Conference paper

Koyamangalath Vadakkepat Abhinav and Mamannamana Vijayan*

Structural diversity and ligand specificity of lectins. The Bangalore effort

Abstract: Structural studies in this laboratory encompass four of the five major classes of plant lectins, including the one discovered by us. In addition to addressing issues specific to individual lectins, the work provided insights into protein folding, quaternary association and generation of ligand specificity. Legume and β-prism fold lectins constitute families of proteins in which small alterations in essentially the same tertiary structure lead to large variations in quaternary structure, including that involving an open structure. Strategies for generating ligand specificity include water bridges, variation in loop length, post translational modification and oligomerization. Three of the structural classes investigated have subunits with three-fold symmetry. The symmetry in the structure is reflected in the sequence to different extents in different subclasses. The evolutionary implications of this observation have been explored. The work on lectins has now been extended to those from mycobacteria.

Keywords: lectin folds; ligand specificity; ICS-27; molecular evolution; mycobacterial lectins; quaternary association.

DOI 10.1515/pac-2014-0607

Based on a plenary lecture dedicated to Professor A. Surolia, a dear friend, close colleague and long term collaborator of the senior author.

Introduction

Structural and related studies on lectins have been a major research programme in our laboratory for more than three decades. Lectins were first identified in plants and their best known property used to be the ability to agglutinate blood. Hence they used to be referred to as phytohemagglutinins. Subsequently, lectins were found in all forms of life. They mediate a variety of biological processes such as cell-cell interactions, plant defense, innate immunity, mutagenesis and serum glycoprotein turn over [1–7]. The basis for all these properties is the ability of lectins to specifically bind to different carbohydrate structures. Despite the common properties, lectins exhibit a variety of folds and quaternary structures.

Although carbohydrates constitute the most abundant class of molecules in the biosphere, their crucial role in biological recognition began to be recognized only comparatively recently. It turns out that protein-carbohydrate interaction is a common mechanism of cell surface recognition. This realization led to a surge in structural and other studies on lectins which are quintessential carbohydrate binding proteins. It is at this stage that we entered the field. At that stage in the late seventies of the last century, the three-dimensional

Article note: A collection of invited papers based on presentations at the 27th International Carbohydrate Symposium (ICS-27), Bangalore, India, 12–17 January 2014.

***Corresponding author: Mamannamana Vijayan,** Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India, Tel.: +080-2360 0765, Fax: +080-2360 0535, E-mail: mv@mbu.iisc.ernet.in
Koyamangalath Vadakkepat Abhinav: Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India

structures of only two lectins, namely, concanavalin A (Con A) [8, 9] and wheat germ agglutinin (WGA) [10], were available. By the time we published our first result, preliminary crystallography studies on peanut lectin [11], the structure of influenza virus hemagglutinin, a viral lectin, also became available [12]. Since then, the structure and interactions of many lectin families have been elucidated. A substantial portion of these investigations has been on plant lectins which belong to six structural classes, five of which, including the one discovered in this laboratory, could be described as major. The five are legume lectins, agglutinins with hevein domain, β-trefoil lectins, β-prism I fold lectins and β-prism II fold lectins. Our efforts have encompassed four of these five major classes. The review presented here deals with these four classes, centered around work in this laboratory. Towards the end, we would also outline our foray into mycobacterial lectins.

Legume lectins

Gal/GalNAc specific lectins: peanut and winged bean agglutinins, rEcorL

Most of the well-characterised legume lectins, can be broadly classified as mannose (Man)/Glucose (Glc) specific or galactose (Gal) specific in terms of affinity at the primary binding site. Most Gal specific lectins bind *N*-acetylgalactosamine (GalNAc) as well. Con A, the most thoroughly studied legume lectin, is Man/Glc specific. Peanut agglutinin (PNA) from *Arachis hypogaea*, the first lectin to be investigated in this laboratory, is Gal specific and is perhaps the second most thoroughly studied legume lectin. PNA, like Con A, is a homotetrameric lectin, M_r 110 000, with 236 amino acid residues in each non-glycosylated subunit. At the disaccharide level, it has high specificity for the T-antigenic disaccharide, Gal β1-3 GalNAc (Thomsen Friedenreich antigen), which is found as O-linked glycans on poorly differentiated and tumor cells, but not on normal cells. The basic winged bean agglutinin (WBAI) and the acidic winged bean lectin (WBAII) from *Psophocarpus tetragonolobus*, two other lectins studied in this laboratory, are Gal/GalNAc specific. Both are dimers, M_r 58 000, with glycosylated subunits. WBAI is specific to blood group A, less so to blood group B and does not bind blood group O substance. WBAII, on the other hand, is specific to blood group O and binds only weakly to blood group substances A and B. Though only to a limited extent, a recombinant form of Gal/GalNAc specific *Erythrina corallodendron* lectin (EcorL), which is highly homologous to winged bean lectins, has also been used in our studies.

Unusual quaternary structure and variability in quaternary association

The work on PNA, which spanned almost 25 years, resulted in a wealth of information on its structure, minor variability in molecular geometry consequent to environmental changes, interactions with sugars, multivalency etc. [13]. However, the most interesting aspects of the structure and interactions of PNA pertain to its quaternary structure and generation of ligand specificity. Each subunit of PNA has a structure very similar to that observed in other legume lectins. It consists of a six-stranded "back β-sheet", a seven-stranded curved "front β-sheet" and a five-stranded "top β-sheet" (Fig. 1). The loops that connect the sheets account for about 50 % of the structure. As in other legume lectins, each subunit contains a manganese and a calcium ion. Although the structure of the subunit in PNA is similar to that in ConA, the quaternary structures in the two are very different.

A well established and almost universally followed paradigm of protein architecture is that multimeritc proteins should have a "closed" structure with point group symmetry [14, 15]. The permissible symmetries for a tetramer is 222 (D_2) or four-fold axis. PNA has neither [16]. Indeed, PNA is the first well characterized tetrameric protein without either of these symmetries. As shown in Fig. 2, the PNA tetramer is made up of two dimers (AD and BC) related to each other by a two-fold axis (P). Each dimer also possesses two-fold symmetry; R1 relates A and D while R2 relates B and C. P, R1 and R2 are inclined with respect to one another; they do not intersect either. The three two-fold axes combine to give rise to an irrational screw axis, Q, perpendicular to all of them and relates C to A (146°, 24 Å) and D to B (–146°, –24 Å). Repeated occurrence of the same

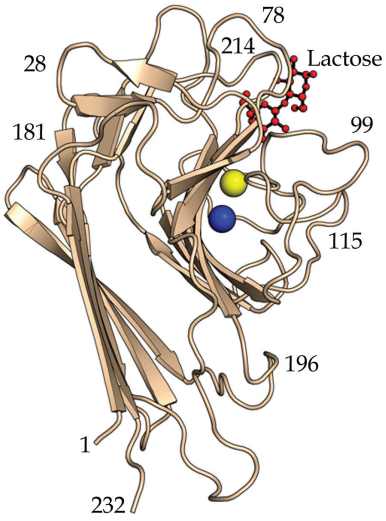

Fig. 1 The legume lectin fold as observed in a PNA-lactose complex, prepared using coordinates available in the Protein Data Bank [80] (PDB Code: 2PEL). The blue and yellow spheres represent manganese and calcium ions, respectively. This and the subsequent figures were prepared using PYMOL [81].

quaternary structure in crystal forms of the lectin grown under widely different conditions [17–24] demonstrates that the observed open conformation is not an artefact of crystallization. Solution studies also showed that PNA is a tetramer and that the tetrameric association in it is perhaps stronger than that in well known tetrameric proteins like ConA and hemoglobin. No ready explanation is available as to why PNA assumes this unusual quaternary association. In any case, the structure of PNA establishes that open quaternary association in multimeric proteins cannot be ruled out.

Con A, the first lectin to be analysed by X-ray crystallography, is also a dimer of dimers. The dimeric association in it, which has subsequently been observed in other legume lectins as well, involves coming together of the two six-stranded back β-sheets with the formation of an extended 12-stranded β-sheet in the dimer. This arrangement is often referred to as the canonical mode of dimerization in legume lectins. PNA also is a dimer of dimers involving subunits similar to those in Con A. The dimeric arrangement in it (as in AD and BC) involves a back-to-back arrangement of the back β-sheets, similar to that found in *Griffonia simplicifolia* lectin IV (GS IV) [25]. It was suggested that the non-canonical mode of dimerization in GS IV was caused by interactions involving covalently bound sugar and to avoid the burial of a glutamic acid residue. In the meantime, a related handshake mode of dimerization was observed in the *Erythrina corallodendron* lectin (EcorL)

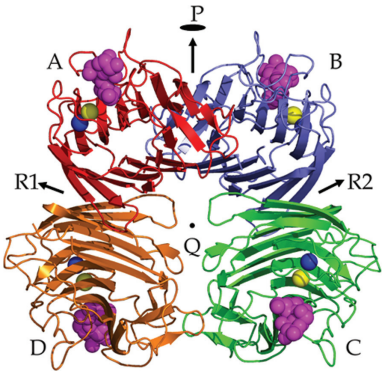

Fig. 2 Structure of PNA (PDB Code: 2PEL). The four subunits are colored differently. The blue and yellow spheres represent manganese and calcium ions, respectively. The bound lactose molecules are shown in magenta space filling representation.

[26]. It was argued that canonical dimerization was prevented in EcorL by the presence of glycosylation sites at the interface of such a dimer. However, PNA is not glycosylated, yet it exhibits a non-canonical mode of dimerization, indicating that the departure from the canonical mode is not necessarily caused by interactions involving covalently bound sugar.

The structures of winged bean lectins provided an ideal system for further exploring this issue. For instance, WBAI has 54% sequence identity with EcorL. However, unlike in EcorL, the glycosylation site in WBAI is far removed from the inter subunit interface of a canonical dimer and hence would not prevent the formation of such a dimer. Thus if the formation of a canonical dimer is prevented by covalently bound sugar, then WBAI would form a canonical dimer. On the other hand, if the non-canonical mode is adopted for reasons intrinsic to the protein, then WBAI would adopt a handshake mode of dimerization as in EcorL. In the event, the crystal structure WBAI exhibits the hand shake mode dimerization (Fig. 3) [27]. The arguments outlined above in relation to WBAI apply to WBAII as well. WBAII also adopts the hand shake mode of dimerization in its crystal structure [28]. Subsequently, the crystal structure of non-glycosylated recombinant EcorL (rEcorL) was determined [29]. Despite the absence of covalently bound sugar, rEcorL adopts the same handshake mode of dimerization as in native EcorL. Thus, departure from the canonical mode of dimerization in legume lectins appears to have been caused by factors intrinsic to the protein. Presumably, once the oligomeric structure is set, glycosylation sites can then appear in regions that are exposed, but not regions that are buried.

Other structural studies on legume lectins also indicated variability in the quaternary association of legume lectins [30]. Our observations and results from other laboratories show that legume lectins are a family of proteins in which small alterations in essentially the same tertiary structure lead to large changes in quaternary association [31]. All the structurally studied legume lectins are dimers or tetramers, each involving two dimers. Dimeric lectins themselves exhibit a variety of modes of quaternary association (Fig. 4). The same is true about the tetrameric lectins (Fig. 5). The back β-sheet is involved in oligomerization in all of them.

The legume lectin fold

In addition to the novelty of the quaternary association, a detailed analysis of the structure of PNA brought to light some important characteristics of the common structure of the legume lectin subunits, which were not appreciated till then. It is through the work on PNA, the existence and the structural role of the small top β-sheets was firmly established. The sheet has a role in holding the two larger sheets together. It was also rec-

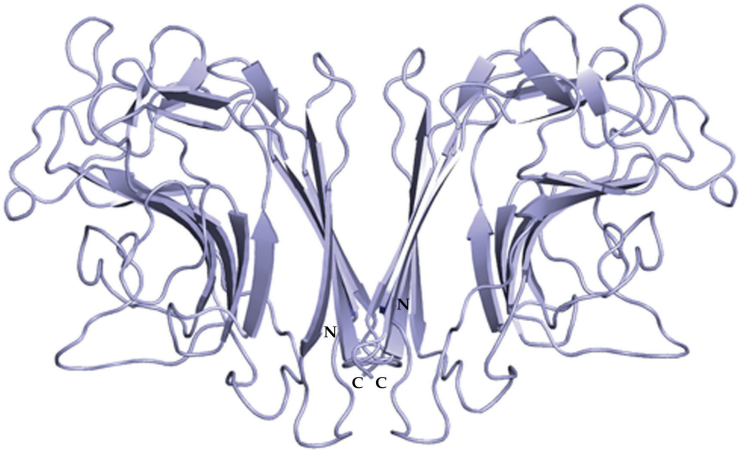

Fig. 3 Structure of WBAI (PDB Code: 1WBL). Figures 4 and 5 present a comparison of the mode of dimerization in WBAI with that in ConA and PNA, each of which is a dimer of dimers.

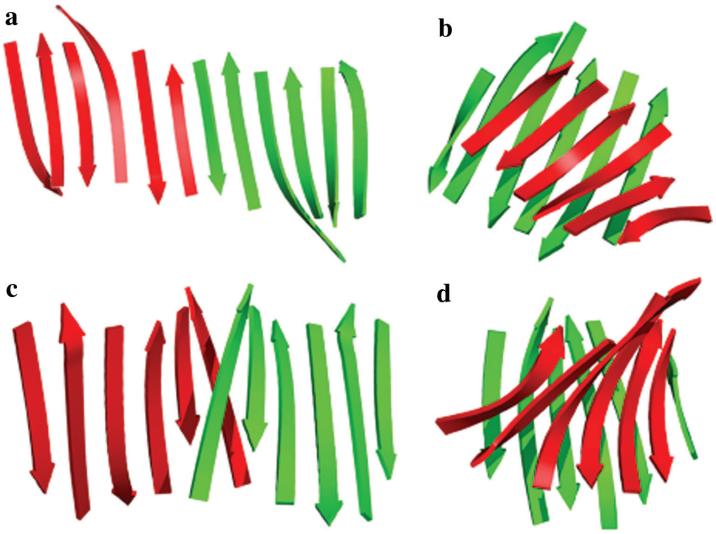

Fig. 4 Schematic representation of the back β-sheets in dimeric legume lectins. (a) pea lectin, (b) GS-IV lectin, (c) EcorL and (d) *Dolichos biflorus* leaf lectin (PDB Codes: 2BQP, 1LEC, 1LTE and 1LUL). Sheets from the two subunits are colored differently. That in pea lectin is the cannonical mode of dimerization, first observed in tetrameric ConA.

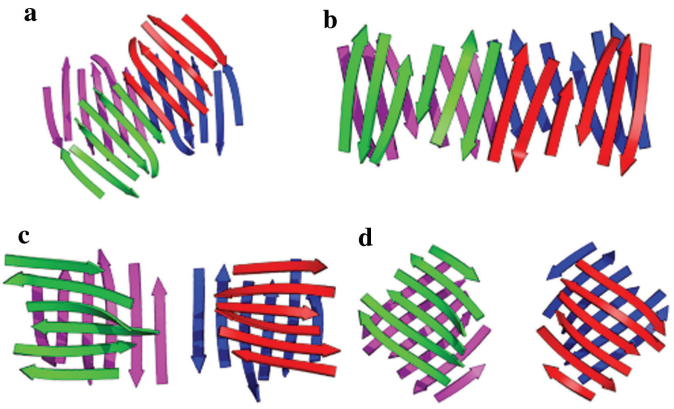

Fig. 5 Schematic representation of the back β-sheets in tetrameric legume lectins. (a) ConA, (b) Soyabean agglutinin, (c) PNA, and (d) *Griffonia simplicifolia* I-B4 lectin (PDB Codes: 3CNA, 2SBA, 2PEL and 1LED). Sheets from the four subunits are colored differently.

ognized that the subunit has two hydrophobic cores, one bounded by the three sheets and the other between the curved front sheet and the long over hanging loops that connect the strands in it.

What is generally described as the legume lectin fold occurs in a variety of proteins other than legume lectins as well. All these proteins appear to have a carbohydrate binding function. A comparative analysis of the relevant proteins from different families which often exhibit hardly any sequence similarly among them, led to the delineation of the minimum structural features that characterize the fold [32]. The most important structural feature turns out to be the presence of the three sheets in specific relative orientations and the topological connectivities involving them. The loops are elaborated to different extents in the proteins. The main hydrophobic core bounded by the three sheets is always present. The presence of the second hydrophobic core is related to the curvature of the front sheet and the lengths of the loops that overhang in front of it. The sequence of the strands in the back sheet and the front sheet do not provide any clue as to why one should be flat and the other curved. When the overhanging loops are absent or short, as in some of the proteins, the front sheet also becomes flat. When the loops are long, the sheet curves such that a secondary hydrophobic core is generated between the sheet and the loops.

Lectin-sugar interactions, generation of ligand specificity and redesign of binding sites

PNA-sugar interactions have been investigated through the X-ray analysis of a number complexes with carbohydrates including a number of disaccharides with different glycosidic linkages between the monosaccharide residues [17–23]. The interactions involve four binding loops, residues 75–83, 91–106, 125–135 and 211–216 in all of them. The primary binding site is always occupied by galactose which invariably has four hydrogen bonds, Gal O3 with Asp 83 OD1, Gly 104 N and Asn 127 ND2 and Gal O4 with Asp 83 OD2 and two water bridges (Fig. 6). As in other Gal specific lectins, an aromatic side chain, that of Tyr 125 in the case of PNA, stacks against the galactose ring. There is considerable variation in the interactions, including water bridges, involving the second sugar residue in the disaccharides. Lactose, Gal β(1–4) Glc, and Gal β(1–3) Gal interacts with the lectin in the same manner except that O3 and O4 of the second residue are interchanged. The affinity of T-antigen, Gal β(1–3) GalNAc, which has an additional acetamido group substituted in the second ring, is 20 times more than that of lactose. This happens only because of the two additional water bridges involving the acetamido group that T-antigen makes with the protein [18]. Thus, PNA demonstrates how water-bridges can be used for generating ligand specificity.

PNA is perhaps the only thoroughly characterized Gal specific legume lectin which does not bind GalNAc at the primary binding site. In the crystal structures of PNA-sugar complexes, O2 forms a water bridge with Glu 129. Substitution of the acetamido group at the 2-position is not thus favoured as that could lead to steric clashes with Gln 129. Subsequently molecular dynamics simulations indicated that O2 often forms a direct hydrogen bond with Glu 129 [33]. This indication further supports the explanation for the inability of PNA to bind GalNAc at the primary binding sites.

Lectin-sugar interactions involving the two winged bean agglutinins, especially WBAI, have also been thoroughly examined [27, 28, 34–37]. The interactions at the primary binding site in them are similar to those in other legume lectins, but with an important difference. These interactions themselves give a partial explanation for the difference in the blood group specificity of WBAI and WBAII. In particular, the length of a sugar binding loop is very different in the two lectins. This has a bearing on their sugar specificities, thus demonstrating variation of loop length as a strategy for generating ligand specificity.

Structural information was also used to redesign the binding sites. In order to improve its utility as a diagnostic tool, exclusivity of PNA binding to T-antigen was sought through protein engineering. The onset of carcinoma is accompanied by the expression of T-antigen on the cell surface. N-acetyllactosamine, whose affinity for PNA is much lower, but significant, also occurs at the termini of the saccharide chain in many

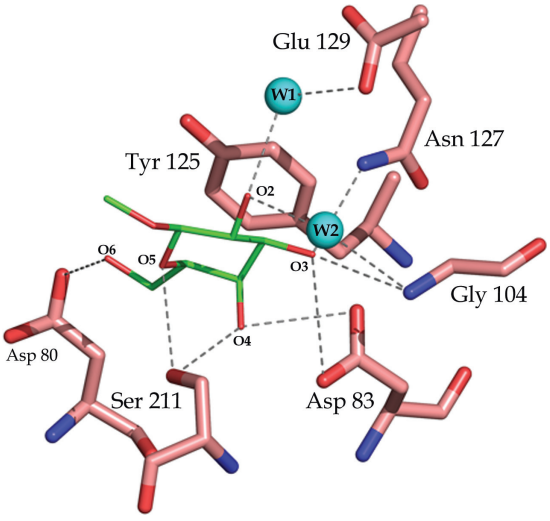

Fig. 6 The binding site of PNA with bound Me-β-Gal (green). W1 and W2 represent water molecules.

glycoproteins. The L212N mutant, produced through site-directed mutagenesis using structural information, binds only T-antigen and not N-acetyllactosamine [38]. L212A however, binds N-acetyllactosamine more strongly than it binds T-antigen. Another mutant E129D binds GalNAc as well as Gal, as the steric interaction of the acetamido group in GalNAc with Glu 129 is now relieved [39]. Another mutant, N41Q, has improved potency for T-antigen, which could be readily explained on the basis of the structure of PNA [40]. Another mutational study involved rEcorL. The lectin has nearly equal affinity for Gal and GalNAc. On the basis of comparisons with similar structures, including WBAI and WBAII, which have higher affinity for GalNAc, a mutant (Y106G) of rEcorL, which has substantially higher affinity for GalNAc than for Gal, was produced and X-ray analysed [41]. The ability to redesign the combining site for predictable properties enhances the confidence in the deductions on carbohydrate binding based on three-dimensional structures.

Plasticity, water structure and multivalency

The crystal structures of PNA and its complexes determined in this laboratory resulted in information on the well defined geometry of 52 subunits. The corresponding number in the case of WBAI and its complexes was 54. That permitted the elucidation of the plasticity of the subunit in the two lectins [23, 37]. The molecules are robust with limited plasticity. In both the cases, the region involving the curved β-sheet which nestles the metal ions is relatively rigid. The carbohydrate binding site is perched on this region. The flat back β sheet is relatively flexible. This sheet is invariably involved in quaternary interactions and its flexibility is perhaps necessary to accommodate the variability of quaternary association exhibited by legume lectins.

Crystal structure analysis results in the location of water molecules around the protein, with varying levels of confidence. The confidence level increases when a water molecule occurs at the same location relative to the protein, in structures determined under different conditions. Twenty seven such “invariant” water molecules are associated with each subunit of PNA. Four of them are involved in metal coordination [23]. These four and another five are involved in stabilizing the carbohydrate binding loops. One of the five water-molecules seen to be involved in protein-sugar interactions is invariant. The other four occur in 80% of the subunits considered. In general, the invariant water molecules have relatively high concentration in the region involving loops, including sugar binding loops. Perhaps, the hydrogen bonds and other direct interactions between protein atoms are not enough to maintain the stability and structural integrity of the region. Additional interactions in the form of water bridges are necessary to stabilize the loops and to maintain their correct juxtaposition with the rest of the molecule.

Multivalency is believed to be important for the activity of lectins, but it is poorly understood in structural terms. Structural and related studies on the complex of tetravalent PNA and a synthetic compound containing two terminal lactose molecules were therefore carried out [24]. Dynamic light scattering indicated aggregation. Poorly diffracting crystals of the complex also could be obtained. Although the resolution was as low as 7.65 Å, the X-ray dataset was good enough to establish the arrangement of extensive cross linked molecules in the crystals (Fig. 7). Extensive modeling studies were also carried out. They indicated that complexation of multivalent lectins and multivalent sugars could lead to an ensemble of a finite number of distinct crosslinked periodic arrays rather than a unique array.

β-prism I fold

Jacalin and artocarpin. New lectin fold and novel strategies for generating ligand specificity

β-Prism I as a lectin fold was discovered in this laboratory through the structure analysis of jacalin, one of the two lectins present in jack fruit (*Artocarpus integrifolia*) seeds [42]. The structure of the second lectin, artocarpin, was also subsequently determined [43]. Both proteins are tetrameric, M_r 66 000, with subunits of

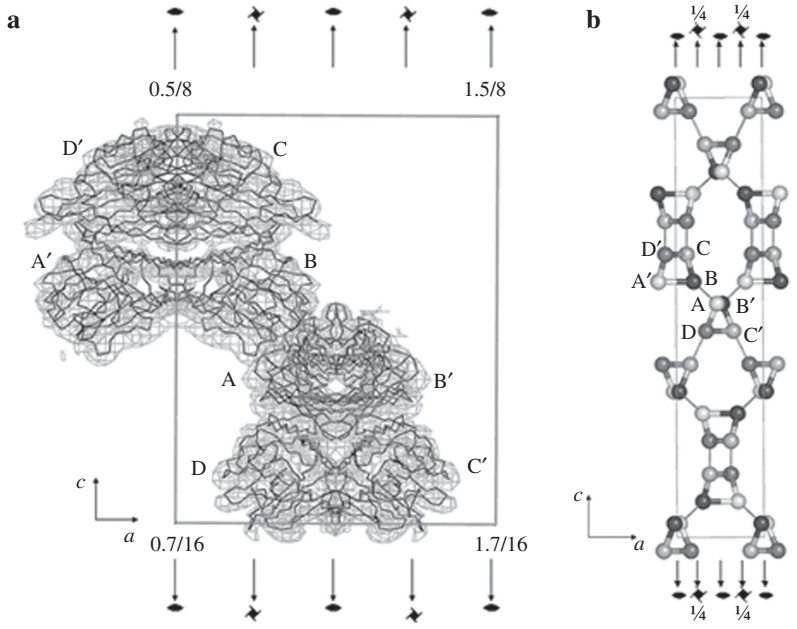

Fig. 7 Electron density for the PNA molecules in its complex with a bidentate ligand and the arrangement of cross linked molecules in the crystal. Reproduced from ref. [24].

similar size. Gal/GalNAc specific jacalin is glycosylated while Man specific artocarpin is not. Artocarpin has a 149 amino acid residues long single chain subunit. The jacalin subunit is made up of two chains, an α-chain consisting 133 amino acid residues and a β-chain comprising 20 amino acid residues. The two chains are produced by the post translational proteolysis of a single chain processor which has 44.2 % sequence identity with the single chain artocarpin subunit.

The structures of the methyl-α-Gal complex of jacalin and the methyl-α-Man complex of artocarpin define a framework for discussing the structure and interactions of β-prism I fold lectins (Fig. 8). In both the cases, each unit is made up of three Greek keys which form three sides of a nearly three-fold symmetric prism. In the case of jacalin, one of the Greek keys (Greek key I) is broken on account of the post translational proteolysis. Much of the shorter β-chain forms one strand of this key. In both the cases, the near three-fold symmetry of the subunit is not reflected in the sequence.

The loops at one end of the prism constitute the carbohydrate binding site. In both the cases, each subunit carries only one binding site. The primary binding sites of jacalin and artocarpin are similar except for a couple of important differences (Fig. 9). Gly 1 in the binding site of jacalin has a positively charged terminal amino group generated by the post translational proteolysis. This group has a strong interaction with O4 which has different orientations in Gal and Man. This interaction, found in all the subsequently analysed Gal specific β-prism I fold lectins, is believed to be important for the specificity of the lectin for Gal. Therefore jacalin-like lectins demonstrate the importance of post-translational modification as a strategy for generating ligand specificity. A peptide amino group is involved in the corresponding, comparatively weaker, interaction with Me-α-Man in artocarpin. As in the case of other galactose specific lectins, the side chain of an aromatic residue, Tyr 78 in the present case, stacks against the galactose ring. The corresponding residue in artocarpin is a threonine. Two more aromatic residues, Tyr 122 and Trp 123, are present in the binding site of jacalin. They are part of a secondary binding site (see later). The primary binding sites of other Gal-specific and Man-specific β-prism I fold lectins are similar to those of jacalin and artocarpin, respectively. Recently, the structure of Ipomoelin, a lectin which has nearly the same affinity for Gal and Man, has also been described [44]. The dual specificity is achieved by subtle changes in the binding site and the location of the ligand.

The crystal structures of several complexes of jacalin, including those involving T-antigen and other disaccharides, have been determined [45–47]. They define the extended binding site of jacalin (Fig. 10). In

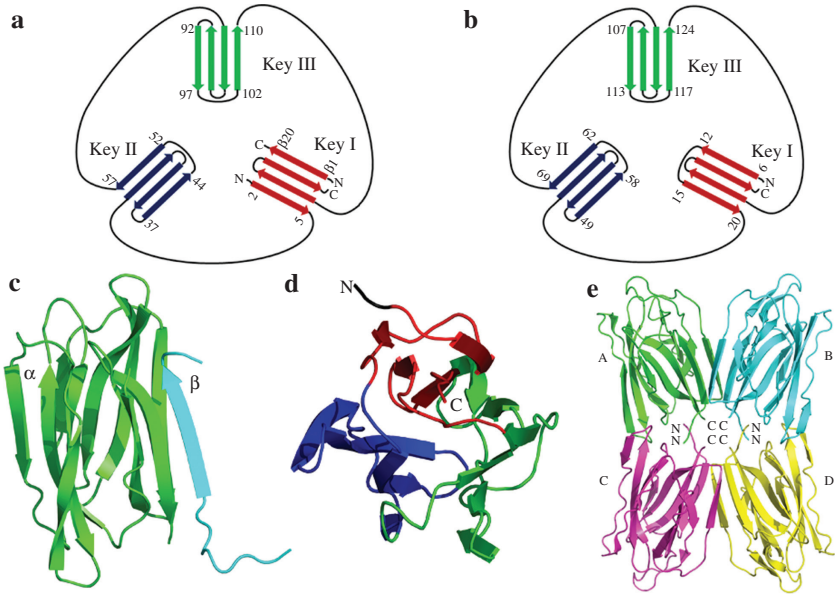

Fig. 8 Topology of the polypeptide chain(s) in (a) jacalin and (b) artocarpin. (c) Side view of the jacalin subunit (PDB Code: 1KUJ) (d) view of the artocarpin subunit down the molecular threefold axis (PDB Code: 1J4S) (e) quaternary association in artocarpin (PDB Code: 1J4S).

addition to the primary binding site, there are two secondary binding sites, secondary site A and secondary site B. Tyr 122, Tyr 123 and Tyr 78, which is also part of the primary binding site, constitute secondary site A. The backbone nitrogen and oxygen atoms of Val 79, Ser 119 OG, Asp 125 OD1 and the carboxyl terminal region of the short β-chain constitute secondary site B. Site A, which is wholly aromatic, exhibits substantial flexibility. All the interactions of the ligands with site B are in the form of water bridges. α-substituents at the anomeric oxygen point towards and interact with secondary site A. The substituent is located at site B in the case of β-substitution. The interaction of α-substituents with secondary site A can be strong. For instance, a methyl group α-substituted to Gal or GalNAc can form C-H...π hydrogen bonds with the aromatic side chain of Tyr 122. Gal binds to jacalin 20 times more strongly than Man does. The affinity of glucose (Glc) is similar to that of Man. However, the Me-α-Gal and Me-α-Man bind to the lectin 27 times more strongly than Gal and Man, respectively. Crystalline complexes of Me-α-Man and Me-α-Glc could be prepared and analysed presumably on account of the strength of the interactions of the methyl group with secondary binding site A [47, 48].

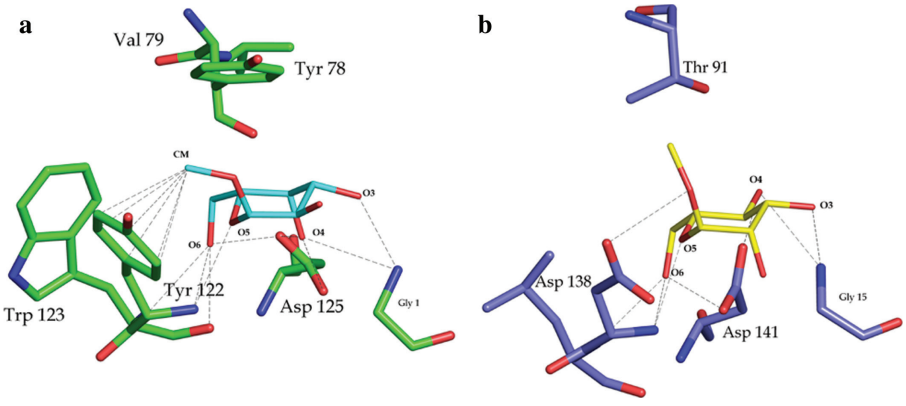

Fig. 9 Sugar binding sites of (a) jacalin (PDB Code: 1JAC) and (b) artocarpin (PDB Code: 1J4U). The bound sugars are Me-α-Gal and Me-α-Man, respectively.

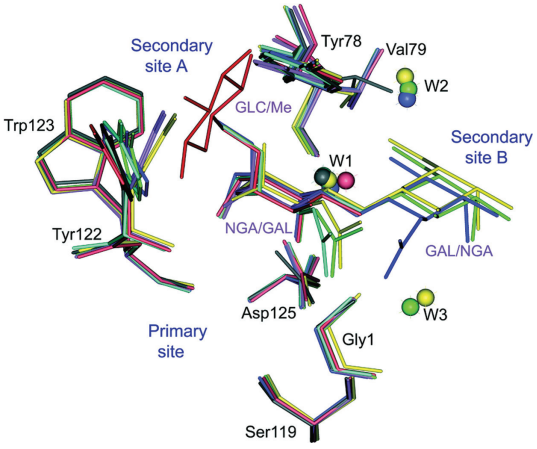

Fig. 10 The extended binding site of jacalin with bound sugars [pink: Gal, black: Me-α-Gal, cyan: Me-α-GalNAc, red: Galβ(1-3) GalNAc, yellow: Galβ(1-3)GalNAc-α-OMe, blue: GalNAcβ(1-3)Gal-α-OMe, green: Gala(1-6)Glc]. Reproduced from ref. [47]. NGA stands for GalNAc and W for water.

The primary binding sites of Gal specific and Man specific β-prism I fold lectins are similar, but the extended sites are very different. This can be seen from the binding sites of artocarpin and heltuba, another well-studied Man specific β-prism I fold lectin, illustrated in Fig. 11 [49, 50]. Loops 14–16 and 137–141 (in artocarpin numbering) constitute the primary binding site. The sites are remarkably similar in the two lectins. The interactions with Man at the primary site are also the same. However, they exhibit different affinities at the oligosaccharide level. This is on account of the difference in a third loop (88–96 in artocarpin). The loop is longer in artocarpin than in heltuba. In the complexes with Man α(1–3) Man α(1–6)-Man (the third residue is disordered in the heltuba complex), for example, the interactions of the second sugar residue onwards with the lectin are very different in the two lectins. The interactions are more extensive and affinity is higher in artocarpin. Thus this comparison shows how change in loop length can be used as a strategy for generating difference in affinity.

Internal sequence similarity and two binding sites in a monocot banana lectin

Mannose specific dimeric banana (*Musa paradisiaca*) lectin, the structure of which was determined in this laboratory and elsewhere [51, 52], carries two binding sites, one primarily on Greek key I and the other on

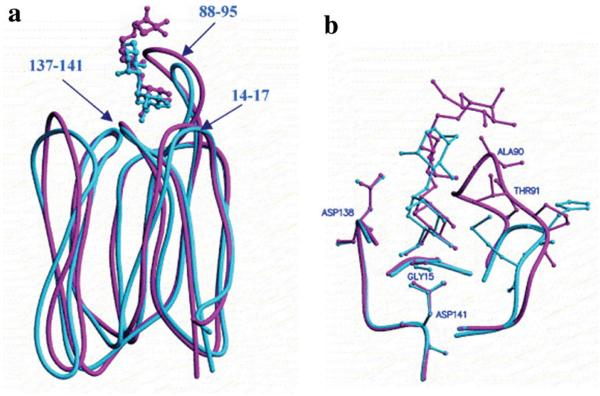

Fig. 11 (a) Structural superposition of heltuba (cyan) on artocarpin (magenta) with bound mannotriose. (b) Sugar-binding sites along with ligands in the two lectins. The third sugar residue is disordered in the heltuba complex. Reproduced from ref. [50].

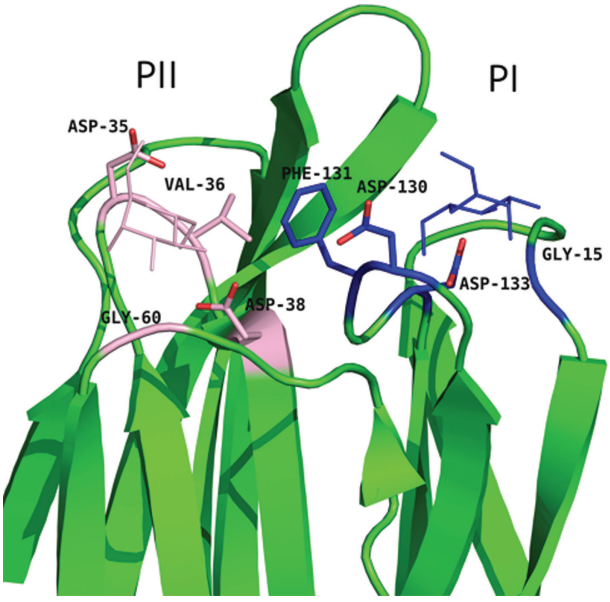

Fig. 12 The primary binding sites PI and PII with bound sugars (Me-α-Man) in banana lectin. Reproduced from ref [82].

Greek key II (Fig. 12). It also turns out that, unlike in the other β-prism I fold lectins studied till then (all from dicots), there is detectable sequence similarly among the three Greek keys. These two observations appear to have evolutionary implications. We had earlier suggested that β-prism I fold lectins could have arisen through successive gene duplication of a primitive carbohydrate binding motif with Greek key topology [42]. Dicot and monocot lectins appear to have adopted somewhat different evolutionary paths. The path of dicot lectins has been such as to obliterate the similarities among the three Greek keys. Vestiges of these similarities remain in monocot lectins.

Anomeric nature of glycosidic linkage and orientation of bound sugar

In the several disaccharide complexes of jacalin, the reducing end occupies the primary binding site when the glycosidic linkage is β(1–3). It was hypothesized that β(1–3) linked disaccharides with the non-reducing end at the primary binding site would lead to steric clashes of the second sugar residue with secondary binding site A in jacalin. The issue came into sharper focus when the complexes of banana lectin with Man α(1–3) Man and Glc β(1–3) Glc (laminaribose) were compared [51, 53]. The non-reducing end occupies the primary binding site in the mannobiose complex while the reducing end does in the laminaribose complex (Fig. 13). The only difference between Man and Glc is in the orientation of O2 which is not involved in any interaction with the lectin molecule. Therefore Man and Glc are expected to behave in the same manner with respect to binding to the lectin. The difference in the behavior of the two disaccharides in their orientation when bound to the lectin could be explained in terms of detailed modeling [53]. The importance of the anomeric nature of the glycosidic linkage on the orientation of the bound sugar could be of significance in determining the affinity of lectins to glyoproteins and other glycoconjugates.

Variability in quaternary association

Although the subunits of all the β-prism I fold lectins have essentially the same structure, they exhibit considerable variability in quaternary association. This variability is most striking in the case of the mannose

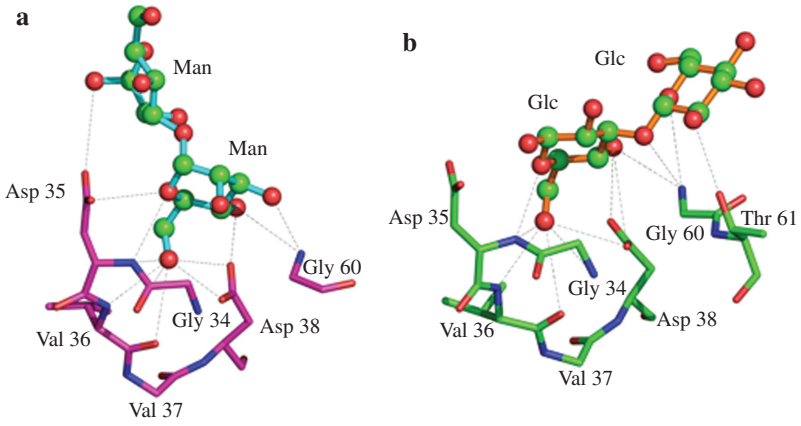

Fig. 13 Interactions of banana lectin with (a) Man α(1-3) Man and (b) Glc β(1-3) Glc at site PII (PDB Codes: 3MIU and 2BNO).

specific lectins. Among them banana lectin and calsepa [54] are dimeric, artocarpin is a tetramer and heltuba is an octamer (Fig. 14). The dimeric association in banana lectin, designated as AB, is different from that in calsepa, A′B′. Both dimers are two-fold symmetric. The N- and the C-terminal stretches, apart from other regions, are involved in dimerization in both the cases. Tetrameric artocarpin is a dimer of two AB type dimers (AB and CD) with 222 symmetry. Four AB type dimers make up the heltuba octamer with 422 symmetry. In addition to the native structures, models involving subunits of one protein with the quaternary structure of another were also considered. For example, banana lectin on artocarpin was constructed by superposing a banana lectin subunit on each of the four subunits of artocarpin. Banana lectin on artocarpin and calsepa on artocarpin were not viable on account on steric clashes. The other models and the native structures were

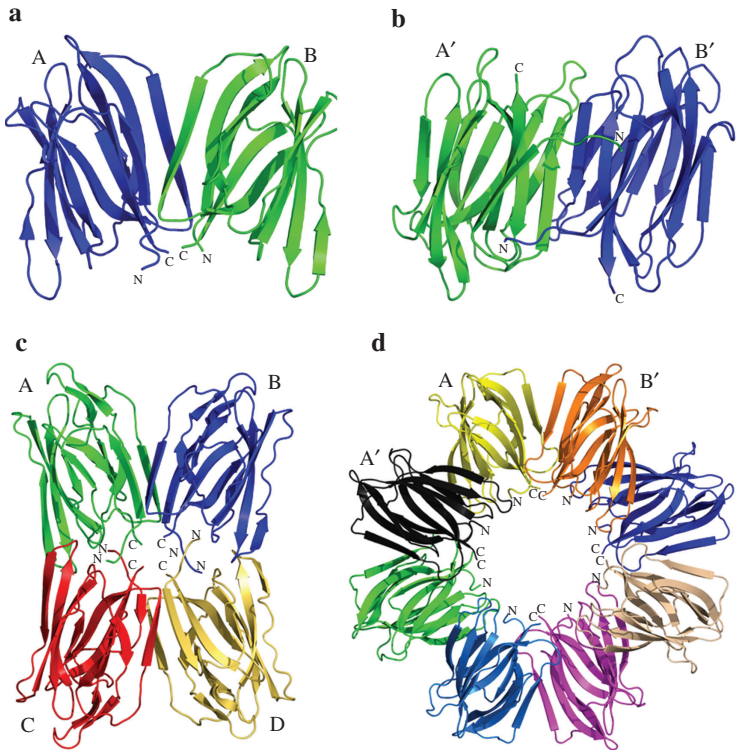

Fig. 14 Quaternary structure in (a) banana lectin (PDB Code: 1X1V), (b) calsepa (PDB Code: 1OUW), (c) artocarpin (PDB code: 1J4S) and heltuba (PDB Code: 1C3K). Adapted from ref. [55].

energy minimized. Molecular dynamics (MD) simulations were carried out on these energy minimized structures and also on the individual subunits of the four lectins [55].

The results of the simulations of the native oligomers in comparison with those resulting from the calculations involving individual subunits provide insights into the intrinsic features of the subunit structure and the changes that occur on oligomerization. They also provide information on the plasticity of the molecule. Interestingly, the internal symmetry of the oligomers is substantially retained during simulations, although no symmetry restraints were applied in the calculations, indicating the robustness of oligomeric arrangement.

A comparison of the simulations of the native structure and the different models show that, for a given quaternary arrangement, the subunit in the native structures yields the minimum energy compared to the subunit of any other lectin, even after adjustments involving energy minimization and molecular dynamics. For instance, the native calsepa dimer turned out to be much more stable than banana lectin on calsepa. Most of the constructed models, such as, for example, banana lectin on heltuba, were untenable on account of severe steric clashes, fell apart during simulation or yielded extremely poor interaction energies among subunits. This provides a rationale for the kind of quaternary association adopted by each type lectin. The only model which remained intact during simulations with reasonable energies is heltuba on artocarpin. This is in consonance with the observation of a tetrameric species of heltuba in solution [56].

Conformational selection and induced fit

The MD simulations carried out on some β-prism I fold lectins, including those on the sugar complexes of jacalin [57] and banana lectin [53], provide insights into the issue relating to conformational selection and induced fit. In the context of ligand binding, conformational selection involves the utilization of those conformations that are most appropriate for the ligand, from an ensemble of possible conformations. Induced fit refers to the creation of a distortion in the existing native conformation to enable the ligand to bind. The relative prevalence of these mechanisms has been addressed in ligand binding to jacalin as well as banana lectin. One example involving jacalin is outlined below to illustrate the situation.

The population distribution in terms r.m.s.ds of atoms in the binding site in MD simulations on free jacalin is shown as a continuous line in Fig. 15a. The distribution of χ^2 in Tyr 122, which defines the orientation of the aromatic ring, is given in Fig. 15b. Also shown in the figure are the corresponding distributions in jacalin complexes with α-anomeric (dashed line) and β-anomeric (dotted line) sugars. The population distribution of atoms in the binding site has a major peak and a minor peak. Each peak corresponds to an ensemble of closely related conformations. Conformations corresponding to the major peak are selected for ligand binding, in a clear case of conformational selection. In the case of Tyr 122, conformations different from those existing in the native structure are adopted in the complexes, in a case of induced fit. χ^2 has different conformations in complexes with α-anomeric and β-anomeric sugars, as the orientation of the aromatic ring required for forming C-H...π hydrogen bonds with anomeric substituents are different in the two cases. On the whole, conformational selection and induced fit appear to operate simultaneously and in a complementary manner on ligand binding. This is true in relation to the conformational changes that occur in the ligand on complexation as well. The work on β-prism I fold lectins also shows how conformational selection can operate in oligomerization and evolution.

β-prism II fold lectins

Dimeric garlic lectin. Oligomerization and ligand specificity

The β-prism II fold in lectins was first characterized through the X-ray crystallographic analysis of snowdrop lectin [58]. Each subunit of this mannose specific tetrameric lectin can be described as made up of three

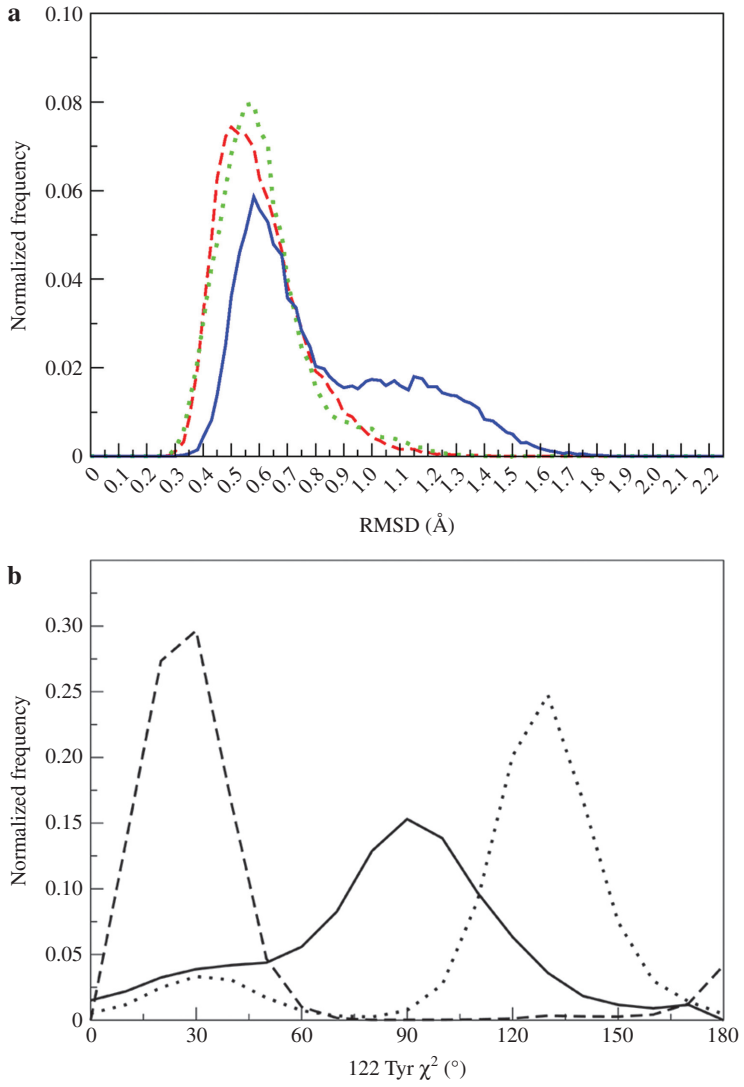

Fig. 15 (a) Population distribution in terms of r.m.s.d. from the starting model in the structures of unliganded jacalin (blue solid line) and complexes involving α-anomeric sugars (red dashed line) and those involving β-anomeric sugars (green dotted line). (b) The population distribution of χ^2 in simulations involving the free lectin (solid line), in complexes with α-anomeric sugars (dashed line) and in those with β-anomeric sugars (dotted line). Adapted from ref. [57].

Greek key motifs arranged as three sides of a prism. The orientation of the strands in the keys is perpendicular to the axis of the prism unlike in the β-prism I fold where they are parallel to the axis. The appropriate three-fold symmetry of each subunit is reflected in the sequence. Each Greek key carries a sugar binding site.

Garlic lectin, analysed in this laboratory, is dimeric with each subunit assuming a β-prism II fold [59, 60]. Technically it is a hetero dimer with the two subunits having molecular weights of 11.5 and 12.5 kDa. The two subunits exhibit a sequence identity of 90 %. As in snowdrop lectin, each subunit carries three biologically significant mannose-binding sites. The reason for the inability of garlic lectin to form tetramers like snowdrop lectin can be explained in terms of differences in critical amino acid residues. Garlic lectin differs from tetrameric lectins like snowdrop lectin in oligosaccharide specificity. Snowdrop lectin binds to the HIV surface glycoprotein of gp120. In the biologically important mode of this binding, the branched mannopentose of gp120 crosslinks two dimers in the tetramer (Fig. 16) [61]. This is obviously not possible in garlic lectin as it is a dimer without a partner to get crosslinked to. This is a situation where oligomerization is used as a strategy for generating ligand specificity.

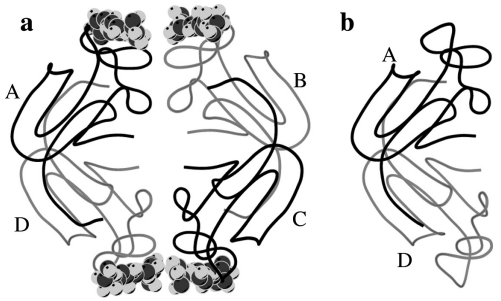

Fig. 16 Quaternary association in (a) snowdrop lectin and (b) garlic lectin. The bound sugar in (a) is a mannose pentasaccharide. Reproduced from ref. [7].

Multivalency

Garlic lectin was used to explore multivalency as well [62]. Solution studies had shown that the binding affinity of the lectin for high mannose oligosaccharides is orders of magnitude higher than that for mannose. This could not be explained on the basis of the binding of the oligosaccharide to a single molecule of the lectin making use of the several sugar binding sites on it. A further exploration showed that oligosaccharides of sufficient length can crosslink lectin dimers, amplifying the protein-oligosaccharide interactions several fold, thus explaining the role of mutivalency in lectins. It turned out that a given dimer pair can be crosslinked simultaneously by two oligosaccharides. The number of crosslinks increases with the size of the oligosaccharide. They often lead to lattices as well (Fig. 17). In addition to explaining the results of solution studies on garlic lectin, these studies are of general interest in relation to the multivalency of lectins.

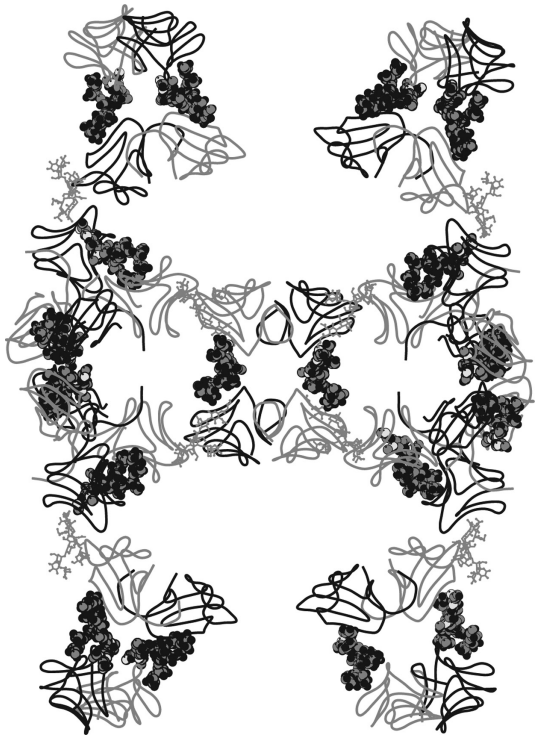

Fig. 17 A lattice-like aggregation starting from the tetramer-like doubly linked pair. The tetramers are linked to each other by a mannose nonasaccharide. Reproduced from ref. [62].

Occurrence and evolution of β-prism fold lectins

β-Prism I and β-prism II fold lectins exhibit substantial similarities as well as differences. For instance, both involve three-fold symmetric arrangement of Greek keys or Greek key-like motifs, although the arrangements are different in the two cases. However, their binding sites are very different. Therefore, a simultaneous genomic search of the two types of lectins was undertaken [63].

The number of binding sites in each subunit can vary between 1 and 3. In β-prism I fold lectins, the number is most often 1, occasionally 2 and rarely 3; it is always 1 in those from dicots. The number is most often 3 in β-prism II fold lectins. The diversity in the number of binding sites is unrelated to the taxonomical position of the organism. However, there is a reasonable correlation between the symmetry in the sequence and the number of binding sites. β-prism I fold lectins predominantly occur in plants. They are found with comparable frequency in dicots as well as monocots. β-prism II fold lectins exhibit more widespread occurrence. Next to plants, their frequency of occurrence is the highest in monera. In plants they mostly occur in monocots.

The genomic analysis suggests that the two types of β-prism lectins perhaps had a common evolutionary origin. Both can be considered to have evolved through successive gene duplication and fusion of a primitive gene corresponding to a Greek key based carbohydrate binding motif. The Greek keys are assembled in different ways in the two lectins (Fig. 18). Since their assembly, they appear to have followed somewhat different

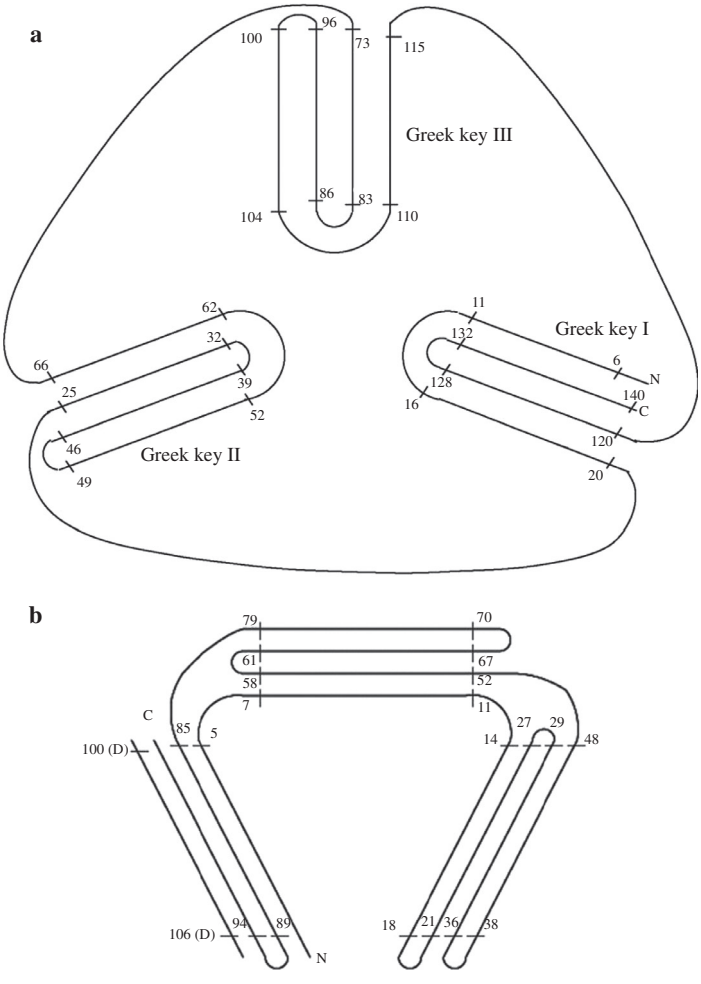

Fig. 18 Schematic representations of the topology of (a) the banana lectin fold and (b) the garlic lectin fold. In (b), the strand from the other subunit involved in swapping to form the complete third Greek key has been labelled as 100(D) to 106(D). Reproduced from ref. [63].

evolutionary paths. In plants, there is perhaps further divergence in the path between dicots and monocots. The number of binding sites tends to be higher in monocots. Dicots have a cork cambium layer to protect them, while such a layer does not exist in monocots. In the absence of this layer, monocots probably have a higher dependence on other defence mechanisms including that involving lectins. The tendency to retain a larger number of binding sites in monocot lectins is perhaps related to this defence requirement.

Non-toxic homologues of type II ribosome inactivating proteins (RIPs)

Type II RIPs typified by ricin [64], contains a catalytic chain and a lectin chain. The lectin chain enables the molecule to bind to the cell. After entry into the cell, the catalytic chain cleaves the N-glycosidic bond of a single adenine adjacent to the universally conserved α-sarcin loop of the rRNA of the large subunit of ribosome. The two chains are linked by a disulphide bond. The lectin chain is made up of two β-trefoil domains, each carrying two galactose binding sites, 1α and 2γ. Interestingly some edible plants are also found to contain homologue of type II RIPs which should obviously have no toxicity. The work on such homologues was taken up to elucidate the structural basis for the absence of toxicity in them.

Snake gourd seed lectin. Structural basis for absence of toxicity

The sequence and structure of snake gourd (*Trichosanthes anguina*) seed lectin (SGSL) were determined by mass spectrometry and X-ray crystallography, respectively [65]. The structure of the molecule is similar to that of other type II RIPs except that the catalytic chain is cleaved into two, without affecting its structural integrity (Fig. 19). A thorough examination of the structure revealed critical differences at the sugar biding sites and at the adenine binding site. At 1α, a residue which is Asn or Asp in other RIPs is replaced by a Gly in SGSL. Likewise, an aromatic stacking residue is replaced by His at 2γ (Figs. 20a and 20b). At the active site of the catalytic chain, an aromatic residue which stacks against adenine is replaced by a Val in SGSL (Fig. 20c).

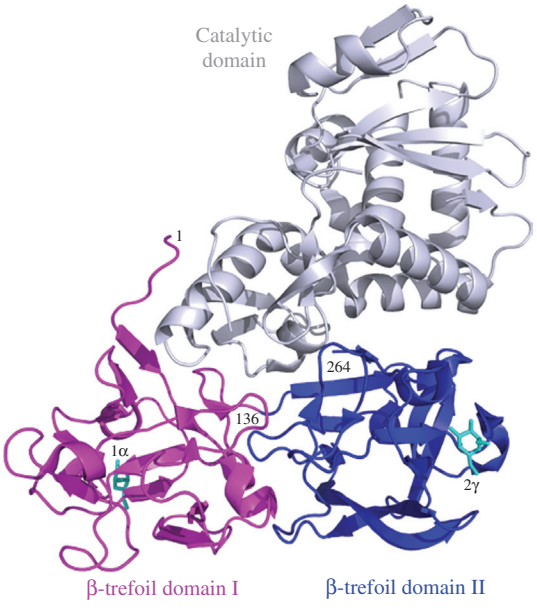

Fig. 19 The structure of SGSL. The catalytic domain and the two lectin domains are colored differently.

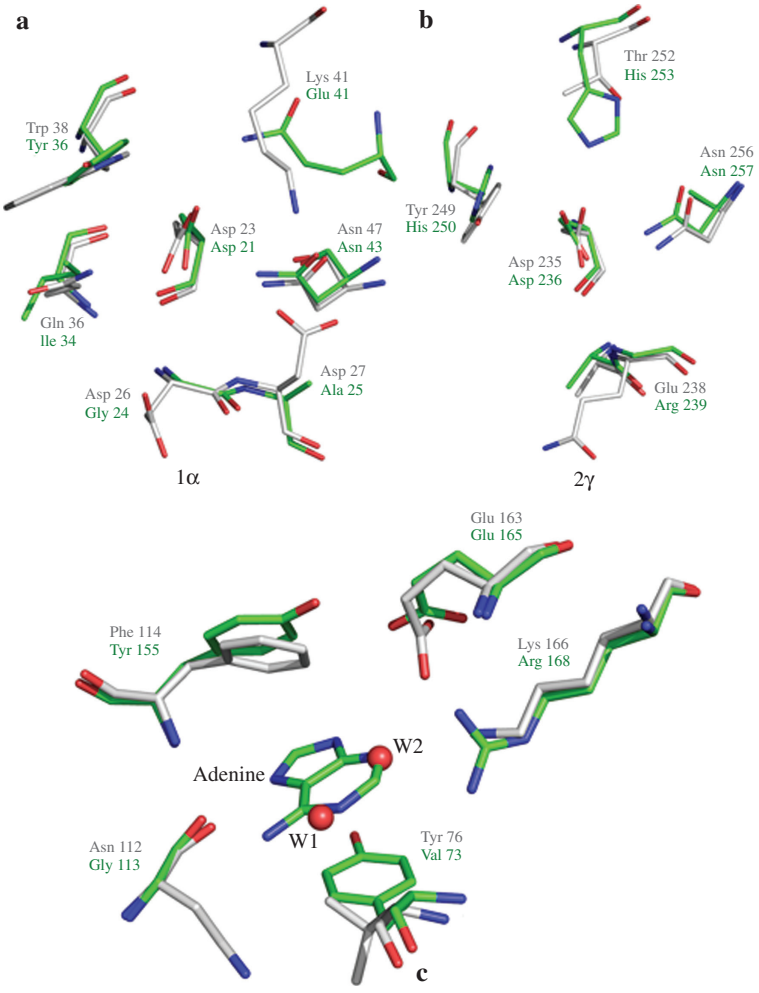

Fig. 20 Superposition of the carbohydrate binding site in (a) 1a and (b) 2γ, of toxic European mistletoe (white) (PDB code: 1PUM) and SGSL (green) (PDB Code: 4HR6). (c) Structural superimposition of the nucleotide-binding site residues of European mistletoe lectin (white) (PDB Code: 1M2T) and SGSL (green) (PDB Code: 4HR6). The mistletoe lectin has a bound adenine. Two water molecules occur in the SGSL structure at sites corresponding to two ring nitrogen atoms of the adenine molecule in the mistletoe lectin structure.

It would thus appear that the non toxicity of SGSL results from a combination of changes at the catalytic and the sugar binding sites. An analysis of sequences on the basis of the known structures of type II RIPs and SGSL, provided valuable insights into the evolution of this class of proteins. The analysis also indicates some variability in sugar binding sites, which could contribute to the different levels of toxicity exhibited by RIP homologues from different sources.

Bitter gourd lectin. Preliminary efforts

Bitter gourd (*Momordica charantia*) lectin exhibits structural homology with SGSL. However, it is a four chain lectin with two RIP homologues connected by a disulphide bridge. Only one structure, that of *Ricinus communis* agglutinin (RCA), of such a four chain lectin has so far been determined [66]. The function of four chain RIPs and the more widely known two chain RIPs is the same. In that context, it is not known why some plants deploy four chain RIPs as well. The structure determination of bitter gourd lectin, which we have initiated [67], might throw some light on this question. It would also further contribute to the understanding of the structural basis for the loss of toxicity of type II RIPs from edible plants.

Mycobacterial lectins. A new initiative

In addition to that on plant lectins, a long range programme of mycobacterial proteins, with special emphasis on TB proteins, is currently under way in this laboratory [68–74]. Taking a common thread from these two programmes, we have recently initiated work on mycobacterial lectins as well. These lectins, which constitute an almost unexplored area, could well be important in many recognitive processes including those involved in infection. The three-dimensional structure of no mycobacterial lectin has so far been determined. On the basis of a preliminary bioinformatics search [63, 75], a lectin form *Mycobacterium tuberculosis* and a lectin domain from *Mycobacterium smegmatis* were cloned, expressed, purified and crystallized. Preliminary X-ray studies on the crystals were carried out [76, 77].

In the meantime, a thorough search was carried out for lectins in all the available mycobacterial genomes [78]. Among the sequences identified as possibly corresponding to lectins or lectin domains, 64 had lectin homologues of known three dimensional structure. They belong to the well-characterised β-prism II, C-type, the *Microcystis viridis* and β-trefoil lectin families. The first three always occur as domains forming part of proteins also containing the LysM, the PI-PLC and the β-grasp domains, respectively, while mycobacterial β-trefoil lectins are unaccompanied by any other domain. Neither the structures of these lectins/lectin domains nor their functions have been characterised. In addition to those identified now, thirty mycobacterial genes, belonging to different species, have been annotated as heparin binding hemagglutinins. Their biological roles have been well characterized, but the structure of none of them has been determined. Thus, mycobacterial proteins constitute a virgin area for structural and related research.

The mycobacterial lectins identified through the genomic search include the two crystallized earlier. The structures of one of them, the lectin domain of MSMEG_3662 from *M. smegmatis*, in complex with mannose and methyl-α-mannose, have now been determined. They have the β-prism II fold found in many plant lectins, but with extensive domain swapping in the dimer [79]. The evolutionary and functional implications of this observation are now being explored. Structural and related studies on other mycobacterial proteins are also in progress.

Summary and conclusions

Tetrameric peanut lectin is perhaps the most thoroughly characterized Gal specific legume lectin. The extensive structural and related work on this lectin, dimeric winged bean lectins and recombinant EcorL has led to a detailed characterization of the structural similarities and differences among them and the nuances of their binding to specific sugars. A β-prism I fold lectin was first characterized through the structure analysis of galactose specific tetrameric jacalin in this laboratory. The other β-prism I fold lectins analysed in this laboratory are mannose specific tetrameric artocarpin and dimeric banana (a monocot) lectin. The structures of these lectins and their complexes have contributed substantially to the full characterization of the structure and interaction of this class of lectins. The work on β-prism II (garlic) and β-trefoil (snake gourd and bitter gourd) lectins have been less extensive, but has led to valuable insights into the structure, sugar-binding and the evolutionary relationships of these lectins. Structural studies on mycobacterial lectins, which constitute a virgin area, have recently been initiated.

In addition to providing information specific to different lectins and lectin families, the work has yielded results that are of general relevance to the structure and action of proteins. The structure of peanut lectin demonstrated that open structures without point group symmetry cannot be ruled out when dealing with quaternary association. The structure analysis of jacalin led to the identification of a new lectin fold. Legume lectins were among the first systems in which it was demonstrated, primarily though the work on peanut and winged beans, that small alterations in essentially the same tertiary structure can lead to widely different quaternary structures. This phenomenon was further elaborated through the work on β-prism I fold lectins. In both the cases, the observed variability in quaternary association could be rationalized to a reasonable extent on structural terms. The extensive work on plant lectins led to the elucidation of several strategies

that could be employed for generating ligand specificity. They include formation of water bridges, variation of loop length, post translational proteolysis and oligomerization. Indeed, the lectin structures elucidated in different laboratories around the world, including ours, constitute a microcosm of the protein world in terms of structural diversity and strategies of generating ligand specificity.

Acknowledgments: The work discussed in this review has resulted from the effort of a large number of students, post doctoral fellows and colleagues. It has all along been supported by the Department of Science and Technology. The graphics facility supported by the Department of Biotechnology has been used for computations. KVA is a C. S. I. R Senior Research Fellow and MV is the Albert Einstein Professor of the Indian National Science Academy.

References

[1] J. M. Rini, Y. D. Lobsanov. *Curr. Opin. Struct. Biol.* **9**, 578 (1999).
[2] T. Feizi. *Immunol. Rev.* **173**, 79 (2000).
[3] N. Sharon, J. Gallagher. *Curr. Opin. Struct. Biol*. **17**, 495 (2007).
[4] M. E. Taylor, K. Drickamer. *Curr. Opin. Struct. Biol*. **19**, 572 (2007).
[5] T. K. Dam, C. F. Brewer. *Glycobiology* **20**, 1061 (2010).
[6] J. Arnaud, A. Audfray, A. Imberty. *Chem. Soc. Rev*. **7**, 4798 (2013).
[7] M. Vijayan, N. R. Chandra. *Curr. Opin. Struct. Biol*. **9**, 707 (1999).
[8] K. D. Hardman, C. F. Ainsworth. *Biochemistry* **11,** 4910 (1972).
[9] G. N. Reeke Jr, J. W. Becker, G. M. Edelman. *J. Biol. Chem*. **250**, 1525 (1975).
[10] C. S. Wright. *J. Mol. Biol*. **111**, 439 (1977).
[11] D. M. Salunke, M. I. Khan, A. Surolia, M. Vijayan. *J. Mol. Biol*. **154**, 177 (1982).
[12] I. A. Wilson, J. J. Skehel, D. C. Wiely. *Nature* **289**, 366 (1981).
[13] S. K. Natchiar, K. Suguna, A. Surolia, M. Vijayan. *Crystallograp. Rev.* **13**, 3 (2007).
[14] J. Monod, J. Wyman, J. P. Changeux. *J. Mol. Biol.* **12**, 88 (1965).
[15] A. J. Cornish-Bowden, D. E. Koshland Jr. *J. Biol Chem.* **246**, 3092 (1971).
[16] R. Banerjee, S. C. Mande, V. Ganesh, K. Das, V. Dhanaraj, S. K. Mahanta, K. Suguna, A. Surolia, M. Vijayan. *Proc. Natl. Acad. Sci., USA* **91**, 227 (1994).
[17] R. Banerjee, K. Das, R. Ravishankar, K. Suguna, A. Surolia, M. Vijayan. *J. Mol. Biol.* **259**, 81 (1996).
[18] R. Ravishankar, M. Ravindran, K. Suguna, A. Surolia, M. Vijayan. *Curr. Sci.* **72**, 855 (1997).
[19] R. Ravishankar, A. Surolia, M. Vijayan, S. Lim, Y. Kishi. *J. Am. Chem. Soc.* **120**, 11297 (1998).
[20] R. Ravishankar, K. Suguna, A. Surolia, M. Vijayan. *Acta Cryst.***D55**, 1375 (1999).
[21] R. Ravishankar, C. J. Thomas, K. Suguna, A. Surolia, M. Vijayan. *Proteins*. **43**, 260 (2001).
[22] S. K. Natchiar, O. Srinivas, N. Mitra, A. Surolia, N. Jayaraman, M. Vijayan. *Acta Cryst.* **D62**, 1413 (2006).
[23] S. K. Natchiar, A. A. Jeyaprakash, T. N. Ramya, C. J. Thomas, K. Suguna, A. Surolia, M. Vijayan. *Acta Cryst.* **D60**, 211 (2004).
[24] S. K. Natchiar, O. Srinivas, M. Nivedita, S. Dev, N. Jayaraman, A. Surolia, M. Vijayan. *Cur. Sci.* **90**, 1230 (2006).
[25] L. T. Delbaere, M. Vandonselaar, L. Prasad, J. W. Quail, K. S. Wilson, Z. Dauter. *J. Mol. Biol*. **230**, 950 (1993).
[26] B. Shaanan, H. Lis, N. Sharon. *Science* **254**, 862 (1991).
[27] M. M. Prabu, R. Sankaranarayanan, K. D. Puri, V. Sharma, A. Surolia, M. Vijayan, K. Suguna. *J. Mol. Biol.* **276**, 787 (1998).
[28] N. Manoj, V. R. Srinivas, A. Surolia, M. Vijayan, K. Suguna. *J. Mol. Biol.* **302,** 1129 (2000).
[29] K. A. Kulkarni, A. Srivastava, N. Mitra, N. Sharon, A. Surolia, M. Vijayan, K. Suguna. *Proteins* **56**, 821 (2004).
[30] T. W. Hamelryck, R. Loris, J. Bouckaert, M. H. Dao-Thi, S. Strecker, A. Imberty, E. Fernandez, L. Wyns, M. E. Etzler. *J. Mol. Biol.* **286,** 1161 (1999).
[31] M. M. Prabu, K. Suguna, M. Vijayan. *Proteins*. **35,** 58 (1999).
[32] N. R. Chandra, M. M. Prabu, K. Suguna, M. Vijayan. *Protein Eng.* **14,** 857 (2001).
[33] J. V. Pratap, G. M. Bradbrook, G. B. Reddy, G. B. A. Surolia, J. Raftery, J. R. Helliwell, M. Vijayan. *Acta Cryst.* **D57,** 1584 (2001).
[34] K. A. Kulkarni, S. Sinha, S. Katiyar, A. Surolia, M. Vijayan, K. Suguna. *FEBS Lett.* **579**, 6775 (2005).
[35] K. A. Kulkarni, S. Katiyar, A. Surolia, M. Vijayan, K. Suguna. *Acta Cryst.* **D62,** 1319 (2006).
[36] K. A. Kulkarni, S. Katiyar, A. Surolia, M. Vijayan, K. Suguna. *Proteins* **68**, 762 (2007).
[37] K. A. Kulkarni, S. Katiyar, A. Surolia, M. Vijayan, K. Suguna. *Acta Cryst.* **D64**, 730 (2008).
[38] V. Sharma, M. Vijayan, A. Surolia. *J. Biol. Chem.* **271,** 21209 (1996).
[39] V. Sharma, V. R. Srinivas, P. Adhikari, M. Vijayan, A. Surolia. *Glycobiology* **8,** 1007 (1998).
[40] P. Adhikari, K. Bachhawat-Sikder, C. J. Thomas, R. Ravishankar, A. A. Jeyaprakash, V. Sharma, M. Vijayan, A. Surolia. *J. Biol. Chem.* **276,** 40734 (2001).

[41] S. Thamotharan, T. Karthikeyan, K. A. Kulkarni, K. N. Shetty, A. Surolia, M. Vijayan, K. Suguna. *Acta Cryst.* **D67.** 218 (2011).
[42] R. Sankaranarayanan, K. Sekar, R. Banerjee, V. Sharma, A. Surolia, M. Vijayan. *Nature Struct. Biol.* **3**, 596 (1996).
[43] J. V. Pratap, A. A. Jeyaprakash, P. G. Rani, K. Sekar, A. Surolia, M. Vijayan. *J. Mol. Biol.* **317**, 237 (2002).
[44] W. C. Chang, K. L. Liu, F. C. Hsu, S. T. Jeng, Y. S. Cheng. *PLoS One*, **7**, e40618 (2012).
[45] A. A. Jeyaprakash, G. P. Rani, G. B. Reddy, S. Banumathi, C. Betzel, K. Sekar, A. Surolia, M. Vijayan. *J. Mol. Biol.* **321**, 637 (2002).
[46] A. A. Jeyaprakash, S. Katiyar, C. P. Swaminathan, K. Sekar, A. Surolia, M. Vijayan. *J. Mol. Biol.* **332,** 217 (2003).
[47] A. A. Jeyaprakash, G. Jayashree, S. K. Mahanta, C. P. Swaminathan, K. Sekar, A. Surolia, M. Vijayan. *J. Mol. Biol.* **347,** 181 (2005).
[48] Y. Bourne, C. H. Astoul, V. Zamboni, W. J. Peumans, L. Menu-Bouaouiche, E. J. Van Damme, A. Barre, P. Rogue. *The Biochemical Journal*. **364**, 173 (2002).
[49] Y. Bourne, V. Zamboni, A. Barre, W. J. Peumans, E. J. Van Damme, P. Rogue. *Structure.* **7**, 1473 (1999).
[50] A. A. Jeyaprakash, A. Srivastav, A. Surolia, M. Vijayan. *J. Mol. Biol.* **338**, 757 (2004).
[51] J. L. Meagher, H. C. Winter, P. Ezell, I. J. Goldstein, J. A. Stuckey. *Glycobiology* **15**, 1033 (2005).
[52] D. D. Singh, K. Saikrishnan, P. Kumar, A. Surolia, K. Sekar, K. M. Vijayan. *Glycobiology* **15**, 1025 (2005).
[53] A. Sharma, M. Vijayan. *Glycobiology* **21**, 23 (2011).
[54] Y. Bourne, V. Zamboni, A. Barre, W. J. Peumans, C. H. Astoul, E. J. Van Damme, P. Rogue. *J. Biol. Chem.* **279**, 527 (2004).
[55] A. Sharma, M. Vijayan. *J. Biosciences.* **36,** 793 (2011).
[56] E. J. M. Van Damme, A. Barre, A. Mazard, P. Verhaert, A. Horman. *Eupb. J. Biochem.* **259**, 135 (1999).
[57] A. Sharma, K. Sekar, M. Vijayan. *Proteins* **77**, 760 (2009).
[58] G. Hester, H. Kaku, I. J. Goldstein, C. S. Wright CS. *Nature Struct. Biol.* **2**, 472 (1995).
[59] N. R. Chandra, G. Ramachandraiah, K. Bachhawat, T. K. Dam, A. Surolia, M. Vijayan. M. *J. Mol. Biol.* **285**, 1157 (1999).
[60] G. Ramachandraiah, N. R. Chandra, A. Surolia, M. Vijayan. *Acta Cryst.* **D58**, 414 (2002).
[61] G. Hester, C. S. Wright. *J. Mol. Biol.* **262,** 516 (1996).
[62] G. Ramachandraiah, N. R. Chandra, A. Surolia, M. Vijayan. *Glycobiology* **13,** 765 (2003).
[63] A. Sharma, D. Chandran, D. D. Singh, M. Vijayan. *J. Biosciences.* **32**, 1089 (2007).
[64] W. Montfort, J. E. Villafranca, A. F. Monzingo, S. R. Ernst, B. Katzin, E. Rutenber, N. H. Xuong, R. Hamlin, J. D. Robertus. *J. Biol Chem.* **262,** 5398 (1987).
[65] A. Sharma, G. Pohlentz, K. B. Bobbili, A. A. Jeyaprakash, T. Chandran, M. Mormann, M. J. Swamy, M. Vijayan. *Acta Cryst.* **D69**, 1493 (2013).
[66] E. C. Sweeney, A. G. Tonevitsky, D. E. Temiakov, I. I. Agapov, S. Saward, R. A. Palmer. *Proteins* **28**, 586 (1997).
[67] T. Chandran, A. Sharma, M. Vijayan. *Acta Cryst.* F**66** *Struct. Biol. Cryst. Commun.* 1037 (2010).
[68] M. Vijayan. *Tuberculosis* **85**, 357 (2005).
[69] R. Krishna, J. R. Prabu, G. P. Manjunath, S. Datta, N. R. Chandra, K. Muniyappa, M. Vijayan. *J. Mol. Biol.* **367,** 1130 (2007).
[70] M. Selvaraj, S. Roy, N. S. Singh, R. Sangeetha, U. Varshney, M. Vijayan. *J. Mol. Biol.* **372**, 186 (2007).
[71] P. S. Kaushal, R. K. Talawar, P. D. Krishna, U. Varshney, M. Vijayan. *Acta Cryst.* D**64**, 551 (2008).
[72] S. Roy, R. Saraswathi, D. Chatterji, M. Vijayan. *J. Mol. Biol.* **375**, 948 (2008).
[73] J. R. Prabu, S. Thamotharan, J. S. Khanduja, N. R. Chandra, K. Muniyappa, M. Vijayan. *Biochim. Biophys. Acta* **1794,** 1001 (2009).
[74] B. Chetnani, P. Kumar, A. Surolia, M. Vijayan. *J. Mol. Biol.* **400,** 171 (2010).
[75] D. D. Singh, D. Chandran, J. Jeyakani, N. R. Chandra. *Protein Pept. Lett.* **14**, 683 (2007).
[76] D. Patra, R. Srikalaivani, A. Misra, D. D. Singh, M. Selvaraj, M. Vijayan. *Acta Cryst.* F**66** *Struct. Biol. Cryst. Commun.* 1662 (2010).
[77] D. Patra, A. Sharma, D. Chandran, M. Vijayan. *Acta Cryst.* F**67** *Struct. Biol. Cryst. Commun*. 596 (2011).
[78] K. V. Abhinav, A. Sharma, M. Vijayan. *Proteins* **81**, 644 (2013).
[79] D. Patra, P. Mishra, A. Surolia, M. Vijayan. *Glycobiology.* Doi:10.1093/glycob/cwu059. (2014).
[80] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne. *Nucl. Acids Res.* **28**, 235 (2000).
[81] The PyMOL Molecular Graphics Systems, Version 1.5.0.4 Schrodinger, LLC.
[82] T. Chandran, A. Sharma, M. Vijayan. *Adv. Protein Chem. and Struct. Biol.* **92**, 135 (2013).