (a) ConA, (b) SBA and (c) PNA.

lengths of the branches represent the approximate evolutionary distance between the sequences. The tree has six major divisions: (1) X1-type, (2) X3-type, (3) II-type, (4) II,X2-type, (5) II,X4-type and (6) X4-type. The first one includes tetrameric lectins SBA, PHAL, DBL, UEAI and MAL and a dimeric lectin DB58. All the lectins in this branch are made up of X1-type dimers and most of them are specific to Gal/GalNAc at the monosaccharide level. One of the sub-branches consists of UEAI and MAL which have unique specificities. The second major branch includes Gal/GalNAc-binding lectins WBAI, WBAA and EcorL. These lectins are dimers of the X3-type. Branch 3 consists of lectins LENL, Favin, LOLI and PSL that are Man/Glc binding and are dimers of the II type. Branch 4 has tetrameric lectins made up of X2-type dimers. Again, all these lectins are Man/Glc specific. DIAB should have come closer to the ConA-type of tetramers but probably having a two-chain subunit unlike the others, is placed slightly away. Branch 5 consists of PNA, a tetramer of a unique

type and is specific to Gal made up of two X4-type dimers. Branch 6 consists of GS4, which binds to complex carbohydrates by accommodating a GalNAc in the primary-binding site and is an X4-type dimer. These two branches are evolutionarily as distant from each other as they are from the others. The major branches are related to the classification of quaternary structures. Further, the branches also appear to reflect the functional divergence of these lectins in terms of their carbohydrate specificity. The clustering we have derived thus shows a strong correlation to the structural classification which most likely would have evolved to reflect biological activity of legume lectins.

A similar alignment and clustering of the sequences corresponding to only the four sugar-binding loops comprising of only approximately 50 residues was performed to examine the extent to which the binding-site loops reflect the respective branch positions in the dendrogram obtained previously by whole length sequence comparisons. Figure 7 shows the dendrogram obtained for the sugar-binding loops.

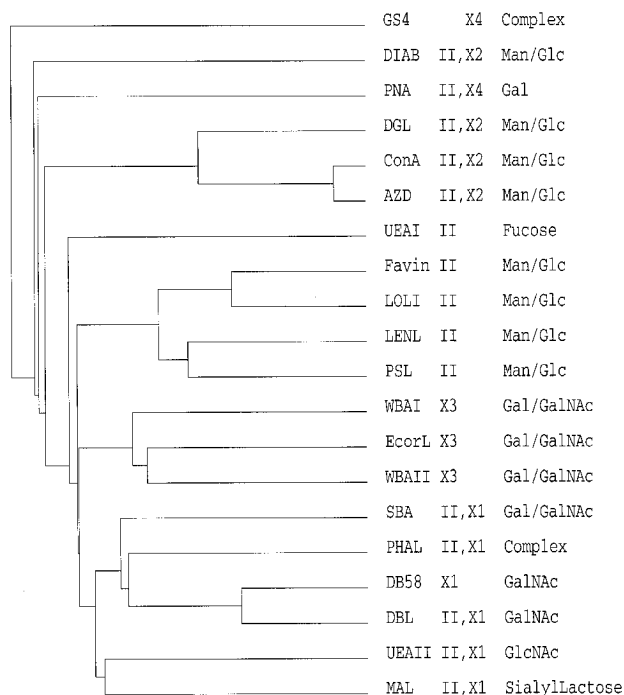


Fig. 6. Dendrogram showing the relationships among the sequences of legume lectins whose structures are available. Branch lengths represent approximate evolutionary distances between the sequences. The type of interface and sugar specificity of the lectins are also indicated in this figure and in Figures 7, 9 and 10.

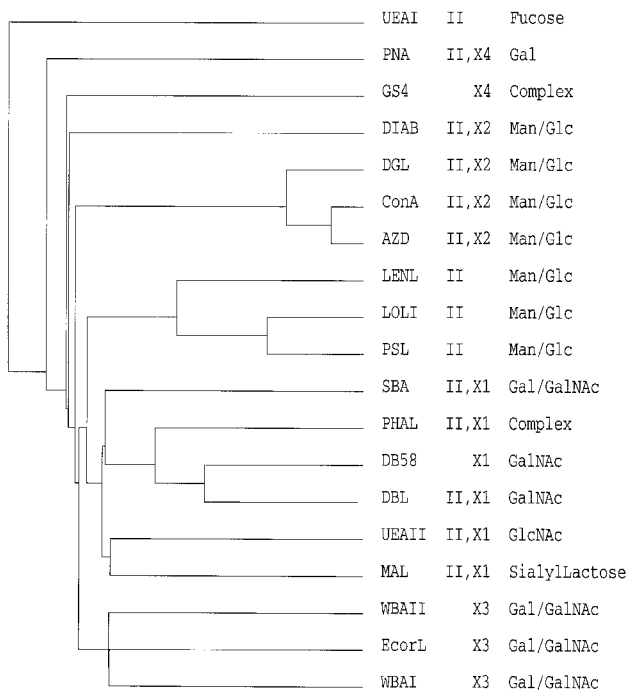


Fig. 7. Dendrogram showing the relationships among the sequences of the sugar-binding loops in legume lectins of known structure.

The tree obtained clearly gave indications of branching that were significant in terms of the quaternary structures and sugar specificity. Surprisingly, branching obtained by using only the binding-site loops preserves the scheme of the clustering obtained using the entire sequences to a large extent. A statistically more significant method of comparing the evolutionary relationships between the sequences is at

their tertiary structure level because most functional restraints on evolutionary divergence operate at the level of tertiary structure (Bajaj and Blundell, 1984). Indeed, protein structures are generally more conserved in evolution than are amino acid sequences. Figure 8 shows the alignment of sequences based on the structures. There are approximately 62 conserved residues in all the sequences out of which 21 are identical. For a typical legume lectin the number of residues in the front, back and top β -sheets are 51, 44 and 19, respectively. The number of conserved residues in the front β -sheet in all the sequences is 25, out of which 10 are identical. The number of residues in the back β -sheet is 13 out of which four are identical while in the top β -sheet there are three conserved residues. From the evolutionary perspective, more changes in the back β -sheet compared to the front β -sheet must have taken place to optimize the interactions between the interfaces to suit their biological function. Figure 9 shows the phylogeny of sequences aligned based on their tertiary structures. The clustering in the tree derived from structures is similar to that obtained from just the sequence comparison. The sequences belong to distinct groups and segregate in a fashion that can be predicted *a priori*. The obvious cluster classification is into groups of lectins having the same quaternary association in terms of dimer formations. The clustering also segregates the sugar specificity groups except for PHAL, MAL and UEAlI that share a common quaternary structure with the GalNac-binding lectins in that group.

Figure 10 shows the phylogenetic tree derived from comparison of the 33 sequences given in Table II. The major branching seems to clearly segregate distinct groups of lectins having similar quaternary structures. Also, as in the previous trees, the clustering appears to be related to its carbohydrate specificity as well although occasionally members of one specificity group occur grouped with members of the other. Within each major branch, the clustering distinguishes the carbohydrate specificity groups (Table II). According to the phylogeny derived here, quaternary structures can probably be predicted for the lectins for which no three-dimensional structure is known. For example, MTA and LSL being Man/Glc specific will probably form II-type dimers like the others in the same branch. The branch containing BPL, a GalNac-binding lectin and GS4, a complex-binding lectin that can accommodate a GalNac in the primary-binding site, suggests that BPL, a dimeric lectin, probably will form an X4-type dimer like GS4. There are two lectins that branch away from all the rest of them, i.e. LTA and PNA. PNA, a strictly Gal-binding lectin, forms a unique tetramer as discussed previously. LTA, a fucose-specific lectin, is expected to form a different kind of quaternary association. Indeed, electron microscopy studies using 19 Å data suggests that LTA forms a novel, tetrameric arrangement of two II-type dimers (Cheng *et al.*, 1998).

An interesting feature that can be observed in all the dendrograms seen so far, is the position of PNA, which is an outlier in all comparisons. The unusual quaternary structure of PNA and probably its strict Gal-binding specificity, could indeed be a reflection of its amino acid sequence. This analysis confirms that legume lectins are an interesting family of proteins in which small alterations consequent to sequence variations, in essentially the same tertiary fold, lead to large changes in quaternary structure.

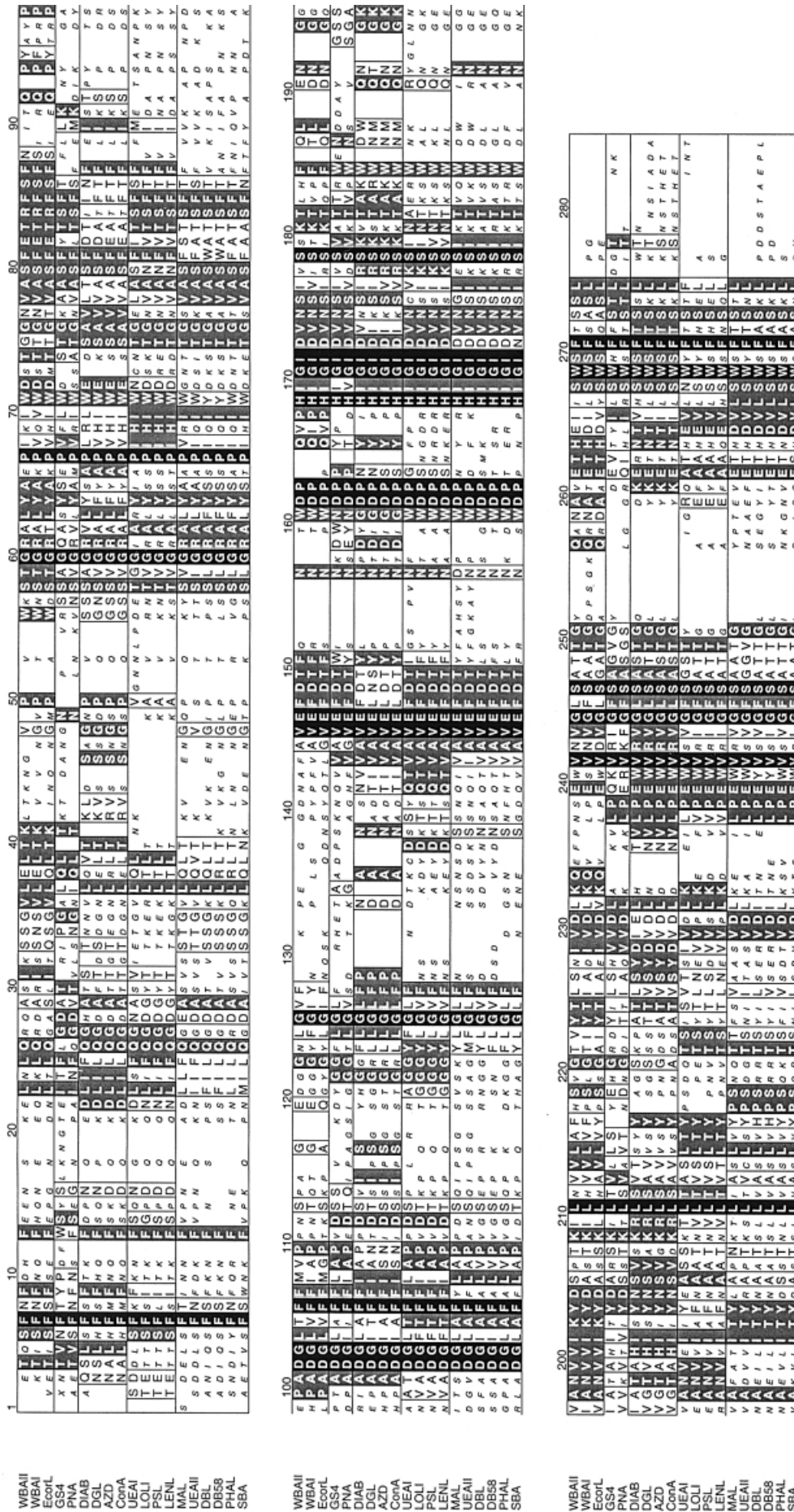


Fig. 8. Alignment of sequences based on structures. The amino acid residues are shaded according to their conservation in the sequences. Also, groups of sequences have been defined as obtained in the phylogenetic trees. The figure was made using the programs AMAS (Livingstone and Barton, 1993) and ALSCRIPT (Barton, 1993).

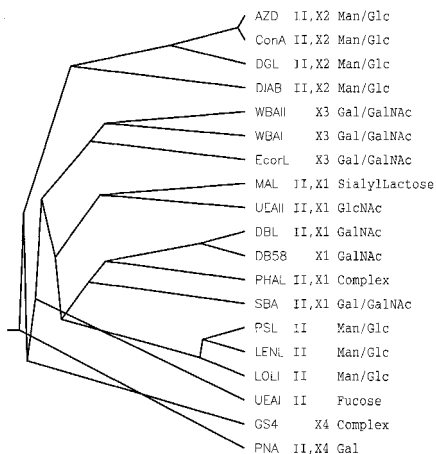


Fig. 9. Phylogenetic tree of the sequences of legume lectins based on superposition of their tertiary structure.

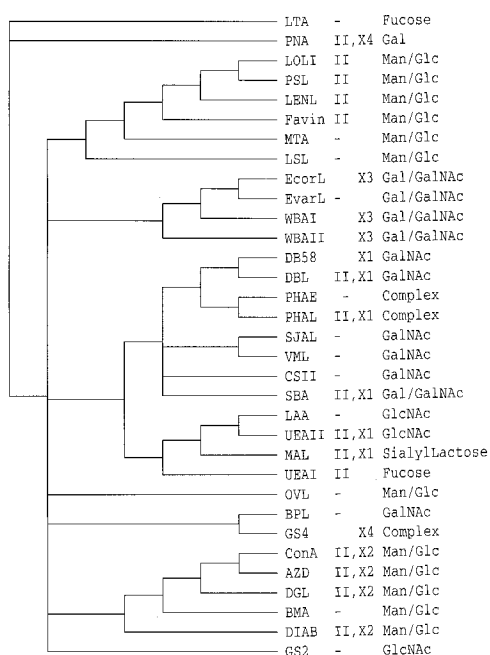


Fig. 10. Phylogenetic tree showing the relationships of some legume lectin sequences. Branch lengths represent approximate evolutionary distances between the sequences.

Role of amino acid residues in determining the oligomerization in legume lectins

As an extension of the above analysis, a detailed examination of the sequences at the various interfaces was performed using the clustering information obtained from the alignment of sequences based on structures, with the objective to pinpoint any residues from the sequences that are conserved within each of the clusters and may probably be crucial either for the formation of certain types of interfaces or prevent the formation of other types of interfaces. The role of the identified residues in the formation of an interface was then examined in the three-dimensional structure by generating the relevant interfaces. Interface residues are defined as those residues whose accessible surface area decreases by greater than 1 Å² on oligomerization. All numbers referred to hereafter correspond to the numbering given in Figure 8. According to this numbering, the residue

ranges in the third, fourth, fifth and sixth strands in the back β-sheet are 80–86, 199–205, 210–216 and 224–230, respectively.

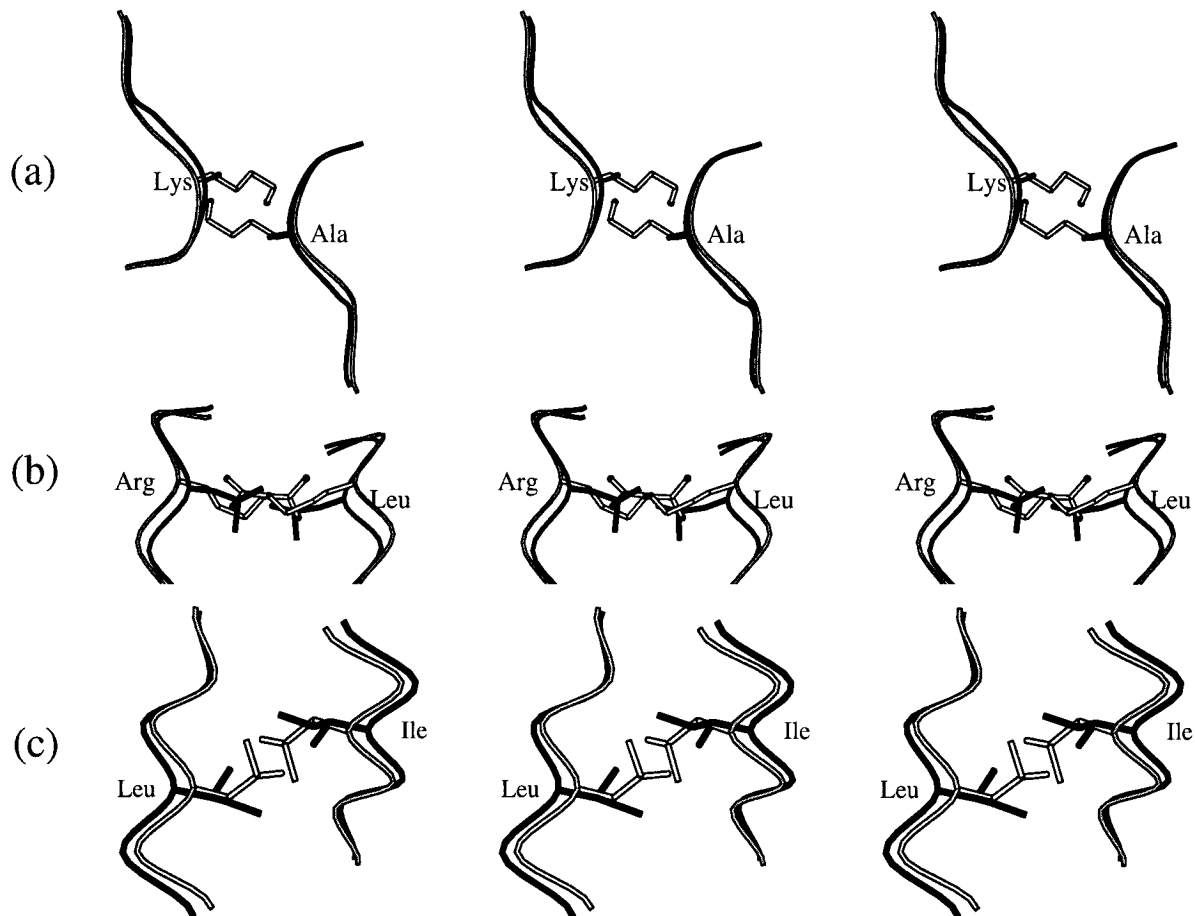
II-Type interfaces

In those lectins that do not form a II-type interface, the amino acid residue at position 66 is charged (Lys, Lys, Glu, Glu in EcorL, WBAl, WBAlI, GS4, respectively). In the case of PNA, which does not have a strict II-type interface, the site is occupied by Met, a large hydrophobic residue. Indeed, modelling of these lectins into a II-type interface shows severe short contacts and burial of these amino acid residues. The amino acids at this position in all the other lectins with a II-type interface are those with small polar or non-polar side chains (Ser, Thr or Ala) (Figure 11a). They are also involved in van der Waals or hydrogen-bonded interactions at the II-type interface. There are other sequence differences between the II-type and X-type classes in this stretch of sequence. A conserved charged residue in the WBA group (with X3-type interface) at position 14 (Glu, His, Glu in WBAlI, WBAl and EcorL, respectively) comes in close contact with the charged amino acid at position 3 (Glu, Lys and Glu in WBAlI, WBAl and EcorL, respectively) making unfavourable interactions in a II-type interface. Similarly in PNA and GS4, charged residues (Arg and Lys, respectively) at position 241 make short contacts and get buried between amino acid residues at positions 17 and 21 in a II-type interface. From the analysis, it appears that the residue at position 66 is completely discriminatory and can act as a switch preventing the formation of a II-type interface. This information can be used to predict whether a II-type interface is possible for a lectin. For example, it can be predicted from a sequence alignment that BPL which has an Arg at this position will not form a II-type interface while LTA or GS2 which have Thr/Ser will probably form a II-type interface.

X-Type interfaces

Prabu *et al.* have shown that the X2-, X3- and X4-type dimers can be generated from each other by a rotation of one subunit with respect to the other about an axis perpendicular to the plane of the dyads (Prabu *et al.*, 1999). A comparison of the residues involved in inter-subunit contacts of each X-type interface and the corresponding regions in the other lectins was performed. The sequence alignment revealed residues which are unique to interfaces of a particular type. For example, in the X3-type interface (WBAl, EcorL and WBAlI), the amino acids Arg and Lys at positions 84 and 203 are unique to this group of lectins and in fact, both these residues make strong hydrogen-bonding interactions across the interface. The Arg84 and Lys203 belong to the third and fourth β-strands of the back β-sheet, respectively, and possibly facilitate the formation of this kind of interface. Alternatively, the presence of an Arg at 210 in all the lectins of the ConA group (X2-type) could prevent the formation of an X3-type of interface by this group. Modelling of an X3-type interface using lectins of this group resulted in short contacts between two Arg210 residues related by a 2-fold axis (Figure 11b).

The X4-type interface provides for a large number of inter-subunit contacts. In fact, residues from all the six β-strands of the back β-sheet participate in the dimer formation. Within this group comprising of PNA and GS4,



the interfaces are not exactly identical; there is a small rotation of the subunits relative to each other (Prabu *et al.*, 1999). A comparison of the relevant stretches of sequences and modelling of a GS4 type of interface using PNA revealed that Leu82 in PNA leads to severe short contacts with Ile223 that gets relieved in the actual PNA interface, while in GS4 the corresponding residues are Tyr and Asp which are involved in good van der Waals interactions. A comparison of the sequences of the SBA group (X1-type) showed that at position 210 a Leu is present in four of the six sequences (SBA, PHAL, DBL and DB58). Generation of an X4-type dimer using this group of lectins showed short contacts of this residue with its 2-fold related one (Figure 11c). The fifth lectin in this group, UEAIL, has a Ser at this position. But an Arg at position 205 that is unique to this lectin gets buried when in an X4-type interface. In the ConA group (X2-type), the residue at position 210 is an Arg that gets buried when an X4-type interface is generated (Figure 11d). This residue of the ConA group also gets buried when an X1-type interface is generated. The corresponding region in the WBA group reveals a unique Lys at position 203 that makes unacceptable steric contacts and gets buried in an X4-type (Figure 11e) or X1-type (Figure 11f) interface. As discussed earlier, this residue is also responsible for making favourable interactions in the X3-type interface of the WBA group. Thus, it appears that the location of Lys at this position could be responsible for the formation of the native dimer and for preventing

the formation of other kinds of interfaces. Similarly, it appears that the location of Arg at 210 for the ConA group could most probably be responsible for this group of lectins not forming an X1-, X3- or X4-type interface. All the residues discussed above belong to one of the three strands of the back β -sheet that are common to X-type interfaces.

Obviously it is not possible to point out from the sequence alignment, the particular amino acid residues that are involved in the formation or prevention of all the four types of X-type interfaces. Although the formation of an interface is the result of the cumulative effect of all the residues present in the interface, the above analysis shows that at least in some cases, crucial residues responsible for oligomerization can probably be identified from the alignment of sequences. This information can provide a basis for mutational studies to evaluate the role of key amino acid residues responsible for variations in modes of oligomerization.

Acknowledgements

Discussions with Professor M.Vijayan, Professor A.Surolia and Professor M.R.N.Murthy are gratefully acknowledged. We thank S.Balaji for help with the structure-based alignment. Facilities at the Supercomputer Education and Research Centre and the Interactive Graphics based facility and Distributed Information Centre (both supported by the Department of Biotechnology) were used in the work. N.M. acknowledges financial support from the University Grants Commission. This work is supported by the Department of Science and Technology.

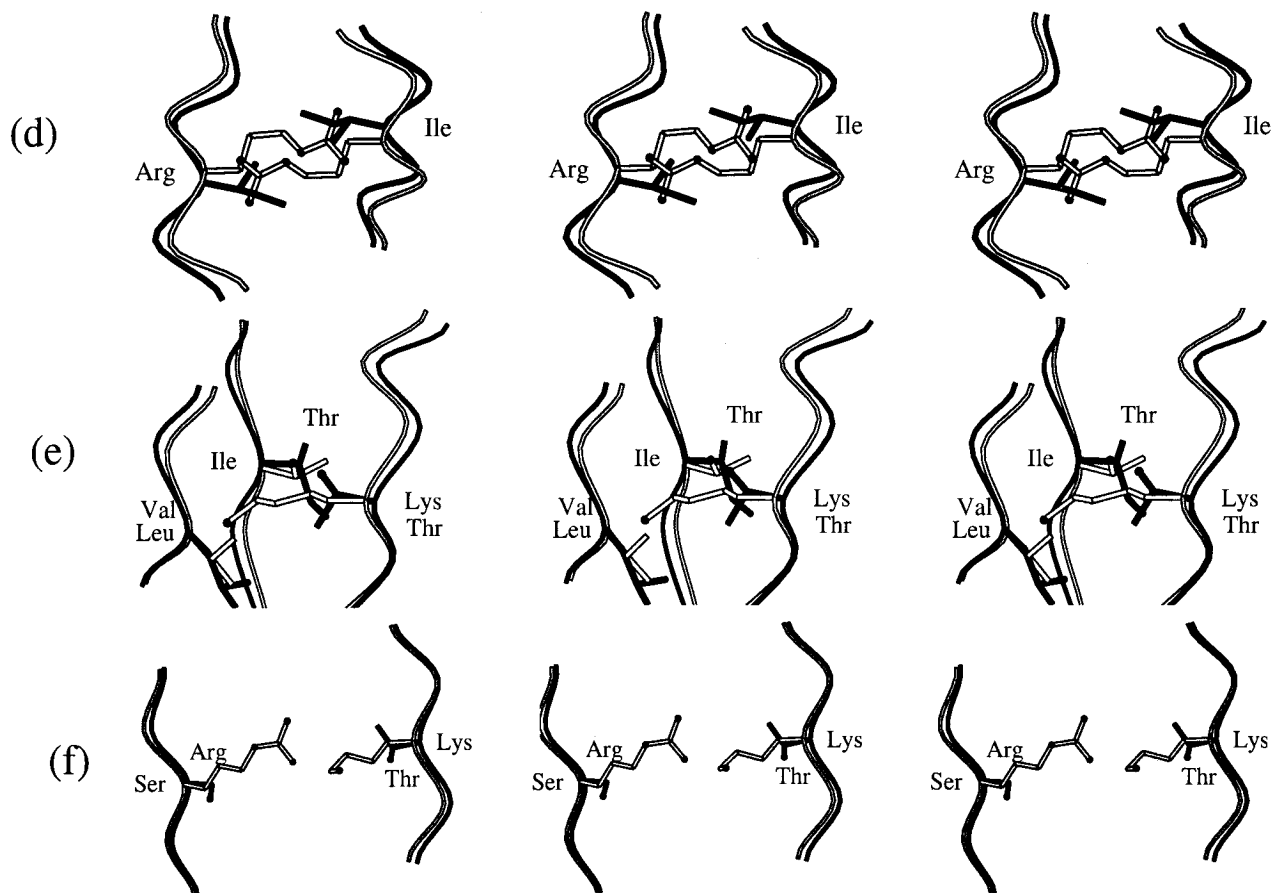


Fig. 11. Stereo views of: (a) A region of the II-type interface in ConA shown in black. Also shown is the modelled region using EcorL structure (grey). (b) A region of the X3-type interface in EcorL (black). Also shown is the modelled region using ConA structure (grey). (c) A region of the X4-type interface in GS4 (black). Also shown is the modelled region using PHAL structure (grey). (d) A region of the X4-type interface in GS4 (black). Also shown is the modelled region using ConA structure (grey). (e) A region of the X4-type interface in GS4 (black). Also shown is the modelled region using EcorL structure (grey). (f) A region of the X1-type interface in PHAL (black). Also shown is the modelled region using EcorL structure (grey). This figure was produced using the program MOLSCRIPT (Kraulis, 1991).

References

- Audette, G.H., Vandonselaar, M. and Delbaere, L.T.J. (2000) *J. Mol. Biol.*, **304**, 423–433.
- Bajaj, M. and Blundell, T. (1984) *Annu. Rev. Biophys. Bioeng.*, **13**, 453–492.
- Bairoch, A. and Apweiler, R. (1997) *Nucleic Acids Res.*, **25**, 31–36.
- Banerjee, R., Mande, S.C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S.K., Suguna, K., Surolia, A. and Vijayan, M. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 227–231.
- Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. and Vijayan, M. (1996) *J. Mol. Biol.*, **259**, 281–296.
- Barton, G.J. (1990) *Methods Enzymol.*, **183**, 403–428.
- Barton, G.J. (1993) *Protein Eng.*, **6**, 37–40.
- Barton, G.J. and Sternberg, M.J. (1987) *Protein Eng.*, **1**, 89–94.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) *Nucleic Acids Res.*, **28**, 235–242.
- Bouckaert, J., Hamelryck, T., Wyns, L. and Loris, R. (1999) *Curr. Opin. Struct. Biol.*, **9**, 572–577.
- Bourne, Y., Abergei, C., Cambillau, C., Frey, M., Rouge, P. and Fontecilla-Camps, J.C. (1990) *J. Mol. Biol.*, **214**, 571–584.
- Cheng, W., Bullit, E., Bhattacharyya, L., Brewer, C.F. and Makowski, L. (1998) *J. Biol. Chem.*, **273**, 35016–35022.
- Dao-Thi, M.H., Rizkallah, P., Wyns, L., Poortmans, F. and Loris, R. (1998) *Acta Crystallogr.*, **D54**, 844–847.
- Dessen, A., Gupta, D., Sabesan, S., Brewer, C.F. and Sacchettini, J.C. (1995) *Biochemistry*, **34**, 4933–4942.
- Delbaere, L.T.J., Vandonselaar, M., Prasad, L., Quail, J.W., Wilson, K.S. and Dauter, Z. (1993) *J. Mol. Biol.*, **230**, 950–965.
- Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E. and Suddath, F.L. (1986) *J. Biol. Chem.*, **261**, 16518–16527.
- Felsenstein, J. (1985) *Evolution*, **39**, 783–791.
- Hardman, K.D. and Ainsworth, C.F. (1972) *Biochemistry*, **11**, 4910–4919.
- Hamelryck, T.W., Dao-Thi, M.H., Poortmans, F., Chrispeels, M.J., Wyns, L. and Loris, R. (1996) *J. Biol. Chem.*, **271**, 20479–20485.
- Hamelryck, T.W., Loris, R., Bouckaert, J., Dao-Thi, M.H., Strecker, G., Imberty, A., Fernandez, E., Wyns, L. and Etzler, M.E. (1999) *J. Mol. Biol.*, **286**, 1161–1177.
- Imberty, A., Gautier, C., Lescar, J., Pérez, S., Wyns, L. and Loris, M. (2000) *J. Biol. Chem.*, **275**, 17541–17548.
- Jones, S. and Thornton, J.M. (1995) *Prog. Biophys. Mol. Biol.*, **63**, 31–65.
- Kraulis, P. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
- Lis, H. and Sharon, N. (1998) *Chem. Rev.*, **98**, 637–674.
- Livingstone, C.D. and Barton, G.J. (1993) *CABIOS*, **9**, 745–756.
- Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. and Wyns, L. (1993) *Biochemistry*, **32**, 8772–8781.
- Loris, R., Hamelryck, T., Bouckaert, J. and Wyns, L. (1998) *Biochim. Biophys. Acta.*, **1383**, 9–36.
- Manoj, N., Srinivas, V.R., Surolia, A., Vijayan, M. and Suguna, K. (2000) *J. Mol. Biol.*, **302**, 1129–1137.
- Needleman, S.B. and Wunsch, C.D. (1970) *J. Mol. Biol.*, **48**, 443–453.
- Prabu, M.M., Sankaranarayanan, R., Puri, K.D., Sharma, V., Surolia, A., Vijayan, M. and Suguna, K. (1998) *J. Mol. Biol.*, **276**, 787–796.
- Prabu, M.M., Suguna, K. and Vijayan, M. (1999) *Proteins: Struct. Funct. Genet.*, **13**, 1–12.
- Reeke, G.N. and Becker, J.W. (1986) *Science*, **234**, 1108–1111.
- Rossmann, M.G. and Argos, P. (1975) *J. Biol. Chem.*, **250**, 7525–7532.
- Rozwarski, D.A., Swami, B.M., Brewer, C.F. and Sacchettini, J.C. (1998) *J. Biol. Chem.*, **273**, 32818–32825.
- Russel, R.B. and Barton, G.J. (1992) *Proteins: Struct. Funct. Genet.*, **14**, 309–323.

- Sanz-Aparicio,J., Hermoso,J., Grangeiro,T.B., Calvete,J.J. and Cavada,B.S. (1997) *FEBS Lett.*, **405**, 114–118.
- Schwartz,R.M. and Dayhoff,M.O. (1978) In Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5. National Biomedical Research Foundation, Washington DC, pp. 353–358.
- Shaanan,B., Lis,H. and Sharon,N. (1991) *Science*, **254**, 862–866.
- Sharma,V. and Surolia,A. (1997) *J. Mol. Biol.*, **267**, 433–445.
- Swamy,M.J., Sastry,M.V.K. and Surolia,A. (1985) *J. Biosci.*, **9**, 203–212.
- Vijayan,M. and Chandra,N. (1999) *Curr. Opin. Struct. Biol.*, **9**, 707–714.
- Weis,W.I. and Drickamer,K. (1996) *Annu. Rev. Biochem.*, **65**, 441–473.
- Young, N.M. and Oomen,R.P. (1992) *J. Mol. Biol.*, **228**, 924–934.

Received November 21, 2000; revised July 19, 2001; accepted July 31, 2001