Protein Science (1999), *8*:1075–1086. Cambridge University Press. Printed in the USA.

Structural bases of lectin-carbohydrate affinities: Comparison with protein-folding energetics

ENRIQUE GARCÍA-HERNÁNDEZ[1] AND ANDRÉS HERNÁNDEZ-ARANA[2]

[1]Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior, Ciudad Universitaria, México D.F., México 04510
[2]Departamento de Química, Universidad Autónoma Metropolitana Iztapalapa, A.P. 55-534, México D.F., México 09340

(RECEIVED October 7, 1998; ACCEPTED January 19, 1999)

Abstract

We have made a comparative structure based analysis of the thermodynamics of lectin-carbohydrate (L-C) binding and protein folding. Examination of the total change in accessible surface area in those processes revealed a much larger decrease in free energy per unit of area buried in the case of L-C associations. According to our analysis, this larger stabilization of L-C interactions arises from a more favorable enthalpy of burying a unit of polar surface area, and from higher proportions of polar areas. Hydrogen bonds present at 14 L-C interfaces were identified, and their overall characteristics were compared to those reported before for hydrogen bonds in protein structures. Three major factors might explain why polar–polar interactions are stronger in L-C binding than in protein folding: (1) higher surface density of hydrogen bonds; (2) better hydrogen-bonding geometry; (3) larger proportion of hydrogen bonds involving charged groups. Theoretically, the binding entropy can be partitioned into three main contributions: entropy changes due to surface desolvation, entropy losses arising from freezing rotatable bonds, and entropic effects that result from restricting translation and overall rotation motions. These contributions were estimated from structural information and added up to give calculated binding entropies. Good correlation between experimental and calculated values was observed when solvation effects were treated according to a parametrization developed by other authors from protein folding studies. Finally, our structural parametrization gave calculated free energies that deviate from experimental values by 1.1 kcal/mol on the average; this amounts to an uncertainty of one order of magnitude in the binding constant.

Keywords: enthalpy; entropy; hydrogen bonding; lectin-carbohydrate association; protein folding

Because of their outstanding ability to encode information stereochemically, carbohydrates have evolved to serve as labels of cellular identity. Lectins are proteins specialized in deciphering the information carried in carbohydrate molecules by offering complementary surfaces where ligands can fit specifically and tightly. A wide variety of processes related to cell recognition and cell communication take advantage of this recognition system. Fertilization, connective tissue regulation, neuronal development, immune response, and pathogen infection are just some processes that have a critical dependence on lectin-carbohydrate (L-C) associations (Sharon & Lis, 1989).

To date, the three-dimensional (3D) structures of several tens of L-C complexes have been solved. This robust body of structural information has led to an improved understanding of the basis of specificity between lectins and carbohydrates (Rini, 1995; Weis & Drickamer, 1996). As expected, hydrogen bonding has been found to be a major determinant of discrimination. Structural comparisons of homologous series of L-C complexes suggest that stereoselectivity is provided by hydrogen bonding to distinctive polar groups of the carbohydrate. Experimentally, it has been shown that the suppression or derivatization of any of these key polar groups leads to a complete inhibition of binding (Leffler & Barondes, 1986; Schwarz et al., 1996). Inhibition is also obtained by mutating residues of the combining site that form hydrogen bonds with the ligand (Hirabayashi & Kasai, 1991; Iobst et al., 1994). In contrast, the specificity of a lectin can be changed to a different carbohydrate configuration by a rational rearrangement of the hydrogen bonding pattern, as shown recently by Iobst and Drickamer (1994) in a site-directed mutagenic study. That study also showed, however, that steric hindrance and stacking of aromatic side chains against sugar rings are factors not to be neglected to obtain biological levels of discrimination and affinity.

Stereochemical complementary is an essential, but not a sufficient, factor to explain why two molecules bind tightly to each other. In the case of L-C complexes, even though the molecular

Reprint requests to: Enrique García-Hernández, Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior, Ciudad Universitaria, México D.F., México 04510; e-mail: egarciah@servidor.unam.mx. Andrés Hernández-Arana, Departamento de Química, Universidad Autónoma Metropolitana Iztapalapa, A.P. 55-534, México D.F., México 09340; e-mail: aha@xanum.uam.mx.

determinants of specificity have been unveiled to a great extent, the energetic basis of association remains to be understood. As a general rule, it has been observed that the binding between a lectin and a carbohydrate occurs enthalpically driven. Given the conspicuous presence of interfacial hydrogen bonds, it is easy to suggest that this kind of interaction is the main source of stability. However, the effective contribution of a hydrogen bond to the formation of an L-C complex depends not only on its intrinsic energy, but also on the energy associated with dehydrating and transferring the hydrophilic pair to an environment of lower polarity (Ben-Naim, 1991). It is the net balance of these contributions that determines whether a hydrogen bond is stabilizing or not, regardless of the relevance of the adduct in the specificity of the reaction. In this respect, Lemieux and colleges have stated that the stability of L-C complexes does not arise from hydrogen bonding, but from dehydration of contact surfaces (Lemieux, 1996). Based on the analysis of enthalpy-entropy compensations, and on Monte Carlo simulations of hydration, the authors have proposed that water molecules that solvate the combining site and the carbohydrate have a higher energy (enthalpy) than free molecules in bulk water. Therefore, they postulate, it is the liberation and return of perturbed waters to the solvent that provides a favorable enthalpy for binding. Recent estimations of dehydration enthalpies have made this picture, however, no longer sustainable. Calculations based either on parameters derived from model compounds or on potential energy functions argue that the dehydration of contact zones in lectins and carbohydrates is an endothermic process, with an absolute magnitude several times higher than the overall binding enthalpy (García-Hernández et al., 1997). In consequence, the stability of a complex must be originated from highly favorable interactions between the lectin and the carbohydrate.

An analysis of the relation between binding enthalpies and structural parameters for three L-C complexes has been done previously (García-Hernández et al., 1997). Results from that study suggest that the parametrization required to account for binding enthalpies differs from those derived from protein folding data. To obtain further insights into the magnitude of the driving forces of L-C associations, we present here an analysis based on the statistical solution of equations relating energetic changes and structural features for a longer database of L-C complexes. Binding entropies and enthalpies are analyzed separately. The analysis shows that L-C complexes form a defined thermodynamic-structural class. More favorable interactions than those involved in other noncovalent reactions such as protein folding or protein–protein binding were found to be a distinguishing feature of the L-C group. Our results help to rationalize in molecular terms the origin of the high energetic efficiency in L-C associations.

Results and discussion

Thermodynamic characteristics of lectin-carbohydrate associations

Table 1 presents changes in free energy (ΔG_b), enthalpy (ΔH_b), and entropy (ΔS_b) for the formation of 10 L-C complexes of known 3D structure. This table also includes changes in polar (ΔASA_{pol}) and apolar (ΔASA_{ap}) accessible surface area for the corresponding association reactions, which were calculated as outlined in Materials and methods. The last columns give values of the thermodynamic functions normalized by the total accessible surface area (ΔASA_t). To compare the energetics of L-C binding with that for

Table 1. *Energetic (25 °C) and surface accessibility changes in the formation of lectin-carbohydrate complexes*

Lectin	Carbohydrate	PDB code	$-\Delta G_b$ (kcal/mol)	$-\Delta H_b$ (kcal/mol)	$-T\Delta S_b$ (kcal/mol)	$-\Delta ASA_{ap}$ (Å^2)	$-\Delta ASA_{pol}$ (Å^2)	$\frac{\Delta G_b}{\Delta ASA_t}$ [a]	$\frac{\Delta H_b}{\Delta ASA_t}$ [a]	$\frac{T\Delta S_b}{\Delta ASA_t}$ [a]	Ref.[b]
Hevein	GlcNAc(β1→4)GlcNAc		3.8	5.3	1.5	308.7	158.5	8.1	11.3	3.2	a,b
Hevein	GlcNAc(β1→4)GlcNAc(β1→4)GlcNAc		5.6	7.0	1.4	342.9	221.7	10.0	12.4	2.4	b
Wheat germ agglutinin	GlcNAc(β1→4)GlcNAc		4.9	13.5	8.6	237.0[c]	277.0[c]	9.5	26.3	16.7	c
Concanavalin A	Manα1Me	5cna	5.3	6.8	1.5	172.6	167.9	14.7	20.9	6.2	d,c
Concanavalin A	Glcα1Me	1gic	4.6	5.3	0.7	200.3	141.7	13.5	15.5	2.0	e,f
Concanavalin A	Man(α1→6)[Man(α1→3)]Man	1cvn	7.8	14.4	6.6	251.3	354.0	12.9	23.8	10.9	g,h
Griffonia simplicifolia lectin	Fuc(α1→2)Gal(β1→3)[Fuc(α1→4)]GlcNAc	1led	6.5	11.9	5.4	312.8	344.9	9.9	18.1	8.2	i,j
Pea lectin	Manα1Me	1rin	4.6	6.5	1.9	144.5	167.0	14.7	20.9	6.2	k,l
Erythrina corallodendron lectin	Gal(β1→4)Glc	1lte	4.5	9.8	5.3	234.0	199.5	10.6	23.2	12.7	m,n
S-lectin	Gal(β1→4)GlcNAc	1slt	6.2	10.0	3.8	223.8	257.8	12.9	20.8	7.9	o,p
Average								11.8	19.2	7.5	
Standard deviation								2.5	4.9	4.8	

[a] cal(mol·Å^2)$^{-1}$.

[b] a, Asensio et al. (1995); b, García-Hernández et al. (1997); c, Bains et al. (1992); d, Naismith et al. (1994); e, Weatherman et al. (1996); f, Harrop et al. (1995); g, Naismith & Field (1996); h, Mandal et al. (1994); i, Delbaere et al. (1993); j, Lemieux (1996); k, Rini et al. (1993); l, Schwarz et al. (1993); m, Shaanan et al. (1991); n, Surolia et al. (1996); o, Liao et al. (1994); p, Ramkumar et al. (1995).

[c] Values were taken from Bains et al. (1992).

Table 2. *Energetic (25 °C) and surface accessibility changes in protein folding and binding*[a]

Protein	PDB code	$-\Delta G$ (kcal/mol)	$-\Delta H$ (kcal/mol)	$-T\Delta S$ (kcal/mol)	$-\Delta ASA_{ap}$ (Å^2)	$-\Delta ASA_{pol}$ (Å^2)	$\frac{\Delta G}{\Delta ASA_t}$ [b]	$\frac{\Delta H}{\Delta ASA_t}$ [b]	$\frac{T\Delta S}{\Delta ASA_t}$ [b]
Cytochrome *c*	5cyt	12.9	22.0	9.1	6,540	3,960	1.2	2.1	0.9
Carbonic anhidrase B	2cab	16.7	49.0	32.3	18,500	9,970	0.6	1.8	1.2
Chymotrypsin	4cha	15.9	100.0	84.1	16,450	8,150	0.6	4.1	3.5
α-Lactalbumin	1alc	20.7	33.0	12.3	8,260	4,260	1.7	2.6	0.9
Lysozyme	1lym	16.5	56.0	39.5	8,560	4,910	1.2	4.2	3.0
Myoglobin	4mbn	12.0	10.0	−2.0	10,700	5,790	0.7	0.6	−0.1
Staphylococcus nuclease	1snc	4.0	8.0	4.0	9,070	5,210	0.3	0.6	0.3
Papain	9pap	27.9	36.5	8.6	15,160	8,390	1.2	1.5	0.3
Parvalbumin	5cpv	18.7	47.5	28.8	6,680	4,030	1.7	4.4	2.7
Pepsinogen	1psg	26.4	19.5	−6.9	25,120	12,880	0.7	0.5	−0.2
Ribonuclease A	7rsa	11.4	62.0	50.6	7,540	4,710	0.9	5.1	4.2
Trypsin	1tld	18.3	84.5	66.2	15,030	7,790	0.8	3.7	2.9
$(\text{Interleukin-8})_2$	1il8	6.7	6.2	−0.5	910	640	4.3[c]	4.0	−0.3[c]
Antibody-angiotensin II		10.9	7.7	−3.2	993	745	6.3[c]	4.4	−1.8[c]
S/lastase-inhibitor		14.5	0.6	−13.9	1,130	660	8.1[c]	0.3	−7.8[c]
Average							1.0	2.7	1.6
Standard deviation							0.4	1.7	1.5

[a] Energetic data were taken from Murphy and Freire (1992), except for the interleukin-8 dimer, and for the antibody-angiotensin II and elastase-inhibitor complexes, which were taken from Burrows et al. (1994), Murphy et al. (1993), and Baker & Murphy (1997), respectively.
[b] $\text{cal}(\text{mol}\cdot\text{Å}^2)^{-1}$.
[c] Value not included in the average of the column.

protein folding, Table 2 shows thermodynamic data that characterize the structural stability of 12 typical, globular proteins. Data in Table 2 pertain to the folding reaction at 25 °C; they were obtained by extrapolation of experimental data at 60 °C (as summarized by Murphy & Freire, 1992) by means of the change in heat capacity for the corresponding folding transitions. Also listed in Table 2 are three well-documented cases of protein–protein associations. A comparison of bottom rows of Tables 1 and 2 reveals notable differences in the thermodynamic behavior of the phenomena being analyzed: a unit of buried area leads, on the average, to a much larger stabilization in the case of L-C adducts than in either protein folding or protein–protein association. This large decrease in free energy originates from a very favorable binding enthalpy for protein–carbohydrate complexes, which overbalances the adverse entropy diminution during the process. It must be recognized, however, that currently available thermodynamic data for L-C interactions are still scarce and biased toward plant (specifically legume) lectins. Thus, a definitive generalization of the results presented here requires a larger body of experimental binding data representing other lectin families. Similarly, the number of protein intersubunit associations that have been characterized calorimetrically is so small, that the inclusion of this type of process in the comparative analysis performed here must be viewed very cautiously. In the following sections we analyze separately enthalpic and entropic contributions to L-C complex formation. The analysis was not extended to consider heat capacity changes due to the reduced number of experimental data available at present.

Enthalpies of lectin-carbohydrate association and protein folding

It is generally accepted that the major contribution to the enthalpy of protein folding comes from two opposite factors (Lazaridis et al., 1995; Makhatadze & Privalov, 1995). On one hand, there is the formation of noncovalent bonds (hydrogen bonds and van der Waals interactions) where regions of the polypeptide come in close proximity, and, on the other, the energetic cost of desolvating these contact surfaces. Furthermore, some authors (Freire, 1995; Bardi et al., 1997) consider that on a first approximation the enthalpy change for the folding of proteins (or, by extension, for protein–protein associations) can be parametrized in terms of ΔASA_{pol} and ΔASA_{ap}:

$$\Delta H = \Delta h_{pol}\,\Delta ASA_{pol} + \Delta h_{ap}\,\Delta ASA_{ap} \qquad (1)$$

where the temperature-dependent parameters Δh_{pol} and Δh_{ap} represent the enthalpy changes (normalized per unit area) for the transfer of polar and apolar surfaces from the aqueous solvent to the protein interior. Both sides of Equation 1 can be divided by any of the two ΔASA variables (say, ΔASA_{ap}) to obtain a graphical representation in two dimensions. Such a plot for the folding data at 25 °C in Table 2 is shown in Figure 1 (open circles). The same type of plot for L-C associations (Table 1) is shown in Figure 1 (solid circles). It is evident that the two data sets notably segregate from each other due to a larger slope and higher $\Delta ASA_{pol}/\Delta ASA_{ap}$ ratios in the case of lectin–sugar complexes. Another marked difference between the two systems is the variation range of $\Delta ASA_{pol}/\Delta ASA_{ap}$. Due to the extremely narrow span in the case of protein folding, ΔH can be indeed parametrized using ΔASA_t instead of ΔASA_{pol} and ΔASA_{ap}; in contrast, the wide span covered by L-C complexes precludes such a reduction to a single independent variable.

Least-squares linear regressions based on Equation 1 gave values of the parameters Δh_{pol} and Δh_{ap} for each of the two data sets being discussed. The solved form of Equation 1 for L-C binding is

$$\Delta H_{lc} = 46.1\ (\pm 6.3)\ \Delta ASA_{pol} - 5.8\ (\pm 6.3)\ \Delta ASA_{ap}. \qquad (2)$$

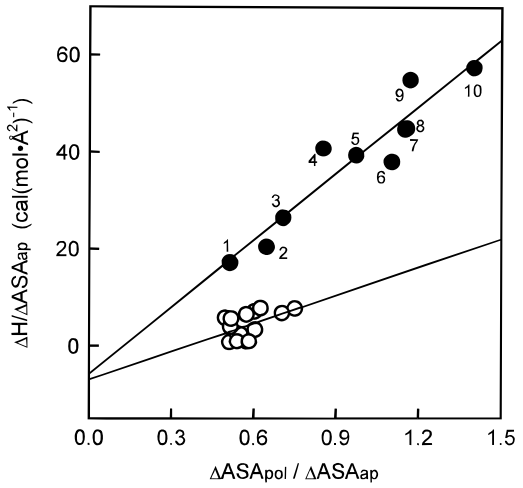

Fig. 1. Enthalpy changes for lectin-carbohydrate binding (solid circles) and protein folding (open circles) as functions of changes in polar (ΔASA_{pol}) and apolar (ΔASA_{ap}) accessible surface areas. The ratio $\Delta ASA_{pol}/\Delta ASA_{ap}$ is used as independent variable to obtain a two-dimensional representation (cf. Equation 1). Data for lectin-carbohydrate binding from Table 1. Complexes are referred to by numerals as follows: 1, hevein-chitobiose; 2, hevein-chitotriose; 3, 1gic; 4, 1lte; 5, 5cna; 6, 1led; 7, 1slt; 8, 1rin; 9, wheat germ agglutinin-chitobiose; 10, 1cvn. Data for protein folding from Table 2. Included in this latter case are the three protein complexes listed in Table 2.

As Figure 2 shows, the parametrization developed in Equation 2 adequately reproduces the experimental binding enthalpies. For this system, the magnitude of Δh_{pol} results more than seven times larger than Δh_{ap}. An analysis of significance of the two independent variables shows a substantially higher score of the *t* statistic for ΔASA_{pol}; furthermore, more than 80% of the variation in enthalpy values can be explained from ΔASA_{pol} variation. These results reveal the important contribution of buried polar areas to the enthalpy of binding. In contrast, less marked differences between coefficients of ΔASA_{pol} and ΔASA_{ap} are observed in the case of protein–protein interactions:

$$\Delta H_{prot} = 19.4\ (\pm 9.4)\ \Delta ASA_{pol} - 7.0\ (\pm 5.5)\ \Delta ASA_{ap}. \qquad (3)$$

A point that deserves special comment is the large magnitude of Δh_{pol} for L-C binding (46.1 cal$(\text{mol}\cdot\text{Å}^2)^{-1}$), which differs with a high degree of significance from the corresponding parameter that characterizes intra- and intermolecular protein contacts (19.4 cal$(\text{mol}\cdot\text{Å}^2)^{-1}$). From the concepts mentioned at the beginning of this section, two main factors (or a combination of them) might be invoked to explain this observation: the average enthalpic cost of desolvating carbohydrate molecules and polar surfaces at the binding site of lectins is smaller than for the bulk of polar groups in proteins; or polar groups at L-C interfaces interact more strongly than they do when a polypeptide chain is folded. Regarding the solvation properties of carbohydrates, results from Monte Carlo simulations suggest that the polyamphiphilic nature of these compounds might lead to special hydration properties of their molecular surfaces (Lemieux, 1996). However, no experimental information is available to support this proposal. Conversely, it has been shown that the magnitudes of heat capacity changes observed in the formation of a reduced number of L-C complexes can be satisfactorily estimated from ASA-dependent parameters developed for protein folding reactions (García-Hernández et al., 1997). Moreover, in the analysis of binding entropies (see below), we found that the approach proposed by D'Aquino et al. (1996) to account for the contribution of solvation to the entropy of protein folding works reasonably well for L-C associations also. This might indicate that, as a first approximation, solvation effects are unique, depending only on the type of surface buried, and irrespective of whether the process is binding or folding. Therefore, the next section of this work focuses on the study of hydrogen bonds, because this type of noncovalent interactions is, by far, the most important between polar groups in both protein folding and protein–carbohydrate association.

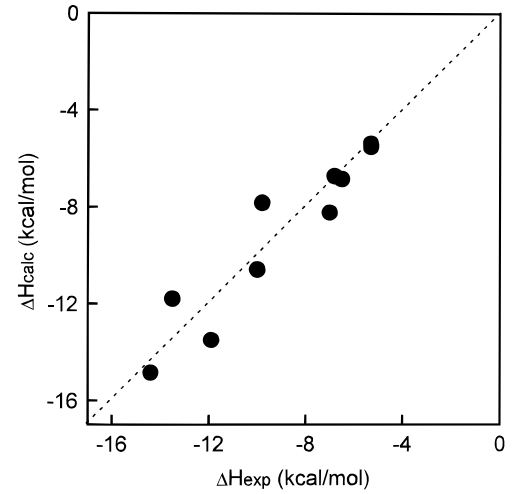

Fig. 2. Calculated binding-enthalpies (ΔH_{calc}) for lectin-carbohydrate complexes plotted vs. the corresponding experimental values (ΔH_{exp}). Calculated enthalpies were obtained from Equation 2.

Hydrogen bonding in lectin-carbohydrate complexes

An analysis of hydrogen bonds at L-C complexes was performed with the aim of comparing their overall geometric characteristics with those observed by Stickle et al. (1992) in the study of the cores of 42 globular proteins. Only structures with a resolution of 2.3 Å or better were considered to keep our study comparable to that of Stickle et al. (1992). Five L-C complexes of the data set used in the thermodynamic analysis satisfying such a stringent requirement were analyzed by means of the program HBPLUS (McDonald & Thornton, 1994) to identify hydrogen bonds (see Materials and methods). To gain significance in the statistical comparison, we included in the analysis other nine L-C complexes whose structures have been solved with high resolution. Consistently, the geometric properties obtained from the extended data set of 14 complexes closely resemble those for the five complexes above mentioned. By using the same geometrical criteria employed by Stickle et al. (1992), we identified a total of 110 hydrogen bonds in the 14 protein–sugar interfaces. Considering the ΔASA_{pol} estimated for the set of L-C interfaces, the average density of hydrogen bonds is 3.45 ($\pm$0.52)/100 Å^2 of polar area buried. In proteins, this density amounts to 3.00 ($\pm$0.57)/100 Å^2,

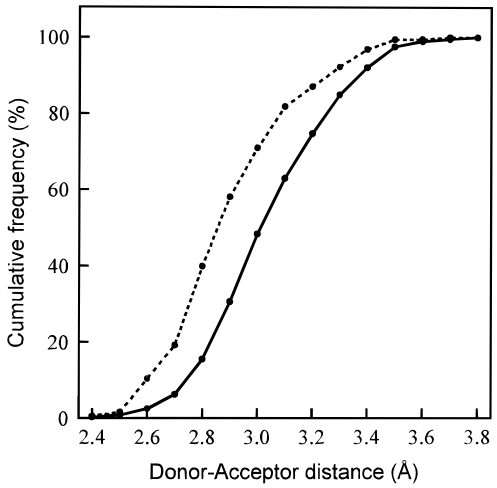

Fig. 3. Cumulative frequency distributions of donor-acceptor atomic distances in hydrogen bonds. Dotted line: hydrogen bonds identified in 14 lectin-carbohydrate complexes. Solid line: hydrogen bonds in 42 proteins (Stickle et al., 1992).

as determined from the data of Stickle et al. (1992) and the ΔASA values calculated by us. According to a test based on the *t* statistic, the difference in hydrogen bond density is significant at a level lower than 1.0% (i.e., the probability that both average densities actually come from the same population is less than 1.0%).

Distances between donor (D) and acceptor (A) atoms forming hydrogen bonds at lectin–sugar interfaces were found distributed in the range 2.5–3.5 Å, with a mean value of 2.94 (±0.25) Å. Figure 3 compares the cumulative frequency distribution of D-A distances for complexes (dotted line) and globular proteins (solid line); it can be seen that the curve for proteins is clearly displaced toward longer distances. Indeed, in this latter case the mean D-A distance is 3.04 (±0.24) Å (Stickle et al., 1992), which is 0.10 Å longer than the corresponding value for L-C interactions. Given the large number of hydrogen bonds analyzed, this difference of the two average D-A distances is highly significant (risk level < 0.05%). A comparison was also made of angular values (at both acceptor and donor atoms) between the two sets of hydrogen bonds (Fig. 4). In the case of the donor antecedent-donor-acceptor (DD-D-A) angle, only minor differences were found, regardless of the hybridization type of the donor atom (Fig. 4A,B). In contrast, values of the acceptor antecedent-acceptor-donor (AA-A-D) angle in complexes are narrowly distributed around 112.3° for sp^3 acceptors, and 120.2° for sp^2 acceptors (Fig. 4C,D); these mean values are close to the ideal values (109.5° and 120° for sp^3 and sp^2 acceptors, respectively). In protein interiors, the distributions are much wider and centered farther from ideal values.

Overall, results from the comparative analysis of hydrogen bonds reveal two major causes that can be responsible for the larger value of Δh_{pol} in L-C complexes: higher density of hydrogen bonds per buried surface area and better hydrogen bonding geometry. A third effect that must be considered is the electronegativity of acceptor and donor atoms, because it is commonly accepted that an important contribution to hydrogen bond strength arises from electrostatic interactions (Creighton, 1993; Lazaridis et al., 1995). When the particular hydrogen-bond densities for complexes and proteins are taken into account, the difference in Δh_{pol} values reduces to an enthalpy (energy) difference of 0.7 kcal/mol in the formation of a hydrogen bond ($\Delta\Delta E_{Hb}$). Of course, it would be interesting to determine whether this more favorable interaction energy in L-C complexes can be explained, on theoretical grounds, by geometric and electronegativity factors. To do this would require, for example, a detailed evaluation of L-C interaction energies by means of the common force fields employed in molecular mechanics simulations. Studies of this type have recently been done for globular proteins (Creighton, 1993; Lazaridis et al., 1995). Here, we have only made a rough estimation of the difference between the energies of formation of a typical hydrogen bond in proteins and in L-C complexes. Largely, the most abundant hydrogen bonds in proteins (68% of the total) are those arising from the interaction of peptide >C=O and >N−H groups (Stickle et al., 1992). The Coulombic energy of this type of hydrogen bond was calculated assigning to each atom the partial charge used in the CVFF force field as implemented in the Discover program (Byosim/Molecular Simulations, San Diego, California); the value of the dielectric constant was 2.5, which is the mean of the range typically used to model protein interiors (Hendsch & Tidor, 1994; Lazaridis et al., 1995). The AA-A-D angle was set as 120°; acceptor, hydrogen, and donor atoms were assumed to be collinear, and the standard N−H bond length (1.00 Å) was used to position the hydrogen atom. Under the above considerations, the interaction energy between an >N−H and a >C=O group is −1.07 kcal/mol at an A-D distance of 3.04 Å and −1.19 kcal/mol at an A-D distance of 2.94 Å. Thus, shortening the A-D distance by 0.10 Å would increase the strength of the >N−H···O=C< interaction by 0.12 kcal/mol, certainly a small value compared to the expected increase in strength for hydrogen bonding in L-C complexes ($\Delta\Delta E_{Hb}$ = 0.7 kcal/mol).

Nevertheless, an analysis of hydrogen bonding by class revealed that in L-C complexes none of the hydrogen bond types constitutes the majority; the three major classes identified are >N−H···O−H, O−H···$(COO)^-$, and O−H···O=C<, which comprise 40, 22, and 14% of the total number of bonds, respectively. It was observed that D-A distances are significantly shorter when the acceptor is a carboxylate oxygen (2.82 ± 0.24 Å) than in those cases which involve uncharged bond-partners (3.00 ± 0.24 Å for >N−H···O−H; 3.06 ± 0.22 Å for O−H···O=C<). Similar observations have been made in the study of hydrogen bonding in small organic molecules, suggesting that charged groups form shorter and stronger hydrogen bonds (Taylor et al., 1983; Creighton, 1993). Interestingly, in proteins only 7% of hydrogen bonds involve charged groups (Stickle et al., 1992). Energies of interaction for each of the three major types of hydrogen bonds in L-C complexes were estimated following considerations similar to those outlined above for the peptide hydrogen bond. Mean D-A distances were used in those calculations. Not surprisingly, the O−H···$(COO)^-$ interaction is the strongest, with a total energy of −4.62 kcal/mol. The other two types have energies (−0.73 kcal/mol for >N−H···C−H; −1.25 kcal/mol for O−H···O=C<) comparable to that for the >N−H···O=C< interaction. After accounting for the relative abundance, the weighted average energy would be −1.95 kcal/mol, which exceeds by 0.9 kcal/mol the energy estimated for a typical >N−H···O=C< interaction in proteins. This energy difference is of the same order of magnitude as the expected value of $\Delta\Delta E_{Hb}$ (0.7 kcal/mol). It must be stressed, however, that the above calculations were only aimed at showing in a semiquantitative way that electronegativity and geometric factors may well explain why hydrogen bonds in L-C complexes are, on the average, stronger than in proteins.

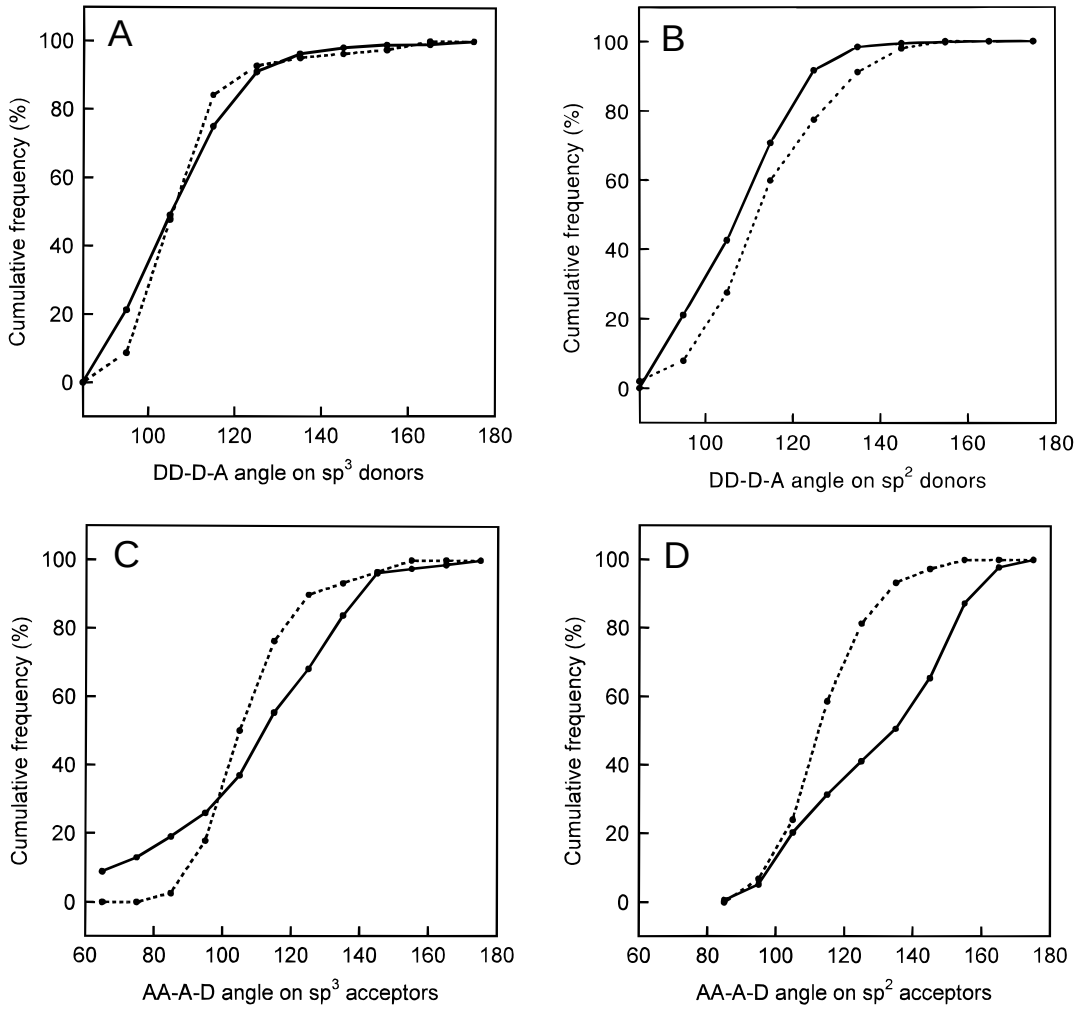

Fig. 4. Cumulative frequency distributions of scalar angles employed to identify hydrogen bonds. The angles considered are DD-D-A (donor antecedent-donor-acceptor) angle at (**A**) sp^3 donor atoms and (**B**) sp^2 donor atoms. AA-A-D (acceptor antecedent-acceptor-donor) angle at (**C**) sp^3 acceptor atoms and (**D**) sp^2 acceptor atoms. In all cases, dotted lines represent hydrogen bonds identified in 14 lectin-carbohydrate complexes, and solid lines represent hydrogen bonds in 42 proteins (Stickle et al., 1992). Angle values in degrees.

Binding entropies

Whenever it has been possible to make a structural comparison between the free and bound states of a lectin, the binding site has been found to undergo virtually no conformational changes upon association to carbohydrate. This has led to the widely accepted picture of "preformed" lectin binding sites (Weis & Drickamer, 1996). In contrast, the flexibility that carbohydrates present in solution keeps controversial. However, the conundrum is confined to establish the extent of torsional oscillations around glycosidic linkages, since sugar rings exhibit the stable chair conformation both free and bound to a lectin (Homans & Rutherford, 1993; Pérez et al., 1994; Siebert et al., 1996). These uncertainties notwithstanding, the facts that in general the conformations of ligands in L-C complexes are close to the predicted global minimum-energy configuration, and that oligosaccharides have been found to present generally similar geometrical arrangements when forming different complexes (Imberty, 1997), support the idea that lectins recognize preferentially the most stable, and presumably the most populated, conformation of the ligand. These considerations are relevant because in the present study the approximation of rigid bodies has been undertaken when calculating ΔASA values. The good fit observed between enthalpy and *ASA* changes in L-C binding (see above) provides an evidence of the suitability of the approximation. However, as stated below, to consider the different sources of entropic changes, this approximation has to be supplemented by taking into account the freezing of the carbohydrate and lectin functional groups with free rotatable bonds involved in binding. Indeed, it has been observed frequently that carbohydrate-binding reduces the mobility of groups with high temperature factors located in the binding site.

In view of the above considerations, the entropy change associated to the binding processes discussed here can be expressed as the sum of four contributions:

$$\Delta S_b = \Delta S_{pol} + \Delta S_{ap} + \Delta S_{rot} + \Delta S_{t-or}. \quad (4)$$

The first two terms in Equation 4 refer to entropy changes due to desolvation of polar and apolar surfaces that are buried upon bind-

ing. ΔS_{rot} originates from single bond rotations, which are frozen in the complex but permitted in the free ligand and in the protein. Finally, $\Delta S_{t\text{-}or}$ is the entropic contribution resulting from restrictions in translation and overall rotation motions of the molecules forming the complex. While entropic effects due to desolvation and hindering of rotatable bonds can be thought to be analogous to those occurring in protein folding, the term $\Delta S_{t\text{-}or}$ in Equation 4 is particular to *binding phenomena*. Thus, a direct comparison of binding entropies with folding entropies is clearly not possible. Nevertheless, it is instructive to compare experimental binding entropies with those calculated by evaluation of each of the terms appearing in Equation 4. For this purpose, we employed two different approaches currently used in protein folding studies to estimate the entropy of desolvation.

Based on experimental data for the transfer of a large number of organic compounds to water, Privalov and Makhatadze (1993) have estimated thermodynamic parameters for the hydration of several chemical groups. Assuming that solvation effects are additive, these authors calculate the contribution of hydration to changes in thermodynamic functions upon unfolding of proteins. In this work, we supplemented the data set of Privalov and Makhatadze by including hydration entropies for cyclic alkane carbon and cyclic ether groups, which are necessary to describe the solvation of carbohydrates. These two group parameters were derived from entropies of transfer from the gas phase to water for relevant organic compounds (Cabani et al., 1981), following the same considerations made by Privalov and Makhatadze. Results obtained were -0.27 and -0.18 cal$(\text{mol}\cdot\text{K}\cdot\text{Å}^2)^{-1}$ for cyclic alkane carbon and cyclic ether surfaces, respectively. With the complete set of hydration entropies (normalized by the accessible surface area of groups involved) and the ΔASA values for the association reaction, we calculated the desolvation term in Equation 4 (column heading $T\Delta S_{solv,ga}$ in Table 3). The other way in which desolvation entropies were calculated was using the parametrizations developed by Freire and coworkers (D'Aquino et al., 1996):

$$\Delta S_{ap} = 0.45\Delta ASA_{ap}\ln(T/385.15) \quad (5)$$

$$\Delta S_{pol} = -0.26\Delta ASA_{pol}\ln(T/335.15) \quad (6)$$

where T is the absolute temperature. At 25 °C, the solvation entropies per unit of apolar and polar area are -0.115 and 0.030 cal$(\text{mol}\cdot\text{K}\cdot\text{Å}^2)^{-1}$, respectively. Results obtained by means of the above relationships are shown under the heading $T\Delta S_{solv,p}$ in Table 3.

The entropic cost of freezing rotations around single bonds is a subject that has been addressed by a number of authors. For example, Finkelstein and Janin (1989) consider that $\Delta S_{rot} \approx -2$ cal$(\text{mol}\cdot\text{K})^{-1}$ per rotatable bond, which approximately corresponds to a reduction from three conformers (rotamers) to a single conformer when a bond rotation is frozen. In a sound analysis of this topic, Doig and Sternberg (1995) recollected and discussed results from many workers. These authors have calculated a mean value for the entropy change that results from restricting the number of rotamers for each type of amino acid side chain. On the average, the entropy loss per rotatable bond is 1.53 cal$(\text{mol}\cdot\text{K})^{-1}$. To calculate the ΔS_{rot} term in Equation 4, we determined which side chains in the protein were restricted from rotational movements upon complex formation; this was accomplished by inspection of the molecular structures of the complex and the protein alone. Side chains, originally exposed to solvent, which become buried upon binding, and in close contact with the ligand, were assumed to be restricted to a single conformer (on analogy with the protein folding case where side chains buried in the protein core are generally restricted to a single conformation (Doig & Sternberg, 1995; D'Aquino et al., 1996)). Knowing which side chains become immobilized, we calculated the entropy loss by using the corresponding values in the entropy scale reported by Doig and Sternberg (1995). The number of rotatable bonds frozen in the carbohydrate part of a complex was similarly determined from molecular structures; their entropic contribution was then calculated considering the average value reported by Doig and Sternberg (1.53 cal$(\text{mol}\cdot\text{K})^{-1}$ per rotatable bond). The total for each of the L-C complexes is shown in the $T\Delta S_{rot}$ column of Table 3.

Regarding the last term in Equation 4, it should be recalled that the entropy associated with the overall molecular movements of translation and rotation in gas phase can be calculated from statistical thermodynamics principles (Finkelstein & Janin, 1989; Holtzer, 1995, and references cited therein). Whereas some authors consider that these calculations also apply to ideal solutions (Finkelstein & Janin, 1989; Janin, 1995), others have made corrections to account for the greater order existing in liquids in comparison to gas phase (Searle & Williams, 1992; Searle et al., 1992). What is clear is that part of the translation-rotation degrees of freedom lost

Table 3. *Dissection of the entropic change accompanying the formation of lectin-carbohydrate complexes*[a]

Complex	$T\Delta S_{solv,ga}$	$T\Delta S_{solv,p}$	$T\Delta S_{solv,lc}$	$T\Delta S_{rot}$	$T\Delta S_{t\text{-}or}$	$T\Delta S_{t\text{-}or,lc}$	$T\Delta S_{calc,ga}$	$T\Delta S_{calc,p}$	$T\Delta S_{calc,lc}$	$T\Delta S_{b,exp}$
H-ch$_2$	23.3	9.2	6.9	−4.2	−6.7	−2.6	12.4	−1.7	−0.1	−1.5
H-ch$_3$	28.4	9.8	7.0	−7.8	−7.2	−2.6	13.4	−5.2	−3.4	−1.4
5cna	17.5	4.4	2.8	−3.0	−6.0	−2.6	8.5	−4.6	−2.8	−1.5
1gic	16.1	5.6	4.0	−2.5	−6.0	−2.6	7.6	−2.9	−1.1	−0.7
1cvn	33.7	5.4	2.8	−6.8	−7.2	−2.6	19.7	−8.6	−6.6	−6.6
1led	34.0	7.6	4.6	−7.8	−7.4	−2.6	18.7	−7.7	−5.8	−5.4
1rin	16.1	3.5	2.0	−1.8	−6.0	−2.6	8.3	−4.3	−2.4	−1.3
1lte	22.1	6.2	4.2	−5.1	−6.7	−2.6	10.3	−5.6	−3.5	−5.3
1slt	34.0	7.6	3.1	−3.5	−6.7	−2.6	14.7	−4.9	−3.0	−3.8

[a] All data correspond to a temperature of 25 °C. H-ch$_2$ and H-ch$_3$ stand for the complexes of hevein with chitobiose and chitotriose, respectively. All other complexes are referred to by their PDB code names shown in Table 1. See text for an explanation of the abbreviations used for different $T\Delta S$ terms. All values in kcal/mol.

in an association reaction reappears as residual movements in the bimolecular complex; therefore, only about 50% of the translation-rotation entropy loss actually contributes to the entropy of binding (Finkelstein & Janin, 1989; Searle & Williams, 1992). Finkelstein and Janin (1989) have estimated this net entropy loss as approximately 50 cal$(\text{mol}\cdot\text{K})^{-1}$, a value that should be, in principle, independent of molecular mass. However, by using a combination of calculation and experimental data for the gas to crystal transfer of diglycine, Brady and Sharp (1997) obtain a lower entropy loss of 14.1 cal$(\text{mol}\cdot\text{K})^{-1}$. From the results presented by Searle et al. (see Fig. 1 in Searle et al., 1992), it is straightforward to estimate the entropy change that occurs when a ligand of a given molecular mass binds to a larger receptor with complete loss of translational and rotational movements. This entropy change is slightly dependent on the molecular mass, and for molecules the size of mono- to tetrasaccharides amounts to 40–50 cal$(\text{mol}\cdot\text{K})^{-1}$. Assuming that 50% of this entropy reappears in the complex as residual movements, the net entropy loss would be 20–25 cal$(\text{mol}\cdot\text{K})^{-1}$. It is also worth mentioning the recent statistical mechanical analysis of this problem done by Amzel (1997) in terms of the cell theory of liquids. The approach used by Amzel seems to overcome major criticisms made to previous methods that estimate the loss of translational entropy upon binding. The resulting formula for this entropy change includes as a major term the so-called "cratic entropy" $(-R \ln 55)$ plus a term representing the "communal" entropy, giving a total change of approximately -10 cal$(\text{mol}\cdot\text{K})^{-1}$. Amzel's results thus lend theoretical support to the use of the cratic entropy in dissecting experimental values, as has been done by some authors (Murphy et al., 1994).

In view of the data presented above, it seems reasonable to expect that for L-C binding the term $\Delta S_{t\text{-}or}$ would lie between -10 to -25 cal$(\text{mol}\cdot\text{K})^{-1}$ (i.e., -3.0 to -7.4 kcal/mol at 298 K). In a first stage of our calculations, we used values of $T\Delta S_{t\text{-}or}$ derived from the work of Searle et al. (1992), which are shown in Table 3.

Columns 8 and 9 of Table 3 show the entropic contributions to binding calculated either with hydration entropies of chemical groups ($T\Delta S_{calc,ga}$) or through the parametrization developed by Freire and coworkers ($T\Delta S_{calc,p}$). For comparison, the experimental $T\Delta S_{b,exp}$ values are listed in last column of the same table. It is immediately evident that use of the group additivity approach to treat desolvation effects leads to a gross overestimation of entropy changes ($T\Delta S_{calc,ga}$ exceeds $T\Delta S_{b,exp}$ by 15.7 kcal/mol on the average). Moreover, in this case calculated and experimental entropies are completely uncorrelated (in fact, there is a negative correlation between these values, as can be seen in Fig. 5A). If the value of $T\Delta S_{t\text{-}or}$ estimated by Finkelstein and Janin (1989) were used in the calculation, the discrepancy between $T\Delta S_{calc,ga}$ and $T\Delta S_{b,exp}$ would be reduced, yet experimental values would still be overestimated by 7.5 kcal/mol on the average, and the lack of correlation would persist. In contrast, binding entropies calculated by using Equations 5 and 6 are much closer to and fairly well correlated with experimental values (Fig. 5B, solid symbols); the distribution of $T\Delta S_{calc,p} - T\Delta S_{b,exp}$ values has a mean of -2.1 kcal/mol (standard deviation: 1.21 kcal/mol). Replacing the $T\Delta S_{t\text{-}or}$ term by the "cratic" entropy (-3.0 kcal/mol) has little effect on the correlation between calculated and experimental binding entropies; on the average, calculated values in this case are higher than the experimental ones by $+2.0$ kcal/mol (standard deviation: 1.78 kcal/mol).

Two major conclusions can be derived from the results presented above. First, desolvation entropies occurring in L-C association are better estimated by means of the parametrization shown in Equations 5 and 6, whereas calculations based on the approach of Privalov and Makhatadze (1993) largely overestimate these effects. It must be mentioned that in protein folding and protein–protein association, the validity of group additivity for polar hydration has been seriously questioned (Lazaridis et al., 1995; D'Aquino et al., 1996). It has been noted, for example, that treat-

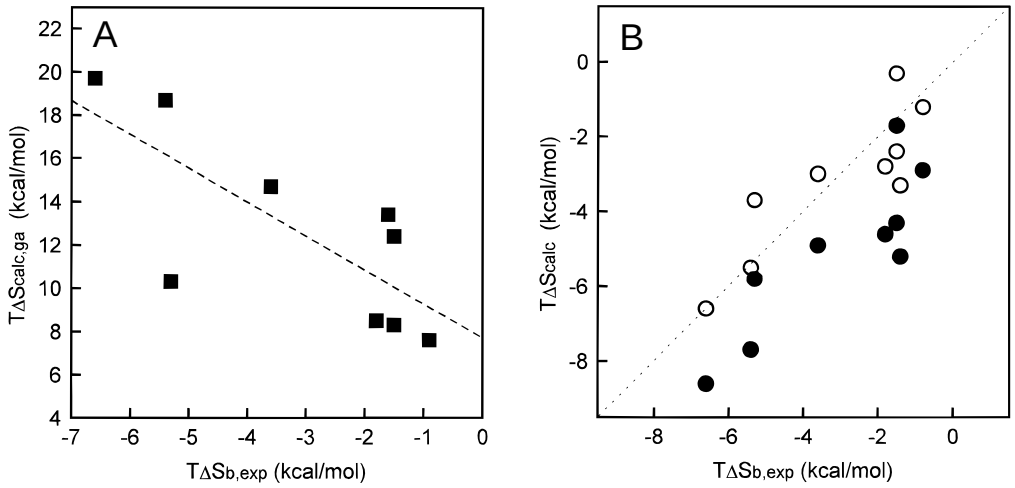

Fig. 5. Calculated ($T\Delta S_{calc}$) vs. experimental ($T\Delta S_{b,exp}$) binding entropies for lectin-carbohydrate complexes. Individual contributions to the binding entropy (see Equation 4 and the section on *binding entropies* for a detailed discussion of each contribution) were estimated and added up to obtain the calculated binding entropy. **A:** Entropic effects due to desolvation were taken into account according to a group additivity approach (Privalov & Makhatadze, 1993). **B:** Solid circles show results obtained when the entropy desolvation was estimated from a parametrization developed by D'Aquino et al. (1996) (cf. Equations 5 and 6); open symbols show results derived from a lectin-carbohydrate *ad hoc* parametrization that accounts for the desolvation entropy and a particular "cratic" entropy term.

ment of solvent-entropy effects by means of the parametrization derived by Freire and coworkers leads to an indirect estimation of the conformational entropy in protein folding that agrees well with values obtained from other direct approaches (Doig & Sternberg, 1995; D'Aquino et al., 1996); on the contrary, use of the group additivity method yields a serious overestimation of conformational entropy changes. Furthermore, Cabani et al. (1981) have pointed out that considerable deviations from additivity can be observed *even* in small organic molecules containing more than one polar group. Second, since the same parametrization seems to properly account for solvent-related entropy changes in protein-folding and in L-C binding, it may be thought that, on a first approximation, the overall thermodynamics of solvation is similar in these two types of events. In this respect, it is interesting to ask whether a set of solvation parameters particular to the L-C case can give better estimates of experimental binding entropies. To explore this possibility, we made a regression analysis of $T(\Delta S_{b,exp} - \Delta S_{rot})$ for the lectin complexes in Table 3 as a linear function of the independent variables ΔASA_{pol} and ΔASA_{ap}, plus a constant term that would represent an ad hoc "cratic" entropy (cf. Equation 4). This analysis yielded entropies per unit of apolar (-0.095 ± 0.029 cal(mol·K·Å^2)$^{-1}$) and polar (0.041 ± 0.025 cal(mol·K·Å^2)$^{-1}$) area that differ slightly from the values obtained from Equations 5 and 6. Solvation entropies calculated with these parameters are presented in Table 3 in the column headed as $T\Delta S_{solv,lc}$. The ad hoc "cratic" entropy ($T\Delta S_{t\text{-}or,lc}$ in Table 3) would be equal to -2.6 kcal/mol at 25 °C. It should be emphasized that this last value is much closer to the estimation derived from Amzel's work (Amzel, 1997) than to estimations obtained from other approaches. Binding entropies calculated with ad hoc parameters are shown in Figure 5B (open symbols) and in Table 3 under the heading $T\Delta S_{calc,lc}$. In this case the estimation of binding entropies is unbiased (i.e., values of $T(\Delta S_{calc,lc} - \Delta S_b)$ average out to 0.0 kcal/mol); otherwise, the accuracy of estimation improves only slightly (standard deviation of $T(\Delta S_{calc,lc} - \Delta S_{b,exp})$ equals 1.31 kcal/mol).

Binding free energies

One of the ultimate goals of structure-based thermodynamic analysis of biomolecular associations is to predict binding affinities from structural considerations. Hence, it is important to determine the accuracy with which L-C binding free energies can be estimated from the parametrizations described above. For this purpose, calculated binding-enthalpies and -entropies (discussed in preceding sections) were used to obtain calculated free energies of association. Figure 6 compares the experimental and calculated (ΔG_{calc}) free energies for the L-C complexes. $\Delta G_{calc} - \Delta G_b$ values have a mean of 0.0 kcal/mol, and a standard deviation of 1.10 kcal/mol, which amounts to an uncertainty in the association constant of one order of magnitude. Encouraging as they are, these results suggest, however, that inclusion of other structural factors not considered in this work might lead to an improvement of predicted L-C affinities.

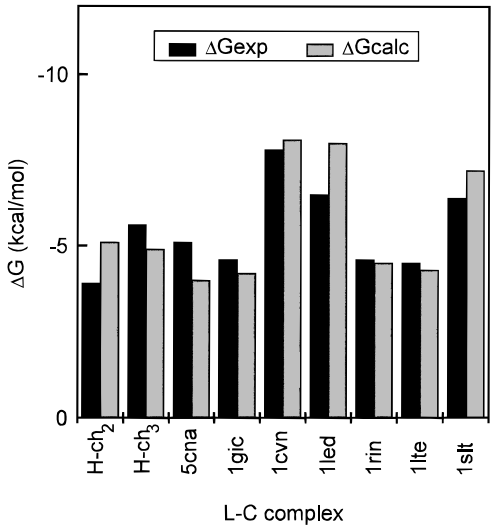

Fig. 6. Comparison of calculated and experimental binding free energies of lectin-carbohydrate complexes. Calculated values were obtained as described in text. All data correspond to a temperature of 25 °C. H-ch$_2$ and H-ch$_3$ stand for the complexes of hevein with chitobiose and chitotriose, respectively. All other complexes are referred to by their PDB code names shown in Table 1.

Finally, the binding of GlcNAc to hevein represents an interesting test case, because of the following reasons: (1) the binding constant, as determined by NMR (Asensio et al., 1995), is only 30 ± 15 (in molar units) at 30 °C, which places this association process at the lower limit of protein–carbohydrate recognition; (2) no calorimetric evidence of binding is found in titrations of hevein with GlcNAc (García-Hernández et al., 1997); and (3) the structure of this complex is unknown, but it can be inferred from the structure of chitobiose, $(GlcNAc)_2$, bound to hevein. In fact, it is reasonable to assume that the binding locus for GlcNAc is one of the sites occupied by either the reducing or nonreducing monosaccharide moieties in hevein-bound chitobiose (Asensio et al., 1995). Both possibilities were tested; results from their structural analysis, together with estimations for the thermodynamic parameters of binding (at 25 °C) are shown in Table 4. Though both calculated association constants are comparable to the experimen-

Table 4. *Energetic estimation of monosaccharide binding to hevein*[a]

Chitobiose-end	ΔASA_{pol} (Å^2)	ΔASA_{ap} (Å^2)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)	ΔG (kcal/mol)	K_a
Nonreducing	−128.3	−210.2	−4.7	−2.1	−2.6	80
Reducing	−88.2	−157.5	−3.1	−1.7	−1.4	11

[a] Structure-based energetic calculations of GlcNAc binding to hevein. Two putative binding loci for GlcNAc can be derived from the hevein-chitobiose structure. Entropic and enthalpic change values for each locus were calculated from the parametrizations developed for lectin-carbohydrate binding, and used to estimate the free energy change. K_a is the equilibrium constant of association.

tal value, the larger constant at the putative site corresponding to the nonreducing end of chitobiose suggests this could be the primary binding site for the monosaccharide. Furthermore, making use of the estimated molar enthalpy change for the reaction (Table 4), the heat evolved in a typical calorimetric titration can easily be calculated. Under conditions employed by García-Hernández et al. (1997) (7 μL of 0.25 M ligand added to 1.4 mL of 0.1 mM hevein solution), the expected liberation of heat ($\sim$20 μcal) would be completely counterbalanced by the heat effect due to ligand dilution.

Materials and methods

The structural-thermodynamic database of lectin-carbohydrate complexes

Only L-C complexes satisfying the following requirements were included in the analysis: (1) known structural information at the atomic level; and (2) calorimetrically-measured binding data. Non-calorimetric data were excluded from the analysis since, as shown recently for different binding systems, van't Hoff enthalpies tend to be significantly higher than those values obtained directly by isothermal titration calorimetry (Naghibi et al., 1995). Table 1 shows 10 L-C complexes satisfying the above mentioned criteria. In this data set, lectin complexes belonging to three different structural families appear (Weis & Drickamer, 1996): one galectin complex, three chitin-binding lectin complexes, and six legume lectin complexes. Homologies in the folding pattern notwithstanding, it is to be noted that lectins of Table 1 show dramatic differences in ligand specificity, subsite multivalency (i.e., number of sugar units that the binding site can accommodate simultaneously), and subunit multivalency (i.e., number of lectin monomers that conform a single binding site) (Rini, 1995).

Estimations of accessibility changes

Structure-based calculations of water-accessible surface areas *ASA* were performed with NACCESS (Hubbard & Thornton, 1993), a version of the Lee and Richards (1971) algorithm. A probe radius of 1.4 Å, a slice width of 0.1 Å, and the van der Waals radii estimated by Chothia (1976) were used. The changes in *ASA* upon binding (ΔASA) were estimated from the difference between the complex and the sum of free molecules. Estimations of ΔASA due to protein folding were obtained as the difference between the native and unfolded states. The X-ray solved structure of the protein was used as the native conformation. The unfolded state was built as a completely extended chain with all dihedral angles φ, ψ, and χ equal to 180°, except the χ_1 angle, which was set at 60°; otherwise, a 180° value for this angle causes the tail of long side chains to be brought near the backbone atoms, introducing a spurious screening effect. In addition to those proteins listed in Table 2, folding ΔASA estimations for the set of 42 proteins analyzed by Stickle et al. (1992) were carried out; name and Protein Data Bank code (Bernstein et al., 1977) of each of these molecules are: cytochrome C551 (351c), actinidin (2act), penicillopepsin (2app), azurin (2aza), phospholipase (1bp2), carbonic anhidrase (2ca2), cytochrome *c*3 (2cdv), α chymotrypsin A (5cha), carboxypeptidase A α (5cpa), crambin (1crn), citrate synthase (2cts), cytochrome *c* (5cyt), dihydrofolate reductase (4dfr), hemoglobin (1ecd), flavodixin (4fxn), γ-II crystallin (1gcr), glyceraldehyde-3-phosphate dehydrogenase (1gd1), glutathione peroxidase (1gp1), glutathione reductase (3grs), hemerythrin (1hmq), amylase inhibitor (1hoe), insulin (1ins), hemoglobin v (2hlb), lysozyme (1lz1), lysozyme (3lzm), myoglobin (1mbo), ovomucoid third domain (2ovo), papain (9pap), apo-plastocyanin (2pcy), pepsin (4pep), avian pancreatic polypeptide (1ppt), trypsin inhibitor (5pti), rubredoxin (1rdg), Ig Bence–Jones protein (2rhe), Lys 25-ribonuclease T1 (3rnt), ribonuclease A (7rsa), scorpion neurotoxin (1sn3), *Staphylococcus* nuclease (1snc), thermolysin (3tln), β-trypsin (1tpp), ubiquitin (1ubq), and Trp repressor (2wrp).

Polar area changes (ΔASA_{pol}) were obtained from the change in accessible area of nitrogen and oxygen atoms, while the apolar area change (ΔASA_{ap}) was computed from contributions of carbon and sulfur atoms. These ΔASA calculations were used to solve Equation 1 and to estimate desolvation entropies according to Equations 5 and 6. To estimate the entropic changes associated to desolvation of contact zones according to the approach of Privalov and Makhatadze (1993), a finer dissection of accessible areas of L-C complexes was computed by explicitly taking into account the following protein chemical groups: aliphatic groups, aromatic groups, peptide bonds, and the polar part of the side chain of each amino acid. For the carbohydrates, ΔASA of linear alkane carbons, cyclic alkane carbons, cyclic ether, and hydroxyl groups were calculated.

Analysis of hydrogen bonding in lectin–carbohydrate interfaces

A geometrical analysis of interfacial hydrogen bonds was performed on 14 L-C complexes pertaining to six different lectin families, as follows. Legume family: peanut lectin binding lactose (2pel, Banerjee et al., 1996), concanavalin A binding α-methyl-mannopyranoside (5cna, Naismith et al., 1994), concanavalin A binding Man(α1$\rightarrow$6)[Man(α1$\rightarrow$3)]Man (1cvn, Naismith & Field, 1996), *Lathyrus ochrus* lectin binding α-methyl-mannopyranoside (1lob, Bourne et al., 1990), *Griffonia simplicifolia* lectin IV binding Fuc(α1$\rightarrow$2)Gal(β1$\rightarrow$3)[Fuc(α1$\rightarrow$4)] GlcNAc (1led, Delbaere et al., 1993). C-type family: mannose-binding protein-C binding α-methyl-fucopyranoside (1rdi, Ng et al., 1996), mannose-binding protein-C binding α-methyl-mannopyranoside (1rdl, Ng et al., 1996), mannose-binding protein-A binding Man(α1$\rightarrow$3)[Man(α1$\rightarrow$6)]Man(3$\leftarrow$1α)Man(2$\leftarrow$1α) Man (2msb, Weis et al., 1992). Galectin family: S-lectin binding N-acetyllactosamine (1slt, Liao et al., 1994). Bacterial family: enterotoxin binding galactose (1lta, Merritt et al., 1994b), cholera toxin binding Gal(β1$\rightarrow$3) GalNAc(β1$\rightarrow$4)[NeuAc(α2$\rightarrow$3)] Gal(β1$\rightarrow$4)Glc (1chb, Merritt et al., 1994a). Chitin-binding family: hevein binding chitobiose (Asensio et al., 1995). β-Prism fold family: snowdrop lectin binding α-methyl-mannopyranoside (1msa, Hester et al., 1995), snowdrop lectin binding Man(α1$\rightarrow$3)[Man(α1$\rightarrow$6)]Man(α1$\rightarrow$6) Man(α1$\rightarrow$3) Man (1jpc, Wright & Hester, 1996). The analysis was based in criteria similar to those used by Stickle et al. (1992) in the study of the interior of 42 globular proteins. (1) Only high-resolution structures ($\leq$2.3 Å, $R \leq 20\%$) were included in the analysis. (2) Suitable hydrogen bonds were identified and analyzed on the basis of the geometric arrangement around acceptor-donor pairs. (3) The sum of expanded radii was used as the maximum allowed distance between the electronegative heavy atoms; in the study of Stickle et al. (1992), an expanded radius corresponds to 1.1 times the van der Waals radius of the corresponding atom. (4) Values for angles at sp^2 and sp^3 donor atoms (i.e., the donor

antecedent-donor-acceptor angle) were accepted in the range of 90–180°. The same range was used as allowed region for the angle at sp^2 acceptors (acceptor antecedent-acceptor-donor). For acceptor atoms with sp^3 hybridization, a range of 60 to 180° was used.

Identification of potential hydrogen bonds was carried out with a slightly modified version of the HBPLUS software (McDonald & Thornton, 1994).

Acknowledgments

This work was supported in part by CONACYT (Grant 473100-5-3771N).

References

Amzel LM. 1997. Loss of translational entropy in binding, folding, and catalysis. *Proteins Struct Funct Genet 28*:144–149.

Asensio JL, Cañada FJ, Bruix M, Rodríguez-Romero A, Jiménez-Barbero J. 1995. The interaction of hevein with N-acetylglucosamine-containing oligosaccharides: Solution structure of hevein complexed to chitobiose. *Eur J Biochem 230*:621–633.

Bains G, Lee RT, Lee YC, Freire E. 1992. Microcalorimetric study of wheat germ agglutinin binding to N-acetylglucosamine and its oligomers. *Biochemistry 31*:12624–12628.

Baker BM, Murphy KP. 1997. Dissecting the energetics of a protein-protein interaction: The binding of ovomucoid third domain to elastase. *J Mol Biol 268*:557–569.

Banerjee R, Das K, Ravishankar R, Suguna K, Surolia A, 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.

Bardi JS, Luque I, Freire E. 1997. Structure-based thermodynamic analysis of HIV-1 protease inhibitors. *Biochemistry 36*:6588–6596.

Ben-Naim A. 1991. The role of hydrogen bonds in protein folding and protein association. *J Phys Chem 95*:1437–1444.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Bourne Y, Rougé P, Cambillau C. 1990. Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: Fine specificity of the monosaccharide-binding site. *Proteins Struct Funct Genet 8*:365–376.

Brady GP, Sharp KA. 1997. Energetics of cyclic crystal packing and solvation. *Biophys J 72*:913–927.

Burrows SD, Doyle ML, Murphy KP, Franklin SG, White JR, Brooks I, McNulty DE, Scott MO, Knutson JR, Porter D, Young PR, Hensley P. 1994. Determination of the monomer-dimer equilibrium of interleukin-8 reveals it is a monomer at physiological concentrations. *Biochemistry 33*:12741–12745.

Cabani S, Gianni P, Mollica V, Lepori L. 1981. Group contributions to the thermodynamics properties of non-ionic organic solutes in dilute aqueous solution. *J Sol Chem 10*:563–595.

Chothia C. 1976. The nature of the accessible and buried surfaces in proteins. *J Mol Biol 105*:1–14.

Creighton TE. 1993. *Proteins structures and molecular properties*. New York: Freeman and Company.

D'Aquino JA, Gómez J, Hilser VJ, Lee KH, Amzel LM, Freire E. 1996. The magnitude of the backbone conformational entropy change in protein folding. *Proteins Struct Funct Genet 25*:143–156.

Delbaere LTJ, Vandonselaar M, Prasad L, Quail JW, Wilson KS, 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.

Doig AJ, Sternberg MJE. 1995. Side-chain conformational entropy in protein folding. *Protein Sci 4*:2247–2251.

Finkelstein AV, Janin J. 1989. The price of lost freedom: Entropy of bimolecular complex formation. *Protein Eng 3*:1–3.

Freire E. 1995. Thermodynamics of partly folded intermediates in proteins. *Annu Rev Biophys Biomol Struct 24*:141–165.

García-Hernández E, Zubillaga RA, Rojo-Domínguez A, Rodríguez-Romero A, Hernández-Arana A. 1997. New insights into the molecular basis of lectin-carbohydrate interactions: A calorimetric and structural study of the association of hevein to oligomers of N-acetylglucosamine. *Proteins Struct Funct Genet 29*:467–477.

Harrop SJ, Helliwell JR, Wan TCM, Kalb(Gilboa) AJ, Tong L, Yariv J. 1995. Structure solution of a cubic crystal of concanavalin A complexed with methyl-α-D-glucopyranoside. *Acta Crystallogr D 52*:143–145.

Hendsch ZS, Tidor B. 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci 3*:211–226.

Hester G, Kaku H, Goldstein IJ, Wright CS. 1995. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nature Struct Biol 2*:472–479.

Hirabayashi J, Kasai K. 1991. Effect of amino acid substitution by site-directed mutagenesis on the carbohydrate recognition and stability of human 14-kDa β-galactoside-binding lectin. *J Biol Chem 266*:23648–23653.

Holtzer A. 1995. The "cratic correction" and related fallacies. *Biopolymers 35*:595–602.

Homans SW, Rutherford T. 1993. Oligosaccharides and recognition: A "shape" problem probed by n.m.r. and molecular modelling. *Biochem Soc Trans 21*:449–452.

Hubbard SJ, Thornton JM. 1993. *NACCESS computer program*. London UK: Department of Biochemistry and Molecular Biology, University College.

Imberty A. 1997. Oligosaccharide structures: Theory versus experiment. *Curr Opin Struct Biol 7*:617–623.

Iobst ST, Drickamer K. 1994. Binding of sugar ligands to Ca^{2+}-dependent animal lectins: II. Generation of high-affinity galactose binding by site-directed mutagenesis and NMR. *J Biol Chem 269*:15512–15519.

Iobst ST, Wormald MR, Weis WI, Dwek RA, Drickamer K. 1994. Binding of sugar ligands to Ca^{2+}-dependent animal lectins: I. Analysis of mannose binding by site-directed mutagenesis and NMR. *J Biol Chem 269*:15505–15511.

Janin J. 1995. Elusive affinities. *Proteins Struct Funct Genet 21*:30–39.

Lazaridis T, Archontis G, Karplus M. 1995. Enthalpic contribution to protein stability: Insights from atom-based calculations and statistical mechanics. *Adv Protein Chem 47*:231–306.

Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static stability. *J Mol Biol 55*:379–400.

Leffler H, Barondes SH. 1986. Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian β-galactosides. *J Biol Chem 261*:10119–10126.

Lemieux RU. 1996. How water provides the impetus for molecular recognition in aqueous solution. *Acc Chem Res 29*:373–380.

Liao DI, Kapadia G, Ahmed H, Vasta GR, Herzberg O. 1994. Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein. *Proc Natl Acad Sci USA 91*:1428–1432.

Makhatadze GI, Privalov PL. 1995. Energetics of protein structure. *Adv Protein Chem 47*:307–425.

Mandal DK, Kishore N, Brewer CF. 1994. Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. *Biochemistry 33*:1149–1156.

McDonald YK, Thornton JM. 1994. Satisfying hydrogen bonding potential in proteins. *J Mol Biol 238*:777–793.

Merritt EA, Sarfaty S, van der Akker F, L'Hoir C, Martial JA, Hol WGJ. 1994a. Crystal structure of cholera toxin B-pentamer bound to receptor G_{M1} pentasaccharide. *Protein Sci 3*:166–175.

Merritt EA, Sixma TK, Kalk KH, van Zanten BAM, Hol WGJ. 1994b. Galactose-binding site in *Escherichia coli* heat-labile enterotoxin (LT) and cholera toxin (CT). *Mol Microbiol 13*:745–753.

Murphy KP, Freire E. 1992. Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv Protein Chem 43*:313–361.

Murphy KP, Xie D, García KC, Amzel LM, Freire E. 1993. Structural energetics of peptide recognition: Angiotensin II/antibody binding. *Proteins Struct Funct Genet 15*:113–120.

Murphy KP, Xie D, Thompson KS, Amzel LM, Freire E. 1994. Entropy in biological binding processes: Estimation of translational entropy loss. *Proteins Struct Funct Genet 18*:63–67.

Naghibi H, Tamura A, Sturtevant JM. 1995. Significant discrepancies between van't Hoff and calorimetric enthalpies. *Proc Natl Acad Sci USA 92*:5597–5599.

Naismith JH, Emmerich C, Habash J, Harrop S, Helliwell JR, Hunter WN, Raftery J, Kalb(Gilboa) AJ, Yariv J. 1994. Refined structure of concanavalin A complexed with methyl α-D-mannopyranoside at 2.0 Å and comparison with the saccharide-free state. *Acta Crystallog Sect D 51*:703–710.

Naismith JH, Field RA. 1996. Structural basis of trimannoside recognition by concanavalin A. *J Biol Chem 271*:972–976.

Ng KKS, Drickamer K, Weis W. 1996. Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J Biol Chem 271*:663–674.

Pérez S, Imberty A, Carver JP. 1994. Molecular modeling: An essential component in the structure determination of oligosaccharides and polysaccharides. *Adv Comp Biol 1*:147–202.

Privalov PL, Makhatadze GI. 1993. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J Mol Biol 232*:660–679.

Ramkumar R, Surolia A, Podder SK. 1995. Energetics of carbohydrate binding by a 14 kDa S-type mammalian lectin. *Biochem J 308*:237–241.

Rini JM. 1995. Lectin structure. *Annu Rev Biophys Biomol Struct 24*:551–557.

Rini JM, Hardman KD, Einspahr H, Suddath FL, Carver JP. 1993. X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 Å resolution. *J Biol Chem 268*:10126–10132.

Schwarz FP, Misquith S, Surolia A. 1996. Effect of substituent on the thermodynamics of D-glucopyranoside binding to concanavalin A, pea (*Pisum sativum*) lectin and lentil (*Lens culinaris*) lectin. *Biochem J 316*:123–129.

Schwarz FP, Puri KD, Bhat RG, 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.

Searle MS, Williams DH. 1992. The cost of conformational order: Entropy changes in molecular associations. *J Am Chem Soc 114*:10690–10697.

Searle MS, Williams DH, Gerhard U. 1992. Partitioning of free energy contributions in the estimation of binding constants: Residual motions and consequences for amide-amide hydrogen bond strengths. *J Am Chem Soc 114*:10697–10704.

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

Sharon N, Lis H. 1989. Lectins as cell recognition molecules. *Science 246*:227–234.

Siebert H-C, Gilleron M, Kaltner H, von der Lieth C-W, Kozár T, Bovin N, Korchagina EY, Vliegenthart JFG, Gabius H-J. 1996. NMR-based, molecular dynamics- and random walk molecular mechanics-supported study of conformational aspects of a carbohydrate ligand (Galβ1-2Galβ1-R) for an animal galectin in the free and in the bound state. *Biochem Biophys Res Comm 219*:205–212.

Stickle DF, Presta LG, Dill KA, Rose GD. 1992. Hydrogen bonding in globular proteins. *J Mol Biol 226*:1143–1159.

Surolia A, Sharon N, Schwarz FP. 1996. Thermodynamics of monosaccharide and disaccharide binding to *Erythrina corallodendron* lectin. *J Biol Chem 271*:17697–17703.

Taylor R, Kennard O, Versichel W. 1983. Geometry of the N—H···O=C hydrogen bond. 1. Lone-pair directionality. *J Am Chem Soc 105*:5761–5766.

Weatherman RV, Mortell KH, Chervenak M, Kiessling LL, Toone EJ. 1996. Specificity of C-glycoside complexation by mannose/glucose specific lectins. *Biochemistry 35*:3619–3624.

Weis WI, Drickamer K. 1996. Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem 65*:441–473.

Weis WI, Drickamer K, Hendrickson WA. 1992. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature 360*:127–134.

Wright CS, Hester G. 1996. The 2.0 Å structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: Evidence for two unique binding modes. *Structure 15*:1339–1352.