Protein Engineering vol.15 no.12 pp.979–986, 2002

Binding free energy calculations of galectin-3–ligand interactions

Tarun K.Mandal and Chaitali Mukhopadhyay[1]

Department of Chemistry, University of Calcutta, 92 A.P.C. Road, Kolkata-700 009, India

[1]To whom correspondence should be addressed.
E-mail: chaitali@cucc.ernet.in, chaitalicu@yahoo.com

Galectins show remarkable binding specificity towards β-galactosides. A recently developed method for calculating binding free energies between a protein and its substrates has been used to evaluate the binding specificity of galectin-3. Five disaccharides and a tetrasaccharide were used as the substrates. The calculated binding free energies agree quite well with the experimental data and the ranking of binding affinities is well reproduced. For all the six protein–ligand complexes it was observed that electrostatic interactions oppose binding whereas the non-polar contributions drive complex formation. The observed binding specificity of galectin-3 for galactosides rather than glucosides is discussed in light of our results.

Keywords: binding free energies/binding specificity/galectin-3/molecular dynamics simulations

Introduction

Mammalian cell surfaces and the extracellular matrix that surround them are rich in glycoconjugates (e.g. glycoproteins and glycolipids). Complex carbohydrates, an integral part of the cell surface glycoconjugates, are now being recognized as molecules with enormous coding capacity of meaningful messages in the form of their monosaccharide components, linkages, branching patterns, etc. They can act as recognition units for indigenous receptors, e.g. lectins (for review see Lis and Sharon, 1998). Lectins are ubiquitous, non-enzymic proteins that bind mono- and oligosaccharides reversibly and with high specificity. The common principles of molecular recognition, i.e. hydrogen bonds, dispersion forces and hydrophobic packing, govern the specificity of lectin–carbohydrate interaction (for reviews see Weis and Drickamer, 1996; Elgavish and Shaanan, 1997). Lectins, apart from specific recognition of carbohydrate moieties, are involved in a wide variety of biochemical processes including intra- and intercellular trafficking, initiation of signal transduction, cell adhesion, etc. In view of the highly versatile functional relevance of lectins, the design of high-affinity ligands to occupy their carbohydrate recognition domains offers the perspective for design of novel drugs for treatment of a wide range of diseases, from microbial infections and inflammatory diseases to cancer.

Amongst the commonly found oligosaccharides on the cell surface the β-galactosides are prominent components of the sugar chains found in both glycoproteins and glycolipids. The receptors for this linkage are galectins which are animal lectins having at least one carbohydrate recognition domain (CRD) of specific β-galactoside binding activity (Barondes *et al*., 1994; Cooper and Barondes, 1999). These galectins are soluble and widely distributed in the vertebrates and play a diverse intra- and extracellular biological function (Perillo *et al*., 1998). In mammals, so far, 12 glycoproteins have been defined as galectins (galectin-1 through galectin-12) (Cooper and Barondes, 1999; Yang *et al*., 2001). One of them, galectin-3, a well-studied and representative member of this family, contains a conserved ~14 kDa carbohydrate recognition domain showing high affinity for β-galactosides. Expression of galectin-3 is highest in activated macrophages, basofils and mast cells (Frigeri *et al*., 1993; Liu, 1993; Sato and Hughes, 1994), some epithelial cells, e.g. intestine, kidney (Foddy *et al*., 1990; Lindstedt *et al*., 1993; Lotz *et al*., 1993) and in some sensory neurons (Regan *et al*., 1986; Cameron *et al*., 1993). It has been shown to activate various cell types through cross linkage of appropriate cell surface glycoproteins, including cell adhesion molecules, to promote neurite growth (Pesheva *et al*., 1998) and induce differentiation and angiogenesis of endothelial cells (Nangia-Makker *et al*., 2000). It acts as a chemoattractant for monocyte (Sano *et al*., 2000) and endothelial cells (Nangia-Makker *et al*., 2000). It has also been shown that galectin-3 is active *in vitro* in inducing pre-mRNA splicing (Dagher *et al*., 1995). Galectin-3 is over-expressed in some types of cancer in which the normal parental cells do not express the protein, including specific types of lymphomas (Hsu *et al*., 1996; Konstantinov *et al*., 1996), thyroid carcinoma (Xu *et al*., 1995; Fernandez *et al*., 1997; Hsu *et al*., 1999), etc. Studies of cells transfected with galectin-3 cDNA or treated with specific antisense oligonucleotide, however, have provided evidence for the involvement of galectin-3 in tumor development and metastasis (Raz *et al*., 1990; Bresalier *et al*., 1998). It is likely that the glycoconjugate-mediated recognition processes of galectin-3 might be the key step in many of these biological processes.

It has been found experimentally that galectin-3 has different binding activity for galactose-containing oligosaccharides. The relative binding affinity of galectin-3 with different β-galactosides is Galβ(1–4)GlcNAc > Galβ(1–3)GlcNAc > Galβ(1–4)Glc > Galβ(1–3)GalNAc > Glcβ(1–4)Glc (Sparrow *et al*., 1987). The binding affinity of galectin-3 for oligosaccharides containing the above disaccharide linkages is stronger. This indicates that the primary binding site in galectin-3 may be specific for galactose but the secondary sites can have more flexibility in terms of the type of ligand. The structure of human galectin-3 carbohydrate recognition domain (CRD) complexed with lactose/*N*-acetyllactosamine has been solved at 2.1 Å resolution (Seetharaman *et al*., 1998). The high binding activity is attained by the galectin-3 through a tightly coordinated combination of hydrogen bonding, hydrophobic aromatic residue–sugar interactions, and a precise steric fit.

To capitalize on knowledge about the subtleties of lectin–carbohydrate interaction for rational marker/drug design, the intimate details of the recognition process need to be understood and exploited. In order to provide mechanistic insight into the ligand–receptor interaction or to explain the binding affinity

of a ligand the individual contributions of various factors which originate from the receptor, the ligand and/or the solvent to the overall free energy change have to be estimated. It is essential to study several receptor–ligand complexes to arrive at a consensus. The X-ray diffraction data, which provides the three-dimensional structure of the complexes, does not provide detailed information about the contributions of ligand and/or receptor flexibility. Also, the number of experimentally known ligand–receptor complexes is limited. Hence to provide mechanistic insights into the origins of ligand binding activity, computer-assisted molecular modeling of the reactants before and after complex formation can be utilized.

We have chosen the relative binding activity data for Galβ-(1–4)Glc, Galβ(1–4)GlcNAc, Galβ(1–3)GalNAc, Galβ(1–3)-GlcNAc, Glcβ(1–4)Glc and an oligosaccharide as test cases. Åqvist *et al*. (Åqvist *et al*., 1994) have proposed a semi-empirical linear interaction energy (LIE) approach for estimating absolute ligand binding free energies which depends on molecular dynamics simulations of the bound and free ligands in solution. Here the binding free energy is approximated as:

$$\Delta G_{bind} = \alpha(\langle E^{ES}\rangle_{pro} - \langle E^{ES}\rangle_{aq}) + \beta(\langle E^{VW}\rangle_{pro} - \langle E^{VW}\rangle_{aq}) \quad (1)$$

where E^{ES} and E^{VW} are the electrostatic and van der Waals interaction energies between ligand and its surroundings in protein (pro) and in aqueous solution (aq), respectively; $\langle\ \rangle$ denotes the ensemble average over molecular dynamics simulations trajectory and α and β are empirical parameters. Åqvist *et al*. (Åqvist *et al*., 1994) have applied this method to a series of test cases and have found that $\alpha \approx 0.5$ and $\beta \approx 0.16$ are suitable for different protein systems. However other values of these two parameters have been reported for different protein systems and it appears that these two parameters may be protein dependent. Regardless of the transferability of these parameters, the LIE calculation method has been shown to be quite successful in predicting relative binding affinities of ligands. There are several advantages to this method. Since the LIE method simulates only the final states, it is quite fast, it takes into account the flexibility of both the reactants, and finally, since solvent molecules are explicitly included, the desolvation free energy can be reasonably handled (W.Wang *et al*., 1999). We report here the results of LIE calculations for galectin-3 and six of its ligands. The results suggest that the number and/or pattern of the hydrogen bonding scheme does not seem to be the ultimate determinant of the binding affinity, the presence of sufficient numbers of non-bonded pairs are also equally important.

Materials and methods

Disaccharide formation

For the molecular modeling components the InsightII Package (version 98.0, Accelrys Inc.) was used. For the molecular mechanics and molecular dynamics calculations DISCOVER Module and CVFF force fields were used. All the disaccharides and the tetrasaccharide were built from monosaccharide templates. The glycosidic dihedral angles were changed from 0° to 360° at 30° intervals. Each conformation thus generated was minimized keeping the dihedral angle fixed. From the grid search the lowest energy conformation was identified and minimized without further constraint. The lowest energy conformation thus obtained was taken as the minimized conformation of the disaccharides. The tetrasaccharide was built joining the minimized disaccharide conformation and further minimized without constraint. The energy was minimized using the conjugate gradient method with no non-bonded cutoff. A constant dielectric of value 1.0 was used for all calculations.

Ligand–water assembly

The minimized conformations of all the disaccharide ligands were layered with about 12 Å thickness of water molecules. The total number of water molecules required was about 359–362 for disaccharides and 454 for the tetrasaccharide. The water layer was divided into two parts. The inner part is constructed defining the interface water molecules around the ligand of about 6 Å radius. The rest of the water molecules were defined as the outer layer. In the inner water layer the number of water molecules was about 134–144 for disaccharides and 186 for the tetrasaccharide. The water layer of the ligand–water assemblies was minimized initially by fixing the inner water layer, and then fixing the outer layer and finally unfixing all the solvent molecules. The minimization was performed by the steepest descent followed by the conjugate gradient method using 15 Å non-bonded cutoff and a constant dielectric of 1.0. All the optimized ligand–water assemblies were subjected to a dynamic run at 300 K constraining the covalent bond length using Rattle. Initial equilibration for 100 ps was followed by a 1.5 ns simulation run. The time step for the dynamic simulation was 2 fs.

Galectin–ligand–water assembly

A recently published crystal structure (Seetharaman *et al*., 1998) of the carbohydrate recognition domain of galectin complexed with Galβ(1–4)GlcNAc (PDB entry A3K) was used as the model of the galectin-3 molecule. The hydrogen atoms are generated in the crystal structure and their positions were optimized by minimizing the structure keeping the heavy atom fixed and removing all implicit water molecules. The minimization was performed using the steepest descent and conjugate gradient method. The minimized conformations of the ligands Glcβ(1–4)Glc, Galβ(1–3)GalNAc, Galβ(1–4)Glc, Galβ(1–3)GlcNAc, Galβ(1–4)GlcNAc, Gal(1–3)GlcNAc(1–3)-Gal(1–4)Glc were docked on the minimized structure of galectin. These docked complexes were further minimized by initially fixing all the protein atoms and then unfixing the side-chain atoms. For the tetrasaccharide the first or the second galactose residue was docked in the binding site and the energy minimized. The complex with the second galactose residue at the primary binding site had ~20 kcal/mol energy less than the complex with the first galactose residue at the primary binding site. We had used the former complex for the MD simulations. Water molecules were layered on the minimized structure of the galectin–ligand complexes with a thickness of 15 Å. A total of about 2800–2900 water molecules was needed. In the simulations reported here, the crystallographic water molecules were not included, as was also the case with Åqvist and Mowbray (Åqvist and Mowbray, 1995). Since the solvent molecules that were used have no bias in their positions, using some of them with defined positions might bias the calculations. The entire system was optimized by initially fixing the complex and then fixing only the backbone atoms of the protein. The minimized conformation was then superimposed on the crystal structure and it was observed that there are water molecules at most of the crystallographic water positions. During the minimization 15 Å non-bonded cutoff and a constant dielectric of 1.0 were used. All the minimized structures of the galectin–ligand–water assemblies were subjected to dynamic simulation

Table I. MD average interaction energies for bound and free ligand and experimental and calculated free energies of binding; energies are in kcal/mol and are an average over the last 400 ps of a 500 ps simulation except for Gal(1–4)GlcNAc which is the last 400 ps of a 1 ns simulation run

Ligand	$<E^{ES}>_{pro}$	$<E^{VW}>_{pro}$	$<E^{ES}>_{aq}$	$<E^{VW}>_{aq}$	ΔG_{expt}	ΔG_{calc} $\beta = 0.81$
Glc[a](1–4)Glc	–82.79 (2.2)	–31.72 (0.04)	–99.17 (0.1)	–19.16 (0.02)	–0.96	–1.98 (1.06)
Gal[a](1–3)GalNAc	–73.61 (0.91)	–35.97 (1.3)	–84.84 (0.35)	–24.04 (0.08)	–3.81	–4.04 (0.5)
Gal[a](1–4)Glc	–71.04 (0.2)	–35.1 (1.4)	–87.18 (0.23)	–20.14 (0.09)	–4.11	–4.04 (1.0)
Gal[a](1–3)GlcNAc	–68.68 (1.6)	–37.74 (0.15)	–75.06 (0.23)	–30.02 (0.17)	–5.215	–3.2 (0.68)
Gal[a](1–4)GlcNAc	–83.4 (1.7)	–41.48 (0.87)	–89.85 (0.15)	–30.76 (0.11)	–5.56	–5.45 (0.14)
Gal(1–3)GlcNAc (1–3)Gal[a](1–4) Glc	–115.17 (1.51)	–62.08 (0.48)	–136.87 (0.05)	–47.87 (0.37)	–5.45	–0.66 (0.36) –5.49 ($\beta = 1.15$)

The error bars are estimated from averaging of the first and second halves of the MD trajectory.
ΔG_{expt} is calculated from the experimental data Sparrow *et al*. (Sparrow *et al*., 1987) using $-RT\ln(1/K_d)$.
[a] Primary binding residue.

at 300 K during which a 10 Å sphere around the ligand molecule containing interface water–protein atoms was allowed to move and the rest of the protein and water atoms were held fixed. The covalent bond lengths were constrained using Rattle (as implemented in DISCOVER). The time step of the dynamic simulation was 2 fs. The initial equilibration of 100 ps was followed by a 500 ps stimulation run.

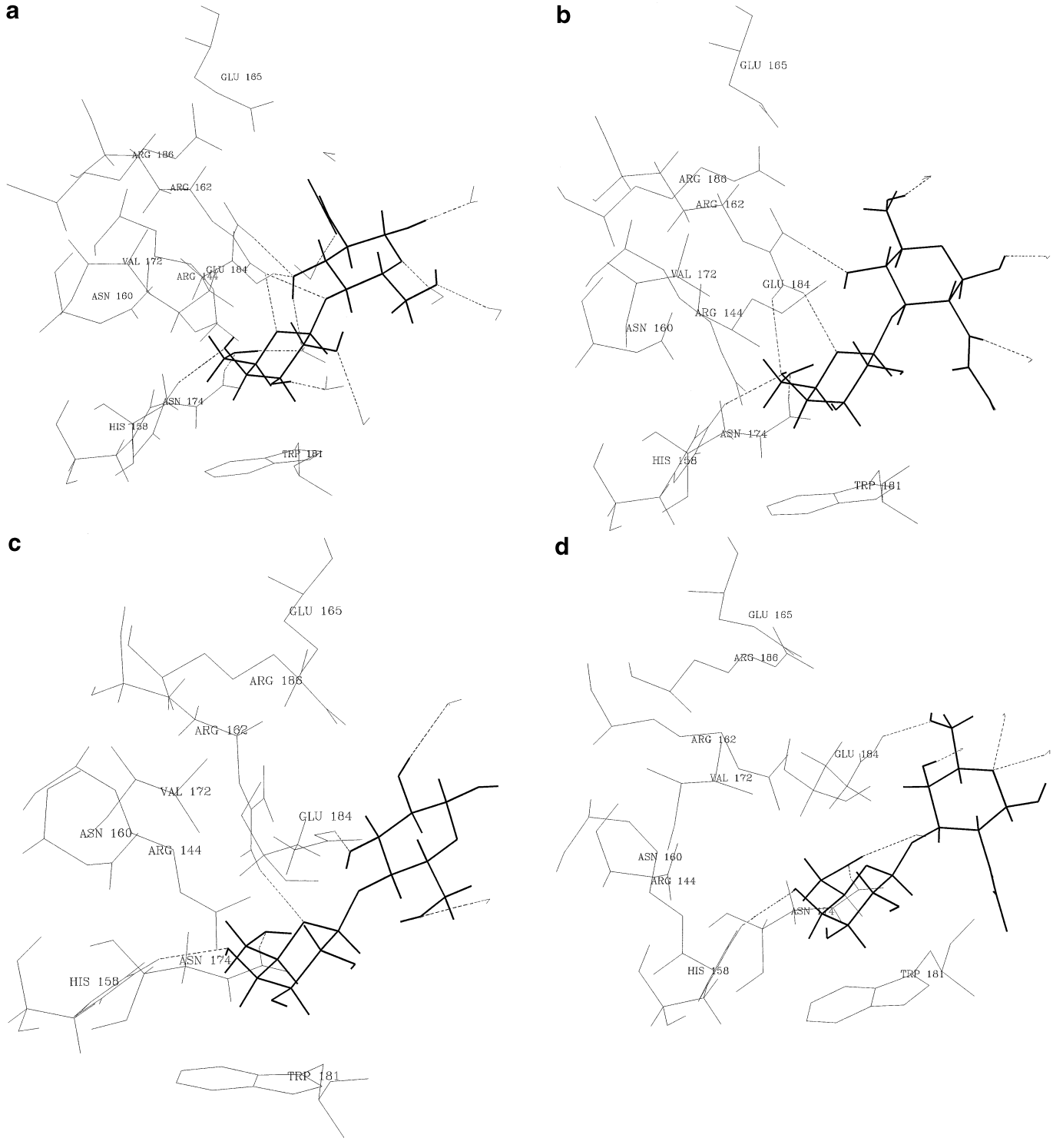
Results and discussion

We have used five disaccharides and a tetrasaccharide ligand to study the interactions with galectin-3. Table I summarizes the average interaction energies between the ligands and their surroundings, e.g. either protein or solvent molecules. All the protein–ligand trajectories were of 500 ps length, except for the natural ligand Galβ(1–4)GlcNAc for which the simulation was extended up to 1 ns. The simulation with Galβ(1–4) GlcNAc was started from the crystal structure (PDB entry A3K). However after 500 ps a shift in the ligand position away from that of the crystallographic position was noted, leading to a slightly altered pattern of the hydrogen bonds (discussed later). This simulation was, therefore, extended to 1 ns in order to get the conformation of the ligand within the binding site properly averaged. For all the ligands, it was observed that the van der Waals interactions are more favourable in the protein-bound form, compared to that in solvent. This is expected since in the protein the number of interacting non-polar groups is far greater in number. On the other hand, we have observed a loss in electrostatic interaction energy in going from free to bound state by all the six sugar ligands. Åqvist and Mowbray (Åqvist and Mowbray, 1995) reported that in the case of monosaccharides binding to the glucose/galactose receptor, an increase in electrostatic interaction in the protein-bound form was obtained from similar free energy calculations. They concluded that for the monosaccharides, electrostatic interactions dominate the binding. In our simulations we have obtained an unfavourable electrostatic component for all the ligands. This indicates that for larger carbohydrates containing several hydroxyl groups, the aqueous environment is preferred electrostatically. Similar reduction in the electrostatic interaction energy upon binding to proteins has been reported by W.Wang *et al*. (W.Wang *et al*., 1999), for ligand binding to P450-cam and thrombin. In another recent report on the binding free energy calculation of MHC class I protein–peptide interactions Froloff *et al*. (Froloff *et al*., 1997) concluded that the electrostatic interactions oppose binding and the protein–peptide association appears to be driven by the hydrophobic effect. Our results indicate that the fine balance between the unfavourable electrostatic energy increase and favourable non-polar interactions determines the binding affinities of these ligands to the protein. These observations are consistent with previous reports (Miyamoto and Kollman, 1993; Åqvist *et al*., 1994).

The coefficient used for the electrostatic interaction energy in Equation 1 has been assigned a value of 0.5 as has been used successfully by others (Åqvist *et al*., 1994; Åqvist and Mowbray, 1995; Paulsen and Ornstein, 1996; W.Wang *et al*., 1999). This value came from the first order approximation of electrostatic contribution to the binding free energy (Kollman, 1993). It has been shown that this first order approximation is reasonable (Still *et al*., 1990). Several workers have optimized the value of the van der Waals coefficient 'β' (referred to as α in the paper by Aqvist *et al*.). The values obtained for different systems varied widely, ranging from 0.16 to 1.043 (Åqvist *et al*., 1994; Åqvist and Mowbray, 1995; Paulsen and Ornstein, 1996; J.Wang *et al*., 1999; W.Wang *et al*., 1999). We have estimated the value of the coefficient from the data of the natural ligand Galβ(1–4)Glc and obtained a value of 0.81. This value was close to the value 0.87 obtained for binding of biotin to avidin (J.Wang *et al*., 1999; W.Wang *et al*., 1999). The value of the coefficient obtained by Åqvist and Mowbray (Åqvist and Mowbray, 1995) for galactose/glucose binding to the galactose/glucose receptor was 0.16. It is seen from Table I that the calculated free energies (using $\beta = 0.81$) are in reasonable agreement with the experimentally observed binding free energies for the disaccharides. The use of a uniform value of β works well with the disaccharides, however, when the ligand is of larger size (i.e. a tetrasaccharide), the use of the same β value, i.e. 0.81, leads to a binding free energy of –0.66 kcal/mol which is considerably different from the observed value (Table I). A value of 1.15 for the coefficient brings the calculated free energy close to the experimental one (Table I). A similar high value of β (1.043) was also used by Paulsen and Ornstein (Paulsen and Ornstein, 1996). The difference in β values for the di- and tetrasaccharides indicates that the exact value of the coefficient might be dependent on the nature of the ligands. This agrees with the observation of W.Wang *et al*. (W.Wang *et al*., 1999). In a recent article while discussing the factors determining the van der Waals coefficient, W.Wang *et al*. suggested that the optimal value of the coefficient depends on the nature of the binding sites for different protein–ligand interactions and may differ from case to case. Our results also suggest that the β values might be case dependent.

The structural agreement between the average MD structures

and the experimental one is very good. The root-mean square deviation of the averaged MD coordinates with respect to the X-ray coordinates of the galectin-3–*N*-acetyllactosamine complex is between 0.63 to 0.76 Å for all the heavy atoms of the protein within 10 Å sphere of the ligand which were kept mobile during the MD simulations. The time-averaged structures of the lectin–ligand complexes obtained from the simulations were examined to understand the atomic details of the binding specificity. Figure 1a–f show the binding site of galectin-3 with the different ligands. The hydrogen bonded partners observed in the MD average conformations have been compared with those observed in the crystal structure of the galectin-3–*N*-acetyllactosamine complex and are given in Table II. The comparison shows that most of the hydrogen bonds are retained during the MD simulations. Figure 1a shows the MD average structure of the galectin-3–*N*-acetyllactosamine complex. It has been seen that the binding is mostly through the hydroxyl groups at the 4, 6 positions and the O5 atom of the galactose residue and the hydroxyl group at position 3 of the GlcNAc residue. The protein residues, which are participating in direct hydrogen bonding, are His158, Arg162, Asn174 and Glu184. In addition to that several water molecules are



also participating in the hydrogen bonding network. Although crystallographic water molecules were not included in the simulations the reported water-mediated hydrogen bonds with O2, O6 of Gal and O6 of Glc/GlcNAc have been reproduced (Table II). In addition, several water-mediated hydrogen bonds have been observed in the solution simulations, which were not observed crystallographically.

It is interesting to note that galectin can bind both Galβ-(1–4) as well as Galβ(1–3) linkages (Sparrow *et al*., 1987). Crystallographic data is available for the Galβ(1–4) linkage only. From the MD simulations hydrogen bonds are reported here. The hydrogen bonding patterns show that the hydrogen bond between the O3 atom of the GlcNAc/GalNAc residue and lectin in the case of the β(1–3) linkage is lost and instead the hydroxyl groups at either the 4 or 6 position form hydrogen bonds with the lectin. We have superimposed the bound conformation of the four galactosides [two containing β(1–4) and two β(1–3) linkages] and have found that the overall three-dimensional envelopes of the sugars are very similar (Figure 2). The position of the CH_2OH group at the C6 position of the Glc/GlcNAc residue in the β(1–4) linkage has been taken up by the *N*-acetyl group at the C2 of the terminal GalNAc/GlcNAc in the case of the β(1–3) linkage. This indicates that the stereochemical requirement at the position C6/C2 of the terminal sugar residue is less stringent than that for the sugar residue non-reducing end. Sparrow *et al*. (Sparrow *et al*., 1987) reported that changes or substitutions at position 4 and 6 of Gal or position 3 of Glc lead to reduced binding affinity, whereas changes at position 2 and 3 of Gal and 1 and 6 of Glc are relatively harmless. In the case of the tetrasaccharide Galβ(1–3)GlcNAcβ(1–3)Galβ(1–4)Glc, the galactose that is located at the primary binding site (the third residue from the non-reducing end) has a substitute at position 3. Encouragingly, the hydrogen bonding scheme shows that all the hydrogen bonds reported between Gal and the lectin in the crystal structure are retained for the tetrasaccharide, indicating that substitutions at position 3 are less harmful. For the tetrasaccharide, apart from the Galβ(1–4)Glc moiety the preceding *N*-acetylglucosamine moiety is also hydrogen bonded to the protein (Figure 1f). This is consistent with the familiar view (Lis and Sharon, 1998) that lectins bind to larger oligosaccharides by extended binding sites. We have also observed that when the galactose is replaced by glucose, i.e. when the axial OH group at position 4 is replaced by an equatorial OH group [as in the case of Glcβ(1–4)Glc)], the hydrogen bonding pattern is quite different. In this case the hydroxyl groups at positions 4, 5 and 6 of the non-reducing glucose moiety form hydrogen bonds with Arg144, Trp181 and Asn174 residues respectively, whereas the O2 and O3 of glucose at the reducing end form hydrogen bonds with the residues Arg186 and Arg162, Glu184 respectively (Figure 1e).

Comparison of the calculated binding free energies with the experimental ones show that the calculations are reasonably successful in reproducing the small observed experimental range, despite the fact that the individual interaction energies are quite large. The correct order of the binding affinities is also reproduced in most of the cases. The specificity of galectins for the galactose-containing disaccharides (β-galactosides) is remarkable, however it is observed that replacement of galactose by glucose at the non-reducing end of the disaccharide leads to a drastic reduction in binding by the lectin (Sparrow *et al*., 1987). From the free energy calculations we also obtained the highest ΔG_{calc} value for the disaccharide Glcβ-(1–4)Glc. In the free form it is observed that Glcβ(1–4)Glc has the most favourable electrostatic interaction with the solvent molecule and binding of lectin to this particular ligand

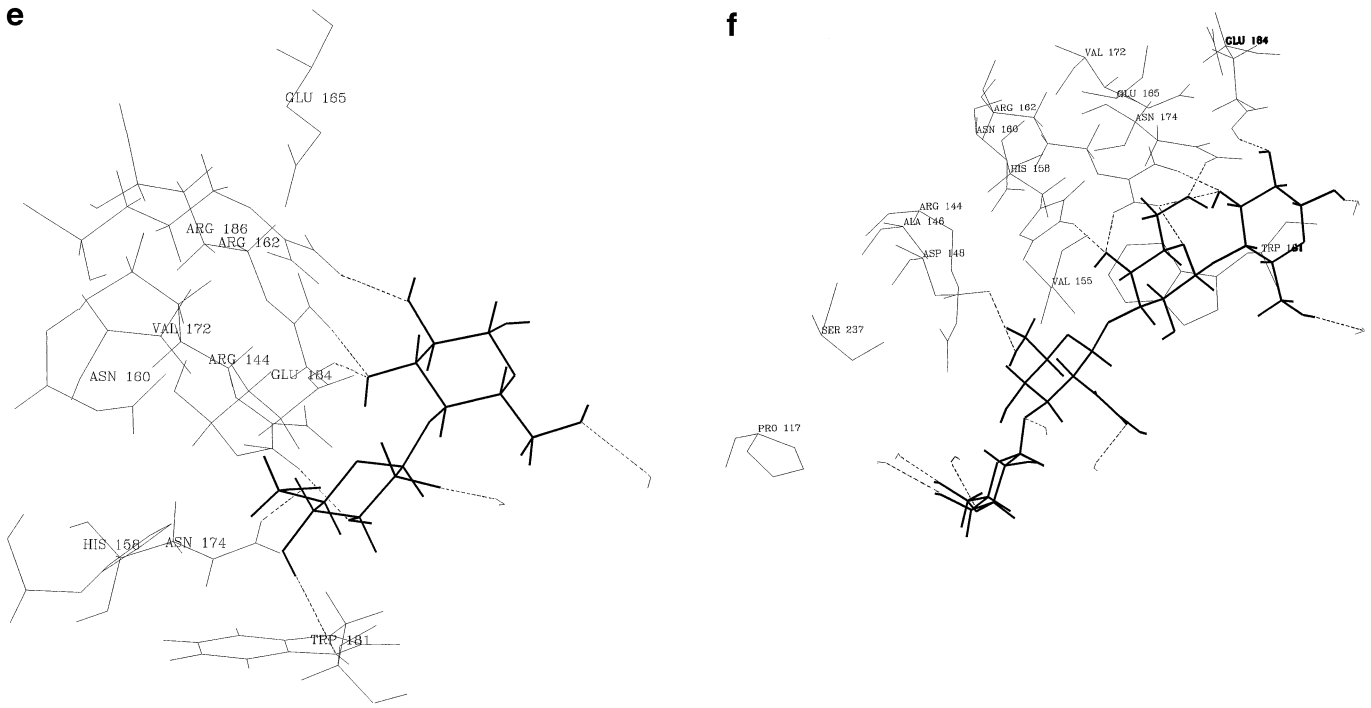


Fig. 1. Average MD structure of the binding site of galectin-3 containing (**a**) Galβ(1–4)GlcNAc, (**b**) Galβ(1–3)GlcNAc, (**c**) Galβ(1–4)Glc, (**d**) Galβ(1–3)GalNAc, (**e**) Glcβ(1–4)Glc, (**f**) Galβ(1–3)GlcNAcβ(1–3)Galβ(1–4)Glc as ligands (thick lines). The protein residues within a 5 Å sphere of the ligand are shown (thin lines). The hydrogen bonds are indicated by dotted lines. The water molecules, which are forming hydrogen bonds with the ligand, are shown. Hydrogen atoms of protein residues, which are not hydrogen bonded, are omitted for clarity.

Table II. Hydrogen bonding pattern in the galectin-3–ligand complexes

Ligand	Residue	O1	O2[a]	O3	O4	O5	O6
Galβ(1–4)GlcNAc (Crystal structure)	Gal				His158, Arg162	Arg162	Asn174, Glu184, Water
	GlcNAc		Water	Arg162, Glu184			Water
Galβ(1–4)GlcNAc	Gal	Arg162, Water	Water	Water	His158	Arg162	Asn174, Water
	GlcNAc	Water	Water	Arg162 Glu184, Water		Water	Water
Galβ(1–3)GlcNAc	Gal			Water	His158, Arg162	Arg162	Asn174
	GlcNAc	Water	Water		Arg162		Water
Gal(1–4)Glc	Gal				His158	Arg162	Asn174
	Glc		Water	Glu184			Water
Gal(1–3)GalNAc	Gal				His158		Asn174, Water
	GalNAc				Water	Water	Glu184
Glc(1–4)Glc	Glc		Water	Arg144	Trp181		Asn174
	Glc		Arg186	Arg162 Glu184			Water
Oligosaccharide Gal(1–4)Glc | (3–1)GlcNAc | (3–1)Gal	Gal			Water	His158, Arg162	Arg162	Asn174
	Glc	Water	Glu184	Arg162			Water
	GlcNAc		Water				Asp148
	Gal			Water	Water		Water

[a] For *N*-acetylated residues instead of O2, NHC=OCH_3 group is present at position 2.

Fig. 2. Superimposed MD average structures of four disaccharides [Galβ(1–4)GlcNAc, Galβ(1–3)GlcNAc, Galβ(1–4)Glc and Galβ(1–3)GalNAc] in the bound form. The first galactose residue has been used for superposition.

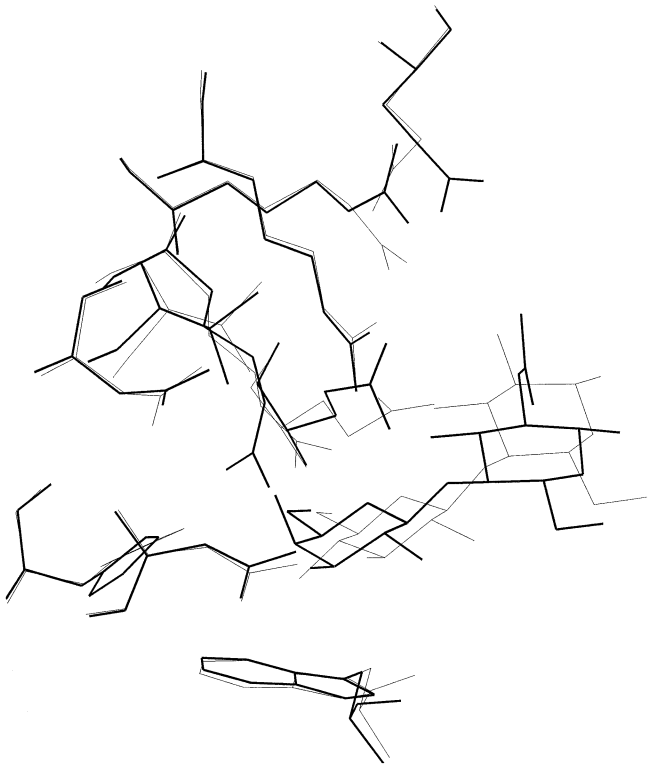


Fig. 3. Superposition of the MD average structure of the galectin-3 complexed with (**a**) Galβ(1–4)GlcNAc (thick lines) and (**b**) Glcβ(1–4)Glc (thin lines). For superimposition of the two complexes, protein atoms were used. Hydrogen atoms have been omitted for clarity.

leads to a large reduction in electrostatic contribution which is not compensated by the small favourable van der Waals component. On the other hand, the Galβ(1–4)GlcNAc ligand with higher experimentally observed binding affinity has a smaller unfavourable electrostatic component and a larger favourable van der Waals interaction. Superimposition of the MD average conformations of the binding sites containing these two ligands is shown in Figure 3. It is seen that most of the binding site residues have a similar conformation in the presence of either of the ligands, except that the three Arg residues (144, 162 and 186) have different side-chain conformations. It is also seen that compared to Galβ(1–4)GlcNAc, the weaker binding ligand Glcβ(1–4)Glc is shifted away from the binding site. The total number of protein atoms coming within a sphere of 4 Å radius drawn around the primary binding Gal/Glc residue of these ligands shows that the number of protein atoms within that sphere is 57 for Galβ(1–4)GlcNAc and 47

Table III. Solvent accessible surface area calculations

Ligand	Ligand–protein complex (MD avg) (Å^2) [Y]	Ligand in water (MD avg) (Å^2) [Z]	SAS difference (Å^2) [X] + [Z] – [Y]
Glc(1–4)Glc	6302	282	31
Gal(1–3)GalNAc	6251	314	114
Gal(1–4)Glc	6231	273	93
Gal(1–3)GlcNAc	6230	300	121
Gal(1–4)GlcNAc	6255	333	129
Tetrasaccharide	6419	541	173

SAS of crystal structure of galectin-3 = 6051 Å^2 [X].

for Glcβ(1–4)Glc. Thus reducing the favourable van der Waals component for weaker binding Glcβ(1–4)Glc compared to the stronger binding Galβ(1–4)GlcNAc. This emphasizes the importance of non-polar interactions in the recognition process of carbohydrates by lectins as has been suggested before (for review see Elgavish and Shaanan, 1997).

In a recent report on calculation of binding free energies for MHC class I protein–peptide interactions using the continuum method, Froloff *et al*. (Froloff *et al*., 1997) used the solvent accessible surface (SAS) as a measure of the non-polar contribution to the binding free energy. They used this term in addition to the electrostatic and loss of entropic (for backbone and side-chain torsional freedom as well as for translational and rotational freedom) terms. When a protein binds to a ligand, the SAS of the complex becomes less than the sum of the individual SAS of the protein and ligand before binding. To check how the change in SAS upon complexation correlates with the binding activity, the solvent accessible surface area of the protein in free and bound forms was calculated. We have used Connolly's method as implemented in the Insight II package for the calculations using a 1.4 Å radius for the water molecule. The surface area of the protein was calculated from the crystal structure (A3K) after including hydrogens and minimizing their positions, and this value was used as the average SAS of the protein. For the ligand, the average MD conformation from the solvent simulation (i.e. in the free state) was used for the SAS calculation. For the complexes the MD average conformations were used. The difference in SAS for various protein–ligand complexes is reported in Table III. It is encouraging to observe that the reduction in SAS upon binding follows the same trend as the binding affinities of the individual ligands, i.e. reduction in SAS is more in the case of a ligand with higher binding affinity. In a recent publication W.Wang *et al*. (W.Wang *et al*., 1999) have shown that the weighted non-polar desolvation ratio, which is derived from SAS is correlated well with the van der Waals coefficient. Our results indicate that tighter binding to ligands will lead to more reduction in the solvent accessible surface area of the protein.

Conclusion

The present study involves the calculation of absolute binding free energies for one tetra- and five disaccharides to galectin-3 using the LIE approximation. The linear response method used here was able to reproduce the experimental data reasonably well. This method is not CPU intensive like free energy perturbation and thermodynamic integration methods. In addition, no ligand mutation is necessary during the simulations and this allows the use of a variety of different ligands. One limitation of this method is the choice of the van der Waals coefficient, for which system dependency has been observed. We have found that this method is able to describe the binding affinity of the lectin satisfactorily. Using this approach it is also possible to suggest modes of binding for a ligand for which crystallographic data is not available.

The results reported here suggest that both hydrogen bonding and hydrophobic interactions are important for determining the affinity of the carbohydrates towards the lectin. While the aqueous environment might favour the solvation of the carbohydrates electrostatically, the binding to protein is favoured by van der Waals interaction. The polar residues of the protein, e.g. Arg, His, Glu, Asn, etc. provide the necessary hydrogen bonding partners and residues like Trp, CH_2 groups of Arg side-chains, etc. can provide the van der Waals neighbours. The results also indicate that the water molecules can compete with protein side-chains to form hydrogen bonds with the ligands. It is interesting to see that in the case of a weakly binding ligand [Glcβ(1–4)Glc], the hydrogen bonding partners, though different from a stronger ligand, are quite high in number, whereas the van der Waals counterparts are fewer. We suggest this indicates that contrary to the popular belief that the sugar binding specificity of the lectins is mostly determined by the number and/or pattern of the hydrogen bonding network, the non-polar interactions are equally crucial. We have also found that better binding ligands lead to a larger burial of solvent accessible surface area of the protein. The intimate details of the recognition process thus obtained can be exploited further for a rational design of mimetics in glycosciences.

Acknowledgements

The work is funded by the Centre for Scientific and Industrial Research, Government of India [No. 01(1500)/98/EMR-II]. The use of the computational facility of the Computer Centre, University of Calcutta is gratefully acknowledged.

References

Åqvist,J. and Mowbray,S.L. (1995) *J. Biol. Chem*., **270**, 9978–9981.
Åqvist,J., Medina,C. and Samuelsson,J.E. (1994) *Protein Eng.*, **7**, 385–391.
Barondes,S.H., Cooper,D.N., Gitt,M.A. and Leffler,H. (1994) *J. Biol. Chem.*, **269**, 20807–20810.
Bresalier,R.S., Mazurek,N., Sternberg,L.R., Byrd,J.C., Yunker,C.K., Nangia-Makker,P. and Raz,A. (1998) *Gastroenterology*, **115**, 287–296.
Cameron,A.A., Dougherty,P.M., Garrison,C.J., Wills,W.D. and Carlton,S.M. (1993) *Brain Res*., **620**, 64–71.
Cooper,D.N. and Barondes,S.H. (1999) *Glycobiology*, **9**, 979–984 and references therein.
Dagher,S.F., Wang,J.L. and Patterson,R.J. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 1213–1217.
Elgavish,S. and Shaanan,B. (1997) *Trends Biochem. Sci*., **22**, 462–467.
Fernandez,P.L., Merino,M.J., Gomez,M., Campo,E., Medina,T., Castronovo,V., Sanjuan,X., Cardesa,A., Liu,F.-T. and Sobel,M.E. (1997) *J. Pathol.*, **181**, 80–86.
Foddy,L., Stamatoglou,S.C. and Haughes,R.C. (1990) *J. Cell Sci*., **97**, 139–148.
Frigeri,L.G., Zuberi,R.I. and Liu,F.T. (1993) *Biochemistry*, **32**, 7644–7649.
Froloff,N., Windemuth,A. and Honig,B. (1997) *Protein Sci*., **6**, 1293–1301.
Hsu,D.K., Hammes,S.R., Kuwabara,I., Greene,W.C. and Liu,F.T. (1996) *Am. J. Pathol*., **148**, 1661–1670.
Hsu,D.K., Dowling,C.A., Jeng,K.C.G., Chen,J.T., Yang,R.Y. and Liu,F.T. (1999) *Int. J. Cancer*, **81**, 519–526.
Kollman,P.A. (1993) *Chem. Rev.*, **93**, 2395–2417.
Konstantinov,K.N., Robbins,B.A. and Liu,F.T. (1996) *Am. J. Pathol.*, **148**, 25–30.
Lindstedt,R., Apodaca,G., Barondes,S.H., Mostov,K. and Leffler, H. (1993) *J. Biol. Chem.*, **268**, 11750–11757.
Lis,H. and Sharon,N. (1998) *Chem. Rev*., **98**, 637–674.
Liu,F.T. (1993) *Immunol. Today*, **14**, 486–490.

Lotz,M.M., Andrews,C.W.Jr, Korzelius,C.A., Lee,E.C., Steele,G.D.Jr, Clarke,A. and Mercurio,A.M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3466–3470.
Miyamoto,S. and Kollman,P.A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8402–8406.
Nangia-Makker,P., Honjo,Y., Sarvis,R., Akahani,S., Hogan,V., Pienta,K.J. and Raz, A. (2000) *Am. J. Pathol.*, **156**, 899–909.
Paulsen,M.D. and Ornstein,R.L. (1996) *Protein Eng.*, **9**, 567–571.
Perillo,N.L., Marcus,M.E. and Baum,L.G. (1998) *J. Mol. Med.*, **76**, 402–412.
Pesheva,P., Kuklinski,S., Schmitz,B. and Probstmeier,R. (1998) *J. Neurosci. Res.*, **54**, 639–654.
Raz,A., Zhu,D., Hogan,V., Shah,N., Raz,T., Karkash,R., Pazerini,G. and Carmi,P. (1990) *Int. J. Cancer*, **46**, 871–877.
Regan,L.J., Dodd,J., Barondes,S.H. and Jessell,T.M. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 2248–2252.
Sano,H., Hsu,D.K., Yu,L., Apgar,J.R., Kuwabara,I., Yamanaka,T., Hirashima,M. and Liu,F.T. (2000) *J. Immunol.*, **165**, 2156–2164.
Sato,S. and Hughes,R.C. (1994) *J. Biol. Chem.*, **269**, 4424–4430.
Seetharaman,J., Kanigsberg,A., Slaaby,R., Leffler,H., Barondes,S.H. and Rini,J.M. (1998) *J. Biol. Chem.*, **273**, 13047–13052.
Sparrow,C.P., Leffler,H. and Barondes,S.H. (1987) *J. Biol. Chem.*, **262**, 7383–7390.
Still,W.C., Tempcyk,A., Hawley,R.C. and Hendrickson,T. (1990) *J. Am. Chem. Soc.*, **112**, 6127–6129.
Wang,J., Dixon,R. and Kollman,P.A. (1999) *Proteins*, **34**, 69–81.
Wang,W., Wang,J. and Kollman,P.A. (1999) *Proteins*, **34**, 395–402.
Weis,W.I. and Drickamer,K. (1996) *Annu. Rev. Biochem.*, **65**, 441–473.
Xu,X.C., El-Naggar,A.K. and Lotan,R. (1995) *Am. J. Pathol.*, **147**, 815–822.
Yang,R.Y., Hsu,D.K., Yu,L., Ni,J. and Liu,F.T. (2001) *J. Biol. Chem.*, **276**, 20252–20260.

Received March 6, 2002; revised September 24, 2002; accepted October 10, 2002