Chemical characteristics of dimer interfaces in the legume lectin family

SHARONA ELGAVISH^{1,2} AND BOAZ SHAANAN^{1,2,3}

¹Department of Biological Chemistry, The Institute of Life Sciences, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel

²The Wolfson Centre for Applied Structural Biology, The Institute of Life Sciences, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel

³Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

(RECEIVED October 18, 2000; FINAL REVISION January 4, 2001; ACCEPTED January 11, 2001)

Abstract

The *Erythrina corallodendron* lectin (EcorL) crystallizes in monoclinic and hexagonal crystal forms. Comparison of the newly determined hexagonal form (PDB code 1fyu) with the monoclinic form shows that the dimeric structure of EcorL reflects the inherent biological structure of the protein and is not an artifact of the crystal packing. To further understand the factors determining the dimerization modes of legume lectins, EcorL, concanavalin A (ConA), and *Griffonia simplicifolia* (GS4) were taken as representatives of the three unique dimers found in the family. Six virtual homodimers were generated. The hydropathy, amino acid composition, and solvation energy were calculated for all nine homodimers. Each of the three native dimers has a distinct chemical composition. EcorL has a dominant hydrophobic component, and ConA has a strong polar component, but in GS4 the three components contribute equally to the interface. This distribution pattern at the interface is unique to the native dimers and distinct from the partition observed in the virtual dimers. Amino acid composition of other members of the family that dimerize like EcorL or ConA maintain the same pattern of amino acids distribution observed in EcorL and ConA. However, lectins that dimerize like GS4 do not show a particularly distinct distribution. In all cases, the calculated solvation energy of the native dimer was lower than that of the virtual dimers, suggesting that the observed mode of dimerization is the most stable organization for the given sequence and tertiary structure. The dimerization type cannot be predicted by sequence analysis.

Keywords: Quaternary structure; hydropathy; solvation energy; protein interface; lectins

The biological importance of lectins, proteins that bind carbohydrates with high specificity, has been increasingly recognized over the last decades. These proteins are involved in numerous cellular processes such as cell–cell and host–

pathogen interactions, targeting proteins within cells, lymphocyte homing, and tissue development (Sharon and Lis 1990; Rini 1995; Weis and Drickamer 1996).

The function of legume lectin is unclear, yet the availability of these lectins and their known three-dimensional structures have turned members of this family into good models for carbohydrate–protein recognition.

All the members of the family are found in an oligomeric state, as dimers or tetramers, which are, in turn, dimers of dimers. In vitro, legume lectins have been shown to be involved in activities that require multiple binding sites, such as stimulation of mitosis, receptor–ligand cross-linking, and cell agglutination. The distinct quaternary structures found in this family dictate well-defined spacing between the carbohydrates' binding sites, and thus, the qua-

Reprint requests to: Dr. Boaz Shaanan, Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel; e-mail bshaanan@bgumail.bgu.ac.il; fax 972-8-647-2890.

Abbreviations: ConA, concanavalin A; DB58, horse gram (*Dolichos biflorus*) stem and leaf lectin; DBL, horse gram (*Dolichos biflorus*) seed lectin; EcorL, *Erythrina corallodendron* lectin; Ecpeg, structure of EcorL in the hexagonal crystal form; GS4, fourth lectin from *Griffonia simplicifolia*; LOL I, *Lathyrus ochrus* isolectin I; NCS, noncrystallographic symmetry; PDB, RCSB Protein Data Bank; PHA-L, Leuckoagglutinating common bean agglutinin; PL, pea lectin; PNA, peanut agglutinin; r.m.s., root-mean-square; SBA, Soybean agglutinin; WBL, winged bean lectin.

Article and publication are at www.proteinscience.org/cgi/doi/10.1110/ps.44001.

ternary organization of these lectins may be of prime importance for their function in vivo.

Legume lectin dimerize in three modes involving the six-stranded β sheets of the two monomers (Prabu et al. 1999): The side-by-side mode, first observed in the structures of concanavalin A (ConA; Becker et al. 1975), and two back-to-back modes, one found in *Erythrina corallodendron* (EcorL; Elgavish and Shaanan 1998) and winged bean lectin (WBL; Prabu et al. 1998) and the second found in *Griffonia simplicifolia* (GS4; Delbaere et al. 1993) and peanut lectin (PNA; Banerjee et al. 1994). Side-by-side mode in which the six-stranded β -sheets form a continuous 12-stranded antiparallel β -sheet were also found in pea lectin (PL; Einspahr et al. 1986), favin (Reeke et al. 1986), *Lathyrus ochrus* isolectin I (LOL I; Bourne et al. 1990), Lentil lectin (Loris et al. 1994), soybean agglutinin (SBA; Dessen et al. 1995), and phytohemagglutinating-L (PHA-L; Hamelryck et al. 1996).

Recently, the structures of two legume lectins from *Dolichos biflorus*, horse gram seed lectin (DBL), and horse gram stem and leaf lectin (DB58) were solved by Hamelryck et al. (1999). Along with the side-by-side dimerization (ConA type) in DBL, a novel quaternary structure, involving a characteristic α helix sandwiched between the two monomers, appears in the dimer of DB58 and the dimer of dimers in DBL. This new type of dimerization is not dealt with in this work because, at the moment, it has no counterparts among other legume lectins of known structure.

Despite the high homology and similarity in their monomeric structure, members of the legume lectin family adopt distinct oligomerization mode. Banerjee et al. (1996) suggested that slight differences in the tertiary structure of the monomer cause distinct quaternary associations.

Comparison of the structure of EcorL in the hexagonal crystal form determined at 2.6 Å (Ecpeg; PDB entry: 1fyu) with the monoclinic form previously published (Elgavish and Shaanan 1998) confirms that the quaternary structure reflects the inherent biologically relevant architecture of the protein and is not an artifact of crystal forces (data not shown). Knowing that the mode of dimerization of EcorL is biologically significant, it is possible to go back and ask how the sequence of each member of the legume family leads to its particular quaternary structure.

It has been pointed out (Prabu et al. 1999) that the question of oligomerization in the legume lectin family relates to the more general and complicated issue of protein–protein interactions in oligomeric proteins, as studied in several works (Miller et al. 1987; Argos 1988; Janin et al. 1988; Janin and Chothia 1990; Jones and Thornton 1995, 1996, 1997; Tsai et al. 1996, 1997; Xu et al. 1997; Lo Conte et al. 1999). Jones and Thornton (1996, 1997) demonstrated that no particular fundamental parameter can distinguish the interface region from other areas on the monomer surface but,

rather, a combination of many parameters is needed for that task. Janin et al. (1988) compared the surface, subunit interface, and interior of oligomeric proteins and found that surfaces involved in subunit contacts have unique characteristics of hydropathy and amino acid composition. Prabu et al. (1999) attempted to indicate discriminative factors in the quaternary assembly of legume lectins, such as buried hydrophobic areas, shape complementarity, and calculated interaction energy.

In this work, we adopted the approach of Horton and Lewis (1992), who calculated solvation energy for several quaternary complexes and obtained values that were highly correlated with the measured dissociation energies. Along with calculating the solvation energy, we characterized the chemical composition at the interfaces of legume lectins that adopt the three distinct dimerization types. The sequence homology of amino acids at the dimer interface in each of the three subgroups (Type ConA, Type EcorL, and Type GS4) was also analyzed and compared to the overall homology. Our analysis deals only with dimers, as dimers have higher representation than tetramers among the known structures of legume lectins.

Results and Discussion

Analysis of dimer interfaces

The question of the mode of dimerization of legume lectins was approached by taking EcorL, ConA, and GS4 (PDB codes: 1ax2, Elgavish and Shaanan 1998; 5cna, Naismith et al. 1994; and 1led, Delbaere et al. 1993, respectively) as prototypes for the known quaternary structures. To determine whether there is a preference for the native dimer over other dimers, each of the native dimers was transformed into the other two types of dimers: EcorL to ConA (EcorL/ConA) and GS4 (EcorL/GS4), ConA to EcorL (ConA/EcorL) and GS4 (ConA/GS4), and GS4 to EcorL (GS4/EcorL) and ConA (GS4/ConA). A few aspects were examined in each of native and virtual dimers: The chemical characteristics of the subunit interface according to the analysis done by Janin et al. (1988), relative thermodynamic stability of the various dimers, and conservation of the interface amino acids within the legume family and the subgroups.

Amino acid composition and hydropathy

To detect potentially unique characteristics, the amino acid composition and hydropathy were compared between the three naturally occurring and six modeled homodimers representing the EcorL, ConA, and GS4 types of dimerization. The calculated composition and hydropathy represent residue and atomic properties, respectively (Table 1). Because the trend in the change between native and virtual dimers is

Table 1. *Hydropathy and amino acid composition of native and virtual dimers*

Dimer	Hydropathy (%) ^a			Amino acid composition (%) ^b			Ratio ^c
	Apolar	Polar	Charge	Apolar	Polar	Charge	
Native EcorL	63	21	16	50.3	28.9	20.8	0.5
EcorL/ConA	42	37	21	19.7	38.9	41.4	3.3
EcorL/GS4	58	24	18	36.1	33.6	30.3	1.1
Native ConA	52	40	8	26.1	54.2	19.7	1.4
ConA/EcorL	41	22	37	19.5	33.2	47.3	3.1
ConA/GS4	42	24	34	18.6	33.1	48.3	2.9
Native GS4	56	22	22	34.1	34.6	31.3	1.0
GS4/EcorL	67	20	13	39.1	36.0	24.9	0.6
GS4/ConA	39	38	23	13.5	51.2	35.3	4.4
Janin et al. (1988) ^d							
Buried interface	65 (4)	22 (7)	13 (5)	45.9	29.9	24.2	
Buried in monomers	58 (5)	39 (5)	4	49.9	28.9	21.2	
Accessible surface in oligomers	57 (3)	22 (3)	21 (5)	28.9	29.7	41.4	
Ratio generated by Lo Conte et al. (1999)							
Protein-protein complexes							
Surface							2.9
Interface							1.3
Oligomeric proteins							
Surface							3.3
Interior							0.6
Interface							0.7

^a Dividing the interface according to the percentage of surface area occupied by apolar, polar, and charged atoms. The categories are according to Janin et al. (1988) and Horton and Lewis (1992): Apolar, carbon atoms; polar, nitrogen, oxygen, and sulfur; charged, oxygen in carboxylate and nitrogen in guanidinium groups.

^b Amino acid composition of interface surface. Amino acids were divided into three groups: Apolar: Ala, Gly, Leu, Met, Phe, Pro, and Val; polar: Asn, Cys, Gln, His, Ser, Thr, Trp, and Tyr; and charged: Asp, Arg, Glu, and Lys.

^c Charged/hydrophobic ratio: Surface percentage occupied by Arg, Lys, Asp, Glu divided by that of Ile, Leu, Val, Phe, Met as defined in Lo Conte et al. (1999).

^d Hydropathy and amino acid composition found by Janin et al. (1988); Values (r.m.s.d) of amino acid composition are the result of summing over the values found for the individual type of amino acids.

the same for both properties, we will refer only to the amino acid composition in what follows. The values for all the dimers are compared to the values calculated by Janin et al. (1988), who calculated the hydropathy and amino acid composition for the buried, accessible, and interface surfaces in 23 oligomers (average values found in this analysis are quoted in Table 1).

EcorL

The hydropathy and amino acid composition at the interface of native EcorL correspond to the average values found

for subunit interfaces (Janin et al. 1988; Jones and Thornton 1995; Table 1). The interface of EcorL/ConA differs significantly from the native one. This is mostly expressed by an increase in the polarity at the interface at the expense of the apolar contribution. The contribution of apolar atoms is 0.4 times that of the native dimer, while the charged and polar contributions increase by 2 and 1.4, respectively.

The EcorL/GS4 amino acid composition at the interface is split into three nearly equal apolar, polar, and charged components. The apolar part is 0.7 times the size of that of native EcorL, and the polar and charged components are 1.2

Fig. 1. Architecture of the three prototypical dimer interfaces of ConA, EcorL, and GS4. The orientation of the six-stranded β -sheet in one monomer (black) is identical in the three interfaces, which are all viewed down a line perpendicular to this β -sheet. The figure was prepared using MOLSCRIPT (Kraulis 1991).

and 1.5 times, respectively, larger than those of native EcorL. Interestingly however, although the hydropathy profile of EcorL/GS4 seems to be similar to that of native EcorL, it is clearly closer to the hydropathy pattern found for accessible surface in oligomers than to that of buried interface, unlike the case of native EcorL (Table 1).

ConA

The hydropathy of the interface of native ConA resembles that of average areas buried in monomers (Table 1, bottom). Generally, the large number of polar atoms in the buried area of the monomers mainly comprises atoms from peptide bonds that form the hydrogen bonds in the α helices and β sheets (Janin et al. 1988). However, at the interface of native ConA, the polarity is mostly caused by the abundance of polar side chains. This interface shows an obvious deficit in charged and apolar residues.

The interfaces of both ConA/EcorL and ConA/GS4 are almost identical in hydropathy and amino acid composition. Both virtual dimers differ, however, from the native ConA, with a tendency to a decrease in the portion of apolar and polar atoms/residues (~ 0.6 of native) and a dominant charged component (2.5 the native).

GS4

The hydropathy of the interface of native GS4 dimer is within the range of average hydropathy found for the accessible surfaces in oligomers (Table 1, bottom). The amino acid composition also resembles values found for the accessible surface area (Table 1). The characteristics of the interface of GS4 also correlate with the analysis done by Jones and Thornton (1995).

GS4/EcorL and native GS4 are similar in their polar components, with $\sim 20\%$ difference in the contribution of apolar and charged residues. The charged constituent at the interface of GS4/ConA is similar to that in the interface of native GS4. Yet the apolar part is 0.4 times smaller and the polar part is 1.5 times larger than in the native dimer.

In the monomers of all the three dimerization types, the region that defines the native interface has hydropathy and amino acid composition that are distinct from the contact regions that define the virtual interfaces. The most prominent characteristic is the decrease in the relative contribution of the apolar component (with the exception of GS4/EcorL compared to native GS4), while changes in the polar and charged components are variable.

The hydropathy of the virtual interfaces is not similar to any hydropathy patterns found to date. This suggests that it is less probable to find the virtual interfaces than their native counterparts. In most cases, the apolar component decreases and the charged part increases, while the trend of the polar part is variable (Table 1). Lo Conte et al. (1999) showed that the charged/hydrophobic ratio (see Table 1 for definition) is higher for protein surface than that of the interior of the

protein (~ 3 for the first and 0.6 for the latter). These authors also showed that the charged/hydrophobic ratio of the oligomeric interfaces is close to those of the protein interior, while those of the protein-protein interfaces is between those of the protein surface and interior of proteins. This correlates with the average hydrophobicity in oligomeric interfaces found by Janin et al. (1988) and Jones and Thornton (1996, 1997). Table 1 shows the charged/hydrophobic ratio of the nine dimers. The interfaces of the three native homodimers cover all the interface ranges: from oligomeric interfaces, found in native EcorL, to the interface of native ConA that is close to that found in protein-protein complexes (Table 1), via the interface of native GS4, which is intermediate between these two interfaces. Thus, within the group of oligomeric interfaces, individuals could vary in their charged/hydrophobic ratio.

Examination of the charged/hydrophobic ratio in the virtual dimers shows that this ratio in GS4/EcorL is almost identical to that of native EcorL and that that of EcorL/GS4 is like native GS4. Hydropathy and the charge/hydrophobic residue ratios do not fully explain why EcorL does not dimerize in a GS4 type of dimerization, and vice versa. One possible explanation is that GS4/EcorL differs in its amino acid composition compared to native EcorL. The apolar portion is smaller, and the charged one is larger. In contrast, the amino acid composition of EcorL/GS4 is almost identical to that of native GS4, and thus, the energetics of the association is probably the parameter that differentiates between them. On the basis of ratio of charged to hydrophobic (Lo Conte et al. 1999; Table 1), it seems that all the other virtual interfaces are within the range of ratios found for surfaces both in oligomeric proteins and in protein-protein complexes, unlike the native interfaces that fall within the ranges of known interfaces. However, the picture is somewhat more complicated, as examination of the charged/hydrophobic ratio of the interface of LOL I and lentil lectin reveals a high ratio that is much closer to that of the protein surface. Thus, other factors, such as residue packing, entropy, and solvation energy probably have an important influence on formation of the dimer and its topology.

Comparison with other representatives

Table 2 shows the amino acid composition of other representative members of each type of dimerization. The dominant apolar component in EcorL persists in WBL. In all the members of the subgroup of ConA, the polar part is the largest, especially in LOL I, lentil, and PL (PL is 83% identity with LOL I), in which the polar part occupies almost the whole surface, leaving the apolar and charged surfaces insignificantly minor.

The prominence of the apolar and polar components in the subgroups of EcorL and ConA, respectively, stems indeed from the numerous apolar and polar amino acids (Table 2) but becomes even more pronounced when the

Table 2. Amino acid composition (percentage) in representatives structures of the Legume lectin family

Protein ^a	Apolar ^b	Polar	Charged
Type EcorL			
EcorL	50 (44) ^c	29 (30)	21 (26)
WBL	48 (48)	31 (29)	21 (24)
Type ConA			
ConA	26 (30)	54 (48)	20 (22)
LOL I	13 (25)	68 (54)	19 (21)
Lentil	10 (20)	77 (60)	13 (20)
SBA	31 (35)	54 (50)	15 (15)
Type GS4			
GS4	34 (28)	35 (48)	31 (24)
Peanut	43 (40)	42 (49)	15 (11)

^a PDB code: 5cna (ConA; Naismith et al. 1994); 1lob (LOLI; Bourne et al. 1990); 1len (lentil; Loris et al. 1994); 1sba (SBA; Dessen et al. 1995); 1ax2 (EcorL; Elgavish and Shaanan 1998); 1wbl (WBL; Prabu et al. 1998); 1led (GS4; Delbaere et al. 1993); 2pel (PNA; Banerjee et al. 1996).

^b The division of the amino acids into apolar, polar, and charged subgroups is given in footnote b in Table 1.

^c Portion of the area occupied by amino acid type. In parenthesis, percentage by count of the amino acid.

contribution to the surface composition as seen in the structure is taken into account (Table 1).

The variability within the subgroup of GS4 is greater than that in the other two groups. Unlike the nearly equal apolar, polar, and charged partitioning of the interface surface in GS4, the apolar and polar surface components are equal in peanut lectin, but the charged one is significantly reduced. The only common characteristic of GS4 and peanut lectin is the nearly equal contribution of the apolar and polar surfaces.

Solvation energy

The thermodynamic stability of the native and virtual dimers was calculated using the method described by Horton and Lewis (1992). The problem of calculating thermodynamical properties based on structure was dealt with by Janin (1995a,b, 1997). However, the method of Horton and Lewis (1992) has provided a significant correlation between the experimental free energies of association and the solva-

tion energy values calculated based on structures (Janin 1995a). Horton and Lewis (1992) emphasize that these calculations only hold when the macromolecules associate as rigid bodies. Because legume lectins are stable only as oligomers, it is probable that monomers isomerize to dimers during assembly. However, because the monomers in the family can be well superimposed, one can assume that the energy of isomerization is also similar.

Overall comparison of native dimers

The first row of Table 3 gives the solvation energies of the three native dimers and the area of the buried interface used in the calculation, including the water at these interfaces. The three dimer structures are in the same range of resolution; therefore, energy differences among them are probably not caused by resolution artifacts. As shown previously (Horton and Lewis 1992), there is no correlation between the buried area and the solvation energy. GS4 does not have the largest buried interface but seems to be the most stable dimer of the three. ConA and EcorL have very

Table 3. Association energies for dimers of EcorL, ConA, and GS4 types calculated according to Horton and Lewis (1992)

<i>Erythrina corralodendron</i> (EcorL)			<i>Concanavalin A</i> (ConA)			<i>Griffonia simplicifolia</i> (GS4)		
Dimer	Association energy ^a (kcal/mole)	Buried area ^b (Å ²)	Dimer	Association energy (kcal/mole)	Buried area (Å ²)	Dimer	Association energy (kcal/mole)	Buried area (Å ²)
EcorL (1.9 ^c)	-37.1	2306	ConA (2.0)	-36.1	3520	GS4 (1.9)	-56.7	3010
EcorL/nowat	-31.0	1633	ConA/nowat	-30.8	2282	GS4/nowat	-44.2	2031
EcorL/ConA	-14.5	1899	ConA/EcorL	-18.6	1125	GS4/EcorL	-19.6	1128
EcorL/GS4	-25.6	1909	ConA/GS4	-23.7	2216	GS4/ConA	-42.4	2126
Ecpeg (2.6)	-36.1	2188						
WBL (2.6)	-32.8	2218	LOL I (1.9)	-26.2	2908	PNA (2.3)	-40.3	3094

^a Association energies in this and other tables were calculated only for the buried area of the dimer and do not include the rotation and translation components (Horton and Lewis 1992) derived from the loss of rotation and translational freedom of one of the monomers.

^b Buried area at the interface between two monomers.

^c Resolution of the crystal structure in Ångstrom.

similar solvation energies, although the buried area of ConA is significantly higher. Energy calculations for peanut lectin, LOL I, and WBL, as additional examples for GS4, ConA, and EcorL types of dimerization, respectively, show that peanut lectin, like GS4, is the most stable among the three, while LOL I is the least stable among all structures. It seems that the members of the family that dimerize like ConA may have highly variable stability as a result of differences in sequence and tertiary structure.

Water contribution

Because the virtual dimers do not include water molecules, native dimers without water molecules were produced for comparison (EcorL/nowat, ConA/nowat, and GS4/nowat for dehydrated EcorL, ConA, and GS4, respectively). For the sake of consistency with the virtual dimers, the dehydrated native dimers were energy minimized before the comparison with the virtual dimers (see Materials and Methods; the energy minimization had negligible effect on

the calculated solvation energy of the native dehydrated dimers). The hierarchy of energetic stability of the dehydrated native dimers is consistent with that found for the native ones (Table 3). Comparison of the solvation energies of the native dimers to those without water gives a rough estimate of the water contribution to the calculated solvation energy, which is ~20% in all three dimer types.

Most of the direct hydrogen bonds that exist in the native dimers are kept in the dehydrated structures. Some of the water-mediated hydrogen bonds are reestablished as direct ones, whereas in other cases, different hydrogen bonds are formed. As a result, the energy term corresponding to oxygen atom is affected more significantly by the modeled dehydration (Table 4).

Table 3 shows the contact areas of dimers in their native and dehydrated structures. Water occupies 29%, 35%, and 33% of the interface surfaces of EcorL, ConA, and GS4, respectively. Water molecules involved in hydrogen bonds stabilize the interface. Overall, water contribution to the total energy is 16%, 15%, and 22% for EcorL, ConA, and

Table 4. Contributions according to atom type to the association energy and buried area in native and virtual dimer interfaces: A-EcorL, B-ConA, C-GS4

Structure and atom type ^a	Energy per atom type (kcal/mole)					E total ^d
	C ^b	N	O	N ⁺	O ⁻	
EcorL	-27.5^c	-1.3	-3.4	-6.1	1.2	-37.1
	1092	119	869	134	92	2306
EcorL/nowat	-26.0	0.9	0.5	-7.0	0.6	-31.0
	1032	135	212	154	100	1633
EcorL/ConA	-20.4	-0.8	1.0	-4.1	9.8	-14.5
	808	138	560	90	303	1899
EcorL/GS4	-27.7	-0.4	-0.9	2.0	1.4	-25.6
	1100	173	283	176	177	1909
Ecpeg	-25.1	-1.0	0.4	-9.4	-1.0	-36.1
	997	143	728	222	98	2188
ConA	-30.7	-2.2	-5.9	0.1	2.6	-36.1
	1218	295	1872	26	109	3520
ConA/nowat	-30.0	-2.6	-2.1	0.1	3.8	-30.8
	1190	342	571	34	145	2282
ConA/EcorL	-11.6	0.8	1.8	-9.7	0.1	-18.6
	462	66	182	213	202	1125
ConA/GS4	-23.4	0.3	0.4	-1.5	0.5	-23.6
	929	161	359	392	375	2216
GS4	-29.3	-0.2	-9.0	-8.3	-9.9	-56.7
	1162	110	1250	190	298	3010
GS4/nowat	-28.7	0.6	0.4	-8.2	-8.3	-44.2
	1138	99	355	180	259	2031
GS4/EcorL	-19.2	0.1	1.6	0.6	-2.7	-19.6
	761	58	164	12	133	1128
GS4/ConA	-21.4	-1.6	-0.8	-7.2	-11.4	-42.4
	848	270	533	162	323	2136

^a The structure for which the energy is calculated.

^b Atom type according to Horton and Lewis (1992): Apolar, carbon (C); polar, nitrogen, oxygen (N, O); charged, oxygen in carboxylate and nitrogen in guanidinium groups (N⁺, O⁻).

^c Parameters for each atom type: energy (bold, kcal/mole); number of atoms (roman)

^d Total energy (bold, Kcal/mole); interface area (italics, Å²).

GS4, respectively, and is slightly smaller than that estimated by Covell and Wallqvist (1997).

Surface water molecules exchange with bulk water much faster than water in narrow crevices and even more so than water molecules in interior cavities (Levitt and Park 1993). It should be borne in mind that only the less mobile water molecules are observed in the electron density map (Levitt and Park 1993). Interface areas are more exposed to solvent than buried interior areas (Janin et al. 1988), and consequently, it is most probable that they are rapidly exchanging water molecules, which are therefore only partly visible in the electron density map. Thus, although the importance of water contribution was emphasized (Larsen et al. 1998; Lo Conte et al. 1999), any conclusions concerning their contribution to the total energy should be treated cautiously. We estimated the contribution of water molecules to the overall free energy, but the lack of confidence in their location makes the comparison between the dehydrated dimers more reliable.

Comparing the stability of the native dehydrated dimers and virtual dimers

Each of the dehydrated native dimers is more stable than the other two virtual ones to which it is transformed, for example, EcorL/nowat is more stable than EcorL/ConA and EcorL/GS4, (Table 3). This is in agreement with the results of Prabu et al. (1999) that have used another algorithm for energy calculation.

This reduction in stability of the virtual dimers is mainly caused by two factors. The first is a decrease in buried area contributed by carbons, leading to decrease in their contribution to the apolar energy term. The second factor is the increased number of charged amino acids not involved in hydrogen bonds. These amino acids contribute positive energy (unfavorable) to the apolar energy term. Specifically, in all the dimers, increasing N or O buried areas either contributes, although insignificantly, to the total energy or does not lead to any change. Buried charged atoms in interfaces that are involved in hydrogen bonding strongly decrease the energy, but nonhydrogen bonded atoms cause an increase of the energy. Overall, if the array of hydrogen bonds is not large enough, extensive polar and charged surfaces at interfaces can, unlike protein surfaces, reduce the interface stability.

Structure–sequence relationship

A natural question that arises is whether the dimerization type can be predicted by inspection of the sequence. Because the sequence provides information on the amino acid composition, irrespective of location in the structure, we first compared the percentage obtained by counting each type of amino acid at the interface to the amino acid percentage corresponding to the surfaces occupied by each

amino acid type at the interface (as presented in Table 1). We found that in the majority of the nine dimers, there is no resemblance between the partition achieved by counting amino acids and that obtained by calculating the surface partition. Counting amino acids usually can not distinguish between the native interface and its corresponding virtual ones: Native EcorL has a different partition from ConA/EcorL and GS4/EcorL, but the partition in Native ConA is similar to that of GS4/ConA, and Native GS4 has an almost identical partition to ConA/GS4. Furthermore, comparing the composition of the three interface regions on each monomer (native and two virtual regions) for each of the three proteins shows no unique pattern in the native one.

Subsequently, amino acids participating in each of the interfaces were taken from the sequence alignment rather than from the structure, and their composition was calculated. Here again, no distinction between the partition in the native interface and the virtual ones can be noticed.

Extensive homology analysis was done to find out whether there is prominent homology among the members of each type of dimerization, either in the sequence of the whole protein or only in the interface residue. No clear correlation was found, and it seems that members of the legume family do not necessarily require extreme homology to adopt the same quaternary structure. This finding supports the uniqueness of the distinct structural characteristics.

It can be concluded that the clear distinction between the native interfaces and the virtual ones is structurally unique and cannot be detected at the sequence level. The contribution of each residue participating in the interface depends on the location of the strand or the loop it belongs to, the length and direction of its side chain, and its spatial relationship to residues in the other monomer. The orientation of the side chain is affected by the immediate environment, including residues that do not belong to the interface but that dictate the preferred energy state. This effect can either be regarded as local, mostly because of the interface residues, or global when interactions with the whole protein are involved. These subtleties can not be predicted simply by counting interface residues, obtained either from the structure or from the alignment, and assigning equal weights to each residue regardless of its location and direction. It also should be kept in mind that not all the residues that precipitate in one interface necessarily coalign with interface residues of another aligned protein.

Conclusions

In this article, the chemical composition of the interfaces of three representative types of dimer found in the legume lectin family was characterized. Each type displays a unique profile of hydropathy and amino acid composition. The interface of ConA is rich in polar atoms and residues but has the smallest relative area of apolar and charged surfaces

among the three representative interfaces. The hydropathy of the EcorL and GS4 interfaces is similar to the average hydropathy found for subunit interfaces and accessible surfaces in oligomers, respectively. This finding demonstrates that although general characteristics were indicated for oligomeric interfaces, individual interfaces within this group can present distinct profiles.

Although each dimerization type has its unique chemical characteristics, there is no parameter that sharply distinguishes the native interfaces from the virtual ones. This is in agreement with the conclusions of Jones and Thornton (1996, 1997) and the results of Prabu et al. (1999). However, the energy of native dimers is lower than the other two virtual ones. This implies that the energy results from combination of parameters that makes the apolar, polar, and charged interactions in the native interface more energetically favorable than in the other two virtual interfaces. The native dimer is the more suitable quaternary organization for the sequence and tertiary structure of each type of monomer and for carrying out the biological activity of those lectins.

Finally, despite the relatively high sequence homology between legume lectins, it is currently impossible to accurately predict their quaternary structure on the basis of sequence homology with the prototypical dimers.

Materials and methods

Hydropathy profiles were calculated by dividing the interface according to the surface area occupied by apolar, polar, and charged atoms. The breakdown to categories follows Janin et al. (1988) and Horton and Lewis (1992): Apolar, carbon atoms; polar, nitrogen, oxygen, and sulfur; charged, oxygen in carboxylate and nitrogen in guanidinium groups.

Amino acid composition was calculated by dividing the interface into apolar, polar, and charged parts, according to the nature of the amino acid at the surface: Apolar: Ala, Gly, Leu, Met, Phe, Pro, and Val; polar: Asn, Cys, Gln, His, Ser, Thr, Trp, and Tyr; and charged: Asp, Arg, Glu, and Lys.

Calculations of solvation energy

Virtual dimers were generated by superimposing on C α and transformation using Align97 (Cohen 1997). A minimization of 500 cycles, using X-PLOR (Brünger 1992), was carried out to avoid steric clashes. The same procedure was applied to get the dehydrated native dimers. To keep the energy-minimized structures as close as possible to the native ones, NCS restraints were applied during minimization when necessary. Solvation energies were calculated according to Equation 7 of Horton and Lewis (1992), using the atomic solvation parameters of Eisenberg and McLachlan (1986). In Equation 7 of Horton and Lewis (1992), the apolar energy term refers to atoms that are not involved in hydrogen bonds or salt bridges. The polar energy term includes atoms that are involved in hydrogen bonds and salt bridges.

The buried interface areas were calculated with the CCP4 (SERC Daresbury Laboratory 1994) programs AREAMOL and DIFFAREA, using probe radius of 1.4 Å.

Multiple sequence alignment for counting apolar, polar, and charged interface amino acids were produced by Macaw (Schuler et al. 1991) and MacVector 6.1 (Oxford Molecular).

Accession number

The final coordinates have been deposited in the RCSB Protein Data Bank with the code 1fyu.

Acknowledgments

We thank D.E. Shalev for her wonderful, generous, and careful help in preparing the manuscript.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Argos, P. 1988. An investigation of protein subunit and domain interfaces. *Protein Eng.* **2**: 101–113.
- Banerjee, R., 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.* **91**: 227–231.
- 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.
- Becker, J.W., Reeke Jr., G.N., Wang, J.L., Cunningham, B.A., and Edelman, G.M. 1975. The covalent and three-dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides. *J. Biol. Chem.* **250**: 1513–1524.
- Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rougé, P., and Fontecilla-Camp, J.C. 1990. X-ray crystal structure determination and refinement at 1.9 Å; resolution of isolectin I from seeds of *Lathyrus ochrus*. *J. Mol. Biol.* **214**: 571–584.
- Brünger, A.T. 1992. *X-PLOR manual version 3.1*. Yale University, New Haven, CT.
- Cohen, G.H. 1997. ALIGN: A program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Cryst.* **30**: 1160–1161.
- Covell, D.G. and Wallqvist, A. 1997. Analysis of protein–protein interactions and the effect of amino acid mutations on their energetic. The importance of water molecules in the binding epitope. *J. Mol. Biol.* **269**: 281–297.
- 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 Å resolution. *J. Mol. Biol.* **230**: 950–965.
- Dessen, A., Gupta, D., Sabesan, S., Brewer, C.F., and Sacchettini, J.C. 1995. X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. *Biochemistry* **34**: 4933–4942.
- Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E., and Suddath, F.L. 1986. The crystal structure of pea lectin at 3.0 Å resolution. *J. Biol. Chem.* **261**: 16518–16527.
- Eisenberg, D. and McLachlan, A.D. 1986. Solvation energy in protein folding and binding. *Nature* **319**: 199–203.
- Elgavish, S. and Shaanan, B. 1998. Structure of the *Erythrina corallodendron* lectin and its complexes with mono- and disaccharides. *J. Mol. Biol.* **277**: 917–932.
- 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.
- Hamelryck, T.W., Loris, R., Bouckaert, J., Dao-Thi, M.H., Strecker, G., Imbert, A., Fernandez, E., Wyns, L., and Etzler, M.E. 1999. Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in the two legume lectin oligomers from *Dolichos biflorus*. *J. Mol. Biol.* **286**: 1161–1177.

- Horton, N. and Lewis, M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci* **1**: 169–181.
- Janin, J. 1995a. Principles of protein–protein recognition from structure to thermodynamics. *Biochimie* **77**: 497–505.
- . 1995b. Elusive affinities. *Proteins* **21**: 30–39.
- . 1997. Angström and calories. *Structure* **5**: 473–479.
- Janin, J. and Chothia, C. 1990. The structure of protein–protein recognition sites. *J. Biol. Chem.* **265**: 16027–16030.
- Janin, J., Miller, S., and Chothia, C. 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**: 155–164.
- Jones, S. and Thornton, J.M. 1995. Protein–protein interactions: A review of protein dimer structures. *Prog. Biophys. Mol. Biol.* **63**: 31–65.
- . 1996. Principles of protein–protein interactions. *Proc. Natl. Acad. Sci.* **93**: 13–20.
- Jones, S. and Thornton, J.M. 1997. Analysis of protein–protein interaction sites using surface patches. *J. Mol. Biol.* **272**: 121–132.
- Kraulis, P.J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**: 946–950.
- Larsen, T.A., Olson, A.J., and Goodsell, D.S. 1998. Morphology of protein–protein interfaces. *Structure* **6**: 421–427.
- Levitt, M. and Park, B.H. 1993. Water: Now you see it, now you don't. *Structure* **1**: 223–226.
- Lo Conte, L., Chothia, C., and Janin, J. 1999. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**: 2177–2198.
- Loris, R., Van Overberge, D., Dao-Thi, M.H., Poortmans, F., Maene, N., and Wyns, L. 1994. Structural analysis of two crystal forms of lentil lectin at 1.8 Å resolution. *Proteins* **20**: 330–346.
- Miller, S., Lesk, A.M., Janin, J., and Chothia, C. 1987. The accessible surface area and stability of oligomeric proteins. *Nature* **328**: 834–836.
- Naismith, J.H., Emmerich, C., Habash, J., Harrop, S.J., Helliwell, J.R., Hunter, W.N., Raftery, J., Kalb (Gilboa), A.J., and Yariv, J. 1994. Refined structure of concanavalin A complexed with α -methyl-D-mannopyranoside at 2.0 angstroms resolution and comparison with the saccharide-free structure. *Acta Crystallog.* **D50**: 847–858.
- 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 Å. *J. Mol. Biol.* **276**: 787–796.
- Prabu, M.M., 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* **35**: 58–69.
- Reeke Jr., G.N. and Becker, J.W. 1986. Three-dimensional structure of favin: Saccharide binding-cyclic permutation in leguminous lectins. *Science* **234**: 1108–1111.
- Rini, J.M. 1995. Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24**: 551–577.
- Schuler, G.D., Altschul, S.F., and Lipman, D.J. 1991. A workbench for multiple alignment construction and analysis. *Proteins* **9**: 180–190.
- SERC Daresbury Laboratory. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallog.* **D50**: 760–763.
- Sharon, N. and Lis, H. 1990. Lectins as cell recognition molecules. *Science* **246**: 227–234.
- Tsai, C.J., Lin, S.L., Wolfson, H.J., and Nussinov, R. 1996. A dataset of protein–protein interfaces generated with a sequence-order-independent comparison technique. *J. Mol. Biol.* **260**: 604–620.
- Tsai, C.J., Lin, S.L., Wolfson, H.J., and Nussinov, R. 1997. Studies of protein–protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Sci.* **6**: 53–64.
- Weis, W.I. and Drickamer, K. 1996. Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**: 441–473.
- Xu, D., Tsai, C.J., and Nussinov, R. 1997. Hydrogen bonds and salt bridges across protein–protein interfaces. *Protein Eng.* **10**: 999–1012.