Proc. Indian natn Sci Acad.(PINSA) B68 No.2 pp205-216 (2002)

Legume Lectins - A Paradigm in Quaternary Structure Variations Arising from Similar Tertiary Structural Fold

NIVEDITA MITRA, V R SRINIVAS and AVADHESHA SUROLIA*
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012

(Received on 12 July 2001; Accepted after revision on 28 August 2001)

Proteins belonging to the legume lectin family are characterized by similarity in their tertiary structures and differences in their modes of quaternary association which are brought about by minor variations in amino acid sequences. It is observed that differences in the modes of their association cause them to adopt different stabilities and folding patterns as studied by differential scanning calorimetry and isothermal denaturation experiments. The proteins with the canonical and the handshake mode of association show a two-state denaturation profile but differ in their calorimetric to van't Hoff enthalpy ratio, thus indicating that these proteins unfold in slightly different ways. In this class of proteins we also see a homotetrameric protein like the peanut agglutinin, which neither shows a D_2 nor a four-fold symmetry. This kind of unusual structure confers upon it an unusual folding pathway where the presence of a partially unfolded and active monomeric intermediate is observed. Thus we see that these proteins are good specimens to study the effects of minor alterations in their sequences on their oligomeric association.

Key Words: Lectins, Proteins, Quaternary associations, Tertiary structural fold

Introduction

Lectins are a ubiquitously distributed class of proteins in nature and are characterized by their carbohydrate-recognizing properties. These are proteins of non-immune origin that bind to mono- and oligosaccharides reversibly with a high degree of stereospecificity in a non-catalytic manner (Sharon & Lis 1989, Lis & Sharon 1998). This specific recognition of complex carbohydrates is implicated in important biological processes such as protein targeting to cellular compartments (Ashwell & Harford 1982), homing of leukocytes (Springer & Laskey 1991), host-pathogen interactions (Sharon & Lis 1989), clearance of glycoproteins from the circulatory system as well as cell interactions in the immune system, in malignancy and metastasis. Although the function of lectins from animals, bacteria and viruses is much better than plant lectins, structural work on them has gathered momentum only recently when compared to plant lectins.

Plant lectins comprise at least seven distinct families of structurally and evolutionarily related proteins (Van Damme et al. 1998a, Van Damme et al. 1998b). Four of these families, namely, the legume lectins, the type 2 ribosome-inactivating proteins (RIPs), the chitin-binding lectins containing hevein domains and the monocot mannose-binding lectins are considered to be 'large' families. The amaranthins, the cucurbitaceae phloem lectins and the jacalin-related lectins comprise at present only a small number of individual lectins and accordingly are considered 'small' families. Of these different classes of proteins the legume lectins are one of the most extensively studied class of plant proteins.

Legume lectins beautifully demonstrate that how with small variations in amino acid sequences, nearly identical tertiary structures give rise to varied modes of quaternary associations (Vijayan & Chandra 1999). The first protein among the legume lectins whose structure was solved was concanavalin A (Hardman et al. 1971). Legume lectins are multimeric proteins, with the final structure being either a dimer or a dimer of dimers. This recurring tertiary motif in these proteins is called the 'jelly roll'

*Fax: 91-80-360 0535, 91-80-360 0683; *E-mail:* surolia@mbu.iisc.ernet.in

motif (Prabu et al. 1999). This is characterized by the presence to 3 anti-parallel β sheet strands. There is a six-stranded flat 'back' β sheet, a curved 7 stranded 'front' β sheet and a short 5 membered β sheet at the 'top' of the molecule. The sheets are connected by several loops (figure 1a).

The jelly roll fold, seen first in legume lectins, is also seen in lectins of non-legume origin and also in proteins with no known lectin activity. Examples include pentraxins, galectins or S type lectins that are of animal origin and are involved in embryonic development, oncogenesis and immune regulation (Srinivasan et al. 1996). The middle component coat protein from Bean pod mottle virus, bacillus 1-3,1-4-beta-glucanase from *Bacillus macerans* also show the 'jelly roll' motif (VAST search), although, the number of strands in each sheet may vary among proteins.

This fold has also been discovered in several bacterial and fungal proteins with no detectable sequence similarity with legume lectins. The two domains flanking the catalytic region of *Vibrio cholerae* sialidase resemble the lectin fold (Srinivasan et al. 1996). This protein plays a role in pathogenesis in the bacterium, by processing higher order gangliosides of the receptor for cholera toxin.

Sugar Binding site

The general features of the carbohydrate binding sites are very similar in all legume lectins. The carbohydrate binding site in these proteins is made of four loops usually referred to as A, B, C and D, all emanating from the seven stranded concave sheet in the top front of the subunit (Sharma & Surolia 1997). The conserved aspartate and glycine (or arginine) residues are contributed by loops A and B, respectively, whereas an asparagine and a hydrophobic residue (Phe, Tyr, Trp or Leu) are present in loop C. Additional interactions, usually with the back-bone atoms, from loop D complete the basic framework of monosaccharide recognition by a legume lectin (Sharma & Surolia 1997). Sugar recognition in legume lectins also depends on the presence and orientation of the metal ions, Ca^{2+} and Mn^{2+} in each of the subunits. Although all the legume lectins are very specific for their substrates, an exception in the form of Ulex europaeus lectin (UEA) has come up (Loris et al. 2000). This fucose specific lectin from the legume family shows a promiscuous primary binding site capable of accommodating both N-acetylglucosamine or galactose. The hydrogen bonding network in this case is suboptimal in agreement with low affinities of these sugars. But when the chitobiose, lactose and fucosyllactose bind to UEA, additional associations occur in the form of hydrophobic interactions.

Functional Aspect

Although the basic feature of a legume lectin is that it binds carbohydrates reversibly, the final outcome of this activity, i.e. *in vivo* function of the protein in the plant is not known very well. They are thought to play a role in defense, encouraging symbionts or in early developmental processes. A lectin, which functions by helping attach symbionts, is the *Pisum sativum* lectin or PSL. Apart from being available in seeds this protein is also present in small amounts in the tip of developing root hairs. The presence of the lectin there facilitates the attachment of Rhizobium, a nitrogen fixing bacteria. Transformation of white clover hairy roots with the *psl* gene confers upon them the ability to be nodulated by pea rhizobia; thus substantiating the biological activity of PSL in roots (van Eijsden R et al. 1995).

Subsequent experiments show that pea lectin-transformed red clover hairy roots form nodule primordium-like structures after inoculation with pea-, alfalfa-, and Lotus-specific rhizobia, which normally do not nodulate red clover (Diaz et al. 2000). It was also shown that chitopentaose, chitotetraose, chitotriose, and chitobiose were able to induce cortical cell divisions and cell expansion in a radial direction in transgenic roots, but not in control roots. This kind of division is also triggered by lipochitin oligosaccharides (LCO's) secreted by the rhizobial microsymbiont. Thus it is seen that the sugar-binding activity of pea lectin is responsible for root nodulation (Diaz et al. 2000).

A similar role is also played by the lectin/nucleotide phosphohydrolase (Db-LNP) in the roots of the legume *Dolichos biflorus*. This protein binds to Nod factor signals produced by rhizobia that nodulate this plant (Kalsi & Etzler 2000).

The role of lectins in the developmental processes in plants has also been demonstrated in the alfaalfa plant. Lines of transgenic alfalfa plants expressing approximately half of the open reading frame of

MsLEC1 or MsLEC2 (putative alfalfa lectin genes), in the antisense or sense orientation, were established and analyzed (Brill et al. 2001). The antisense plants displayed severe abnormalities in embryogenesis, and both vegetative and reproductive development was perturbed. Thus suggesting that MsLEC1 and MsLEC2 gene products, in addition to being important for embryogenesis, are required throughout alfalfa for development.

Quaternary Associations

The most interesting aspect of legume lectins as mentioned earlier is the variability they have achieved in their quaternary associations. This kind of variation is a direct result of small differences in the amino acid sequence of these lectins (Srinivas et al. 2001). The first lectin whose structure was solved was concanavalin A (Hardman et al. 1971). This lectin formed a dimer with the coming together of the six-stranded back β sheets to form a 12-stranded contiguous sheet (figure 1b, figure 2(i)a). Subsequently, many other legume lectins were found to have a similar dimeric association and hence this interaction came to be known as the 'canonical' mode of dimerization. Subsequently, two other kind of dimeric associations have been found. The 'back to back' mode as seen in GSIV and the 'handshake type' as seen in the Erythrina corallodendron lectin (ECorL) and the winged bean basic (WBA I) and acidic (WBAII) lectins (figure 2(i)b) (Delbaere 1993, Shaanan et al. 1991, Manoj et al. 2000, Prabu et al. 1998). The first among these whose structure was solved was ECorL (Shaanan et al. 1991). The fact that glycosylation would cause stearic hindrance in the protein if it assumed a canonical structure was the reasoning given to justify this alternate structure. But subsequently when structures of proteins WBA I and WBA II from winged bean were solved it was seen that even these proteins had the ECorL like 'handshake' structure, although they didn't have the 'strategically' positioned glycosylation that would prevent the canonical association. This lead to the suggestion that this form of alternative structure was not due to the presence of glycosylation but due to an intrinsic property (i.e. amino acid sequence) of the protein itself.

Another lectin whose structure was found to be different from that of con A was the lectin from

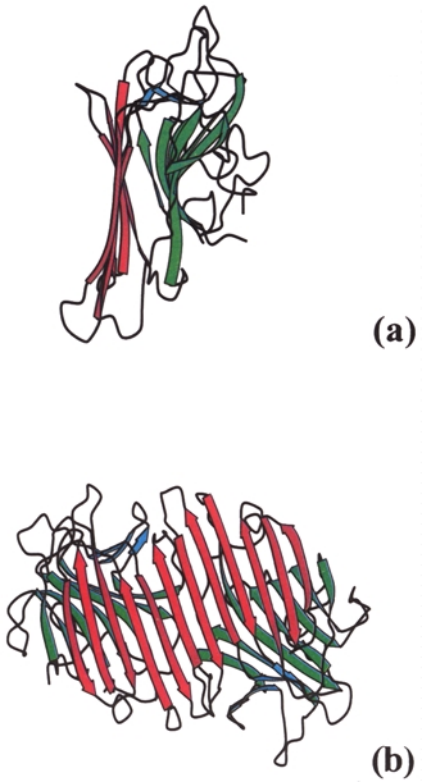

Figure 1 (a): Tertiary structure of a legume lectin monomer is called the "jelly roll" fold which consists of a flat six stranded anti-parallel "back" sheet (dark gray), a curved seven stranded "front" sheet (medium gray) and a five stranded "top" sheet (very light gray) linked by loops of varying lengths. Monomers of different lectins essentially vary in the length and the shape of their loops. (b): Dimerization in the canonical arrangement involves antiparallel alignment of their flat six stranded "back" sheets giving rise to the formation of a contiguous twelve stranded sheet. Subsequently similar mode of dimerization was seen in many other legume lectins, hence it is also described as the "canonical" mode of dimerization.

Griffonia simplicifolia (GSIV) (figure 2 (i)(c) (Delbaere 1993). In this lectin the subunits interact through the 'back to back' mode. It is a heterodimer with both monomers differing in their covalently bound carbohydrates. Asparagine 5 is glycosylated in one and Asparagine18 in both the monomers. In this case the deviation from the canonical mode was ascribed to the presence of a charged residue glutamate 58, which would get buried in the interface in the case of a canonical association.

The next step in the quaternary association is the tetramer formation. Con A forms a tetramer above *p*H 5 (figure 2 (ii) (a) (Hardman et al. 1971). Here the dimers associate with their back sheets. Because of the concave nature of the back β sheets, the center of the molecule is formed by a large water filled cavity

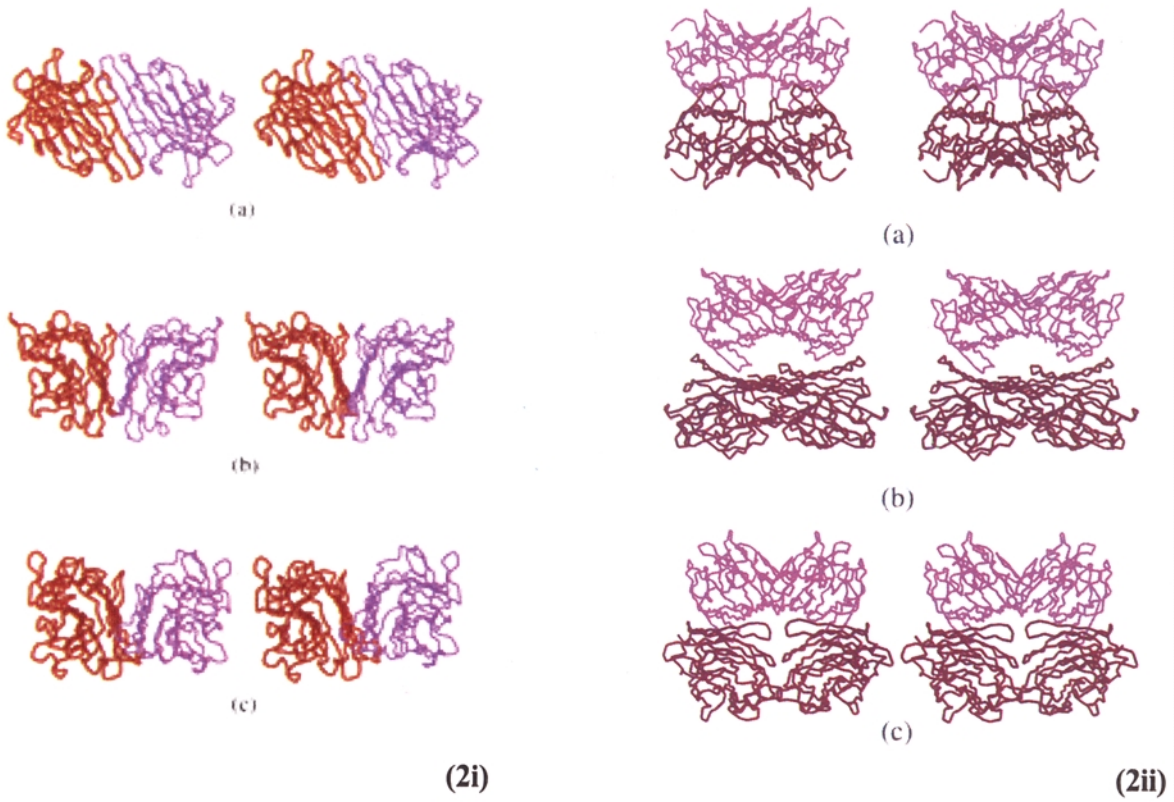

Figure 2 (i) (a): Stereo view of the canonical type of dimer in legume lectins as observed in Con A, pea lectin, lentil lectin etc. (b): Back-to-back "handshake" kind of dimer observed in ECorL, WBA I and WBA II. The two sheets are inclined at about 47.6-49.4° with respect to each other. (c): Back-to-back dimer seen in GS IV and in the two physiologic dimers of PNA i.e. 1-4 and 2-3 dimers. The two sheets are mutually inclined at 78°. (ii)Stereo diagrams of the tetramers in (a): Con A, (b): SBA and PHA-L and, (c): PNA. In Con A the two dimers associate in a back-to-back fashion perpendicular to the canonical dimers at an angle of roughly 82°. In SBA and also in PHA-L the two canonical dimers associate in a back-to-back fashion but which are nearly parallel to each other. PNA, however, shows an open quaternary structure as it neither exhibits C_4 nor D_2 symmetry expected of a tetrameric protein and shows 3 distinct kinds of interfaces. Interface across subunits 1 and 2 is like Con A etc. but the two "back" sheets are bridged by 6 water molecules, the interface between subunits 1-4 and 2-3 is akin to GS IV dimer while that spanning subunits 3-4 is an open one and has not been observed, so far, in any other tetrameric protein. These figures are partly adapted from Prabu et. al 1999, with permission from the authors and publishers.

of about 25 X 25 X 8 Å^3. At low *p*H, the number of van der Waals and hydrogen bonding contacts are drastically reduced which gives rise to the dimeric form in solution (Bouckaert et al. 1995). This reduction can largely be explained by the different protonation states of aspartates and glutamates at the interface.

Another variation of the con A kind of legume lectin tertramerization is seen in soyabean agglutinin (SBA) and phytohemagglutinin (PHA-L) (figure 2 (ii) (b) (Hamelryck 1996, Dessen et al. 1995). Both these proteins form a canonical dimer but tetramerization pattern is distinctly different from that in con A. The tetramerization pattern is not *p*H dependent. Here the two dimers associate with their two outermost strands creating a large channel in the middle of the tetramer. This channel probably contains the disordered C-terminus of these lectins and protects it from proteolytic degradation. The interface formed by the two outermost strands, consist mainly of a number of relatively short side-chains that intercalate in a zipper like fashion.

The most remarkable tetramerization pattern observed till date belongs to the lectin from *Arachis hypogaea*, the peanut agglutinin (PNA) (Banerjee et al. 1996). Unlike the other tetrameric proteins PNA exhibits neither a D2 nor a four-fold symmetry. Consequently, it gives rise to an open quaternary structure (figure 2 (ii) (c). Each PNA tetramer may be considered to be a dimer of two dimers, where subunits 1 and 4 and subunits 2 and 3 associate in a GS IV kind of back-to-back fashion. These dimers in turn come together in such a way that monomers 1 and 2 associate in a canonical fashion, except that the

```
ConA    TNALHFMFNQFSKDQKDLILQGDATTGTDGNLELTRVSSNGSPQGSSVGRALFYAPVHIW 182
PNA     AETVSFNFNSFSEGNPAINFQGDVTVLSNGNIQLTNLN-----KVNSVGRVLYAMPVRIW 55
Pea     TETTSFLITKFSPDQQNLIFQGDGYT-TKEKLTLTKAV------KNTVGRALYSSPIHIW 53

ConA    ESS-AVVASFEATFTFLIKSP-DSHPADGIAFFISNIDSSIPSGSTG-RLLGLFPDANAD 2
PNA     SSATGNVASFLTSFSFEMKDIKDYDPADGIIFFIAPEDTQIPAGSIGGGTLGVSDTKGAG 115
Pea     DRETGNVANFVTSFTFVINAPNSYNVADGFTFFIAPVDTKPQTGGGYLGVFNSAEYDKTT 113

ConA    TIVAVELDTYPNTDIGDPSYP-HIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVDKRL 61
PNA     HFVGVEFDTYSNSEYNDPPTD-HVGIDVNSVDSVKTVPWNSVSGAVVKVTVIYDSSTKTL 174
Pea     QTVAVEFDTFYNAAWDPSNRDRHIGIDVNSIKSVNTKSWKLQNGEEANVVIAFNAATNVL 173

ConA    SAVVSYPNADSATVSYDVDLDNVLPEWVRVGLSASTGLY-KETNTILSWSFTSKLKSNST 120
PNA     SVAVTNDNGDITTIAQVVDLKAKLPERVKFGFSASGSLGGRQIHLIRSWSFTSTLITTTR 234
Pea     TVSLTYPNVTSYTLSDVVSLKDVVPEWVRIGFSATTGAE-YAAHEVLSWSFHSELS---- 228

ConA    HE- 122
PNA     RSN 237
Pea     ---
```

a) Residues interacting in the canonical mode.

```
WBA1    -KTISFNFNQFHQNEEQLKLQRDARISSNSVLELTKVVN-GVPTWNSTGRALYAKPVQVW 58
WBA2    -ETQSFNFDHFEENSKELNLQRQASIKSSGVLELTKLTKNGVPVWKSTGRALYAEPIKIW 59
ECoRL   VETISFSFSEFEPGNDNLTLQGASLITQSGVLQLTKINQNGMPAWDSTGRTLYAKPVHIW 60

WBA1    DSTTGNVASFETRFSFSIRQPFPRPHPADGLVFFIAPPNTQTGEGGGYFGIYNPLSP--- 115
WBA2    DSTTGNVASFETRFSFNITQPYAYPEPADGLTFFMVPPNSPGGEDGGNLGVFKPSNPE-G 118
EcoRL   DMTTGTVASFETRFSFSIEQPYTRPLPADGLVFFMGPTKSKPAQGYGYLGIFNQSKQDNS 120

WBA1    YPFVAVEFDTFRNTWDP-QIPHIGIDVNSVISTKTVPFTLDNGGIANVVIKYDASTKILH 174
WBA2    DNALAVEFDTFQNTWDP-QVPHIGIDVNSIVSSKTLHFQLENGGVANVVIKYDSPTKILN 177
ECoRL   YQTLGVEFDTFSNPWDPPQVPHIGIDVNSIRSIKTQPFQLDNGQVANVVIKYDASSKLLH 180

WBA1    VVLVFPSLGTIYTIADIVDLKQVLPES--VNVGFSAATGDPSGKQRNATETHDILSWSFS 232
WBA2    VVLAFHSVGTVYTLSNIVDLKQEFPNSEWVNVGLSATTG----YQKNAVETHEIISWSFT 233
ECoRL   AVLVYPSSGAIYTIAEIVDVKQVLPEW--VDVGLSGATG----AQRDAAETHDVYSWSFQ 234

WBA1    ASLPGTNEF 241
WBA2    SSLQETN-- 240
ECoRL   ASLPE---- 239
```

b)Residues interacting in the handshake mode.

```
PNA     AETVSFNFNSFS----EGNPAINFQGDVTVLSNGNIQLTNLNKVN-----SVGRVLYAMP 51
GS4     XNTVNFTYPDFWSYSLKNGTEITFLGDATRIP-GALQLTKTDANGNPVRSSAGQASYSEP 59

PNA     VRIWSSATGNVASFLTSFSFEMKDIKDYDPADGIIFFIAPEDTQIP--AGSIGGGTLGVS 109
GS4     VFLWDS-TGKAASFYTSFTFLLKNYG-APTADGLAFFLAPVDSSVKDYGGFLGLFRHETA 117

PNA     DTKGAGHFVGVEFDTYSNSEYNDPPTDHVGIDVNSVDSVKTVPWN---SVSGAVVKVTVI 166
GS4     ADPSKNQVVAVEFDTWINKDWNDPPYPHIGIDVNSIVSVATTRWENDDAYGSSIATAHIT 177

PNA     YDSSTKTLSVAVTNDNGDITTIAQVVDLKAKLPERVKFGFSASGSLGGRQIHLIRSWSFT 226
GS4     YDARSKILTVLLSYEHGRDYILSHVVDLAKVLPQKVRIGFSAG--VGYDEVTYILSWHFF 235

PNA     STLITTTRRSN 237
GS4     STLDGTNK--- 243
```

c) Residues interacting in the back to back mode.

Figure 3 Residues at the intersubunit interface(s) in various legume lectins. The interacting residues were determined by their loss of accessible surface areas (ASA) upon dimerization. The residue-wise ASA was calculated using NACCESS (Hubbard 1996). All the interacting residues are shown as bold and italicized. The dimerization in the case of WBA1, WBA2 and ECorL is through the handshake mode (b). Pea lectin and ConA have a canonical interaction. GS4 dimerizes through the back-to back-mode (c). In peanut lectin, three different interactions take place, the back-to-back which gives rise to the dimer (c). The dimers come together to give rise to two kinds of interfaces, the canonical (a) and the open interface (not shown). Because a circular permutation occurs in conA, the sequence used for alignment starts from residue Thr123. Residues 1 through 122 have been added after the last residue i.e. residue 237. Residue 1 in the protein which is an alanine is underlined.

sheets do not interact directly, but by means of 6 water bridges. The interface between subunits 3 and 4 is an open one and has not been observed in any other protein so far. The residues involved in the intersubunit interactions in some of the representative lectins are shown in figure 3.

Folding Studies

This kind of varied quaternary interactions between molecules which have nearly identical tertiary structures, makes legume lectins a good system to study the effect of subunit oligomerization on the stability, folding, function as well as the evolution of multimeric structures. Towards this end our lab has been working for the past few years to elucidate the folding pathways and stability differences between these proteins. These studies have been done, by inducing denaturation, either chemically or thermally, and then studying the unfolding pattern in lectins with different kinds of association modes. Thermal denaturation is followed using differential scanning calorimetry (DSC) which measures the heat absorbed during unfolding. Chemical denaturation experiments are followed spectroscopically, usually by changes in circular dichroism and fluorescence measurements as the protein progressively unfolds.

The differential scanning calorimeter (DSC) consists of two cells, one containing the sample and the other an inert reference solution. These solutions are heated at a programmed rate by heaters, controlled to give a zero temperature difference between the cells (Freire 1995). When a thermally induced change takes place in the sample, the control system either supplies more or less energy to the sample cell, as the case may be, to maintain a zero temperature difference. The output of the calorimeter is this change in supplied heat as a function of temperature from which the van't Hoff enthalpy (ΔH_v), the calorimetric enthalpy (ΔH_c), the temperature at the peak maximum, T_p and the temperature at half the peak area, T_m are obtained (figure 4). These four parameters together define the stability of a protein. Another important parameter describing the unfolding process is the ratio of calorimetric to van't Hoff enthalpy. For monomeric single domain proteins, the van't-Hoff enthalpy is expected to be equal to the calorimetric enthalpy. When these values do not match it

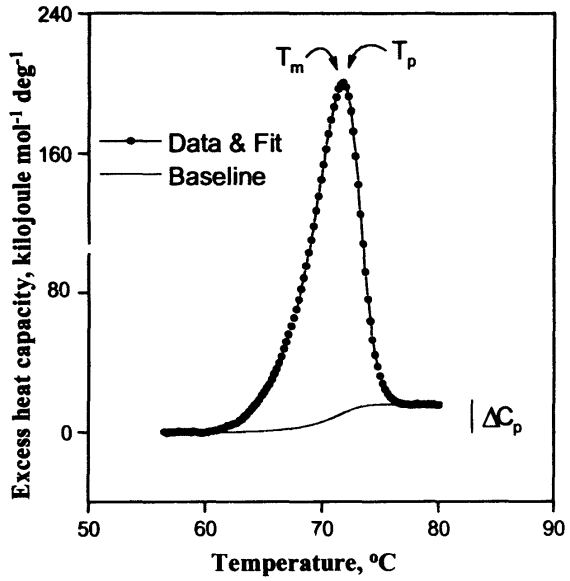

Figure 4 This is a typical single peak DSC scan. The parameters indicated are Tm, the temperature at half peak area. Tp, the temperature at the transition peak maxima. The ΔCp is the heat capacity differences between the folded and the unfolded protein. The area under the peak gives the calorimetric enthalpy of the transition, ΔH_c. The van't Hoff enthalpy is given by the ratio of the transition peak height to the peak area. The ratio of these two enthalpies gives the transition co-operativity of the unfolding process.

indicates the participation of more than one molecular entity in the transition. The ratio of calorimetric to van't-Hoff enthalpy, H_c/H_v, is hence a measure of co-operativity of the transition, that can provide information on the number and organization of domains in a protein that participate in the unfolding reaction. In our studies on legume lectins, we observe distinct differences in this ratio between the three groups of legume lectins. These differences correlate well with the general organization of the monomers in the oligomeric structure (Schwarz et al. 1993, Surolia et al. 1996, Schwarz et al. 1991, Srinivas et al. 1998, Reddy et al. 1999) and, in fact, in more than one instance led to the correct prediction of the mode of monomer association in this class of proteins (Schwarz et al. 1991, Srinivas et al. 1998).

The extensive buried area of 1000 Å^2 per monomer at the interface of con A seems to give it considerable structural stability. In the case of the lectins with this mode of association, differential scanning calorimetric measurements show that the dimers unfold and dissociate into respective monomers at the denaturation temperature. The unfolding profiles are distinctly symmetric and the ratio of

calorimetric to van't-Hoff enthalpies is close to one (Schwarz et al. 1993). Considering that these are dimeric proteins this is quite striking, since it suggests that these proteins unfold as though they were single molecular entities, despite the presence of non-covalently interacting monomers. The extensive sheet interactions between the monomers appears to be responsible for their highly cooperative unfolding. The strongly stabilizing intersubunit interactions results in the entire dimeric structure behaving as a single molecular species, that dissociates and unfolds together (figure 5).

Another lectin with the canonical mode of dimerization whose folding has been studied is pea lectin (PSL) (Ahmad et al. 1998). Isothermal denaturation and DSC studies show that this protein has a relatively high ΔCp. Similar studies on con A show that ΔG at the temperature of maximum stability and the ΔCp of the protein are comparable to pea lectin (Srinivas et al., unpublished

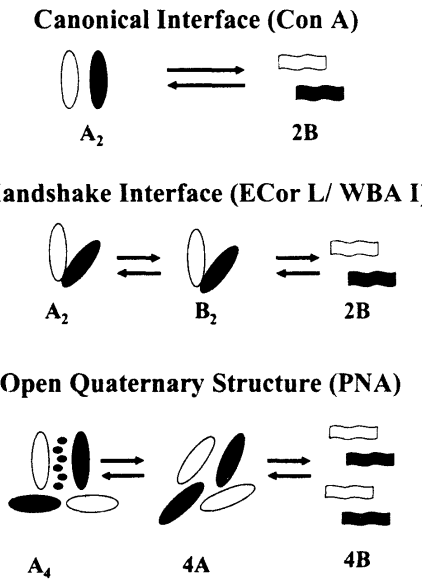

Figure 5 In the canonical type of oligomerization (Top Panel), as seen in Con A, the dimeric interface is extensive. Consequently, it unfolds as a single entity i.e. the interface accounts for the stability of the dimer entirely, such that the dissociation of the dimer and unfolding of the constituent polypeptide chains occur simultaneously. The handshake type of interface (Middle Panel) seen in EcorL etc., is less extensive. Consequently, its two subunits unfold independently, and dissociate at the denaturation temperature. In the case of PNA (Bottom Panel), because of the less tenacious nature of interactions between the subunits in the tetramer, it dissociates rather easily giving rise to the formation of a stable partially unfolded monomeric intermediate, which, interestingly, retains the carbohydrate binding property of the native protein. The folded monomer subsequently unfolds completely.

observations). Gel filtration studies, on progressively denatured protein show the presence of only two kinds of species, the folded dimers and the unfolded monomers. This tells us that the folded monomer is probably incapable of independent existence and hence dimerization is essential for maintaining protein stability.

In the ECorL dimer, the two monomers dock to build an interface of an "handshake" type. The interface is stabilized by hydrophobic as well as polar interactions. The area buried at the interface is about 700 Å^2 per monomer (Shaanan et al. 1991). These monomers thus appear to be less tightly attached when compared to the canonical proteins such as con A. This is reflected in the scanning calorimetric studies on ECorL. The ratio of calorimetric to van't-Hoff enthalpy, H_c/H_v, that is the unfolding co-operativity, is close to two, implying the participation of two unfolding domains in the transition. But the DSC transition is both, single and asymmetric, indicating that the domains unfold at the same temperature and are thus identical in nature. These domains correspond to the two monomers of the protein (Surolia et al. 1996). As a consequence of the less tenacious dimerization, with a smaller contact region (700 Å^2 per monomer), the monomers of the protein unfold and dissociate as independent domains, with each domain contributing independently and significantly to the measured enthalpy (figure 5). The thermal unfolding parameters of WBA I (winged bean basic agglutinin) unexpectedly yielded two peaks (Schwarz et al. 1991). It thus appeared that WBA I possessed relatively lesser dimeric integrity than the "canonically" dimerized lectins (Schwarz et al. 1991, Sharma et al. 1996, Puri & Surolia 1994). Subsequent determination of its crystal structure has shown that WBA I has the same handshake kind of dimeric association, with similar surface area buried at the subunit junction, as in ECorL (Prabu et al. 1998). WBA II (winged bean acidic agglutinin) exhibits a DSC transition very similar to that of ECorL, viz. a single asymmetric transition peak with the H_c/H_v close to two. Given that WBA II is a homodimer like ECorL, this again suggested that just as in ECorL, there existed two independent unfolding domains in this case too. This was again proven to be the case when its structure was solved recently (Manoj

et al. 2000). It must be noted that in general, lectins with canonical mode of dimerization have higher transition temperatures (T_p and T_m) and transition enthalpies than the lectins with the handshake quaternary structure. This again suggests greater stabilization of the former proteins as a result of more extensive contacts between their subunits.

As mentioned above, the peanut agglutinin has a very different and distinct kind of quaternary association, described as an open quaternary structure. Thus while PNA is made up of two GS IV type dimers, both 2-fold symmetric, and related to each other by a dyad, these different dyads fail to intersect (Banerjee et al. 1994). The thermal denaturation of peanut agglutinin was found to be distinctly different from that observed for the other legume lectins. There appears to be two transitions, each with a different T_m and co-operativity ratio. The first peak has an unfolding co-operativity (H_c/H_v) close to 4 (calculated for tetramer concentration) while the second peak has a H_c/H_v ratio close 0.25 (Reddy et al.1999a). The first transition is identified with the dissociation of the tetramer to monomeric intermediates, while the second transition describes the unfolding of these intermediates (figure 5). It may be mentioned that solution denaturation of PNA using urea and guanidine hydrochloride, where the unfolding changes are monitored spectroscopically, exhibited distinct biphasic and non-overlapping profiles, also suggesting the presence of an intermediate in the unfolding equilibrium (Reddy et al. 1999b). The intermediate was monomeric as deduced from gel-filtration studies. It had lost a considerable amount of its tertiary structure but with significant secondary structure as seen in CD studies (Reddy et al. 1999b). This is in striking contrast to other legume lectins, which exhibit typical two-state monophasic denaturation profiles. Thus in the case of PNA, it seems that the back-to-back association between the two subunits, 1-4 and 2-3 is intrinsically less stable, than for instance the canonical dimeric interfaces of con A, pea and lentil lectins. Likewise the 1-2 interface, though resembling the canonical association of these legume lectins, would also be intrinsically less stable due to the presence of six water bridges. Since the 1-2 interface is common to both the 1-4 and 2-3 dimer, the whole tetramer perhaps dissociates quite readily into monomers.

Table 1 *Relative sugar binding ability of the native and the monomeric PNA unfolding intermediate using isothermal titration calorimetry*

Sugar	nt PNA	mPNA(intermediate)
Lactose	1	1
T-antigen	11.9	12.2
Methyl-a-galactose	1.5	1.4
Methyl-b-galactose	0.94	0.91

*The table has been derived from the binding constant values reported in Reddy et. al. 1999b. The values provided are relative to the binding affinities to lactose. K_b for lactose for native PNA and monomeric PNA are 1990 and 1620 M^{-1} respectively.

Discussion

As stated quite often in the above discussion, legume lectins are oligomeric proteins. But, is the oligomeric state a requirement of structure or function of the protein. For the basic function of carbohydrate binding oligomerization is not essential. This is seen in case of the monomeric intermediate in the PNA unfolding pathway. This intermediate despite having lost most of its tertiary structure is able to bind to sugar with affinities comparable to that of the native protein (table 1) (Reddy et al. 1999b). But for biological functions of lectins to be fulfilled this monovalent existence is not helpful. This is seen in case of con A. Here it is observed that sugars containing 1, 2 or 5 mannose groups bind to dimeric con A with affinities equal to that of tetrameric con A, but this similarity is not observed in higher order sugars with 7, 8 or 9 mannose groups where the affinities in the tetrameric protein are much higher as compared to the dimeric molecule (Mandal & Brewer 1998). Thus oligomerization appears to impart to these lectins the finer specificity needed for the recognition of oligosaccharides and complex glycoconjugates that is responsible for their specific biologic activities. But the role of oligomerization in structural stabilization is also evident from the fore-going space discussion.

The legume lectin family of proteins can provide considerable scope for discussions on the evolution of oligomeric proteins. The jelly-roll fold common to all legume lectins, is also found in several non-legume lectins, as well as non-lectin proteins, thus showing that this structure is conserved in

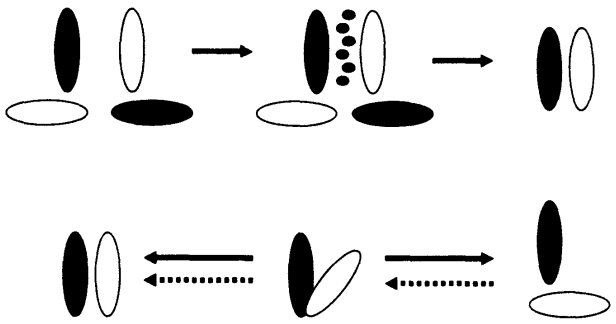

Figure 6 Schematic representation of the plausible evolution of different states of dimerization in legume lectins. Top panel: To begin with, a GSIV type of dimer could give rise to a tetramer like that of PNA wherein because of the reinforcement of side-by-side interface by expulsion of water the bottom sheets will experience steric clash. This together with intrinsically low stability of back to back dimers will allow their dissociation from PNA type assembly and facilitate the formation of Con A type of dimer. Bottom panel: Beginning with the handshake type of dimer (ECorL). The divergence in the mode of dimerization could occur either by a clockwise rotation of the sheets with respect to each other leading to back-to-back type of dimerization (e.g. GS IV, PNA) or by an anti-clockwise rotation of the same leading to canonical dimerization. Alternatively, a convergent mode of dimerization from GS IV type of dimer to "canonical" dimer involving "handshake" kind of interface can also be envisaged. However, the reverse is ruled out because of the intrinsically high stability of "canonical" dimers.

evolution. The domain-swapping hypothesis provides a fascinating mechanism of oligomer assembly and higher order protein evolution. In 3D domain swapping, one domain of a monomeric protein is replaced by the same domain from an identical protein chain (Schlunegger et al. 1997,Bennett et al. 1994a, 1994b, 1995). This results in the formation of dimers and higher order oligomers. The dimers of diptheria toxin, RNase A, IFN-gamma and IL-5 are excellent examples of domain swapped proteins (Bennet et al. 1994b). In the case of legume lectins, one can imagine a modified model, producing the observed array of oligomer associations (figure 6). For instance, to consider the peanut agglutinin structure that exhibits features common to the canonical lectins (the 1-2 interface of PNA) as well as the back-to back GS IV dimers (1-4 and 2-3 interface). It is now easy to imagine the development of the canonical structure (such as in pea and lentil lectin) from the back-to-back one (or vice versa), through a PNA kind of tetrameric intermediate. Two pairs of back-to-back dimers may first associate with one another across one of their "back" sheets. This intermediate would be PNA like tetramer. This could then be followed by a gradual disruption of the back-to-back dimer (to recall, the 1-4 or 2-3 dimers of PNA) itself while strengthening the side by side interface thus resulting, finally in the canonical dimer. The peanut lectin protein may represent a final well established, and stabilized, oligomer in its own right. Also beginning with a back-to-back associated dimer of the GS IV type, the canonical dimer could have evolved through the "handshake" kind of interface as an intermediate. Alternatively, a divergent mode of the oligomeric evolution could ensue from the lectins with "handshake" type of association in opposite directions giving rise to both the back-to-back and the canonically associated dimers respectively. Thus as "natural mutants" of quaternary association, legume lectin family can serve as a paradigm for studies addressing the effects of quaternary association on the stability, folding and function of oligomeric proteins, as well as the evolution of multimeric structures.

References

Ahmad N, Srinivas V R, Reddy G B and Surolia A 1998 Thermodynamic characterization of the conformational stability of the homodimeric protein, pea lectin; *Biochemistry* 37 16765-16772

Ashwell G and Harford J 1982 Carbohydrate-specific receptors of the liver; *Annu. Rev. Biochem.* 51 531-534

Banerjee R, Das K, Ravishankar R, Suguna K, Surolia A and Vijayan M 1996 Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex; *J. Mol. Biol.* 259 281-296

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

Bennett M J, Choe S and Eisenberg D 1994a Refined structure of dimeric diphtheria toxin at 2.0 Å resolution; *Protein Sci.* 3 1444-1463

_____, Choe S and Eisenberg D 1994b Domain swapping: entangling alliances between proteins: *Proc. Natl. Acad. Sci. U S A.* 91 3127-3131

_____, Schlunegger M P and Eisenberg D 1995 3D domain swapping: a mechanism for oligomer assembly; *Protein Sci.* 4 2455-2468

Bouckaert J, Loris R, Poortmans F and Wyns L 1995 Crystallographic structure of metal-free concanavalin A at 2.5 A resolution; *Proteins : Struct. Funct. Genet.* 23 510-524

Brill L M, Evans C J and Hirsch A M 2001 Expression of MsLEC1- and MsLEC2-antisense genes in alfalfa plant lines causes severe embryogenic, developmental and reproductive abnormalities; *Plant J.* 25 453-461

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

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

Diaz C L, Spaink H P and Kijne J W 2000 Heterologous rhizobial lipochitin oligosaccharides and chitin oligomers induce cortical cell divisions in red clover roots, transformed with the pea lectin gene; *Mol. Plant Microbe Interact.* 13 268-276

Freire E 1995 Differential scanning calorimetry; *Methods Mol. Biol.* 40 191-218

Hamelryck T W, Dao-Thi M H, Poortmans F, Chrispeels M J, Wyns L and Loris R 1996 The crystallographic structure of phytohemagglutinin-L; *J. Biol. Chem.* 271 20479-20485

Hardman K D, Wood M K, Schiffer M, Edmundson A B and Ainsworth C F 1971 Structure of concanavalin A at 4.25-angstrom resolution; *Proc Natl Acad Sci USA* 68 1393-1397

Hubbard S J, NACCESS, Computer Program, Department of Biomolecular Sciences, UMIST, Manchester, UK 1996 Version 2.1.1

Kalsi G and Etzler M E 2000 Localization of a Nod factor-binding protein in legume roots and factors influencing its distribution and expression; *Pl. Physiol.* 124 1039-1048

Loris R, De Greve H, Dao-Thi M H, Messens J, Imberty A and Wyns L 2000 *J. Mol. Biol.* 301 987-1002

Mandal D K and Brewer C F 1993 Differences in the binding affinities of dimeric concanavalin A (including acetyl and succinyl derivatives) and tetrameric concanavalin A with large oligomannose-type glycopeptides; *Biochemistry 32* 5116-5120

Manoj N, Srinivas V R, Surolia A, Vijayan M and Suguna K 2000 Carbohydrate specificity and salt-bridge mediated conformational change in acidic winged bean agglutinin; *J. Mo. Biol.* 302 1129-1137

Prabu M M, Sankaranarayanan R, Puri K D, Sharma V, Surolia A, Vijayan M and Suguna K 1998 Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 A resolution; *J. Mol. Biol.* 276 787-796

_____, Suguna K and Vijayan M 1999 Variability in quaternary association of proteins with the same tertiary fold: a case study and rationalization involving legume lectins; *Proteins: Struct. Funct. Genet.* 35 58-69

Puri K D and Surolia A 1994 Amino acid sequence of the winged bean (Psophocarpus tetragonolobus) basic lectin. Adenine binding and identification of the active-site tryptophan residue; *J. Biol. Chem.* 269 30917-30926

Reddy G B, Bharadwaj S and Surolia A 1999a Thermal stability and mode of oligomerization of the tetrameric peanut agglutinin: a differential scanning calorimetry study; *Biochemistry* 38 4464-4470

_____, Srinivas V R, Ahmad N and Surolia A 1999b Molten globule-like state of peanut lectin monomer retains its carbohydrate specificity. Implications in protein folding and legume lectin oligomerization; *J. Biol. Chem.* 274 4500-4503

Schlunegger M P, Bennett M J and Eisenberg D 1997 Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly; *Adv. Protein Chem.* 50 61-122

Schwarz F P, Puri K D, Bhat R G and Surolia A 1993 Thermodynamics of monosaccharide binding to concanavalin A, pea (Pisum sativum) lectin, and lentil (Lens culinaris) lectin; *J. Biol. Chem.* 268 7668-7677

_____, Puri K D and Surolia A 1991 Thermodynamics of the binding of galactopyranoside derivatives to the basic lectin from winged bean (Psophocarpus tetrogonolobus); *J. Biol. Chem.* 266 24344-24350

Shaanan B, Lis H and Sharon N 1991 Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose; *Science* 254 862-866

Sharma V and Surolia A 1997 Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity; *J. Mol. Biol.* 267 433-445

_____, Srinivas V R and Surolia A 1996 Cloning and sequencing of winged bean (Psophocarpus tetragonolobus) basic agglutinin (WBA I): presence of second glycosylation site and its implications in quaternary structure; *FEBS Lett.* 389 289-292

Sharon N and Lis H 1989 Lectins as cell recognition molecules; *Science* **246** 227-234

_____ and_____1998 Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition; *Chem. Rev.* **98** 637-674

Srinivasan N, Rufino S D, Pepys M B, Wood S P and Blundell T L 1996 A superfamily of proteins with the lectin fold; *Chemtracts-Biochemistry and Molecular Biology* **6** 149-164

Springer T A and Laskey L A 1991 Cell adhesion. Stickv sugars for selectins; *Nature* **349** 196-197

Srinivas V R, Singha N C, Schwarz F P and Surolia A 1998 Differential scanning calorimetric studies of the glycoprotein, winged bean acidic lectin, isolated from the seeds of Psophocarpus tetrogonolobus; *FEBS Lett.* **425** 57-60

Surolia A, Sharon N and Schwarz F P 1996 Thermodynamics of monosaccharide and disaccharide binding to Erythrina corallodendron lectin; *J. Biol. Chem.* **271** 17696-17703

Van Damme W J M, Peumanns W J, Barre A and Rouge P 1998a Plant Lectins: A Composite of Several Distinct Families of Structurally and Evolutionary Related Proteins with Diverse Biological Roles; *Crit. Rev. Plant Sci.* **17** 575-692

_____, _____, Pusztai A and Bardocz S 1998b *Handbook of Plant Lectins: Properties and Biomedical Applications* (Chichester, UK: John Wiley & Sons)

van Eijsden R, Diaz C L, de Pater B S and Kijne J W 1995 Sugar-binding activity of pea (Pisum sativum) lectin is essential for heterologous infection of transgenic white clover hairy roots by Rhizobium leguminosarum biovar viciae; *Plant Mol Biol.* **29** 431-439

VAST structure neighbours search, Entrez, NCBI, http://www.ncbi.nlm.nih.gov/Structure/vast/vastsrv.cgi?uid=3247 & chaindom=32470100

Vijayan M and Chandra N 1999 Structural basis of carbohydrate recognition by lectin II from Ulex europaeus, a protein with a promiscuous carbohydrate-binding site; *Curr. Opin. Struct. Biol.* **9** 707-714