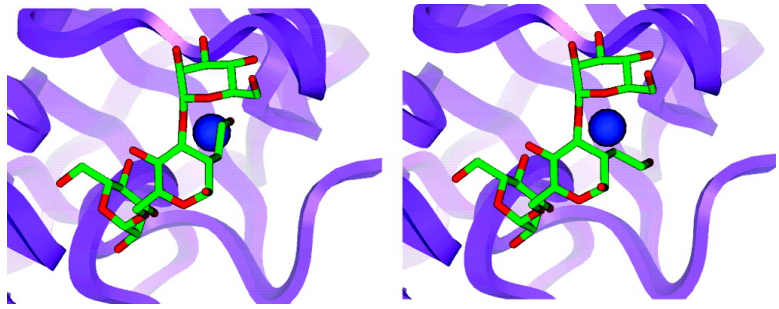
Article

Involvement of Water in Carbohydrate#Protein Binding: Concanavalin A Revisited

Renuka Kadirvelraj, B. Lachele Foley, Jane D. Dyekjær, and Robert J. Woods

J. Am. Chem. Soc., **2008**, 130 (50), 16933-16942 • DOI: 10.1021/ja8039663 • Publication Date (Web): 18 November 2008



J|A|C|S
ARTICLES

Published on Web 11/18/2008

Involvement of Water in Carbohydrate−Protein Binding: Concanavalin A Revisited

Renuka Kadirvelraj, B. Lachele Foley, Jane D. Dyekjær,† and Robert J. Woods*

Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, Georgia 30602

Received May 27, 2008; E-mail: rwoods@ccrc.uga.edu

Abstract: Ordered water molecules bound to protein surfaces, or in protein−ligand interfaces, are frequently observed by crystallography. The investigation of the impact of such conserved water molecules on protein stability and ligand affinity requires detailed structural, dynamic, and thermodynamic analyses. Several crystal structures of the legume lectin concanavalin A (Con A) bound to closely related carbohydrate ligands show the presence of a conserved water molecule that mediates ligand binding. Experimental thermodynamic and theoretical studies have examined the role of this conserved water in the complexation of Con A with a synthetic analog of the natural trisaccharide, in which a hydroxyethyl side chain replaces the hydroxyl group at the C-2 position in the central mannosyl residue. Molecular modeling earlier indicated (Clarke, C.; Woods, R. J.; Glushka, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J. *J. Am. Chem. Soc.* 2001, *123*, 12238−12247) that the hydroxyl group in this synthetic side chain could occupy a position equivalent to that of the conserved water, and thus might displace it. An interpretation of the experimental thermodynamic data, which was consistent with the displacement of the conserved water, was also presented. The current work reports the crystal structure of Con A with this synthetic ligand and shows that even though the position and interactions of the conserved water are distorted, this key water is not displaced by the hydroxyethyl moiety. This new structural data provides a firm basis for molecular dynamics simulations and thermodynamic integration calculations whose results indicate that differences in van der Waals contacts (insertion energy), rather than electrostatic interactions (charging energy) are fundamentally responsible for the lower affinity of the synthetic ligand. When combined with the new crystallographic data, this study provides a straightforward interpretation for the lower affinity of the synthetic analog; specifically, that it arises primarily from weaker interactions with the protein via the positionally perturbed conserved water. This interpretation is fully consistent with the experimental observations that the free energy of binding is enthalpy driven, that there is both less enthalpic gain and less entropic penalty for binding the synthetic ligand, relative to the natural trisaccharide, and that the entropic component does not arise from releasing an ordered water molecule from the protein surface to the bulk solvent.

Introduction

The interactions between carbohydrates and proteins impact important biological functions like cell−cell interactions,[1] signal transduction,[2] host−pathogen recognition,[3] and inflammation[4] even though carbohydrate−protein interactions are weak, with dissociation constants typically in the range of 10^{-3} to 10^{-6} M. The low affinities of these interactions reflect inherent features of oligosaccharides, like their lack of potential to form strong hydrophobic contacts, internal mobility and absence of highly polar functional groups. Oligosaccharide affinities therefore are a result of relatively weak interactions such as hydrogen bonds involving the carbohydrate hydroxyl groups, and van der Waals contacts. These essentially enthalpic terms are accompanied by entropic factors like receptor and ligand surface desolvation (which may or may not be net favorable) and unfavorable configurational entropy arising from restricted internal motion. It is not possible to characterize carbohydrate binding without considering the balance of enthalpic and entropic interactions between the free ligand and receptor with water, and the bound complex.

Water molecules in carbohydrate−protein crystal structures are often observed in and around the carbohydrate binding site, bridging the interaction between the protein and ligand. The interaction between the xenoantigen and lectin 1B4 from *Griffonia simplicifolia*[5] is a typical example. When multiple structures are available for a carbohydrate binding protein, bound waters may be observed to consistently form similar interactions with a range of related ligands.[6−8] It has been proposed that the displacement of such conserved water molecules would

† Current address: Evotec Ltd, 114 Milton Park, Abingdon, Oxfordshire OX14 4SA, United Kingdom.

(1) Geijtenbeek, T.; Torensma, R.; van Vliet, S.; van Duijnhoven, G.; Adema, G.; van Kooyk, Y.; Figdor, C. *Cell* **2000**, *100*, 575–585.
(2) Sacchettini, J. C.; Baum, L. G.; Brewer, C. F. *Biochemistry* **2001**, *40*, 3009–3015.
(3) Karlsson, K. A. *Biochem. Soc. Trans.* **1999**, *27*, 471–474.
(4) Kansas, G. S. *Blood* **1996**, *88*, 3259–3287.
(5) Tempel, W.; Tschampel, S.; Woods, R. J. *J. Biol. Chem.* **2002**, *277*, 6615–6621.
(6) Ravishankar, R.; Suguna, K.; Surolia, A.; Vijayan, M. *Acta Crystallogr.* **1999**, *D55*, 1375–1382.
(7) Elgavish, S.; Shaanan, B. *Trends Biochem. Sci.* **1997**, *22*, 462–467.

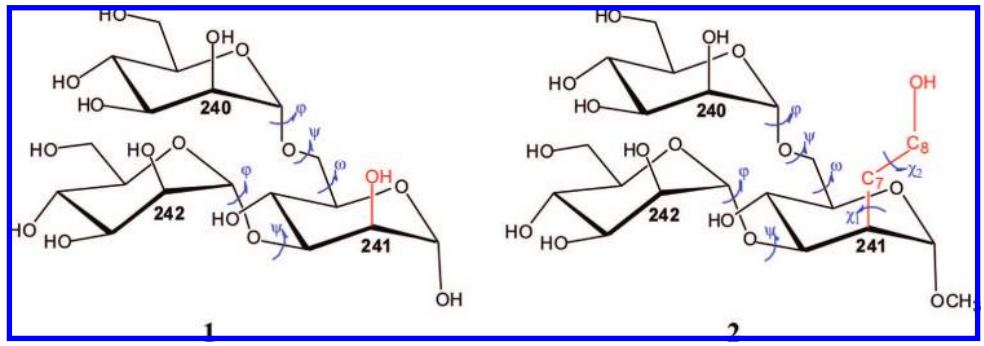


Figure 1. Con A natural ligand α-D-Man*p*(1→6)[α-D-Man*p*(1→3)]α-D-Man*p* (**1**) and synthetic analog α-D-Man*p*(1→6)[α-D-Man*p*(1→3)]α-D-2-hydroxyethyl-Man*p*(α1-O)Me (**2**). The inter-residue dihedral angles as well as those of the synthetic hydroxyethyl moiety are defined as follows: α(1→6) linkage: φ = O5−C1−O6′−C6′; Ψ = C1−O6′−C6′−C5′; ω = O6′−C6′−C5′−O5′, α(1→3) linkage: φ = O5−C1−O3′−C3′; Ψ = C1−O3′-C3′-C4′. Side chain dihedral angles: χ_1 = C1−C2−C7−C8 and χ_2 = C2−C7−C8−O8.

result in an increase in the entropy of binding, and binding free energy.[9] For example, the displacement of a conserved water molecule in HIV-1 protease[10] has been exploited to enhance protease inhibitor activity. The displacement was achieved by integrating a carbonyl group into the inhibitor, which was able to take the place of the conserved water, while retaining the former network of hydrogen bonds.[11] The potential role and impact of such conserved water molecules in carbohydrate-protein binding has been the subject of considerable study[12−18] and discussion.[19,20] We[18] as well as others[13,14,16] have used the legume lectin concanavalin A (Con A) as a model system for exploring the molecular basis of water-mediated carbohydrate-protein interactions.

Con A is a storage and defense lectin[21] isolated from the legume *Canavalia ensiformis.*[22] Crystal structures of Con A have been determined in its ligand-free state,[23−25] as well as bound to a series of mannose-containing carbohydrate ligands.[13,15,24,26−30] These studies have provided information about the specificity of the interaction of Con A with carbohydrate ligands and furthered its value in histochemical staining.[31] The changes in thermodynamic binding parameters, which occur during the association of Con A with various carbohydrate ligands, have also been of interest to the scientific community.[12,14,16,18,32−34] The body of structural data indicates that a conserved water molecule plays an important role in anchoring mannoside ligands to Con A by mediating the interaction between the ligand and the protein. The quantification of the thermodynamic contribution of this water to binding affinity is of interest as it could impact the design of inhibitors of carbohydrate-processing enzymes[35] and provide further insight into the role of desolvation free energy in carbohydrate-protein binding.[16,36]

The role and impact of the putative displacement of this conserved water molecule has been studied previously by Clarke et al.[18] through the complexation of Con A with a synthetic analog of the trimannoside epitope of *N*-linked carbohydrates α-D-Man*p*(1→6)[α-D-Man*p*(1→3)]α-D-Man*p*, (**1**). The hydroxyl group at C-2 of the central mannosyl residue of the native epitope was replaced with a hydroxyethyl moiety designed to occupy the position of the conserved water (Figure 1, ligand **2**). The thermodynamics of binding were measured for this synthetic ligand by isothermal titration calorimetry (ITC), and a putative structure for the complex generated by the extrapolation of the native ligand in the crystal structure of Con A,[13] to the synthetic analog. The model of the complex with the synthetic ligand indicated that the hydroxyl group in the hydroxyethyl side chain could occupy essentially the same position as the conserved water. The observed thermodynamic data were subsequently interpreted under the assumption that

(8) Casset, F.; Hamelryck, T.; Loris, R.; Brisson, J.-R.; Tellier, C.; Dao-Thi, M.-H.; Wyns, L.; Poortmans, F.; Pérez, S.; Imberty, A. *J. Biol. Chem.* **1995**, *270*, 25619–25628.
(9) Sharrow, S. D.; Edmonds, K. A.; Goodman, M. A.; Novotny, M. V.; Stone, M. J. *Protein Sci.* **2005**, *14*, 249–56.
(10) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. *Science* **1989**, *245*, 616–621.
(11) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C.-H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. *Science* **1994**, *263*, 380–384.
(12) Williams, B. A.; Chervenak, M. C.; Toone, E. J. *J. Biol. Chem.* **1992**, *267*, 22907–22911.
(13) Naismith, J. H.; Field, R. A. *J. Biol. Chem.* **1996**, *271*, 972–976.
(14) Swaminathan, C. P.; Surolia, N.; Surolia, A. *J. Am. Chem. Soc.* **1998**, *120*, 5153–5159.
(15) Bouckaert, J.; Hamelryck, T. W.; Wyns, L.; Loris, R. *J. Biol. Chem.* **1999**, *274*, 29188–29195.
(16) Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2005**, *109*, 662–670.
(17) Lu, Y.; Yang, C.-Y.; Wang, S. *J. Am. Chem. Soc.* **2006**, *128*, 11830–11839.
(18) Clarke, C.; Woods, R. J.; Glushka, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J. *J. Am. Chem. Soc.* **2001**, *123*, 12238–12247.
(19) Lemieux, R. U. Carbohydrate Antigens; ACS Symposium Series 519; American Chemical Society: Washington, DC, 1992; 6−18.
(20) Li, Z.; Lazaridis, T. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.
(21) Peumans, W. J.; van Damme, E. J. *Histochem. J.* **1995**, *27*, 253–271.
(22) Hardman, K. D.; Ainsworth, C. F. *Biochemistry* **1972**, *11*, 4912–4919.
(23) Weisgerber, S.; Helliwell, J. R. *J. Chem. Soc. Faraday Trans.* **1993**, *89*, 2667–2675.
(24) Naismith, J. H.; Emmerich, C.; Habash, J.; Harrop, S. J. *Acta Crystallogr.* **1994**, *D50*, 847–858.
(25) Deacon, A.; Gleichmann, T.; Kalb, A. J.; Price, H.; Raftery, J.; Bradbook, G.; Yariv, J.; Helliwell, J. R. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 4305–4312.
(26) Derewenda, Z.; Yariv, J.; Helliwell, J. R.; Kalb, A. J.; Dodson, E. J.; Papiz, M. Z.; Wan, T.; Campbell, J. *Embo J.* **1989**, *8*, 2189–2193.
(27) Loris, R.; Maes, D.; Poortsmans, F.; Wyns, L.; Bouckaert, J. *J. Biol. Chem.* **1996**, *271*, 30614–30618.
(28) Moothoo, D. N.; Naismith, J. H. *Glycobiology* **1998**, *8*, 173–181.
(29) Moothoo, D. N.; Canan, B.; Field, R. A.; Naismith, J. H. *Glycobiology* **1999**, *9*, 539–545.
(30) Dimick, S. M.; Powell, S. C.; McMahon, S. A.; Moothoo, D. N.; Naismith, J. H.; Toone, E. J. *J. Am. Chem. Soc.* **1999**, *121*, 10286–10296.
(31) Ree, H. J. *Cancer* **1983**, *51*, 1639–1646.
(32) Loris, R.; Stas, P. P. G.; Wyns, L. *J. Biol. Chem.* **1994**, *269*, 26722–26733.
(33) Mandal, D. K.; Kishore, N.; Brewer, C. F. *Biochemistry* **1994**, *33*, 1149–1156.
(34) Gupta, D.; Dam, T. K.; Oscarson, S.; Brewer, F. C. *J. Biol. Chem.* **1997**, *272*, 6388–6392.
(35) Kawatkar, S. P.; Kuntz, D. A.; Woods, R. J.; Rose, D. R.; Boons, G. J. *J. Am. Chem. Soc.* **2006**, *128*, 8310–8319.
(36) Bryce, R. A.; Hillier, I. H.; Naismith, J. H. *Biophys. J.* **2001**, *81*, 1373–1388.

the conserved water had been displaced. Based on this assumption, the favorable entropy of binding for the modified ligand complex was attributed to the return of the highly ordered conserved water molecule to bulk solution, while the unfavorable enthalpy term was attributed to destruction of the strong hydrogen bonding network associated with the conserved water. Thus, the fact that ligand **2** had lower binding free energy with Con A than did ligand **1** was explained by favorable entropy (from water displacement) being offset by unfavorable enthalpy effects.[18]

To address the lack of experimental structural evidence for the displacement of this conserved water molecule, thermodynamic and modeling data notwithstanding, we undertook a crystallographic analysis of Con A complexed with ligand **2**. We report here the resulting crystal structure, which clearly indicates that although the synthetic ligand perturbs the position of the conserved water, it does not displace it. In light of the crystallographic data, a new and straightforward explanation for the measured thermodynamic binding data emerges.

Given the availability of crystal structures of the Con A complexes with **1** and **2**, we have performed molecular dynamics (MD) simulations of each, and computed ligand binding properties using thermodynamic integration (TI) calculations.

Results and Discussion

Structural Topology of the Con A:2 Complex. The overall β-sheet topology of the lectin structure is similar to the native fold described by Edelman et al.[37] The asymmetric unit consists of four 25 kDa, monomeric subunits, as does the biological form of the protein. Each monomer is a sandwich consisting of two antiparallel β-sheets, thus retaining the primary topology of other Con A:carbohydrate structures[13,15,27] (Supporting Information Figure S1). Each monomeric subunit of the lectin contains two metal binding sites; site S1 binds Mn and other transition metal ions like Ni, Co, Zn and Cd while site S2 binds only Ca and Cd ions.[22] The two sites are located approximately 12.5 Å and 8.5 Å respectively from the center of the carbohydrate binding site.[24,26] The presence of a divalent cation in S1 in turn induces the formation of S2 and it is well-known that both metal sites have to be occupied for the binding of the saccharide to take place.[38] The interaction of the protein residues with the metal ions molds the amino acid side chains into a shallow binding site where the carbohydrate consequently binds.[39] In all four chains of the ConA:**2** structure, the manganese ion is octahedrally coordinated to residues Glu9, Asp10, Asp19, His24 and two water molecules. The calcium ion is seven coordinate and interacts with two oxygen atoms from Asp10 as well as with residues Tyr12, Asn14, Asp19 and two water molecules. A similar coordination sphere for the two metal ions has been observed earlier in ConA:**1**.[13] The loop areas consisting of residues 117−122, 183−187 and 202−205 exhibit a high degree of mobile flexibility as illustrated by their high B-factors (Supporting Information Figure S1 and CIF).

Bound Ligand Conformation. The ligand, **2**, was modeled into well-defined difference electron density at the carbohydrate binding site. The conformation of **2** and its interactions with the protein were analyzed and compared with the structure of ConA:**1** complex (protein data bank, PDB, code 1CVN).[13] The specificity-determining interactions between the terminal 1→6-linked mannosyl residue in **1** or **2**, and the Con A protein surface are similar to those observed with other mannose-containing mono[26] and disaccharide[27] ligands. The central mannosyl residue (Man241) and the 1→3-linked residue (Man242) bind in a shallow secondary site bounded by loops formed by residues Tyr12, Asn14, Leu99, Tyr100, Asp208 and Arg228.[40]

Table 1. Average Dihedral Angles (deg) for Glycosidic Linkages of Di and Trimannosides in Complex with Con A[a]

linkage	angle	Con A:**2** (Crystal)	Con A:**2** (20 ns MD)	Con A:**1** (Crystal)[13,c]	Con A:**1** (10 ns MD)[18]
α(1→3)	φ	70.0 ± 1.5[b]	77 ± 12[b]	66.6	67 ± 9[b]
	Ψ	142.1 ± 1.8	112 ± 16	125.7	127 ± 17
α(1→6)	φ	65.7 ± 1.0	76 ± 10	66.0	64 ± 10
	Ψ	−169.7 ± 1.7	−174 ± 9	−171.0	−172 ± 9
	ω	−57.8 ± 2.0	−69 ± 14	−45.5	−58 ± 10

[a] The average values were computed from all crystallographically independent monomers in the asymmetric unit. [b] Errors are ± 1 standard deviation. [c] The disordered mannosides from chains A, D are not included in the calculation.

Table 2. Average Dihedral Angles (deg) and Rotamer Populations for the Hydroxyethyl Moiety in **2** in Solution and in Con A:**2**

linkage	Free **2** (20 ns MD)	Con A:**2** (Crystal)	Con A:**2** (20 ns MD)	Con A:**2** (10 ns MD)[18]
χ_1	−70 ± 14 (74%)	−74.4 ± 3.0	−69 ± 15 (92%)	−69 ± 14 (14%)
	−153 ± 12 (26%)		−146 ± 12 (8%)	−168 ± 13 (86%)
χ_2	179 ± 9 (46%)	179.9 ± 5.0	178 ± 11 (32%)	180 ± 14 (100%)
	71 ± 14 (24%)		77 ± 17 (15%)	−
	−63 ± 12 (30%)		−67 ± 13 (53%)	−

The conformation of the ligand may be conveniently described by the dihedral angles of the α(1→3) and α(1→6) glycosidic linkages (Table 1). The relevant dihedral angles from the ConA:**1** complex,[13] from a 20 ns MD simulation of the ConA:**2** complex and an earlier 10 ns MD simulation of the ConA:**1** complex[18] are included for comparison. The conformational preferences of the hydroxyethyl moiety may be defined by the χ_1- and χ_2-torsion angles (Table 2). Included for comparison are the values computed for the free and bound ligands from 20 ns MD simulations.

Conformations of the exocyclic groups in carbohydrates are driven by several factors, including steric hindrance (the χ_1- and χ_2-angles for the hydroxyethyl group in **2**, and the Ψ-angle), the *exo*-anomeric effect (the φ-angle), and the *gauche* effect (the ω-angle), see Figure 1. For the β-linkages in **1** and **2**, a value of approximately 60° for the φ-angle is consistent with expectations based on the *exo*-anomeric effect[41,42] (Table 1). For two-bond glycosidic linkages (1−2, 1−3 or 1−4) the global energy minimum for the Ψ-angle has been calculated to be approximately 120°.[43,44] In 1−6 linkages, this angle is more typically approximately 180°.[45] Thus, each of the glycosidic φ and Ψ-angles in **1** and **2** falls within expected ranges either free in solution[18] or bound to Con A (Table 1).

(37) Edelman, G. M.; Cunningham, B. A.; Reeke, G. N.; Becker, J. W.; Waxdal, M. W.; Wang, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2580–2584.
(38) Kalb, A. J.; Levitzki, A., *Biochem. J.* **1968**, *109*, 669–672.
(39) Becker, J. W.; Reeke, G. N.; Wang, J. L.; Cunningham, B. A.; Edelman, G. M. *J. Biol. Chem.* **1975**, *250*, 1513–1524.
(40) Mandal, D. K.; Bhattacharyya, L.; Koenig, S. H.; Brown, R., III; Oscarson, S.; Brewer, F. C. *Biochemistry* **1994**, *33*, 1157–1162.
(41) Juaristi, E.; Cuevas, G. *Tetrahedron* **1992**, *48*, 5019–5087.
(42) Wolfe, S.; Whangbo, M.-H.; Mitchell, D. J. *Carbohydr. Res.* **1979**, *69*, 1–26.
(43) Imberty, A.; Gerber, S.; Tran, V.; Pérez, S. *Glycoconjugate J.* **1990**, *7*, 27–54.
(44) Lemieux, R. U.; Koto, S. *Tetrahedron* **1974**, *30*, 1933–1944.
(45) Dowd, M. K.; Reilly, P. J.; French, A. D. *Biopolymers* **1994**, *34*, 625–638.

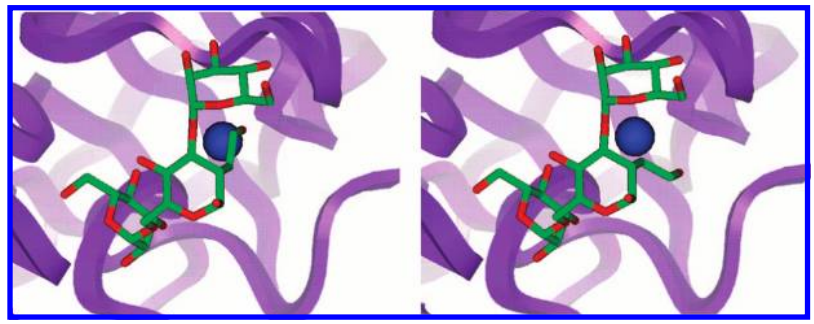


Figure 2. Illustration of the orientation of the hydroxyethyl side chain in ligand **2** bound to Con A required to displace the conserved water (left),[18] and that observed in the crystal structure of ConA:**2** (right). Protein shown as purple ribbon, conserved water as blue sphere and ligand as stick structure.

Three-bond linkages, such as the α(1−6) linkage in **1** and **2,** are relatively flexible due to the additional rotatable bond (*ω*-angle), which can potentially adopt three distinct rotamers referred to as *gauche*−*trans* (*gt*, $\omega \approx 60°$), *gauche*−*gauche* (*gg*, $\omega \approx -60°$), and *trans-gauche* (*tg*, $\omega \approx 180°$). In pyranoses that contain an equatorial hydroxyl group at the C-4 position (mannose, glucose, etc.), the gauche effect preferentially stabilizes the *gg* and *gt* rotamers over the *tg* rotamer.[46,47] The rotameric preferences associated with the gauche effect, such as the instability of the *tg* rotamer in these structures, has been shown to result from synperiplanar electrostatic repulsion between the O4 and O6 atoms, which dominates when the favorable O4 − O6 intramolecular hydrogen bonds are disrupted by water.[48,49] MD[18] simulations and solution NMR studies[50,51] of **1** and **2**, confirm the preference for the *gg* and *gt* rotamers. In the crystal structures of Con A complexed to either **1**,[13] or **2**, the *ω* torsion angle of the α(1,6) linkage adopts only the *gg* rotamer, which maximizes contact with the protein surface (Table 1).

MD simulations of free ligand **2** show that the dihedral angles of the hydroxyl ethyl side chain (χ_1, χ_2) populate five of the six possible rotamers in solution. These arise from two possible values for χ_1 (major and minor conformations, the 60° rotamer is unfavorable as it places the side chain directly below the pyranoses ring), and three possible values for χ_2 (Table 2). In contrast to the earlier prediction,[18] the crystal structure of ConA:**2** clearly indicates that the χ_1 angle of the hydroxyethyl moiety populates only the major conformation ($\chi_1 = -74.4°$). The current MD simulations also demonstrate a pronounced preference for this rotamer (>90%). The protein is able to accommodate the hydroxyethyl moiety in this -*gauche* orientation (of χ_1) without the need to displace the conserved water, although it does perturb the position of the water and its coordination sphere. Earlier modeling had suggested that if the side chain were to adopt the antiorientation (Figure 2), the hydroxyl group would be well positioned to replace the conserved water.[18] The fact that this did not occur provides insight into the stability of the interactions associated with the conserved water. It is significant that the bound conformation of the χ_1 angle in the crystal structure is also the dominant conformation of the ligand in solution. A -*gauche* orientation for χ_1, with an *anti*-orientation for χ_2, allows hydroxyl group O8 to form a hydrogen bond with the conserved water instead of displacing it (Table 2). However, the relatively free rotation of the χ_2-angle seen in the MD data for the ConA:**2** complex, suggests that the hydrogen bond between O8 and the conserved water hydrogen bond may be weak. This observation is consistent with the fact that the O8−water hydrogen bond is longer (2.81 Å) than the corresponding O2−conserved water hydrogen bond (2.45 Å) in the ConA:**1** complex (Table 3).

Table 3. Hydrogen Bond Distances (Å) Involving the Conserved Water Molecule (Wat_{Con}) in the Con A Complexes and Details of the Salt Bridge between Asp16 and Arg228

H-bond details	Con A (1GKB)	Con A:**1** (1CVN)	Con A:**2**
Wat_{con}···Asn14(Nδ2)	2.86	2.70	2.73 ± 0.08
Wat_{con}···Asp16(Oδ1)	2.64	2.74	2.59 ± 0.05
Wat_{con}···Asp16(Oδ2)	3.23	3.19	3.18 ± 0.08
Wat_{con}···Arg228(NH2)	3.09	3.16	3.03 ± 0.11
Wat_{con}···Man241(O2)		2.45	
Wat_{con}···Hem241(O8)			2.81 ± 0.13
Arg228(NH1)···Asp16 (Oδ2)	2.82	2.95	3.18 ± 0.28
Side chain torsion angles for Arg228	$\chi_1 = -176.0$	$\chi_1 = 94.0$	$\chi_1 = -175.0 \pm 3.4$
	$\chi_2 = 176.3$	$\chi_2 = 136.8$	$\chi_2 = -177.5 \pm 3.4$
	$\chi_3 = -178.9$	$\chi_3 = 175.5$	$\chi_3 = -177.9 \pm 3.2$

Coordination Geometry of the Conserved Water. The conserved water molecule acts as the bridging moiety between the protein and ligand by hydrogen bonding to Asn14, Asp16, Arg228 and to the central mannosyl residue (Man241) of ligand **2**. These interactions were analyzed by comparing the unliganded crystal structure of Con A[52] with the structures of Con A:**1**[13] and ConA:**2**. Similar interactions were seen in both Con A:**1** and Con A:**2** structures, in which the conserved water molecule hydrogen bonds to Asn14(Nδ2), Asp16(Oδ2) and Arg228(NH1) (Table 3). All three structures display a salt bridge between NH1 of Arg228 and Oδ2 of Asp16.

In Con A:**1**, the conserved water makes strong interactions with the ligand's Man241(O2) of **1** and with Arg228(NH2). The side chain of Arg228 is drawn more closely into the binding site in ConA:**1** than it is in the free protein (Figure 3), adopting altered χ_1 and χ_2 side chain angles relative to the free lectin (Table 3). In Con A:**2**, the conserved water hydrogen bonds to hydroxyl group O8 of the hydroxyethyl side chain and with Arg228(NH2) (Table 3). Compared to the position of this water in the ConA:**1** complex, the hydroxyethyl moiety distorts the position of the conserved water by approximately 1.5 Å. The interaction between the conserved water and the central residue in the synthetic ligand is weaker than the corresponding interaction with the natural trisaccharide as seen by the lengthening of the hydrogen bond between the conserved water

(46) Kroon-Batenburg, L. M. J.; Kroon, J. *Biopolymers* **1990**, *29*, 1243–1248.
(47) Wolfe, S. *Acc. Chem. Res.* **1972**, *5*, 102–111.
(48) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541–10545.
(49) Gonzalez-Outeiriño, J.; Kirschner, K. N.; Thobhani, S.; Woods, R. J., *Can. J. Chem.* **2006**, *84*, 569–579.
(50) Brisson, J.-R.; Carver, J. P. *Biochemistry* **1983**, *22*, 1362–1368.
(51) Sayers, E. W.; Prestegard, J. H. *Biophys. J.* **2000**, *79*, 3313–3329.

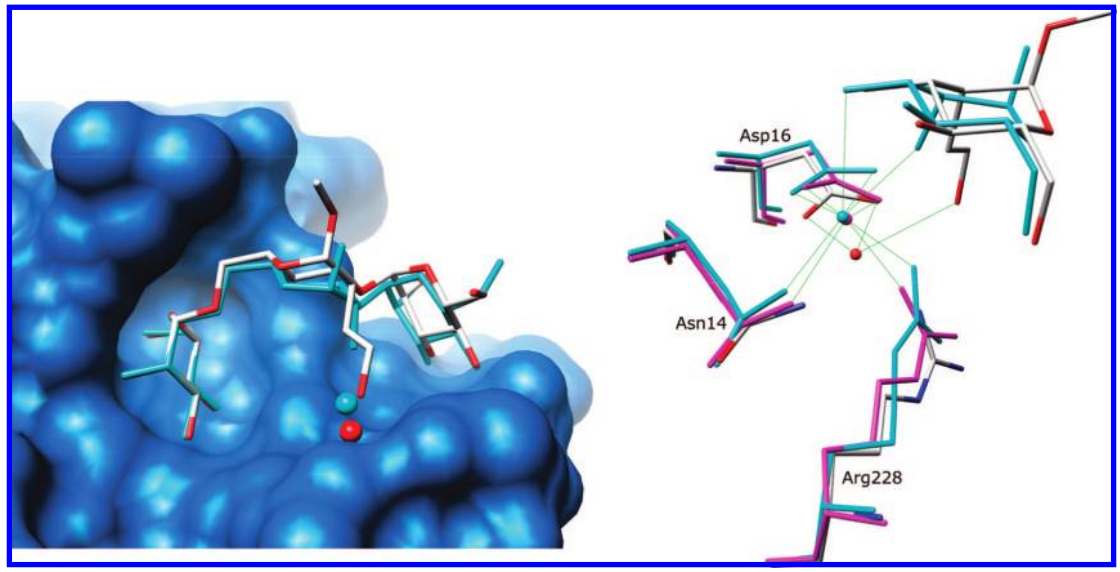


Figure 3. (Left) Distortion of the conserved water position by the hydroxyethyl moiety in ConA:**2** (atomic coloring) compared ConA:**1** (cyan). The solvent accessible protein surface is shown in light blue and the conserved waters as large spheres. (Right) The orientation of the side chain of Arg228 in free Con A (pink), Con A:**1** (cyan) and Con A:**2** (atomic coloring) with hydrogen bonds involving the conserved water molecule highlighted in green.

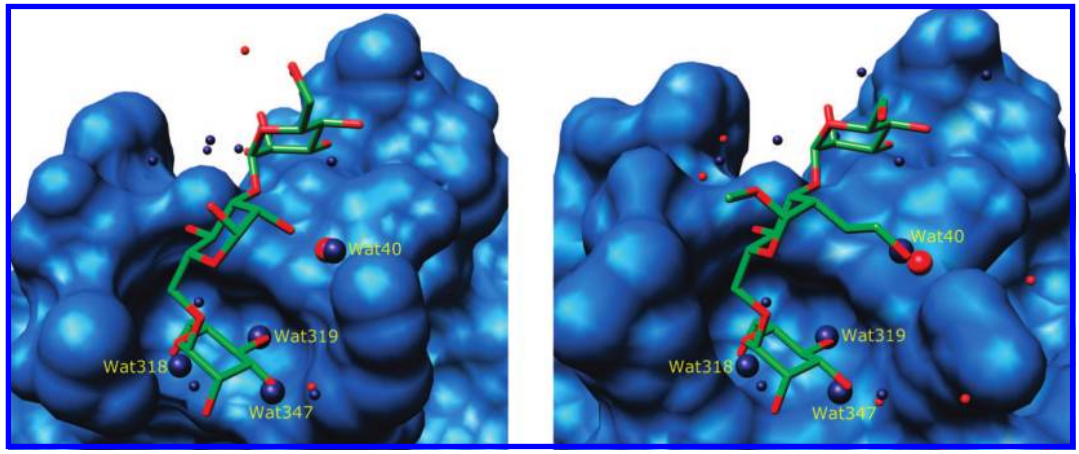


Figure 4. (Left) Solvent accessible surface of Con A (light blue) in complex with the natural trisaccharide ligand (Con A:**1**). (Right) Solvent accessible surface of Con A in complex with the synthetic trisaccharide ligand (Con A:**2**). The waters of crystallization in Con A:**1** and Con A:**2** are shown as red spheres. Superimposed on both structures are the waters of crystallization (dark blue) from the ligand-free structure of Con A. Key waters, which appear to be displaced by the ligand, as well as the conserved water molecule, are shown as larger spheres.

and O8 in the ConA:**2** complex (2.81 Å), relative to that with O2 in the ConA:**1** complex (2.45 Å). Since all other hydrogen bonds and nonbonded contacts are essentially equivalent between the two ligands and Con A (Supporting Information Tables S1 and S2), the distortion in the conserved water position is likely to be the primary source of the lower free energy of binding of **2** to Con A.

In contrast to ConA:**1**, in ConA:**2** the side chain of Arg228 maintains the position it displays in the free lectin,[52] possibly to facilitate the distortion of the conserved water position upon binding of the synthetic ligand. These structural alterations also lead to lengthening of the salt bridge between Arg228 and Asp16 in ConA:**2** (Table 3, Figure 3). Terminal residues maintain equivalent positions relative to the protein in ConA:**1** and ConA:**2** but, the central mannosyl residue in ConA:**2** is displaced from that in the natural trisaccharide. Thus, binding of the synthetic ligand introduces perturbations of the positions of both the conserved water (relative to its location in free Con A and in ConA:**1**), as well as of the core residue in the ligand (relative to its position in ConA:**1**).

Bound Waters in the Ligand Binding Site of Free Con A. An examination of the binding site in the superimposed structures of Con A:**1** and ligand-free Con A indicated that three water molecules in the carbohydrate binding site occupy positions that are critical for ligand binding (Figure 4), in addition to the conserved water (Wat40). Water-347 (Wat347) is situated in the periphery of the binding site and hydrogen bonds to the backbone nitrogen of Arg228. Wat319 hydrogen bonds to the side chain atoms N-δ of Asn14 and O-δ of Asp208, while Wat318 is stabilized by three interactions, namely, with the backbone nitrogen atoms of Leu99 and Tyr100 and the side chain O-δ atom of Asp208. Upon ligand binding, hydroxyl groups of the terminal mannosyl residue (Man240 in both **1** and **2**) displace these three water molecules (Figure 4). O3 in Man240 displaces Wat347, O4

(52) Kantardjieff, K. A.; Hochtl, P.; Segelke, B. W.; Tao, F. M.; Rupp, B. *Acta Crystallogr.* **2002**, *D58*, 735–743.

Table 4. Interaction Energies (ΔG) of **1** Relative to **2** with Water and with Con A, Computed Using TI

	Coulomb contribution[a]		van der Waals contribution[b]		total energies	
	100 ps/λ	500 ps/λ	100 ps/λ	500 ps/λ	100 ps/λ	500 ps/λ
			12 point Gaussian quadrature[c]			
Water	62.1 ± 0.4	62.3 ± 0.5	-4.6 ± 0.6	-4.6 ± 0.7	57.5 ± 0.7	57.7 ± 0.9
Con A	60.8 ± 0.5	60.5 ± 0.5	-6.3 ± 1.0	-6.6 ± 0.7	54.5 ± 1.1	53.9 ± 0.9
Relative Affinities	$\mathbf{-1.3 \pm 0.6}$	$\mathbf{-1.8 \pm 0.7}$	$\mathbf{-1.7 \pm 1.1}$	$\mathbf{-2.0 \pm 1.0}$	$\mathbf{-3.0 \pm 1.3}$	$\mathbf{-3.8 \pm 1.2}$
			21 point splined fit (100 ps/λ)[c]			
Water	62.2 ± 0.3	-4.5 ± 0.5	57.7 ± 0.6			
Con A	62.6 ± 0.4	-6.6 ± 0.5	56.0 ± 0.6			
Relative Affinities	$\mathbf{0.4 \pm 0.5}$	$\mathbf{-2.1 \pm 0.7}$	$\mathbf{-1.7 \pm 0.8}$			

[a] $k_\lambda = 1$. [b] $k_\lambda = 6$. [c] See Materials and Methods for details.

displaces Wat319, and O6 displaces Wat318 but they maintain the same hydrogen bonding pattern that Wat347, Wat319 and Wat318 exhibited with the protein, namely, Man240(O3)···R228(N) (2.9 Å), Man240(O4)···N14(Nδ2) (2.89 Å), Man240(O4)···D208(Oδ2) (2.66 Å), Man240(O6)···L99(N) (3.03 Å), Man240(O6)···Y100(N) (2.99 Å), and Man240(O6)···Asp208(O δ1) (2.99 Å). The binding of the native ligand induces only a modest displacement in the position of the conserved water (Wat40) relative to the ligand-free structure, but the synthetic ligand **2** significantly distorts this position (Figure 4).

Energetics of Ligand Binding. The difference in free energy between two states of a system may be computed by MD simulation using free energy perturbation methods, in which a continuous reaction coordinate path is defined (in terms of internal coordinate λ) between the initial ($\lambda = 0$) and final states ($\lambda = 1$). In thermodynamic integration (TI), the system is mutated from the initial to final states through a series of small accumulative perturbations in λ; the potential at any point along this path is defined by a mixing rule such as eq 1.

$$V(\lambda) = (1-\lambda)^{k} V_0 + [1-(1-\lambda)_k] V_1 \quad (1)$$

The change in free energy associated with the total perturbation is shown in eq 2.

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda \approx \sum w_i \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda_i} \quad (2)$$

The TI calculations employed the new crystallographic data for ConA:**2** and were performed by converting or "disappearing" the hydroxyethyl side chain in **2** to the corresponding hydroxyl group in **1**. While both forward and reverse directions of this conversion sequence should in principle result in the same value of ΔG, there is often an element of hysteresis, which may be used to estimate the error in the calculation.[53,54] Forward and reverse mutations were attempted, but the mutation from hydroxyl to hydroxyethyl proved to be problematic due to steric collisions that developed with the protein surface. An examination of the experimental structures indicated that these collisions were relieved in ConA:**2** by distortion of the position of the conserved water, which was accommodated by a difference in the orientation of the side chain of Arg228 (Figure 3). Such complex reorganizations are difficult to achieve in TI simulations and therefore we present the results for the disappearing mutation of the hydroxyethyl group along with estimates of the error from a statistical analysis of the intervening energies ($\partial V/\partial \lambda$) during the mutation (Table 4, and Materials and Methods). These mutations can be more precisely designated as "charging" and "insertion" components,[55] but for this discussion we have retained the terms "coulombic" and "van der Waals". In the latest version of AMBER,[56] soft-core potentials provide an additional way to perform thermodynamic integration calculations. In contrast to AMBER8 employed here, the system setup has been simplified so that appearing and disappearing atoms can be present at the same time and no dummy atoms need to be introduced.

Table 4 shows that the difference between the coulomb contributions to binding for each ligand are statistically insignificant at the highest sampling level (21 point), whereas the van der Waals interactions consistently favor the natural ligand (**1**) over the synthetic analog (**2**). For the most accurate computation (21 point sampling in λ), the favorable binding free energy of **1** relative to **2** of 1.7 kcal/mol compares reasonably with the corresponding difference in the experimental binding energies for **1** and **2** of 1.2 ± 0.08 kcal/mol.[18] Increased sampling had little effect on the average energies computed for the mutation in water, and led to only a slight decrease in their standard deviations, indicating that the mutation in water has already converged with 12 point sampling. In contrast, when the mutation was performed in the presence of the protein, both the average values of the energies (especially for the coulomb component) and the magnitude of their standard deviations were sensitive to the level of sampling. This sensitivity reflects the fact that the time scale for protein side chain motions is longer than that for water reorientation, and proportionally greater simulation and sampling are required to reach convergence. The side chain of Arg228 adopts slightly different orientations in each of the complexes, and during these simulations this side chain did not reorient from that present in ConA:**2** to that in ConA:**1**. It is likely that this feature contributes to the slight (0.5 kcal/mol) overestimation of the difference in binding free energies between **1** and **2**. Notably, increasing the data collection period from 100 to 500 ps per λ value did not alter the average ΔG values significantly, indicating convergence had been reached with 100 ps per λ of data collection. Convergence was further confirmed by plotting $\partial V/\partial \lambda$ versus data collection time for each λ value and by computing the autocorrelation functions, for both the coulomb and van der Waals perturbations (Supporting Information Figures S2−S11). This feature suggests that finer sampling along the perturbation pathway (21 point) is more important than increased sampling (500 ps) at a lower number of λ values (12 point).

(53) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.
(54) Pathiaseril, A.; Woods, R. J. *J. Am. Chem. Soc.* **2000**, *122*, 331–338.
(55) Shirts, M. R.; Pitera, J.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740–5761.
(56) Case, D. A.; et al. *AMBER 10* ; University of California: San Francisco, 2008.

Both trisaccharides participate in similar hydrogen bond interactions with the bound water and the protein surface (Table 3 and Supporting Information Tables S1 and S2) so the small difference in their coulomb interactions with ConA is not unexpected. The differential van der Waals contribution is estimated to be greater than that from Coulombic interactions and may be attributed to the larger atomic volume associated with the hydroxyethyl side chain in **2** (relative to the hydroxyl group in **1**), which requires a proportionally greater free energy to insert into water (4.5−4.6 kcal/mol). An even larger van der Waals insertion energy (6.3−6.6 kcal/mol) is required to accommodate the hydroxyethyl group ConA:**2**, resulting in the net preference for the natural trisaccharide. This feature is consistent with the experimental observation that the synthetic ligand distorts the geometry of the binding site in the vicinity of the conserved water, relative to the natural ligand.

Materials and Methods

Crystallization and Data Collection. Con A (Sigma-Aldrich, USA) was dissolved in 20 mM Tris buffer at pH 7 containing x10 molar fold concentration of the synthetic ligand (**2**), 50 mM NaCl, 1 mM $MnCl_2$, and 1 mM $CaCl_2$. Diffraction quality crystals of Con A:**2** complex were grown at 18 °C from hanging drops composed of 2 μL of the lectin/ligand solution and 5 μL of precipitant. The drops were equilibrated against 10−16% polyethylene glycol 6000 and 100 mM sodium cacodylate, pH 6.8. Low temperature diffraction data were collected at the A1 station at the Cornell High Energy Synchrotron Source (CHESS) using a Quantum-210 detector (Area Detector Systems Corporation, Poway, CA). A single crystal was cryoprotected with the precipitant containing 5% PEG 4000 and flash-cooled in a stream of nitrogen. Data were recorded as 269 nonoverlapping images with a 0.75° oscillation step and a 20-s exposure time. The HKL suite of programs[57] was used for data processing and scaling. Intensities were converted to structure factors using the program TRUNCATE from the CCP4 (Collaborative Computational Project Number 4, 1994)[58] suite of packages. The space group was $P2_1$ and the unit cell edges and angles at 120 K were $a = 60.015$ Å; $b = 63.547$ Å; $c = 126.215$ Å, $\beta = 86.91°$ respectively. The Matthew's coefficient was 2.4 Da/Å^3 and the solvent content in the asymmetric unit was 48%. The data collection statistics are summarized in Table 5.

Structure Solution and Refinement. The structure was solved using molecular replacement and the program AMoRe[60] with the Con A:Man(α1→3)Man-α1-OCH_3 complex (PDB code 1QDO) as the search model. Metal ions, carbohydrates and water molecules were removed from the search model and a single monomer used as input into AMoRe.[60] The molecular replacement solution was a tetramer. A rigid body refinement of the solution at 3Å resolution gave an R-factor of 0.363. A 10% subset of data was set aside to provide an unbiased assessment of the structure refinement.[61] The structure was refined using the CNS suite of programs.[62] Cross-validated sigma-A weighted[63] difference density maps were used to find the positions of the metal atoms, ligand **2**, water molecules and glycerol (Figure 5) which were then successively fitted into the electron density using O.[64] Each step of structure rebuilding was followed by simulated annealing and monitoring of refinement cycles using R_{free}. The identification of water molecules was subject to the following criteria: difference density ($F_o - F_c$) of 3σ or above, hydrogen bonding and the reappearance of the water molecule in the supporting 2Fo-Fc map at the 1σ level or more (Figure 5). Initial noncrystallographic symmetry restraints with a force constant of 200 kcal mol^{-1} Å^{-2} were gradually dropped to zero in the final cycles of refinement. The final model comprises of the Con A tetramer, eight metal atoms, four molecules each of ligand **2** and glycerol and 577 water molecules in the asymmetric unit. The structure has an *R*-factor of 0.205 and a R_{free} of 0.238 for data up to 1.8 Å. A Ramachandran plot[65] showed the residues tightly packed into the core and allowed regions of the map. Stereochemical parameters plotted using PROCHECK[66] showed that the quality of the model was better than a test set of 118 structures with a resolution of at least 2 Å and an average *R* factor of 0.20. A summary of the refinement statistics are in Table 5.

The atomic coordinates and structure factor amplitudes for ConA:**2** have been deposited in the RCSB Protein Data Bank[67] as entry 3D4K.

Table 5. X-Ray Data and Refinement Statistics

Data Collection	
Space Group	$P2_1$
Wavelength (Å)	0.935
Unit cell parameters *a, b, c* (Å)	60.015, 63.547, 126.215
β (°)	86.91
Resolution (Å)	30 to 1.8
Data completeness (%)	99.2 (98.7)[a]
Observed Reflections	358491
Unique Reflections	87269
Multiplicity	4.2
R_{sym}	0.066 (0.488)[a]
$\langle I/\sigma(I)\rangle$	19.7 (3.14)[a]

Refinement	
Number of protein atoms	7236
Number of water molecules	577
Number of metal and ligand atoms	156
rmsd bond length[b] (Å)	0.005
rmsd bond angles[b] (°)	1.53
Average temperature (B) factors (Å^2)	
All atoms	24.55
Main chain	22.43
Side Chain and water molecules	26.48
Ramachandran statistics (%)	
Core, additionally allowed regions	100
Disallowed, generously allowed regions	0
Crystallographic *R*-Factor	0.205
Crystallographic R_{free}	0.238

[a] Values in parentheses are for the highest resolution shell (1.86−1.80 Å). [b] Root mean square deviations from restraint targets[59]

(57) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307–326.
(58) Collaborative Computational Project, N. *Acta Crystallogr.* **1994**, *D50*, 760–763.
(59) Engh, R. A.; Huber, R. *Acta Crystallogr.* **1991**, *A47*, 392–400.
(60) Navaza, J. *Acta Crystallogr.* **1994**, *A50*, 157–163.
(61) Brunger, A. T. *Acta Crystallogr.* **1993**, *D49*, 24–36.
(62) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr.* **1998**, *D54*, 905–921.
(63) Read, R. J. *Acta Crystallogr.* **1986**, *A42*, 140–149.
(64) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. *Acta Crystallogr.* **1991**, *A47*, 110–119.
(65) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283–438.
(66) Bakina, E.; Wu, Z.; Rosenblum, M.; Farquhar, D. *J. Med. Chem.* **1997**, *40*, 4013–4018.
(67) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.

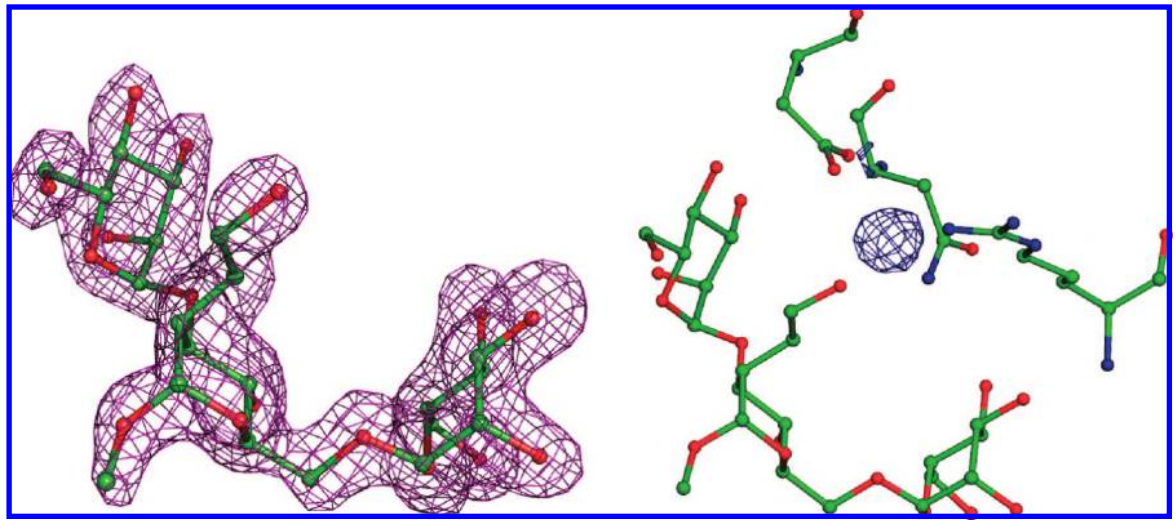


Figure 5. (Left) Ligand **2** with supporting $2F_o - F_c$ density (violet). Carbon atoms are colored green and oxygen atoms are shown in red. (Right) Difference density ($F_o - F_c$) map contoured at 3σ level with density corresponding to the conserved water shown in blue mesh.

Energy Minimization and Molecular Dynamics. MD simulations were performed under periodic boundary conditions at constant pressure and temperature (300 K) using the SANDER module of the AMBER 8 program,[68] employing the PARM99 all-atom force field for proteins[69] and the GLYCAM06 parameters for carbohydrates.[70] Partial charges and atom types employed with **1** and **2** are presented in Supporting Information Table S3 and Figure S12, respectively. For solvated simulations, each protein or protein-carbohydrate complex was placed in a theoretical box of TIP3P[71] water molecules with a minimum distance between the solute surface and the box face of 8 Å. For each system, initial solvent configurations were subjected to 20,000 cycles of energy minimization, composed of 15 000 cycles of steepest descent (SD) and 5000 cycles of conjugate gradient (CG) minimization, with the solute coordinates frozen. Subsequently, simulated annealing was performed on the solvent by heating the system from 5 to 300 K in 25 ps steps, maintaining the system at 300 K for 25 ps and cooling the system back to 5 K over 25 ps steps. Following simulated annealing the entire system was energy minimized (15 000 SD, 5000 CG). Finally, the system was reheated from 5 to 300 K over 25 ps steps and maintained there for an equilibration period of 250 ps. The 1−4 electrostatic and van der Waals interactions were scaled by the recommended values for carbohydrates (SCEE = 1.0, SCNB = 1.0).[48] These scale values differ from the standard values for protein simulations in AMBER of 1.2 and 2.0, respectively, [68] and in order to ensure the overall stability of the protein fold, a weak harmonic restraint was applied to all Cα positions (restraint weight = 22.0) during the MD simulations. Bond lengths involving hydrogens were constrained to equilibrium values using SHAKE.[72] A 2 fs time step was employed to integrate the equations of motion. Long range electrostatic interactions were computed using the particle mesh Ewald option,[73] with a 10 Å cutoff for nonbonded interactions. All MD and TI simulations were performed at constant pressure (1 atm) and temperature (300 K) as an nPT ensemble.

Initial Lectin Configurations. Con A contains 237 residues and structurally significant Mn238 and Ca239 ions. Due to the lack of force field parameters for Mn^{2+}, it was replaced by Ca^{2+}, an approximation employed in a recent simulation of Con A.[16] Explicit bonds were not defined between the ions and any coordinating groups. Histidines were assumed to be neutrally charged with the labile proton located at N-ϵ, all other ionizable side chains were assumed to be ionized. Water molecules coordinated to the metal ions or proximal to the ligand (in free protein, waters proximal to the equivalent position of the ligand) were retained; protons were added to them using the GWH package in AMBER 8 and they were treated as explicit TIP3P water molecules in the fully solvated simulations. The net charge on Con A (−3 au) was neutralized by 3 Na^+ ions, placed in electrostatically optimal locations using tLEaP.[68] In each system only a monomer of Con A was extracted from the crystal structures for analysis and simulation.

System-Specific Details. Con A with Synthetic Trisaccharide 2 in Water (Chain A): Waters 250 and 447 (coordinated to Ca^{2+} 239), 307 and 365 (coordinated to Mn^{2+} 238), and all waters within 5 Å of the carbohydrate ligand (368, 383, 390*, 568, 607, 700, 707, 759, 765) were included in the initial configuration. The asterisk indicates the conserved water molecule (which corresponds to water 85 in the Con A:1, 1CVN[13]). The fully solvated system contained 7967 TIP3P water molecules for a total of 27 545 atoms in a periodic box with dimensions 65.717 × 80.972 × 65.734 Å.

Con A with Native Trisaccharide 1 (Extracted from 1CVN,[13] Chain A): Waters 1 and 20 (coordinated to Ca^{2+} 239), 4 and 32 (coordinated to Mn^{2+} 238), and all waters within 5 Å of the carbohydrate ligand (35, 55, 85*, 112, 146, 198, 221, 257) were included in the initial configuration. The system contained 7086 TIP3P water molecules for a total of 24 893 atoms in a periodic box with dimensions 65.385 × 77.306 × 63.403 Å.

(68) Case, D. A.; et al. *AMBER 8*; University of California: San Francisco, 2004.
(69) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
(70) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.
(71) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Phys. Chem.* **1983**, *79*, 926–935.
(72) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.
(73) Darden, T. A.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

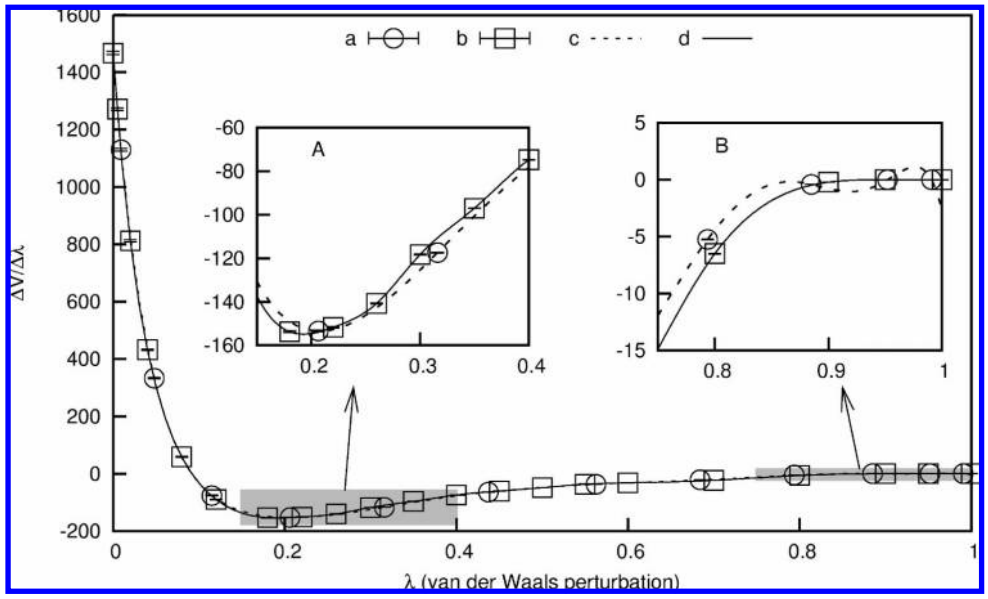


Figure 6. Comparison of results from two integration techniques. (a) 12-point Gaussian quadrature integration; (b) 21-point custom curve; (c) 11-degree polynomial fit to data in a; (d) cubic spline fit to data in b. (The following λ values and integration weights (w) were used[68] for 12 point quadrature (λ/w: 0.00922/0.02359, 0.04794/0.05347, 0.11505/0.08004, 0.20634/0.10158, 0.31608/0.11675, 0.43738/0.12457, 0.56262/0.12457, 0.68392/0.11675, 0.79366/0.10158, 0.88495/0.08004, 0.95206/0.05347, 0.99078/0.02359). For the Coulombic integration via cubic spline, λ values were spaced at intervals of 0.05 from 0 to 1. For the van der Waals integration via cubic spline, λ values were chosen to reflect integrand curvature: 0, 0.005, 0.02, 0.04, 0.08, 0.12, 0.18, 0.22, 0.26, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8, 0.9, 0.95, 1).

Synthetic Trisaccharide Ligand 2 in Water: Crystallographic waters included in the simulation of ConA:**2** were included here for consistency. The system contained 829 TIP3P water molecules for a total of 2562 atoms in a periodic box with dimensions 35.010 × 34.259 × 34.229 Å.

Thermodynamic Integration (TI). Charge and van der Waals perturbations were performed with mixing exponents of $k = 1$ (coulomb) and 6 (van der Waals), as recommended.[68] General protocols were equivalent to MD simulations except that an integration step size of 1 fs was employed to maintain equilibrium. Coulomb mutations were performed prior to van der Waals.

Equilibration and Data Collection: Prior to TI calculations, systems were subjected to pre-equilibration as follows: fully relaxed minimization of the solvent, counterions, ligand and protein (1500 SD, 2500 CG) followed by simulated annealing of the noncrystallographic waters (solvent) and counterions only (heating from 5 K → 300 K in 50 ps, maintenance at 300 K for 100 ps, cooling to 5 K in 50 ps) and a fully relaxed minimization of the solvent, crystallographic waters, counterions, ligand and protein (1500 SD, 2500 CG). During the TI phase of each simulation, all restraints, other than on the Cα positions, were removed and a sequential protocol was followed, in which equilibration (20 ps at 300 K for each value of λ) was performed employing the coordinate and velocity information from either the preceding λ value, or from the initial equilibration stage. After each equilibration stage, data collection was performed (100 ps at 300 K for each value of λ) employing the coordinate and velocity information from the preceding equilibrated λ value. To assess the effect of a longer data collection window on the computed free energies, $\partial V/\partial\lambda$ values were also collected by extending the data collection period by 500 ps, in a nonsequential protocol, from the end of each of the 120 ps runs (Supporting Information Figures S2−S11).

Numerical Integration of $\partial V/\partial\lambda$: Gaussian quadrature[74] applies λ values and corresponding integration weights w_i optimal for the integration of the polynomial function. But, many $\partial V/\partial\lambda$ curves can be approximated by a fit to a polynomial of appropriate degree. Figure 6 illustrates two integration techniques for a perturbation in the van der Waals component of the system's energy (for the Rha2 2-OH → 2-H mutation in water). The data were collected at positions appropriate to 12-point Gaussian quadrature integration at a set of 21 customized integration points so that the point density increases in areas of greatest curvature and near the end points. Points were so chosen to produce well-behaved splines.

In Figure 6, the 12-point set has been fit to an 11-degree polynomial and the 21-point set smoothed by a cubic spline. The polynomial fit introduces ripples near $\lambda = 1$ that is not directly implied by the data (Figure 6, inset B). The cubic spline does not introduce such unexpected structure and so we used trapezoidal integration of a cubic spline fit to a customized set of integration points. For this integration, the difference in ΔG between the 12-point quadrature and the spline fit was approximately 5%.

Error Estimation in TI Calculations: Errors in average $\partial V/\partial\lambda$ values were determined based on statistical efficiency, s, of the simulations in order to obtain statistically significant error estimate in the ΔG values.[75] The overall error in the integrations is given by:

$$\sigma^2_{\Delta G} = \sum_i w_i * \sigma_i^2(\langle \partial V/\partial\lambda \rangle_{\text{tot}})$$

(74) Hummer, G.; Szabo, L. *J. Chem. Phys.* **1996**, *105*, 2004–2010.
(75) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford Science Publications: Oxford, 1987.

weight$_i$ is the fraction of the integration range for which the measurement is relevant.

Conclusions

The current work investigates the role of water molecules in the binding of Con A to saccharide ligands and addresses the presumed displacement of an ordered, conserved water molecule from the Con A binding site. The crystal structure of Con A bound to the modified trimannoside **2** revealed the irrefutable presence of this conserved water molecule in the lectin binding site (Figure 5), and enabled MD and TI calculations to assess the influence of this water on ligand affinity.

The results from the TI calculations at 298 K ($\Delta\Delta G_{\text{binding}} \approx$ 2kcal/mol, in favor of ligand **1**) are in qualitative agreement with experiment ($\Delta\Delta G_{\text{binding}} = 1.2 \pm 0.9$ kcal/mol). The experimental data[18] indicated that the enthalpic component of the binding free energy preferred ligand **1** over ligand **2** ($\Delta\Delta H_{\text{binding}} = 2.3 \pm 0.9$ kcal/mol), while the entropic component slightly preferred ligand **2** ($\Delta\Delta S_{\text{binding}} = -1.1 \pm 0.9$ kcal/mol). The current results provide a straightforward interpretation for these enthalpic and entropic differences in binding. In order to accommodate the hydroxyethyl moiety in ligand **2**, the location and coordination of the conserved water is distorted from the equilibrium position it occupies in unliganded Con A (1GKB) and the ConA:**1** complex (Figures 3 and 4). This distortion leads to less optimal contacts between the central mannosyl residue of ligand **2** and the protein (Figure 3), relative to the case of **1**, which is seen experimentally as an unfavorable $\Delta\Delta H_{\text{binding}}$ component. The difference in the entropic components may then be assigned to the relief of enthalpic constraints, that is, the difference is a manifestation of enthalpy−entropy compensation. Overall, the complexation of Con A with ligand **2** is driven by enthalpy rather than entropy in the same manner as that of ligand **1**. In the ConA:**1** complex, this conclusion has been supported by recent computational results employing inhomogeneous fluid solvation theory.[16] The displacement of the conserved water from the binding site of Con A is not trivial, and its presence in all crystal structures of Con A complexes suggests that it is tightly bound. While there may be an entropic gain obtained from displacing this water from the protein surface, the maximum entropic benefit from releasing a tightly bound water has been estimated to be approximately 2 kcal/mol.[76] Thus, in the case of Con A, it is likely that the enthalpic loss from displacement of the conserved water would outweigh any entropic gain.

Acknowledgment. We thank the National Institutes of Health (GM55230, RR05357) for financial support, and Professor G.-J. Boons (UGA) for generously providing a sample of the synthetic ligand **2**. J.D.D. thanks the Carlsberg Foundation for a postdoctoral fellowship.

Supporting Information Available: CIF for **2**. Figure S1: Structure of ConA:**2** with details of the carbohydrate binding site and secondary structural elements colored according to B factors of amino acid residues. Figures S2−S7: Plots of $\partial V/\partial\lambda$ for the coulomb and van der Waals mutations. Figures S8−S11: Autocorrelation plots for the coulomb and van der Waals mutations. Tables S1 and S2: Hydrogen bond details for interactions between Con A and **2** for all chains in Con A. Table S3 and Figure S12: The GLYCAM partial charges and atom types for **1** and **2**. Complete refs 56 and 68. This material is available free of charge via the Internet at http://pubs.acs.org.

JA8039663

(76) Dunitz, J. D. *Science* **1994**, *264*, 670.