Carbohydrate-Protein Recognition: Molecular Dynamics Simulations and Free Energy Analysis of Oligosaccharide Binding to Concanavalin A

R. A. Bryce,* I. H. Hillier,† and J. H. Naismith‡
*School of Pharmacy and Pharmaceutical Sciences and †Department of Chemistry, University of Manchester, Manchester M13 9PL and ‡Centre for Biomolecular Sciences, The University, St. Andrews KY13 BST, United Kingdom

ABSTRACT Carbohydrate ligands are important mediators of biomolecular recognition. Microcalorimetry has found the complex-type N-linked glycan core pentasaccharide β-GlcNAc-(1→2)-α-Man-(1→3)-[β-GlcNAc-(1→2)-α-Man-(1→6)]-Man to bind to the lectin, Concanavalin A, with almost the same affinity as the trimannoside, Man-α-(1→6)-[Man-α-(1→3)]-Man. Recent determination of the structure of the pentasaccharide complex found a glycosidic linkage ψ torsion angle to be distorted by 50° from the NMR solution value and perturbation of some key mannose–protein interactions observed in the structures of the mono- and trimannoside complexes. To unravel the free energy contributions to binding and to determine the structural basis for this degeneracy, we present the results of a series of nanosecond molecular dynamics simulations, coupled to analysis via the recently developed MM-GB/SA approach (Srinivasan et al., *J. Am. Chem. Soc.* 1998, 120:9401–9409). These calculations indicate that the strength of key mannose–protein interactions at the monosaccharide site is preserved in both the oligosaccharides. Although distortion of the pentasaccharide is significant, the principal factor in reduced binding is incomplete offset of ligand and protein desolvation due to poorly matched polar interactions. This analysis implies that, although Concanavalin A tolerates the additional 6 arm GlcNAc present in the pentasaccharide, it does not serve as a key recognition determinant.

INTRODUCTION

The interaction of complex sugars with proteins dictates a broad spectrum of physiological processes, ranging from leukocyte adhesion and tumor metastasis to host-pathogen recognition (Springer and Lasky, 1992; Sharon and Lis, 1998). To explore the associated possibilities for therapeutic intervention, it is therefore imperative to understand the salient factors governing carbohydrate-protein recognition. Recent advances in organic synthesis, high resolution spectroscopy, and microtitration calorimetry have combined to yield valuable insights into the structural and energetic principles underlying these important receptor–ligand interactions, with particularly extensive work on the archetypal carbohydrate-binding protein, the legume lectin concanavalin A (Con A). Con A is a plant storage and defense protein (Peumans and van Damme, 1995), which can be isolated from the legume, jack bean *Canavalia ensiformis*. The natural substrates of the lectin are thought to be high-order N-linked glycans, constituent components of many cell membrane proteins. A number of the components of these complex sugars are illustrated in Fig. 1. Glycans are linked to the nitrogen of an asparagyl residue and are usually branched, as illustrated by the biantennary fucosylated saccharide of bovine immunoglobulin G, labeled sugar V in Fig. 1 (Fujii et al., 1990). Complex-type N-linked glycans also share a common pentasaccharide sequence (sugar IV). Molecular-level details of Con A in the native saccharide-free state (Weisgerber and Helliwell, 1993) and bound to a variety of saccharide ligands have been determined by x-ray crystallography (Naismith et al., 1994; Naismith and Field, 1996; Moothoo and Naismith, 1998; Dimick et al., 1999). The overall topology of saccharide-free Con A is found to be an α_2 homodimer above pH 7 and an α_4 tetramer below pH 7. Each subunit is a 26-kDa monomer of 237 residues,

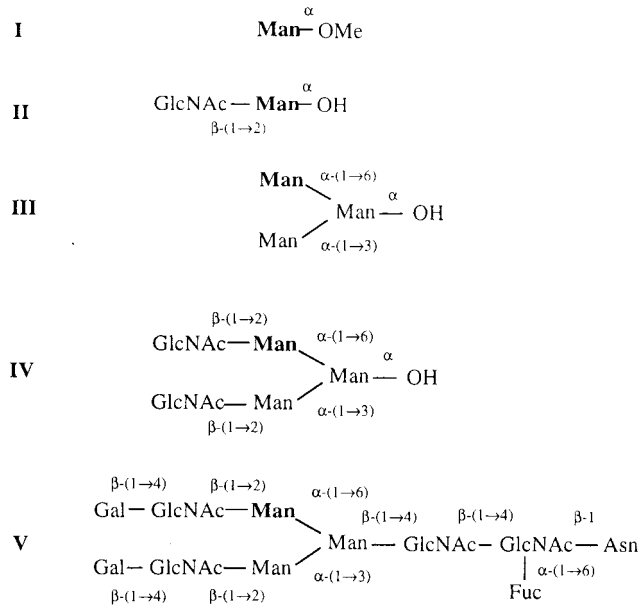

FIGURE 1 Substrates recognized by Con A: mannoside (I), disaccharide (II), trimannoside (III), pentasaccharide (IV) and a fucosylated biantennary glycan (V).

Received for publication 6 March 2001 and in final form 16 May 2001.

Address reprint requests to Richard Bryce, School of Pharmacy & Pharmaceutical Sciences, University of Manchester, Oxford Rd., Manchester M13 9PL, UK. Tel.: 44-161-275-2401; Fax: 44-161-275-2481; E-mail: r.a.bryce@man.ac.uk.

0006-3495/01/09/1373/16 $2.00

TABLE 1 Thermodynamic energies of binding from microtitration calorimetry (kcal/mol)

Model	Ligand	ΔG_{bind}	ΔH_{bind}	$-T\Delta S_{bind}$
I	α-MeO**Man**	−5.3*	−8.2	2.9
		−5.3†	−6.6	1.3
		−5.2‡	−6.8	1.6
II	β-GlcNAc-(1→2)-α-**Man**	−5.2*	−5.3	0.1
III	**Man**-α-(1→6)-[Man-α-(1→3)]-Man	−7.5*	−14.1	6.6
IV	β-GlcNAc-(1→2)-α-Man-(1→3)-[β-GlcNAc-(1→2)-α-**Man**-(1→6)]-Man	−8.4*	−10.6	2.2

Common mannose that occupies the monosaccharide binding site is in bold.
*Mandal et al. (1994)
†Williams et al. (1992)
‡Schwartz et al. (1993)

forming a β-sandwich of two β-sheets. The amino acid sidechains of the lectin subunit are preorganized into a shallow binding site, partly through interactions with the two proximal divalent cations, Ca^{2+} and Mn^{2+}. With the modularity of one binding site per subunit, the multimeric arrangement of Con A permits recognition of multiple sugars with enhanced avidity, a cooperative phenomenon known as multivalency (Weis and Drickamer, 1996).

The first Con A complex to be crystallographically determined was a tetramer of Con A with methyl-α-D-mannopyranoside (mannoside or α-MeOMan) at each of the four binding sites (Derewenda et al., 1989; Naismith et al., 1994). The overall fold was found to be identical to the native structure, with a considerable number of direct contacts to the protein forged by α-MeOMan. Subsequent determination of Con A in complex with the disaccharides, Man-α-(1→6)-Man-α-OMe and Man-α-(1→3)-Man-α-OMe (Bouckaert et al., 1999), found the O-1-linked mannose to occupy the monosaccharide binding site in an analogous fashion to α-MeOMan. Similarly, the crystal structure of a trisaccharide, Man-α-(1→6)-[Man-α-(1→3)]-Man (Naismith and Field, 1996), found the 1→6 terminal mannose to occupy the monosaccharide site. The central reducing sugar in the core and the 1→3 terminal mannose occupied an adjacent extended region of the binding groove, also making direct polar and apolar contacts with the protein, albeit fewer in number. The O_2 of the reducing sugar was observed to participate in a hydrogen bond with a structural water. Interestingly, an alternative binding mode for the trimannoside to one of the four Con A subunits was observed from a separate crystallographic study (Loris et al., 1996). In this structure, the α-(1→3)-linked mannose is rotated with respect to the other saccharide units, but interacts with the same amino acid residues as before. This duality in binding mode has also been observed for the Man-α-(1→2)-Man-α-OMe/Con A crystal structure (Moothoo et al., 1999) and clearly highlight the subtle balance of forces governing lectin–carbohydrate interactions.

The interpretation of this structural information has been greatly enhanced by thermodynamic parameters of association from microcalorimetry (Table 1). Although there is variation in measured entropies and enthalpies of binding, the free energies converge to a value of ~ -5.2 kcal/mol for α-MeOMan, corresponding to a modest affinity constant of $0.82 \times 10^5\ M^{-1}$. As expected from the greater number of interactions with the lectin, the trimannosyl ligand binds with a higher affinity, in fact almost 50-fold greater than the mannoside, consistent with the notion of the three mannose residues as an essential recognition determinant in the glycan core. Thermodynamic measurements have additionally been obtained for the β-GlcNAc-(1→2)-α-Man-(1→3)-[β-GlcNAc-(1→2)-α-Man-(1→6)]-Man pentasaccharide, a carbohydrate fragment common to complex-type N-linked glycan core regions. Given the increased magnitude in binding free energy of 2 kcal/mol of the trimannoside over the mannoside, it is perhaps of some surprise that the pentasaccharide binds with only a slightly higher affinity of 0.9 kcal/mol relative to the trimannoside (Table 1). In an effort to understand the structural basis of the lectin specificity, the structure of the Con A/pentasaccharide complex was recently reported at a resolution of 2.7 Å (Moothoo and Naismith, 1998). It was found that the 1→6 mannose occupying the monosaccharide binding site was slightly displaced in orientation relative to the mannoside and trimannoside conformations, with a loss in the number of van der Waals but not polar contacts (Fig. 2). This was thought to arise from an unfavorable orientation of the GlcNAc six arm required to alleviate steric hindrance by lectin residues Thr^{226}, Gly^{227} and Arg^{228}, where destabilization lay principally in a distortion of the ψ dihedral by $\sim 50°$ from the observed solution minimum, as determined by NMR studies (Homans, 1995). Ancillary evidence is provided from calorimetric studies of the disaccharide β-GlcNAc-(1→2)-α-Man (sugar II of Table 1), which only liberates 5.2 kcal/mol upon binding to Con A. This is essentially the same affinity as for the mannoside, thereby implying that binding of the GlcNAc unit is energetically neutral.

Although experimentally inaccessible, determination of the free energy components associated with these structural features is feasible through theoretical analysis. Because

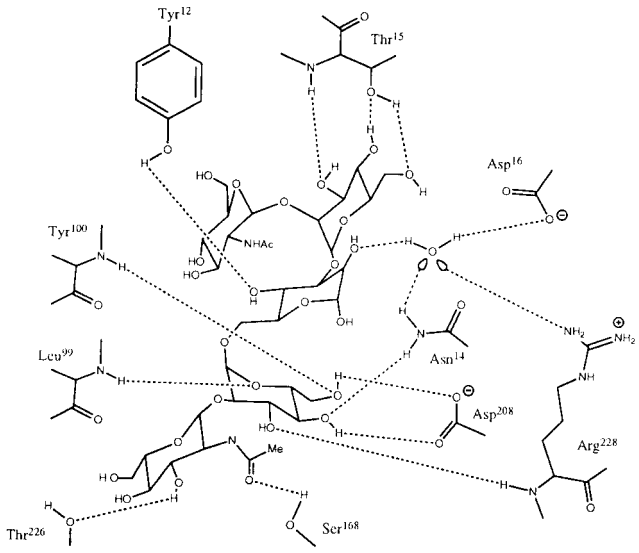

FIGURE 2 Interactions of pentasaccharide bound to Con A.

carbohydrates are dynamic in nature and interact intimately with aqueous solvent, we chose to use a recently proposed computational model that performs free energy analysis on molecular dynamics (MD) trajectories of a receptor–ligand complex and the ligand free in solution. The molecular mechanics—Poisson–Boltzmann/surface area (MM-PB/SA) approach (Srinivasan et al., 1998) has previously proved useful in the study of sequence-dependent solvation of DNA helices (Cheatham et al., 1998), RNA–protein interactions (Reyes and Kollman, 2000), protein folding (Lee et al., 2000), protein–protein interfaces (Massova and Kollman, 1999) and protein–ligand interactions (Kuhn and Kollman, 2000a,b). Similar performance has been afforded by the more computationally expedient generalized Born/surface area (MM-GB/SA) variant of the MM-PB/SA approach (Srinivasan et al., 1998). Therefore, we chose to use the MM-GB/SA model in seeking to capture the important effects of conformational plasticity and aqueous solvent on carbohydrate–protein binding thermodynamics and structure. Consequently, the objective of this study was to explore sugar–protein interactions in the lectin complexes of sugars I, III, and IV, analyzing nanosecond trajectories of the bound and free ligands by the MM-GB/SA method to quantify the possible intrinsic energetic penalty of the distorted glycosidic linkage; the concomitant effect of this orientation on the interactions of the sugar moieties with the lectin; and to consider other possible contributions to the observed free energy of binding, notably the influence of solvent. Comparison with spectroscopic and calorimetric studies also permits assessment of the quality of the MM-GB/SA potential used for dynamics and free energy calculations in its application to carbohydrate–protein recognition.

MATERIALS AND METHODS

Molecular dynamics simulations

Dynamical studies were performed for saccharide ligands I, III, and IV, both bound to the lectin receptor Con A and in aqueous solution.

Sugar–protein complexes

Initial geometries for the sugar–protein complexes were constructed from the crystallographically determined atomic coordinates: for I, PDB entry code 5cna, resolution 2.0 Å (Naismith et al., 1994); for III, PDB entry code 1cvn, resolution 2.3 Å (Naismith and Field, 1996); and for IV, PDB entry code 1tei, resolution 2.7 Å (Moothoo and Naismith, 1998). The complexes crystallize as α_4 tetramers. However, the four binding sites exhibit a near identical shape and binding affinity (Dimick et al., 1999), with an absence of cooperativity allowing the simulation studies to focus on the receptor–ligand interaction of a single subunit (subunit A). Crystallographic waters within 3 Å of the protein were included in the model, leading to 92 waters for I, 62 for III, and 15 for IV, a reflection of the resolution of the respective crystal structures. An analysis of water structure in the crystals of 11 legume lectins by Loris et al. (1994) has highlighted seven conserved hydration sites, with two waters bound to each metal ion (Ca^{2+} and Mn^{2+}), a water associated with a β-bulge, another with an α-hairpin and importantly, a structural water at the binding pocket, which serves as a bridge from Asn^{14}, Asp^{16} and Arg^{228} of Con A to the O_2 oxygen of the reducing mannose. In the pentasaccharide/Con A structure, however, the second water coordinated to Ca^{2+} was absent in subunits A, C, and D, although present in B. Consequently, an appropriate water was inserted at the coordination site; without this water, energy minimization led to gross distortion of the crystallographic geometry. Hydrogens were added to the protein using the *edit* module of AMBER (Case et al., 1999), and to the carbohydrate ligands using MAVIS (Parkinson et al., 1998). A number of hydroxyl rotamers required adjustment to reproduce the crystallographically observed polar contacts in subsequent calculations. Before minimization and molecular dynamics, the three complexes were immersed in further solvent cubes, previously equilibrated, leading to an average total of 7623 water molecules. Three sodium counterions were also added at minima in the protein electrostatic potential to ensure an electroneutral unit cell for calculation of the periodic Coulomb potential.

All minimization and dynamics were conducted using AMBER 6.0 (Case et al., 1999), combined with the Cornell et al. (1995) force field for the protein, the TIP3P water potential (Jorgensen, 1983) and the GLYCAM_93 force field for carbohydrates (Woods et al., 1995). The carbohydrate van der Waals parameters were updated for use with AMBER 6.0, by analogy to Woods et al. (1995) using for GLYCAM_93 atom types OG and AC/EC, the OS and CT values from the Cornell et al. (1995) force field. The GLYCAM_93 atomic point charges were retained as they had been derived from fitting to the HF/6–31G* electrostatic potential (ESP). The efficacy of the GLYCAM potential in describing sugar conformation and energetics has been evidenced by a number of studies (Woods et al., 1998; Bradbrook et al., 1998, 2000). Parameters for the metal cations were taken from the crystallographic and simulation study of Con A complexes by Bradbrook et al. (1998), which were derived to reproduce the coordination geometry about the two ions found in the crystal structures of Con A, taking values of R* and ϵ of (1.79 Å, 0.014 kcal/mol) for Ca^{2+} and (1.69 Å, 0.014 kcal/mol) for Mn^{2+}. Periodic boundary conditions were used in conjunction with a particle-mesh Ewald summation (Essman et al., 1995) to treat long-range electrostatics. To our knowledge, the present study is the first study to provide an accurate treatment of long-range electrostatics in the simulation of protein–carbohydrate systems. Van der Waals interactions were truncated at 9 Å. The SHAKE algorithm (van Gunsteren et al., 1977) was used to constrain all covalent bond lengths involving hydrogen. A time step of 2 fs was used. The equilibration protocol involved rounds of energy refinement and simulated annealing, with progressively decreas-

ing harmonic restraints on the ligand–protein geometries. This was followed by a 200-ps simulation in the isothermal-isobaric ensemble, in the absence of constraints, at 300 K and 1 atm using a heat bath coupling of 1.0 ps (Berendsen et al., 1984) and a pressure coupling of 1.0 ps. Subsequently, the equilibrated trajectories of the system were allowed to evolve within the canonical ensemble for 1.0 ns. Every picosecond, structures were archived for analysis.

Sugars in aqueous solution

Corresponding aqueous solution simulations of ligands I, III, and IV were conducted, using the potential energy function and parameters outlined above. In detailing the construction of representative solution phase conformations of the flexible sugar solutes, geometries of the carbohydrate ligands are subsequently discussed using the crystallographic definition of glycosidic torsions: φ is O5-C1-O1-C*i*, ψ is C1-O1-C*i*-C(*i*-1), and ω is O6-C6-C5-O5. C*i* and C(*i*-1) are aglyconic atoms. For comparison with NMR results, we also use the International Union of Pure and Applied Chemistry definition for ω, which is ω_H as O6-C6-C5-H5. For the methyl mannoside (I), an initial GG rotamer was considered for the primary hydroxymethyl group, conventionally defined by the O5-C5-C6-O6 and C4-C5-C6-O6 dihedral angles. Solution structures for the methyl trimannoside (III) have been previously determined by NMR studies (Brisson and Carver, 1983; Cummings et al., 1986; Carver and Cummings, 1987) The α-(1→6) linkage has been shown by NMR and semi-empirical studies to exhibit considerable flexibility (Homans et al., 1987): two conformers were found to be populated at room temperature, varying in the value of the α-(1→6) interresidue dihedral angle ω_H: conformers with ω_H of $-60°$ (*g*) and 180° (*t*) were observed experimentally, with approximately equal populations of both (Brisson and Carver, 1983). Terminal hydroxymethyl orientations were assigned the GG conformation, as found in both NMR and crystallographic studies (Imberty et al., 1991). Because interconversion between the two conformers was expected to be slow (Brisson and Carver, 1983), both conformers were modeled, denoted III/*g* ($\omega_H = -60°$) and III/*t* ($\omega_H = 180°$). Similarly, an NMR geometry of the pentasaccharide (IV) was made available (Homans, 1995), with the CH_2OH groups in the GG conformation and a *trans* orientation of ω. Alternative conformers of IV were not considered. As before, hydrogens were added to the protein using the *edit* module of AMBER (Case et al., 1999), and to the carbohydrate ligands using MAVIS (Parkinson et al., 1998), selecting appropriate hydrogen-bonding hydroxyl orientations. Addition of pre-equilibrated solvent to the sugars led to an average total number of water molecules of 506. Energy refinement was followed by simulated annealing and 200 ps of equilibration via NPT dynamics. One nanosecond of data was acquired at 300 K in the canonical ensemble, with similar simulation details to the protein–sugar complex studies. Every picosecond, structures were archived for analysis.

Molecular mechanics and generalized Born/surface area calculations

Energetic analysis of equidistant points in phase space along the dynamical trajectories was performed using the MM-GB/SA method of Srinivasan et al. (1998). This hybrid approach calculates a gas-phase contribution to binding using an all-atom force field and incorporates the influence of solvent via the Generalized Born continuum solvent method (Still et al., 1990; Jayaram et al., 1998). For computational expediency, an energy difference between bound and free protein and ligand is calculated at points along the bound and aqueous trajectories (in this case, 100 snapshots due to sampling at 10-ps intervals). The implicit assumption here is that the ensemble generated by the molecular mechanics potential used in the MD simulations overlaps substantially with the distribution of configurations that would be generated by a hybrid MM-GB/SA potential. According to this method, we may calculate binding affinity of ligand L to receptor R in aqueous solution, $\Delta G_{bind}^{RL(aq)}$, as (Massova and Kollman, 1999)

$$\Delta G_{bind}^{RL(aq)} = (\Delta H_{bind}^{RL(g)} - T\Delta S_{bind}^{RL(g)}) + \Delta\Delta\bar{G}_{solv}^{RL}, \quad (1)$$

where $\Delta H_{bind}^{RL(g)}$ and $\Delta S_{bind}^{RL(g)}$ is the enthalpy and entropy of R-L binding respectively, without solvent contributions, and T is temperature. $\Delta\Delta\bar{G}_{solv}^{RL}$ is the ensemble-averaged difference in solvation free energy between R and L, bound and free, calculated by the GB/SA continuum model. The enthalpy of binding, $\Delta H_{bind}^{RL(g)}$, can be approximated by the ensemble-averaged differences, denoted by a bar in Eq. 2, in bond-angle-dihedral, electrostatic and van der Waals potential energies upon binding,

$$\Delta H_{bind}^{RL(g)} = \Delta\bar{E}_{bad}^{RL(g)} + \Delta\bar{E}_{elec}^{RL(g)} + \Delta\bar{E}_{vdw}^{RL(g)}. \quad (2)$$

Calculation of the solute entropy of binding ($\Delta S_{bind}^{RL(g)}$ in Eq. 1) is problematic due to integration over the large number of receptor and ligand degrees of freedom. Approximate approaches have used normal mode calculations on a small number of structures. However, it has been noted that these harmonic calculations, using gas-phase or solvent reaction field minimized structures, may lead to inaccurate estimates of solute entropy (Srinivasan et al., 1998). An alternative approach is to directly extract the solute entropy from a complete molecular dynamics trajectory (Karplus and Kushik, 1981). Here, the quasi-harmonic configurational entropy is related to the covariance matrix for fluctuations in degrees of freedom,

$$S = \tfrac{1}{2}\, nk_B + \tfrac{1}{2}\, k_B \ln[(2\pi)^n \boldsymbol{\sigma}], \quad (3)$$

where n is the number of degrees of freedom, k_B is the Boltzmann constant and $\boldsymbol{\sigma}$ is the determinant of the covariance matrix evaluated over a given trajectory. Due to the large size of the protein, we estimate the quasi-harmonic entropy of binding of the carbohydrate to the protein using only the covariance of the ligand torsional degrees of freedom, i.e., set $\Delta S_{bind}^{RL(g)} = \Delta S_{bind}^{L(g)}$ and n to be the number of rotatable bonds in the ligand. The carbohydrate torsions are expected to exhibit the most significant restriction on binding of all the protein and ligand geometric degrees of freedom. For comparison, we also present $\Delta S_{bind}^{RL(g)}$, calculated within the harmonic approximation using representative trajectory snapshots. After Kuhn and Kollman (2000b), we calculate the entropy of binding for the ligand from solution to a subregion of protein within 8 Å of the bound ligand, using the *nmode* module of AMBER. To include the effect of bulk solvent, minimizations were performed using a distance-dependent dielectric of 4r. We note that the solvent entropy contribution to binding is implicitly included within the theoretical framework of the GB/SA model.

Ensemble-averaged absolute solvation free energies ($\Delta\bar{G}_{solv}^{RL}$) were used to calculate the relative binding energy via Eq. 1. At each timestep, ΔG_{solv}^{RL} was determined using the pairwise additive generalized Born (GB) model (Still et al., 1990; Hawkins et al., 1996), parameterized to be consistent with AMBER (Jayaram et al., 1998). ΔG_{solv}^{RL} is composed of Coulombic (ΔG_{pol}^{RL}) and nonpolar (ΔG_{np}^{RL}) components, which may be further resolved into contributions for each atom i of an N-atom solute:

$$\Delta G_{solv}^{RL} = \Delta G_{pol}^{RL} + \Delta G_{np}^{RL} = \sum_{i=1}^{N} (\Delta G_{pol,i}^{RL} + \gamma A_i). \quad (4)$$

Here, A_i is the solvent-accessible surface area, and γ is a surface tension parameter, which we take to be 0.0072 kcal/(mol Å). Due to the sum over individual solute atoms, it is straightforward to consider contributions to ΔG_{solv}^{RL} from particular chemical groups, in this instance, from specific saccharide or amino acid residues. Nonpolar contributions are calculated using the MSMS program to compute solvent-accessible surface areas (Sanner et al., 1996). PARSE atomic radii (Sitkoff et al., 1994) and Cornell/GLYCAM 93 charges were used to calculate electrostatic solvation free energies with a solvent dielectric ϵ of 80. Because electrostatic solvation parameters were unavailable for Ca^{2+} and Mn^{2+} ions, we adopted the van der Waals radii from Bradbrook et al. (1998). An effective

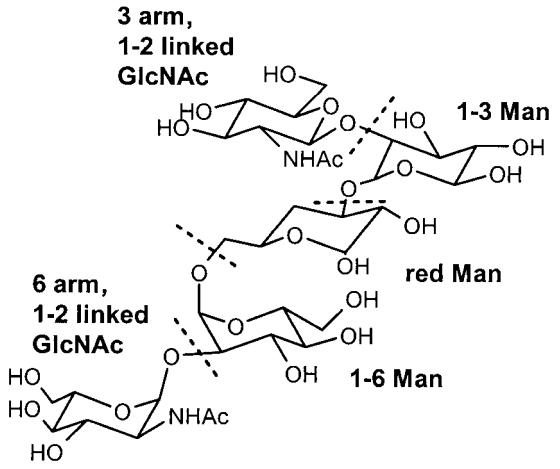

FIGURE 3 Partitioning of five saccharide units of IV for analysis: 6 arm, a GlcNAc unit, β-(1→2)-linked to 1→6 Man, α-(1→6)-linked to red Man; 3 arm, a GlcNAc unit, β-(1→2)-linked to 1→3 Man, α-(1→3)-linked to red Man.

Born radius is then obtained by scaling after the method of Hawkins et al. (1996). Following Hawkins et al. (1996) and Jayaram et al. (1998), we set the ions' screening parameters to unity. We then calibrated the scale factors against the experimental ion solvation free energies (Marcus, 1991). In combination with a small nonpolar contribution, scale factors of 1.028 (Mn^{2+}) and 1.145 (Ca^{2+}) were found to best reproduce the metal ion experimental solvation free energies.

RESULTS

Before energetic postprocessing by the MM-GB/SA approach, we report structural and dynamic aspects of the molecular dynamics simulations of the carbohydrate, bound and free in solution, in relation to available NMR spectroscopic and x-ray data. To facilitate discussion, the saccharide residues that comprise I, III, and IV, are labeled according to the scheme in Fig. 3. The central saccharide unit of pentasaccharide β-GlcNAc-(1→2)-α-Man-(1→3)-[β-GlcNAc-(1→2)-α-Man-(1→6)]-Man is the reducing mannose and thus labeled red Man. This residue is linked at O3 to C1 of a second mannose, which we label the 1→3 Man. In turn, connected to the O_2 of this 1→3 Man is the terminal GlcNAc arm, which we denote as the 3 arm GlcNAc, referring to which branch from the reducing mannose the β-(1→2)-linked GlcNAc is connected. The central reducing mannose is also linked at O6 to C1 of an adjacent mannose, labeled 1→6 Man. A terminal GlcNAc residue is β-(1→2)-linked to O_2 of this 1→6 Man, and is called the 6 arm GlcNAc. The 1→6 Man residue is common to all three ligands (mono-, tri-, and pentasaccharide). III and IV share the central trimannoside core. Correspondingly, we partition the carbohydrate binding site on Con A into five regions: the 3 arm binding site, the 1→3 Man site, the red Man site, the 1→6 Man (monosaccharide) site, and the 6 arm site. These sites overlap, sharing several protein residues.

Carbohydrate–solution trajectories

We first consider the conformation of the unbound carbohydrate ligands in aqueous solution, discussed in terms of chemical group.

Glycosidic linkages

Ensemble-averaged glycosidic angles are compared with experiment for III and IV (Table 2). For all linkages, there is good agreement between simulation and NMR average values. No interconversion is observed between the *g* and *t* forms of the α-(1→6) linkage between 1→6 Man and red Man in III. This is as expected, given that interconversion occurs slowly in comparison to the tumbling time of the ligand of 150 ± 20 ps measured by NMR (Brisson and Carver, 1983). Interestingly, increased flexibility is found in the α-(1→3) ψ angle of III/*g* relative to the *trans* conformer, possibly as a consequence of its more linear shape, with fewer internal hydrogen bonds. This ψ angle displays considerable flexibility in III and IV, which samples *trans* and *gauche* orientations, reflected in the large standard deviations (Table 2). Motional averaging around the α-(1→3) linkage has been noted from NMR and simulation studies of a number of carbohydrates (Homans, 1995). In general, greater flexibility is observed in the glycosidic linkages of IV with respect to III and I. Multiple rotamer sampling is reflected in the torsional standard deviations (Table 2), particularly for the β-(1→2) linkages of the GlcNAc arms. This increased flexibility at the molecular extremities is reminiscent of the fraying of oligonucleotides observed in aqueous solution. Of particular note is the behavior of the ψ angle of the 6 arm/1→6 Man linkage, which populates the bound rotamer ($\psi = -132°$) to a limited degree (Fig. 4). The distribution of (φ, ψ) angles for the GlcNAc arms indicates a broad spread in ψ relative to φ (Fig. 4). This is expected from stereoelectronic control exerted by the exoanomeric effect on φ of the β-(1→2) linkage absent on ψ (Lemieux, 1971; Juaristi and Cuevas, 1994).

Hydroxymethyl groups

NMR studies indicate approximately equal populations of GT and GG (Nishida et al., 1984; Hori et al., 1990). In these simulations, the carbohydrate exocyclic hydroxymethyl groups remain stable in their initial GT or GG conformation over the timescale of the simulations. This stability has been observed elsewhere (Simmerling et al., 1998) and is a reflection of the sizeable free energy barrier to rotation in solution, given by potential of mean force calculations to be 4–6 kcal/mol for glucose in solution.

Hydroxyl groups

Primary and secondary carbohydrate hydroxyls are fully engaged in hydrogen bonds, both to the solvent and to other

TABLE 2 Dihedral angles (degrees) of glycosidic linkages from crystallography (XRD), NMR, and MD simulation

	β-GlcNAc-(1→2)-**Man** (6 arm/1→6 Man)		α-**Man**-(1→6)-Man (1→6 Man/red Man)			α-Man-(1→3)-Man (red Man/1→3 Man)		β-GlcNAc-(1→2)-Man (1→3 Man/3 arm)	
	φ	ψ	φ	ψ	ω	φ	ψ	φ	ψ
Sugar I									
aq/MD			68 (16)						
ConA/XRD*			63						
ConA/MD			65 (10)						
Sugar III									
aq/NMR†			60	180	−60/60	75	−135		
aq/*t*/MD			66 (11)	−179 (12)	−66 (10)	65 (10)	−145 (15)		
aq/*g*/MD			65 (11)	180 (15)	59 (12)	66 (12)	−144 (25)		
ConA/XRD‡			64	−170	−45	67	−110		
ConA/XRD§			65	−177	−42	133	−109		
ConA/MD			69 (8)	−169 (7)	−63 (8)	73 (8)	−135 (12)		
Sugar IV									
aq/NMR¶	−92	−83	70	−170	−58	80	−121	−92	−83
aq/MD	−70 {13}	−92 (37)	72 (23)	−178 (17)	−62 (13)	70 (12)	−134 (31)	−73 (18)	−92 (43)
ConA/XRD‖	−71	−132	69	176	−43	62	−99	−80	−82
ConA/MD	−78 (7)	−125 (7)	71 (8)	−176 (7)	−62 (8)	65 (7)	−109 (13)	−80 (16)	−77 (9)

Residue names are also given for glycosidic linkages.
The 1→6 Man sugar is shown in boldface type.
Standard deviations for MD structures are given in parentheses.
*Naismith et al. (1994)
†Brisson and Carver (1983)
‡Naismith and Woods (1996)
§Loris et al. (1996)
¶Homans (1995)
‖Moothoo and Naismith (1998)

regions of the solute. Conformational transitions are frequent, highlighting the dynamic nature of the water hydrogen bond network encompassing the sugar.

N-acetyl groups

The C-C-N-C dihedrals that define the rotation of the terminal N-acetyl groups attached to the terminal glucose moieties display restricted rotation in solution, with infrequent transitions between low energy conformers. Inspection of the trajectories reveals frequent hydrogen bonds between the N-acetyl group of the terminal GlcNAc arms and the adjacent mannose residues, mediated by two, one, or no bridging water molecules (for example, direct hydrogen bonds between the 6 arm amide group and the 6-hydroxyl group of the neighboring mannose). Direct or indirect intersaccharide hydrogen bonding explains the limited conformational freedom of the N-acetyl group. Water-bridged hydrogen bonds are also observed between other adjacent residues in sugar IV: for example, O4 of the 1→3 Man with O_2 of the reducing Man.

Therefore, we conclude that the three sugars adopt conformations consistent with NMR-derived parameters, and can, in fact, sample their bound conformations in aqueous solution, albeit infrequently. The backbone φ torsions are fairly rigid, whereas ψ can show larger torsional librations. Evidence of frequent water-mediated interresidue interactions is observed, particularly involving the N-acetyl arms of IV.

Carbohydrate–protein trajectories

Structural stability is demonstrated by the low-protein backbone root mean square deviation (RMSD) of ~1 Å over the 3-ns trajectories of Con A in complex with I, III, and IV (Fig. 5). A slight drift observed in the RMSD is not uncommon for large protein/solvent systems (Tara et al., 1999). The seven structural waters identified by Loris et al. (1994) remain integral to the protein throughout the three trajectories. To compare with solution trajectories, we again discuss the dynamics of the carbohydrate ligands in terms of chemical groups.

Glycosidic linkages

As with carbohydrate-solution simulations, good agreement is observed between experimental and MD-derived torsion angles (Table 2). Bound ligands III and IV do not explore the alternative crystallographic conformation at the α-(1→3) linkage found by Loris et al. (1996). All torsions experience rigidification upon binding, although the effect is small for angle φ linking 3 arm and 1→3 Man in IV. This motion is illustrated by the (φ, ψ) plot in Fig. 4, where we

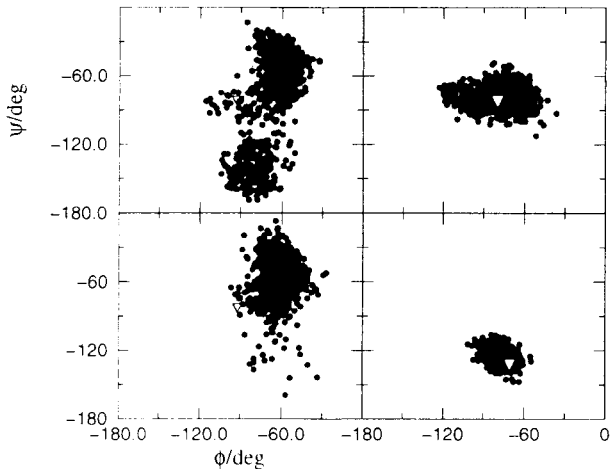

FIGURE 4 Glycosidic torsion angle sampling for 1→3 Man/arm (*top*) and 1→6 Man/arm (*bottom*) linkages of IV during molecular dynamics trajectories for IV/Con A complex (*right*) and IV in aqueous solution (*left*). Spectroscopically determined geometries are also indicated (∇).

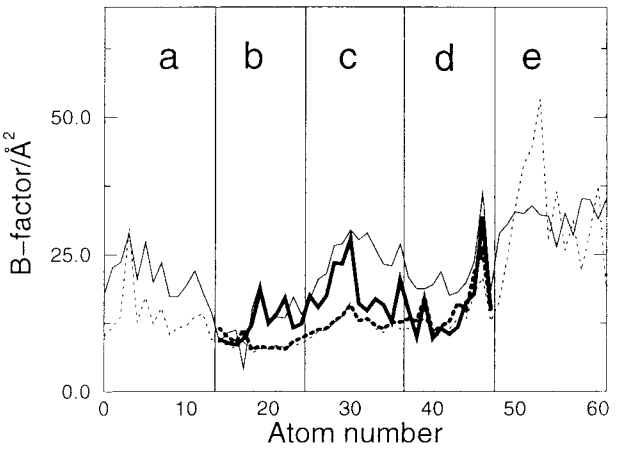

FIGURE 6 Comparison of calculated (*dotted line*) and crystallographic (*solid line*) B-factors (Å^2) for III/Con A (*bold*) and IV/Con A (*standard*). Saccharide residues are labeled (*a*) 6 arm, (*b*) 1→6 Man, (*c*) red Man, (*d*) 1→3 Man, and (*e*) 3 arm.

can see that both GlcNAc arms explore a subset of conformations that are found in solution, but that the 3 arm explores a larger proportion of these, perhaps due to the higher degree of solvent exposure of the 3 arm in the lectin complex. B-factors also provide information on the motion of the protein and carbohydrate (Bradbrook et al., 1998). Therefore, we can observe the enhanced flexibility of the 3 arm relative to the rest of ligand IV by high average x-ray (32.0 Å^2) and MD-calculated (31.5 Å^2) B-factors (atoms 48–61 of Fig. 6). Nevertheless, the 6 arm average x-ray B-factor (21.4 Å^2) is not negligible (Moothoo and Naismith, 1998). The lowest mobility and tightest torsional fluctuations are exhibited by the 1→6 Man (Table 2, Fig. 6), with an average x-ray B-factor of 13.0 Å^2. Although the trends in atomic crystallographic B-factors are well reproduced by the simulation, we note that the B-factors calculated from the trajectories are a lower limit, due in part to neglect of global rotational and translational motion in the MD, and averaging that may occur over different crystallographic geometries.

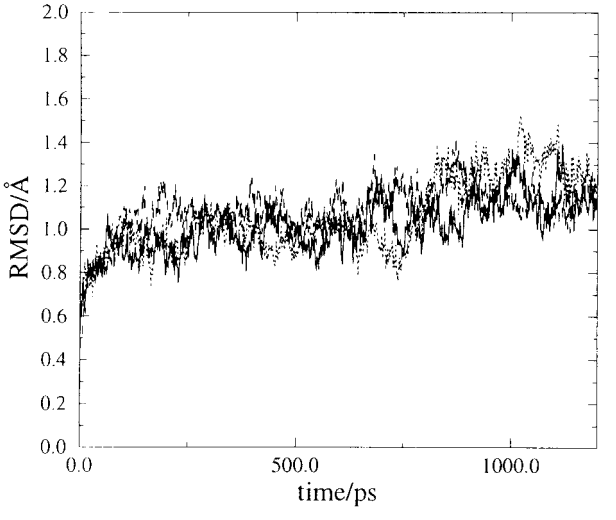

FIGURE 5 Backbone RMS deviations (Å) from crystal structure of Con A during simulation of complex with ligands I (*dotted*), III (*dashed*), and IV (*solid*).

Hydroxymethyl groups

In complex with Con A, the exocyclic CH_2OH groups of I, III, and IV exhibit oscillations about a stable GG conformation, comparable in magnitude to those found in solution. The exception is the CH_2OH group of the 6 arm, which samples all three rotamers, with 43.7% GT, 55.7% GG, and a trace amount of TG. This contrasts with a single GG conformer observed in solution.

N-acetyl groups

In both arms of IV, the N-acetyl groups are conformationally restricted. For the 3 arm, the constraint is by van der Waals contacts with Tyr^{12}, Pro^{13} and His^{205}, and for the 6 arm GlcNAc, it is via a hydrogen bond to Oγ of Ser^{168}.

Hydroxyl groups

In the majority of hydroxyls, rotamer sampling is considerably restricted upon binding. The frequency of sampling for the C6–O6 torsion of the 6 arm reduces by an order of magnitude. For each of the 1→6 Man hydroxyls, nearly unhindered rotation in solution reduces to a single rotamer,

TABLE 3 Average carbohydrate–protein interatomic distances (Å) from X-ray crystallography (d_{xrd}) and molecular dynamics (d_{MD}), with fractional (f_{occ}) occupancy in parentheses

Protein Atom	Sugar Atom	Ligand IV		Ligand III		Ligand I	
		d_{xrd}*	d_{MD} (f_{occ})	d_{xrd}†	d_{MD} (f_{occ})	d_{xrd}‡	d_{MD} (f_{occ})
	6 arm						
Thr^{226} OG1	O3	2.6	3.0 (0.9)				
Gly^{224} O§	O4	2.8	3.4 (0.5)				
Thr^{226} O§	O6	3.4	3.6 (0.6)				
Ser^{168} OG	O7	2.6	2.7 (1.0)				
	1→6 Man						
Arg^{228} N	O3	3.0	3.1 (1.0)	2.9	2.9 (1.0)	2.8	3.1 (1.0)
Arg^{228} HE¶	O3	3.8	3.1 (0.9)	4.0	4.4 (0.1)	4.3	3.9 (0.4)
Asn^{14} ND2	O4	2.9	2.9 (1.0)	2.9	3.1 (0.9)	3.0	2.9 (1.0)
Arg^{228} N§	O4	3.3	3.7 (0.2)	3.5	3.4 (0.7)	3.2	3.4 (0.6)
Asp^{208} OD1	O4	2.8	2.6 (1.0)	2.7	2.7 (1.0)	2.8	2.6 (1.0)
Leu^{99} N	O5	3.0	3.2 (0.9)	2.9	2.9 (1.0)	3.0	3.1 (0.9)
Leu^{99} N§	O6	2.9	3.1 (1.0)	3.1	3.1 (1.0)	3.1	3.1 (1.0)
Tyr^{100} N	O6	3.0	3.2 (1.0)	3.1	3.2 (1.0)	3.0	3.2 (0.9)
Asp^{208} OD2	O6	3.0	2.6 (1.0)	2.9	2.7 (1.0)	2.9	2.6 (1.0)
	red Man						
OW	O2	2.7	2.8 (1.0)	2.6	2.9 (1.0)		
Asp^{16} OD2§	O2	3.3	4.1 (0.2)	3.1	2.8 (0.9)		
Tyr^{12} OH	O4	2.8	2.8 (1.0)	2.8	2.8 (1.0)		
	1→3 Man						
Tyr^{12} OH‖	O3	—	5.4 (0.0)	2.8	6.1 (0.0)		
Pro^{13} O§	O3	2.9	2.7 (1.0)	2.9	3.0 (0.9)		
Pro^{13} O‖	O4	—	5.3 (0.0)	2.8	5.8 (0.0)		
Thr^{15} N	O3	2.9	3.2 (0.9)	2.8	3.0 (1.0)		
Thr^{15} N‖	O4	—	3.5 (0.5)	3.4	2.9 (0.2)		
Thr^{15} OG1	O3	3.0	3.3 (0.8)	2.9	3.2 (0.9)		
Thr^{15} OG1	O4	2.6	2.7 (1.0)	3.1	2.8 (1.0)		
Asp^{16} N	O4	3.0	3.2 (0.9)	3.0	3.3 (0.8)		
Asp^{16} N‖	O6	—	6.4 (0.0)	2.8	6.8 (0.0)		

*Moothoo and Naismith (1998)
†Naismith and Field (1996)
‡Naismith et al. (1994)
§Large deviation from linearity.
¶Not observed in crystal structure.
‖Loris et al. (1996) alternative binding geometry of 1→3 Man.

reflecting strong, specific polar carbohydrate–protein hydrogen bonds at the monosaccharide (1→6 Man) binding site.

Carbohydrate–protein interactions

Accordingly, short average interatomic distances for the key protein and carbohydrate groups at the monosaccharide binding site, for I, III, and IV, are observed in the MD simulations (Table 3). The specificity of the interactions is further highlighted by the high hydrogen bond occupancy over the trajectory (defined as a heavy atom interatomic distance of <3.5 Å). The notable exception to this is the small fraction (0.2–0.7) of occupied Arg^{228} N···O4 hydrogen bonds. The non-ideal geometry of this polar interaction has been noted elsewhere (Naismith et al., 1994; Naismith and Field, 1996; Moothoo and Naismith, 1998). Inspection of the trajectory indicates that, although the Arg^{228} backbone N···O3 interaction is well preserved in I, III, and IV (Table 3), the long, flexible side-chain of Arg^{228} explores a range of conformations and, consequently, a range of interactions with the carbohydrate. We may classify these interaction geometries into three modes: a single hydrogen bond from Arg^{228} NE to O3 (Fig. 7); a bifurcated interaction with the terminal guanidinyl nitrogen atoms of Arg^{228} to O3 (Fig. 7); and a mode without a side-chain hydrogen bond to O3 (Fig. 8). The distribution of these interactions can be mapped using as a coordinate, the Arg^{228} HE···O3 interatomic distance, with characteristic values 2.0, 3.6, and 4.9 Å, respectively, for the three modes. Each complex explores two of the three possible modes (Fig. 9). All three modes are consistent with an Arg^{228} salt-bridge interaction with Asp^{16}, as observed in the crystal structure, although any of the three guanidinyl Ns may act as donor. Thus, we can see an approximately inverse correlation in population of Arg^{228} N···O4 and Arg^{228} NE···O3 interactions (Table 3).

Fewer contacts are observed for the remaining residues (Table 3), indicating the less specific nature with which the

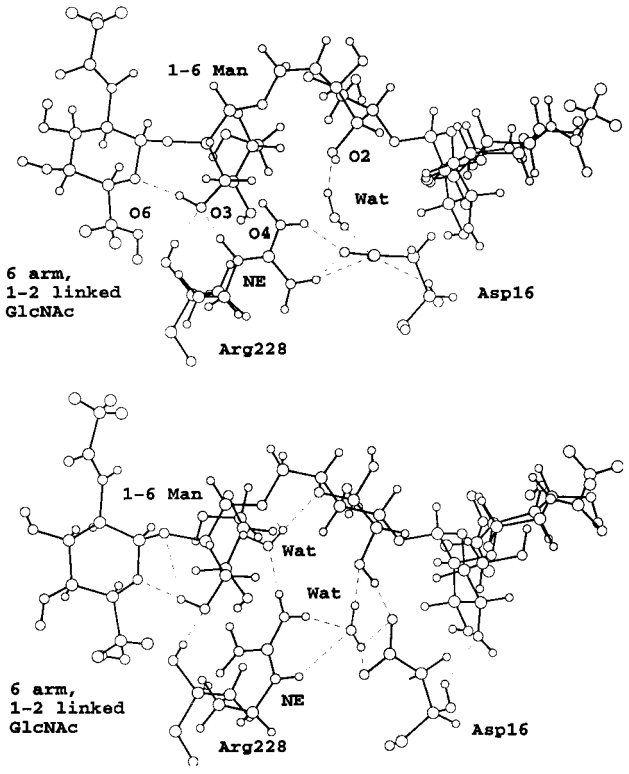

FIGURE 7 Interactions of ligand IV with Arg^{228}: (*top*) single hydrogen bond mode; (*bottom*) bifurcated hydrogen bond mode.

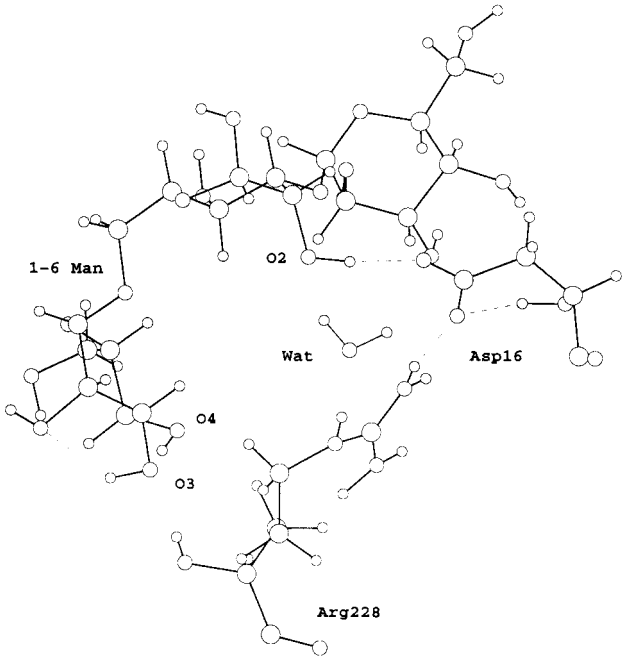

FIGURE 8 Interactions of ligand III with Arg^{228}: no hydrogen bond mode.

residues are recognized. Occupancy of polar contacts is generally high, although low for the nonideal Asp^{16} OD2···O_2 geometry of red Man and low for the two protein backbone hydrogen bonds formed with the 6 arm in IV/Con A (Table 3). In the 6 arm GlcNAc, the MD average interatomic distances are all greater than in the crystal. We may attribute these observations to the exploration of different interaction modes, due in part to Arg^{228}. With respect to Thr^{226}, and, indirectly, Gly^{224}, this is due to the multiple rotamer sampling of the 6 arm exocyclic hydroxymethyl group, reported above (Fig. 7). We also note that there is sampling of a hydrogen bond between Thr^{15} N···O4 of the 1→3 Man, that is observed in the Loris et al. (1996) geometry of III (Table 3). The values of this distance in the III/Con A and IV/Con A x-ray structures are 3.7 (Naismith and Field, 1996) and 3.6 Å (Moothoo and Naismith, 1998), respectively. During the MD, however, periodic N···O4 distances of ~3 Å are found.

Finally, we may superimpose the 237-amino acid polypeptide backbones of the ensemble-averaged structures of I, III, and IV in complex with Con A from the three trajectories (Fig. 10). There is a clear consensus in the primary importance of the conserved structure of the 1→6 Man and the monosaccharide binding site thereof. However, there is some distortion with respect to the mannoside. Superimposing the protein backbone atoms of the MD-average structures of I/Con A with III/Con A gives an RMSD of 0.42 Å for the nonhydrogen sugar atoms of the 1→6 Man residue. For overlay of I/Con A with IV/Con A, we find an RMSD of 0.81 Å. The higher RMSD for the pentasaccharide is centered around distortion at the O3 and β-(1→2)-linked O_2 atoms. Perturbation at the O_2 site in IV/Con A has been observed previously in the static x-ray structures (Moothoo and Naismith, 1998).

The carbohydrate–protein trajectories demonstrate a stable protein conformation, good agreement with crystallo-

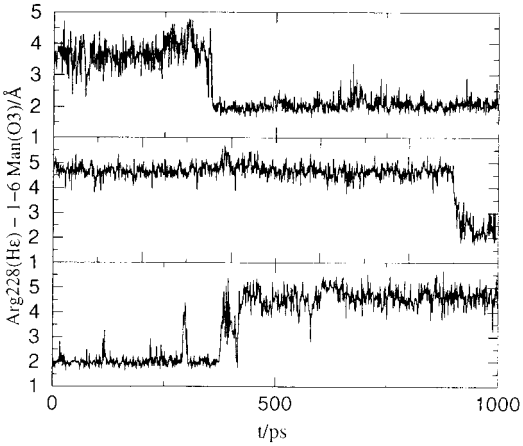

FIGURE 9 Variation of Arg^{228} HE···O3 distance (Å) indicating sampling of Arg^{228} modes over molecular dynamics trajectories of I (*bottom*), III (*middle*), and IV (*top*) bound to Con A.

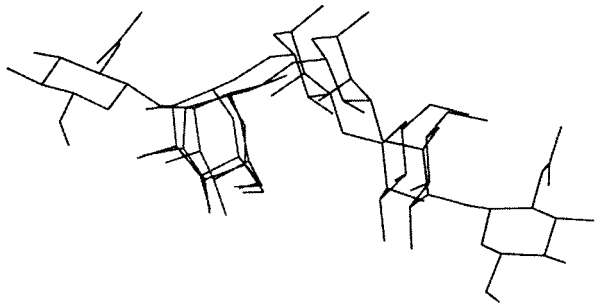

FIGURE 10 Overlay of average MD structures of I, III, and IV in complex with Con A by superposition of all protein backbone atoms.

graphic B-factors and geometric parameters for the ligand, and near complete occupancy of the dominant carbohydrate–protein interactions observed from static x-ray structures. The dynamic ensemble also contains a fraction of alternative interactions arising from ligand and protein contacts not observed from the static crystal structure, including water-mediated interactions. Based upon these structurally validated dynamics simulations of the carbohydrates, bound and in solution, we may analyze the trajectories using the MM-GB/SA approach to obtain a thermodynamic interpretation of binding affinity.

Binding free energy

We now turn to consider energetic analysis of binding by applying Eq. 1 using the solution and protein-bound carbohydrate trajectories. Free energies of binding, $\Delta G_{\text{bind}}^{\text{RL(aq)}}$, are reported in Table 4. The trends exhibited by the calorimetrically determined binding free energy (Table 1) are reproduced by the calculated $\Delta G_{\text{bind}}^{\text{RL(aq)}}$: the oligosaccharide ligands IV and III bind to Con A with equal affinity, and both about 50% more tightly than the mannoside I. The binding free energy of III/*t* and III/*g* are almost identical, implying that the absolute free energies of the ligands in solution are the same. This observation is in agreement with NMR studies that indicate the two conformers to be essentially equally populated. In subsequent discussion of the energetics of III, we will use a Boltzmann average based upon the relative binding free energy of the two ligands in solution (*trans* is favored by 0.6 kcal/mol). This averaging has a small bearing on the calculated binding free energies for III, and we do not consider multiple conformational states of IV in solution. Subsequent incorporation of the entropic contribution to $\Delta G_{\text{bind}}^{\text{RL(aq)}}$, via calculation of normal modes within the harmonic approximation, leads to a decrease in binding by 4.2 kcal/mol of IV over III. To consider absolute binding free energies, we may use this estimate of the entropy with $\Delta G_{\text{bind}}^{\text{RL(aq)}}$, and incorporate the free energy change due to loss of global translational and rotational degrees of freedom of ligand and receptor on binding, of approximately 7–11 kcal/mol (Ajay and Murcko, 1995). From this, we observe that the absolute binding energies are overestimated for IV and III, and underestimated for I, reflecting the difficulty in calculating free energies of complex systems using a summation of components approach, and highlights, in particular, limitations in accurately estimating the solute entropy of binding. One might consider the ligand torsional degrees of freedom to be the most salient entropic discriminator for binding. Using only the quasi-harmonic ligand torsional entropy, we predict a decrease in binding of IV relative to III of 4.5 kcal/mol (Table 4), perhaps fortuitously close to the estimate from harmonic calculations. We now proceed to use the MM-GB/SA model as a tool to identify and quantify the physical contributions to binding of ligands I, III, and IV, in particular seeking to explain the observed and calculated low affinity of IV for Con A.

Ligand and receptor contributions to binding

From Table 4, we observe increasing magnitude in enthalpies of binding, $\Delta H_{\text{bind}}^{\text{RL(g)}}$, with increasing ligand size and number of protein–ligand contacts, from −77.8 kcal/mol for

TABLE 4 Energy contributions to sugar–protein binding (kcal/mol)

	IV	III/*t*	III/*g*	I
$\Delta\bar{E}_{\text{elec}}^{\text{RL(g)}}$	−121.9 (9.6)	−90.7 (8.1)	−92.6 (8.1)	−61.2 (8.3)
$\Delta\bar{E}_{\text{vdw}}^{\text{RL(g)}}$	−47.8 (3.4)	−27.6 (3.1)	−28.9 (3.1)	−17.5 (3.5)
$\Delta\bar{E}_{\text{bad}}^{\text{RL(g)}}$	3.2 (5.8)	0.1 (6.1)	0.9 (6.1)	0.9 (5.5)
$\Delta H_{\text{bind}}^{\text{RL(g)}}$	−166.5 (9.6)	−118.2 (8.6)	−120.4 (7.2)	−77.8 (9.2)
$\Delta\Delta\bar{G}_{\text{np}}^{\text{RL}}$	−6.8 (0.1)	−4.9 (0.1)	−5.1 (0.1)	−3.0 (0.2)
$\Delta\Delta\bar{G}_{\text{pol}}^{\text{RL}}$	123.4 (7.9)	74.1 (7.1)	75.8 (7.3)	48.4 (6.8)
$\Delta\Delta\bar{G}_{\text{solv}}^{\text{RL}}$	116.6 (8.0)	69.2 (7.3)	70.8 (7.3)	45.4 (6.8)
$\Delta G_{\text{bind}}^{\text{RL(aq)}}$*	−49.9 (6.5)	−49.0 (6.2)	−49.6 (6.2)	−32.4 (6.5)
$-T\Delta S_{\text{bind}}^{\text{RL(g)}\dagger}$	20.7	16.4	16.9	30.4
$-T\Delta S_{\text{bind}}^{\text{L(g)}\ddagger}$	11.5	6.2	9.1	2.9

See text for definition of energetic quantities.

Standard errors are given in parentheses.

*Without solute entropy contribution.

†Solute entropy obtained via harmonic approximation.

‡Solute entropy obtained via quasi-harmonic approximation.

TABLE 5 Gas-phase sugar distortion enthalpies ($\Delta H^{L(g)}_{distort}$) and distortion free energies ($\Delta G^{L(aq)}_{distort}$) in solution (kcal/mol)

Energy	IV	III/*t*	III/*g*	I
$\Delta\bar{E}^{L(g)}_{elec}$	10.8 (1.0)	12.1 (0.8)	10.2 (0.8)	7.9 (1.2)
$\Delta\bar{E}^{L(g)}_{vdw}$	1.8 (0.3)	0.0 (0.3)	−1.3 (0.3)	0.2 (0.5)
$\Delta H^{L(g)}_{distort}$	15.7 (1.0)	12.2 (0.9)	10.0 (0.9)	9.1 (1.3)
$\Delta\Delta\bar{G}^{L}_{np}$	0.2 (0.1)	0.0 (0.1)	−0.1 (0.1)	0.0 (0.0)
$\Delta\Delta\bar{G}^{L}_{pol}$	−7.0 (0.8)	−6.7 (0.7)	−5.0 (0.7)	−4.9 (1.0)
$\Delta\Delta\bar{G}^{L}_{solv}$	−6.8 (0.8)	−6.7 (0.7)	−5.2 (0.7)	−4.9 (1.0)
$\Delta G^{L(aq)}_{distort}$	9.0 (0.7)	5.4 (0.6)	4.8 (0.6)	4.2 (0.9)

Standard deviations are given in parentheses for averaged energies.

I to −166.5 kcal/mol for IV. The enthalpy of binding also incorporates the cost of distortion on binding, and, as a result, is not a direct measure of the receptor–ligand interaction. We can separate $\Delta H^{RL(g)}_{bind}$ into a receptor–ligand interaction enthalpy ($\Delta H^{RL(g)}_{inter}$) and ligand ($\Delta H^{L(g)}_{distort}$) and receptor ($\Delta H^{R(g)}_{distort}$) distortion components:

$$\Delta H^{RL(g)}_{bind} = \Delta H^{RL(g)}_{inter} + \Delta H^{R(g)}_{distort} + \Delta H^{L(g)}_{distort}. \quad (5)$$

Here, we obtain $\Delta H^{L(g)}_{distort}$ by restricting the application of the MM-GB/SA method to the ligand in its bound and solution trajectories (Table 5). We neglect distortion of the protein (setting $\Delta H^{R(g)}_{distort}$ to zero), which would require a further separate MD trajectory of the native lectin in aqueous solution. The ligand distortion enthalpies increase with ligand size. $\Delta H^{RL(g)}_{bind}$ is 15.7 kcal/mol for IV, which is 4.1 kcal/mol larger than the Boltzmann-averaged value for III. This enthalpy difference agrees well with the results of a study using the MM2CARB force field to map the glycosidic linkage of a disaccharide, β-GlcNAc-α-(1→2)-Man. The map suggests that the energetic penalty of altering the ψ angle from the solution to bound value is of the order of 3 kcal/mol (Imberty et al., 1991).

We can subsequently calculate a receptor–ligand interaction enthalpy, $\Delta H^{RL(g)}_{inter}$, from Eq. 5. For the three ligands (Table 6), this quantity follows the same trends as $\Delta H^{RL(g)}_{bind}$ (Table 4). Through utilization of the extended binding sites by the two additional mannose sugar residues, III achieves a 43.6 kcal/mol increase in binding over the mannoside I. Additional contacts formed by IV lead to a 51.9 kcal/mol gain in enthalpy over III. We can evaluate the contribution of each sugar residue to the overall binding enthalpy (Table 6). From this decomposition, we observe the 1→6 Man saccharide units in I, III, and IV to interact most strongly with Con A, with an enthalpy of between −80.4 and −86.8 kcal/mol (Table 6). This agrees well with an enthalpy of −83.8 kcal/mol calculated from previous simulation of the methyl mannose/Con A complex by Bradbrook et al. (1998). The 1→6 Man saccharide unit of I binds with the highest affinity of the three ligands, in agreement with the suggestion (Moothoo and Naismith, 1998) that the methyl mannoside/Con A complex represents the optimal geometry for binding at the monosaccharide site. However, the associated standard errors for the 1→6 Man/Con A interaction enthalpies are large (Table 6), and, statistically, we cannot distinguish between the energies. It has also been postulated that the reducing mannose has the least interaction energy with Con A, and our calculations are in line with this, with an interaction enthalpy of between −13.3 and −20.0 kcal/mol for the three ligands. In III/Con A, the 1→3 Man is thus the major contributor to the additional affinity arising from the lectin extended binding site. Analysis also reveals that the source of the increased $\Delta H^{RL(g)}_{inter}$ of IV/Con A over III/Con A is principally protein contacts with the 6 arm GlcNAc (−43.5 kcal/mol), and a weaker interaction with the 3 arm (−14.2 kcal/mol). These favorable interactions are accompanied by a slight weakening by 6.7 kcal/mol and by 2.9 kcal/mol at the red Man and 1→3 Man sites respectively (Table 6).

The saccharide unit–protein interaction enthalpy can be separated further into individual amino acid contributions to identify key amino acid residues for sugar binding. Performing this decomposition for the 1→6 Man unit (Fig. 11) clearly highlights the crucial role of Asp^{208} in tethering the sugar, with a significant, mainly electrostatic contribution to $\Delta H^{RL(g)}_{inter}$, of between −36.5 and −45.5 kcal/mol for the three complexes. Asp^{208} forms two short, strong hydrogen bonds to the saccharide O4 and O6 atoms (Table 3). Significant contributions are also made by hydrogen bonds

TABLE 6 Per residue average energy contributions to binding (kcal/mol)

	I/Con A			III/Con A			IV/Con A		
Residue	$\Delta H^{RL(g)}_{inter}$	$\Delta\Delta G^{RL}_{pol}$	$\Delta\Delta G^{RL}_{np}$	$\Delta H^{RL(g)}_{inter}$	$\Delta\Delta G^{RL}_{pol}$	$\Delta\Delta G^{RL}_{np}$	$\Delta H^{RL(g)}_{inter}$	$\Delta\Delta G^{RL}_{pol}$	$\Delta\Delta G^{RL}_{np}$
6 arm	—	—	—	—	—	—	−43.5 (0.9)	10.0	−1.3
1→6 Man	−86.8 (0.5)	6.9	−2.0	−80.4 (0.4)	6.1	−1.6	−84.2 (0.5)	5.8	−1.2
red Man	—	—	—	−20.0 (0.5)	−1.1	−0.5	−13.3 (0.6)	0.4	−0.5
1→3 Man	—	—	—	−30.0 (0.4)	5.1	−0.9	−27.1 (0.5)	2.5	−0.8
3 arm	—	—	—	—	—	—	−14.2 (0.4)	5.5	−0.6
total	−86.8 (0.5)	6.9	−2.0	−130.4 (0.8)	10.1	−3.0	−182.3 (1.4)	24.2	−4.4

$\Delta H^{RL(g)}_{inter}$ is the sugar–protein interaction energy using a single trajectory.
Electrostatic ($\Delta\Delta G^{RL}_{pol}$) and nonelectrostatic ($\Delta\Delta G^{RL}_{np}$) solvation energies correspond to difference between solution and protein complex trajectories.
Boltzmann average is used for III.
Standard errors for interaction energies are given in parentheses, but, for clarity, are omitted for solvation energies.

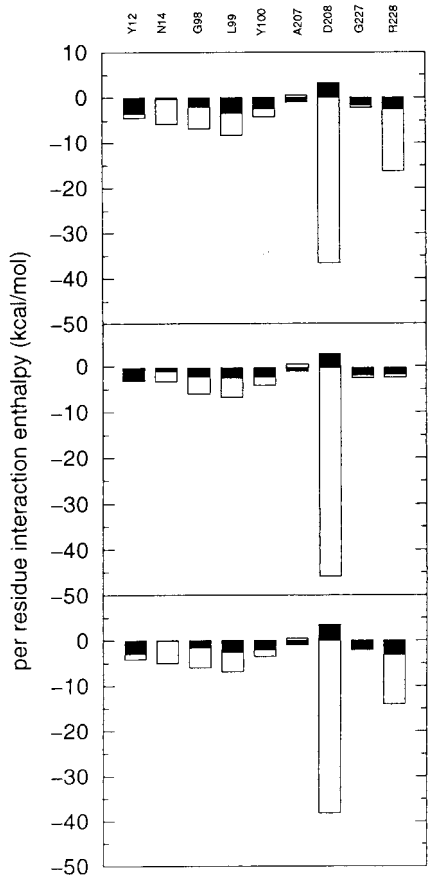

FIGURE 11 Per amino acid energy decomposition (kcal/mol) for interaction of Con A with 1→6 Man of I (*top*), III (*middle*) and IV (*bottom*); van der Waals contribution (*shaded*) and total interaction energy (*unshaded*).

with Gly[98], Leu[99] and Tyr[100]. Interestingly, despite a favorable van der Waals interaction with the 1→6 Man, Ala[207] makes a net unfavorable contribution. The greatest variation is seen for Arg[228], arising from the previously noted conformational sampling. The low average interaction in III/Con A arises from sampling principally of the "no hydrogen bond" mode with respect to O3 of 1→6 Man (Fig. 8), relative to other modes found for I/Con A and IV/Con A. In III/Con A, the −2.1 kcal/mol interaction with 1→6 Man comes from the Arg[228] backbone N···O3 interaction, which is conserved (Table 3). A similar analysis can be applied to the −43.5 kcal/mol interaction enthalpy of the 6 arm in the pentasaccharide IV with Con A. Here, the major contributors are residues Thr[226] (−12.3 kcal/mol) and Ser[168] (−9.4 kcal/mol), which make hydrogen bond contacts (Table 3). A more modest contribution is made by the less sampled Gly[224] contacts (−4.6 kcal/mol). We also note a large net contribution to $\Delta H_{\text{inter}}^{\text{RL(g)}}$ of −22.2 kcal/mol by various charged aspartyl residues clustered around the binding site (Asp[16], Asp[203], Asp[208], and Asp[235]), providing a favorable electrostatic potential for ligand binding.

TABLE 7 Ligand entropy of binding, $\Delta S_{\text{bind}}^{\text{L(g)}}$, from dihedral covariance calculations (cal/(K mol))

Residue	I/Con A $\Delta S_{\text{bind}}^{\text{L(g)}}$	III/*t*/Con A $\Delta S_{\text{bind}}^{\text{L(g)}}$	III/*g*/Con A $\Delta S_{\text{bind}}^{\text{L(g)}}$	IV/Con A $\Delta S_{\text{bind}}^{\text{L(g)}}$
6 arm	—	—	—	1.0
1→6 Man	−9.7	−8.0	−12.1	−17.6
red Man	—	−4.4	−5.2	−6.7
1→3 Man	—	−3.9	−7.0	−5.4
3 arm	—	—	—	−3.3
subtotal	−9.7	−16.5	−24.3	−32.1
total	−9.7	−20.8	−29.8	−38.2

Individual saccharide contributions that neglect interresidue cross-correlation, their sum (subtotal), and total contribution of ligand including interresidue cross-correlation (total) are given.

To complete the consideration of receptor and ligand contributions to binding, in the absence of a solvent field, we here consider the entropies of binding. The global ligand entropies of binding are reported in Table 4. Entropy contributions calculated from harmonic analysis ($-T\Delta S_{\text{bind}}^{\text{RL(g)}}$) are of similar magnitude to binding enthalpies, leading to small binding free energies. For III and IV, we note that these entropy calculations provide a lower limit to the entropy decrease on binding, due to limited sampling about the α-(1→3) linkage in solution. However, ligand I appears to lose the greatest free energy on binding due to entropy (30.4 kcal/mol), compared to 20.7 kcal/mol for IV and then 16.5 kcal/mol for III. The origin of this trend is unclear. It has been found for the avidin system that the entropy contribution from averaged normal mode calculations can vary by 5 kcal/mol in unfavorable cases (Kuhn and Kollman, 2000a). The quasi-harmonic entropy calculations, although smaller in magnitude because we consider only the dihedral contribution to the entropy, show a clear progression in magnitude of entropy loss with increasing ligand size. This entropy can be factored into saccharide residue contributions, partitioning the covariance matrix into residue blocks and applying Eq. 3 to the determinant of each block. The assumption is neglect of inter-residue torsional correlation, which may be estimated from the difference between the residue total and the total using the complete ligand covariance matrix, and is ~6 cal/(K mol) for binding of I, III, and IV (Table 7). Given this approximation, we observe that, as for $\Delta H_{\text{inter}}^{\text{RL(g)}}$ (Table 6), the largest residue contribution to $\Delta S_{\text{bind}}^{\text{L(g)}}$ is from the 1→6 Man; this quantity is, on average, −12.8 cal/(K mol). Thus, the most tightly bound fragment has the largest entropy decrease on binding. In general, however, the residue ordering for is not quite the same as for $\Delta S_{\text{bind}}^{\text{L(g)}}$. With regard to binding of IV, $\Delta S_{\text{bind}}^{\text{L(g)}}$ for the reducing mannose is slightly greater in magnitude than 1→6 Man. As expected, the 3 arm GlcNAc experiences very little change in entropy (−3.3 cal/(K mol)) on binding, as it remains effectively free in solution. Interestingly, the 6 arm yields a small increase of 1.0 cal/(K mol) in entropy on binding. Examination of the covariance matrix shows the largest contributor to be the C6–O6 torsion, which, in the bound state, explores a variety of rotamers, but in

solution, due possibly to a restrictive solvent network noted earlier, samples only the GG conformer. It is unclear how appropriate, however, the application of a multivariate Gaussian distribution assumed by Eq. 3 is in this instance, where multiple minima are sampled. Perhaps a bimodal or higher moment distribution function is more suitable (Fisher, 1993). Nevertheless, there does appear to be some residual entropy in both GlcNAc arms of IV when bound, that results in a smaller than anticipated $\Delta\Delta S_{\text{bind}}^{\text{L(g)}}$ between III and IV. This observation is supported by the calorimetrically determined entropies of binding, which indicate a smaller value of $-T\Delta S_{\text{bind}}$ for IV over III by 4.4 kcal/mol (Table 1). This is in contrast to the total harmonic and quasi-harmonic calculations that predict a smaller value for III of the same magnitude. However, we note that the calorimetric energies also incorporate solvent entropy effects, which these entropy calculations do not (Bradbrook et al., 2000).

Having examined binding free energy components of the ligands to Con A without considering the energetic influence of solvent, we find a strong preferential $\Delta H_{\text{bind}}^{\text{RL(g)}}$ for IV over III by -47.7 kcal/mol, due principally to additional contacts formed by the 6 arm GlcNAc. This is despite small unfavorable contributions from a change in ligand internal energy. We also observe that entropy calculations favor III over IV by 4.2–4.5 kcal/mol.

Solvent contributions to binding

We now consider the effect of solvent, via the GB/SA model, on binding profiles. All three ligands experience an unfavorable desolvation penalty upon binding, comprising 58% of the binding enthalpy for I, 59% for III, and 70% for IV (Table 4). The nonpolar contribution to solvation is small, favoring binding due to burial of solvent-accessible surface area, and is largest for ligand IV (-6.8 kcal/mol), which buries the greatest ligand and protein area (944 Å^2). The dominant contribution to solvation is electrostatic, due to interaction with polar protein and carbohydrate surfaces. We find that ligand IV incurs a 47.0 kcal/mol greater desolvation penalty upon binding relative to III. This cost is nearly double the 24.2 kcal/mol larger penalty for III over I. What is the origin of this effect?

Given the pairwise nature of the GB/SA model (Eq. 4), we can perform a residue-based decomposition of the solvation energy analogous to the interaction enthalpies (Table 6). Considering first the saccharide units of the ligands, we can see that nonpolar solvation makes a small contribution to $\Delta\Delta\bar{G}_{\text{solv}}^{\text{RL}}$. Burial of the 1→6 Man is deepest, occluding the most solvent-accessible surface area for binding of I (278 Å^2) and III (222 Å^2). Concomitantly, the largest electrostatic desolvation cost is incurred for this residue (6.1 to 6.9 kcal/mol). With regard to binding of ligand IV, however, the 1→6 Man is less completely buried (167 Å^2), with a larger decrease in area for the 6 arm of 181 Å^2. $\Delta G_{\text{pol}}^{\text{RL}}$ is 5.8 kcal/mol for the 1→6 Man unit, compared to 5.5 kcal/mol for the 3 arm GlcNAc, and provides a large contribution of 10.0 kcal/mol due to burial from solvent of the 6 arm. This 10.0 kcal/mol desolvation penalty is principally the result of occlusion of the carbonyl group of the N-acetyl function (4.8 kcal/mol) and the O6 hydroxyl group (2.8 kcal/mol). In contrast, there is almost no desolvation penalty for binding the reducing mannose in III and IV (Table 6).

Loss of carbohydrate–solvent interaction accounts for 13.2 kcal/mol of the 47.0 kcal/mol greater desolvation penalty upon binding to Con A of ligand IV over III (Table 4). The balance of this penalty arises from solvent occlusion effects on the lectin. The most significant $\Delta\Delta\bar{G}_{\text{solv}}^{\text{RL}}$ contributions can be linked to desolvation of amino acids around the 6 arm binding site: Thr^{226} (5.6 kcal/mol), Ser^{168} (4.3 kcal/mol), Gly^{227} (3.2 kcal/mol), and Gly^{224} (1.4 kcal/mol). Two other groups make important contributions: the metal cofactors, Ca^{2+} (4.2 kcal/mol) and Mn^{2+} (4.5 kcal/mol); and, secondly, the adjacent Arg^{228} residue (6.0 kcal/mol). We note that, for binding of IV to Con A, Arg^{228} buries 53 Å^2 as it forges hydrogen bonds to the 1→6 Man and 6 arm. However, only 32 Å^2 of solvent-accessible surface area is buried by Arg^{228} on binding of III to Con A. In particular, for most of the III/Con A trajectory, one of the guanidinyl NH_2 groups projects out into solution (Fig. 8). The increased burial of Arg^{228} in IV/Con A and the presence of the 6 arm serves to shield the metal cofactors from the solvent reaction field more in IV/Con A than in III/Con A, leading to a larger desolvation penalty on binding of IV. We note that the remaining protein contributions to desolvation are less than 3.5 kcal/mol.

Therefore, we can trace the larger proportion of desolvation on binding of IV, relative to III, to the substantial burial of the 6 arm GlcNAc, with resulting occlusion of protein residues, shielding of highly charged metal cofactors from solvent and the neighboring Arg^{228} group. This must also be weighed against perhaps the smaller than expected $\Delta H_{\text{inter}}^{\text{RL(g)}}$ of the 6 arm with Con A, due to a low number of specific polar interactions of poor geometry and occupancy. Thus, despite burying a nearly equal surface area as 1→6 Man, the 6 arm has a lower density of compensating, mainly polar, interactions, implying a lack of specific recognition, leading to an approximately equal free energy of binding for III and IV.

DISCUSSION AND CONCLUSIONS

Carbohydrates are characterized by their considerable conformational flexibility and dense polar functionality. These features combine to challenge standard methods of phase space sampling in simulations and the potential energy functions that they use. Calculations of relative binding free energies have been attempted previously for carbohydrate–protein complexes using the free energy perturbation method (Liang et al., 1996; Pathiaseril and Woods, 2000). However, these studies have been limited to ligands differing by only a hydroxyl or methyl group and limited in simulation length, due to the numerous intermediate win-

dows required by the method. A further approximation of carbohydrate simulations in general has been the treatment of solvent, which is modeled either as a harmonically restrained solvent sphere (Bradbrook et al., 2000; Pathiaseril and Woods, 2000) or using a periodic box without accurate treatment of long-range electrostatic forces (Liu and Brady, 1996; Simmerling et al., 1998). The application of the particle-mesh Ewald method to highly polar systems such as DNA and RNA in solution has led to improved accuracy in simulation of these biomolecules. We also note the success with which the MM-PB/SA and MM-GB/SA methods have been applied to the energetic analysis of nucleic acids, allowing for solvent polarization within the continuum solvent approximation (Srinivasan et al., 1998; Cheatham et al., 1998; Reyes et al., 2000). These new methods for accurately treating flexible, polar molecules seem ideally suited to a study of carbohydrates in aqueous solution and dynamically bound to protein receptors. In particular, the MM-GB/SA approach requires fewer, and therefore longer and more accurate, simulations; and permits the comparison of quite disparate ligands, such as the mono-, tri- and pentasaccharide treated here.

An adequate treatment of water has been shown to be particularly important for carbohydrates, which are involved in complex networks of hydrogen bonds with solvent (Liu and Brady, 1999). In this study, we preserve some memory of the directional solvent effect on solute structure through use of the explicit TIP3P solvent model in ensemble generation, and subsequently incorporate solvent polarization by post-processing the ensemble using the GB/SA model. This approach successfully interprets the trends in free energies of binding for a series of three sugar–protein complexes. The MM-GB/SA method correctly ranks the ligands, identifying key sugar and protein residues for binding. In this case, predictions derived from considerations of the x-ray structure are confirmed: the 1→6 Man residue is most tightly bound; the 1→3 Man unit provides the additional binding affinity for III; and the 6 arm GlcNAc makes a number of protein contacts, although not optimal in geometry. As a result, an ensemble of ligand and receptor interactions are explored at the 6 arm binding site. Sufficient sampling of these interactions is required, and it appears that even longer simulations may be necessary to fully sample these long time-scale events.

Our energetic analysis also indicates that a 4.1-kcal/mol greater distortion energy is required for ligand IV over III to adopt the bound conformation in solution. The x-ray structure of IV/Con A (Moothoo and Naismith, 1998) shows the ψ glycosidic angle between the 6 arm and the 1→6 mannose is $-129 \pm 3°$ in the complex with Con A, a considerable deviation from the solution minimum of $-83 \pm 14°$. Although the φ glycosidic dihedral, because of its implication in the endo- and exo-anomeric effects, has been thoroughly explored by a wide range of theoretical approaches (Wiberg and Murcko, 1989), there has been a relative paucity of high-level theoretical investigations of the conjugate ψ angle. A molecular mechanical study of the glycosidic linkage found in the disaccharide β-GlcNAc-(1→2)-Man suggests that the cost of altering the ψ angle from the solution to bound value is of the order of 3 kcal/mol (Imberty et al., 1991). However, an ab initio study (Odelius et al., 1995) of a more fundamental model, 2-cyclohexoxytetrahydropyran, appears to indicate a flat potential energy surface, finding the bound value of ψ only 0.2 kcal/mol higher in energy than the solution geometry at the MP2 level with a double-ζ basis set. A more recent HF/6–31G* investigation (Tvaroska and Carver, 1997) explored energy as a function of ψ in the 1→2, 1→3, and 1→4 glycosidic linkages of carbohydrate model compounds. Of particular interest is the rotational profile of ψ in (2S, 3R)-2,3-dimethoxytetrahydropyran, which most closely resembles the β-GlcNAc-(1→2)-Man linkage under consideration here. Three minima were found. The two low-energy minima correspond to the solution and bound conformations of the disaccharide, and, interestingly, their bound value was found to be lower in energy by 0.2 kcal/mol. The third minimum is less thermally accessible, being 1.6 kcal/mol higher in energy than the bound conformation. Again, the evidence suggests that a distortion of ψ intrinsically contributes only marginally, if at all, to the less favorable $\Delta G_{\text{bine}}^{\text{RL(aq)}}$ of the pentasaccharide. Accordingly, analysis of the force-field torsional potential for ψ over our pentasaccharide trajectory reveals a very modest increase in internal (bond-angle-dihedral) energy on binding, from 3.4 to 3.5 kcal/mol. Therefore, it appears from quantum and classical calculations that the intrinsic distortion of the ψ dihedral contributes no more than 0.2 kcal/mol of unfavorable binding energy.

Entropy calculations, by means of the harmonic and quasi-harmonic approximation, disfavor the binding of IV over III by 4.2–4.5 kcal/mol. Due to its global nature, the calculation of entropies from molecular simulation is problematic and represents an active area of development. Indeed, the restrictive nature of an assumed multivariate Gaussian probability distribution has been already noted. Nevertheless, with the development of algorithms to describe large systems, we can expect extension of the quasi-harmonic method to study more ambitious biomolecules, such as the protein component of this study, Con A.

We can conclude that binding of the 6 arm GlcNAc of pentasaccharide IV is energetically neutral, due to a proportionally large desolvation penalty rather than to a distorted ψ angle. This arises from a combination of transient polar interactions between the residue and the protein and masking from solvent of hydrogen bonding groups on the sugar and protein binding surfaces. Therefore, other than in a steric sense dictated by the lectin fold, the 6 arm GlcNAc is not specifically recognized by Con A. Rather, the residue is tolerated by Con A. We can describe this effect as a lack of hydropathic complementarity. Indeed, calorimetry has shown that multivalent interactions are required to increase

significantly further oligosaccharide/Con A affinity beyond that of the trimannoside anchor (Dimick et al., 1999).

We have also shown that the MM-GB/SA method is a useful tool in the analysis of carbohydrates, bound and in aqueous solution. Here, the method has correctly ranked the binding affinity of a series of carbohydrate–protein complexes. More quantitative estimates of relative binding free energies were obtained using the MM-PB/SA method for the binding of a set of diverse ligands to the protein, avidin; moderate success was found in the calculation of absolute affinity constants (Kuhn and Kollman, 2000a,b). Given the adequate generation and sampling of trajectories, systematic errors can enter into the method from the force field, the solvation model and the entropy calculation. Recently, it has been shown that the AMBER force field, using RESP charges, describes the conformational behavior of 55 organic and biological molecules with an average absolute error of 0.28 kcal/mol (Wang et al., 2000). A considerably larger source of error is the Generalized Born solvation model, which has been shown to reproduce experimental absolute solvation free energies to about 1% accuracy (Jayaram et al., 1998). Due to the large solvation energies associated with polyvalent proteins, the error in the solvation contribution to absolute binding affinity can be of the order of tens of kcal/mol. Due to cancellation effects, however, this error is expected to have less impact on calculated relative binding free energies. We also note that efforts are underway to improve the GB model for biological systems (Onufriev et al., 2000). Finally, considering the errors in entropy calculation, it has been observed that these calculations can have an associated error of up to 5 kcal/mol (Kuhn and Kollman, 2000b). As indicated earlier, improvement is anticipated by extension of the quasi-harmonic method to study more ambitious biomolecules. Development of solvation and entropy contributions will lead to improvements in the calculation of quantitative relative and absolute free energies of binding by the MM-GB/SA method. Indeed, the hybrid molecular dynamics/continuum solvent approach continues to show great promise in the thermodynamic analysis of biomolecular structure and function.

We thank Biotechnology and Biological Sciences Research Council for support of this research. R.A.B. thanks the Ramsay Trust for a fellowship, Steve Homans for provision of the NMR structure of IV, and Darren Green and Gail Bradbrook for helpful discussions.

REFERENCES

Ajay, and M. A. Murcko. 1995. Computational methods to predict binding free energies in ligand–receptor complexes. *J. Med. Chem.* 38: 4953–4967.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

Bouckaert, J., T. W. Hamelryck, L. Wyns, and R. Loris. 1999. The crystal structures of Man(α1–6)Man(α1-O)Me in complex with concanavalin A. *J. Biol. Chem.* 274:29188–29195.

Bradbrook, G. M., T. Gleichmann, S. J. Harrop, J. Habash, J. Raferty, J. Kalb (Gilboa), J. Yariv, I. H. Hillier, and J. R. Helliwell. 1998. X-ray and molecular dynamics studies of concanavalin A glucoside and mannoside complexes. *J. Chem. Soc. Faraday Trans.* 94:1603–1611.

Bradbrook, G. M., J. R. Forshaw, and S. Perez. 2000. Structure/thermodynamics relationships of lectin–saccharide complexes. The *Erythrina corallodendron* case. *Eur. J. Biochem.* 267:4545–4555.

Brisson, J. R., and J. P. Carver. 1983. Solution conformation of α-D-(1→3)-linked and α-D-(1→6)-linked oligomannosides using proton nuclear magnetic resonance. *Biochemistry.* 22:1362–1368.

Carver, J. P., and D. A. Cummings. 1987. Reevaluation of rotational populations for 1,6-linkages—reconciliation with potential energy calculations. *Biochemistry.* 26:6676–6683.

Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, Jr., R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. 1999. AMBER 6, University of California, San Fransisco.

Cheatham, T. E., J. Srinivasan, D. A. Case, and P. A. Kollman. 1998. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dynam.* 16:265–280.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

Cumming, D. A., D. S. Dime, A. A. Grey, J. J. Krepinsky, and J. P. Carver. 1986. Specific deuteration of a trimannoside confirms the existence of a disputed interresidue nuclear overhauser enhancement. *J. Biol. Chem.* 261:3208–3213.

Derewenda, Z., J. Yariv, J. R. Helliwell, A. J. Kalb (Gilboa), E. Dodson, M. Paiz, T. Wan, and J. W. Campbell. 1989. The structure of the saccharide-binding site of concanavalin A. *EMBO J.* 8:2189–2193.

Dimick, S. M., S. C. Powell, S. A. McMahon, D. N. Moothoo, J. H. Naismith, and E. J. Toone. 1999. On the meaning of affinity: cluster glycoside effects and concanavalin A. *J. Am. Chem. Soc.* 121: 10286–10296.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

Fisher, N. I. 1993. Statistical Analysis of Circular Data. Cambridge University Press, New York.

Fujii, S., T. Nishiura, A. Nishikawa, R. Miura, and N. Taniguchi. 1990. Structural heterogeneity of sugar chains in immunoglobulin G—conformation of immunoglobulin G molecule and substrate specificities of glycoyl transferases. *J. Biol. Chem.* 265:6009–6018.

Hawkins, G. D., C. J. Cramer, and D. G. Truhlar. 1996. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* 100:19824–19839.

Homans, S. W. 1995. Conformational studies on oligosaccharides. *In* Biomolecular NMR spectroscopy. J. N. S. Evans, editor. Oxford University press, Oxford, UK.

Homans, S. W., R. A. Dwek, and T. W. Rademacher. 1987. Solution conformations of N-linked oligosaccharides. *Biochemistry.* 26: 6553–6560.

Hori, H., Y. Nishida, H. Ohrui, and H. Meguro. 1990. Conformational analysis of hydroxymethyl group of D-mannose derivatives using (6S)-(6-H-2)-D-mannose and (6R)-(6-H-2)-D-mannose. *J. Carb. Chem.* 9:601–618.

Imberty, A., M. M. Delage, Y. Bourne, C. Cambillau, and S. Perez. 1991. Databank of 3-D structures of disaccharides. 2. N-acetyllactosamine type N-glycans. *Glycoconjug. J.* 8:456–483.

Jayaram, B., D. Sprous, and D. L. Beveridge. 1998. Solvation free energy of biomacromolecules: parameters for a modified generalized Born model consistent with the AMBER force field. *J. Phys. Chem.* 102: 9571–9576.

Jorgensen, W. L., J. Chandrasekhar, J. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Juaristi, E., and G. Cuevas. 1994. The Anomeric Effect. CRC Press, Boca Raton, FL.

Karplus, M., and J. N. Kushick. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules.* 14:325–332.

Kuhn, B., and P. A. Kollman. 2000a. A ligand that is predicted to bind better to avidin than biotin: insights from computational alanine scanning. *J. Am. Chem. Soc.* 122:3909–3916.

Kuhn, B., and P. A. Kollman. 2000b. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* 43:3786–3791.

Lee, M. R., Y. Duan, and P. A. Kollman. 2000. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins.* 39:309–316.

Lemieux, R. U. 1971. Effects of unshared pairs of electrons and their solvation on conformational equilibria. *Pure Appl. Chem.* 25:527–548.

Liang, G., R. K. Schmidt, H. A. Yu, D. A. Cumming, and J. W. Brady. 1996. Free energy simulation studies of the binding specificity of mannose-binding protein. *J. Phys. Chem.* 100:2528–2534.

Liu, Q., and J. W. Brady. 1996. Anisotropic solvent structuring in aqueous sugar solutions. *J. Am. Chem. Soc.* 118:12276–12286.

Loris, R., P. P. G. Stas, and L. Wyns. 1994. Conserved waters in legume lectin crystal structures. *J. Biol. Chem.* 269:26722–26733.

Loris, R., D. Maes, F. Poortmans, L. Wyns, and J. Bouckaert. 1996. A structure of the complex between concanavalin A and methyl-3,6-di-O-(α-D-mannopyranosyl)- α-D-mannopyranoside reveals two binding modes. *J. Biol. Chem.* 271:30614–30618.

Mandal, D. K., N. Kishore, and C. F. Brewer. 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.

Marcus, Y. 1991. Thermodynamics of solvation of ions. *J. Chem. Soc. Faraday Trans.* 87:2995–2999.

Massova, I., and P. A. Kollman. 1999. Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* 121:8133–8143.

Moothoo, D. N., and J. H. Naismith. 1998. Concanavalin A distorts the β-GlcNAc-α-(1→2)-Man linkage of β-GlcNAc-(1→2)-α-Man-(1→3)-[β-GlcNAc-(1→2)-α-Man-(1→6)]-Man upon binding. *Glycobiology.* 8:173–181.

Moothoo, D. N., B. Canan, R. A. Field, and J. H. Naismith. 1999. Man α1→2 Man α-OMe-concanavalin A complex reveals a balance of forces involved in carbohydrate recognition. *Glycobiology.* 9:539–545.

Naismith, J. H., C. Emmerich, J. Habash, S. J. Harrop, J. R. Helliwell, W. N. Hunter, J. Raferty, A. J. Kalb (Gilboa), and J. Yariv. 1994. Refined structure of concanavalin A complexed with methyl α-D-mannopyranoside at 2.0 Å resolution and comparison with the saccharide-free structure. *Acta. Cryst*. D50:847–858.

Naismith, J. H., and R. A. Field. 1996. Structural basis of trimannoside recognition. *J. Biol. Chem.* 271:972–976.

Nishida, Y., H. Ohrui, and H. Meguro. 1984. H-1 NMR studies of (6R)-deuterated and (6S)-deuterated D-hexoses. *Tett. Lett.* 25:1575–1578.

Odelius, M., A. Laaksonen, and G. Widmalm. 1995. A model glycosidic linkage—an ab initio geometry optimisation study of 2-cyclohexoxytertrahydropyran. *J. Phys. Chem.* 99:12686–12692.

Onufriev, A., D. Bashford, and D. A. Case. 2000. Modification of the Generalized Born model suitable for macromolecules. *J. Phys. Chem. B.* 104:3712–3720.

Parkinson, C. I., M. D. Cooper, W. T. Hewitt, and I. H. Hillier. 1998. MAVIS: an interactive visualization tool for computational chemistry in a distributed networked environment. *Pac. Symp. Biocomp.* 98:189–200.

Pathiaseril, A., and R. J. Woods. 2000. Relative energies of binding for antibody–carbohydrate–antigen complexes computed from free energy simulations. *J. Am. Chem. Soc.* 122:331–338.

Peumans, W. J., and E. J. van Damme. 1995. The role of lectins in plant defence. *Histochem. J.* 27:253–271.

Reyes, C. M., and P. A. Kollman. 2000. Structure and thermodynamics of RNA–protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.* 297:1145–1158.

Sanner, M. F., A. J. Olson, and J. C. Spehner. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers.* 38:305–320.

Schwarz, F. P., K. D. Puri, R. G. Bhat, and A. Surolia. 1993. Thermodynamics of monosaccharide binding to concanavalin A, pea lectin and lentil lectin. *J. Biol. Chem.* 268:7668–7677.

Sharon, N., and H. Lis. 1998. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* 98:637–674.

Simmerling, C., T. Fox, and P. A. Kollman. 1998. Use of locally enhanced sampling in free energy calculations: testing and application to the $\alpha \rightarrow \beta$ anomerization of glucose. *J. Am. Chem. Soc.* 120:5771–5782.

Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Springer, T. A., and L. A. Lasky. 1992. Selectins—interpreters of cell-specific carbohydrate information during inflammation. *Science.* 258:964–969.

Srinivasan, J., T. E. Cheatham, III, P. Cieplak, P. A. Kollman, and D. A. Case. 1998. Molecular dynamics and continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* 120:9401–9409.

Still, W. C., A. Tempczyk, R. C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.

Tara, S., T. P. Straatsma, and J. A. McCammon. 1999. Mouse acetylcholinesterase unliganded and in complex with huperzine A: a comparison of molecular dynamics simulations. *Biopolymers.* 50:35–43.

Tvaroska, I., and J. P. Carver. 1997. Ab initio molecular orbital calculations of carbohydrate model compounds. 4. Flexibility of ψ-type glycosidic bonds in carbohydrates. *J. Mol. Struct.* 395–396:1–13.

Van Gunsteren, W. F., and H. J. C. Berendsen. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34: 1311–1327.

Wang, J., P. Cieplak, and P. A. Kollman. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.* 21:1049–1074.

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

Weisgerber, S., and J. R. Helliwell. 1993. High resolution crystallographic studies of native concanavalin A using rapid laue data-collection methods and the introduction of a monochromatic large-angle oscillation technique (LOT). *J. Chem. Soc. Faraday Trans.* 89:2667–2675.

Wiberg, K. B., and M. A. Murcko. 1989. Rotational barriers. 4. Dimethoxymethane—the anomeric effect revisited. *J. Am. Chem. Soc.* 111: 4821–4828.

Williams, B. A., M. C. Chervenak, and E. J. Toone. 1992. Energetics of lectin–carbohydrate binding—a microcalorimetric investigation of concanavalin A–oligomannoside complexation. *J. Biol. Chem.* 267: 22907–22911.

Woods, R. J., R. A. Dwek, C. J. Edge, and B. Fraser-Reid. 1995. Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM 93 parameter development. *J. Phys. Chem.* 99:3832–3846.

Woods, R. J., A. Pathiaseril, M. R. Wormald, C. J. Edge, and R. A. Dwek. 1998. The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. *Eur. J. Biochem.* 258:372–386.