Relative Energies of Binding for Antibody−Carbohydrate-Antigen Complexes Computed from Free-Energy Simulations

Ahammadunny Pathiaseril and Robert J. Woods*

Contribution from the Complex Carbohydrate Research Center, The University of Georgia, 220 Riverbend Road, Athens, Georgia 30602

Received May 6, 1999. Revised Manuscript Received October 28, 1999

Abstract: Free-energy perturbation (FEP) simulations have been applied to a series of analogues of the natural trisaccharide epitope of *Salmonella* serotype B bound to a fragment of the monoclonal anti-*Salmonella* antibody Se155-4. This system was selected in order to assess the ability of free-energy perturbation (FEP) simulations to predict carbohydrate−protein interaction energies. The ultimate goal is to use FEP simulations to aid in the design of synthetic high affinity ligands for carbohydrate-binding proteins. The molecular dynamics (MD) simulations were performed in the explicit presence of water molecules, at room temperature. The AMBER force field, with the GLYCAM parameter set for oligosaccharides, was employed. In contrast to many modeling protocols, FEP simulations are capable of including the effects of entropy, arising from differential ligand flexibilities and solvation properties. The experimental binding affinities are all close in value, resulting in small relative free energies of binding. Many of the $\Delta\Delta G$ values are on the order of 0−1 kcal mol^{-1}, making their accurate calculation particularly challenging. The simulations were shown to reasonably reproduce the known geometries of the ligands and the ligand−protein complexes. A model for the conformational behavior of the unbound antigen is proposed that is consistent with the reported NMR data. The best agreement with experiment was obtained when histidine 97H was treated as fully protonated, for which the relative binding energies were predicted to well within 1 kcal mol^{-1}. To our knowledge this is the first report of FEP simulations applied to an oligosaccharide−protein complex.

Introduction

An ever increasing number of biological processes, ranging from cell−cell interactions necessary for fertilization of mammalian eggs to the immune system's reaction to foreign antigens, are being reported that depend on the recognition of specific carbohydrates by other molecules.[1−3] Despite the significance of these processes, the underlying mechanisms remain largely undetermined. A detailed understanding of carbohydrate recognition requires the ability to perform structure−function studies. Unfortunately, these studies have been hampered in part by the extremely time-consuming nature of carbohydrate synthesis. Computational methods have a long history of application to carbohydrate conformational analysis,[4−10] and more recently to the study of carbohydrate−protein complexes.[11−13] Although the ability to predict the effects of chemical modification on ligand−receptor binding affinity remains one of the most challenging areas in computational chemistry, free-energy perturbation (FEP) is perhaps the most promising computational approach.[14] In contrast to computational approaches based on energy minimization, which are enthalpy driven, the dynamics-based FEP approach includes both entropic and enthalpic contributions to the binding energy. Entropic contributions play a significant role in carbohydrate-binding free energies.[15] The accuracy of the computed free energy depends in part on the similarity of the initial and final states of the simulation.[16] This similarity extends not only to the structures of the solutes, but to the solvent as well. By computing the free energy relative to a reference state, it is possible to minimize some of the systematic errors. The application of relative free-energy simulations to protein−ligand binding studies, with groups of closely related ligands, can lead to predicted relative binding energies that are accurate to within approximately 1 kcal mol^{-1}.[17−20] To our knowledge, only two

* Author to whom correspondence may be addressed. Telephone: (706) 542-4454. Fax: (706) 542-4401. E-mail: rwoods@ccrc.uga.edu.

(1) Lis, H.; Sharon, N. *Chem. Rev.* **1998**, *98*, 637−674.
(2) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683−720.
(3) Varki, A. *Glycobiology* **1993**, *3*, 97−130.
(4) Woods, R. J. In *The Application of Molecular Modeling Techniques to the Determination of Oligosaccharide Solution Conformations*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1996; Vol. 9, Chapter 3.
(5) Pérez, S.; Imberty, A.; Carver, J. P. *Adv. Comput. Biol.* **1994**, *1*, 147−202.
(6) Brady, J. W. *Curr. Opin. Struct. Biol.* **1991**, *1*, 711−715.
(7) French, A. D.; Brady, J. W. *Computer Modeling of Carbohydrate Molecules*; American Chemical Society: Washington, DC, 1990; Vol. 430.
(8) Pincus, M. R.; Scheraga, H. A. *Macromolecules* **1979**, *12*, 633−644.
(9) Pincus, M. R.; Burgess, A. W.; Scheraga, H. A. *Biopolymers* **1976**, *15*, 2485−2521.
(10) Pincus, M. R.; Burgess, A. W.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 468.
(11) Liang, G.; Schmidt, R. K.; Yu, H. A.; Cumming, D. A.; Brady, J. W. *J. Phys. Chem.* **1996**, *100*, 2528−2534.
(12) Imberty, A.; Pérez, S. *Glycobiology* **1994**, *4*, 351−366.
(13) Zacharias, M.; Straatsma, T. P.; McCammon, J. A.; Quiocho, F. A. *Biochemistry* **1993**, *32*, 7428−7434.
(14) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(15) Chervenak, M. C.; Toone, E. J. *J. Am. Chem. Soc.* **1994**, *116*, 10533−10539.
(16) Pearlman, D. A.; Kollman, P. A. In *Free Energy Perturbation Calculations: Problems and Pitfalls Along the Gilded Road*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: The Netherlands, 1989; pp 101−119.
(17) Wlodek, S. T.; Antosiewicz, J.; McCammon, J. A.; Straatsma, T. P.; Gilson, M. K.; Briggs, J. M.; Humblet, C.; Sussman, J. L. *Bioploymers* **1996**, *38*, 109−117.

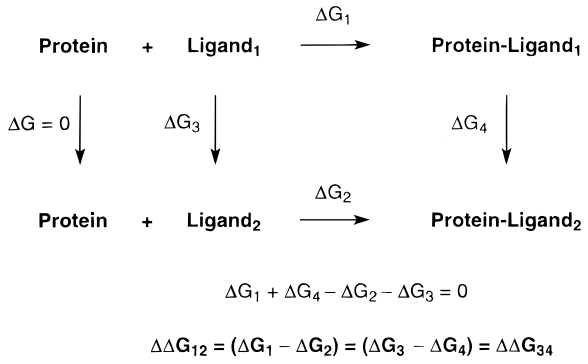
10.1021/ja9914994 CCC: $19.00
Published on Web 12/23/1999

reports exist of the application of FEP simulations to carbohydrate−protein complexes. In a report by Zacharias et al., FEP simulations were applied to study the differential binding between arabinose and fucose with arabinose binding protein (ABP).[13] More recently, Liang et al. examined the binding of mannose versus galactose with a mannose binding protein (MBP).[11] Both systems involved the binding of monosaccharides to lectins and led to data that were consistent with experiment. ABP and MBP are well-studied carbohydrate-binding lectins, which display extensive networks of intermolecular hydrogen bonds.[21,22] The binding between MBP and mannose is also mediated by coordination to a calcium ion.

In contrast to ABP and MBP, the interaction between the carbohydrate *O*-antigen of *Salmonella* serogroup B and monoclonal antibody (mAb) Se155-4 is remarkably hydrophobic. The hydrophobic character arises from the presence of a 3,6-dideoxypyranosyl residue (abequose, Abe) in the carbohydrate epitope. This system has been studied in solution by NMR and in the solid state with X-ray diffraction and is perhaps the most extensively characterized carbohydrate−antibody complex.[23,24] As part of an epitope-mapping study, several analogues of the natural ligand have been synthesized and their free energies of binding to the Fab fragment of Se155-4 reported.[25,26] To date, none of the synthetically modified antigens displays significantly greater affinity for the antibody than the naturally occurring antigen, thus the ligands can be coarsely classified as either binding, all with similar affinities, or not binding. Because of the biological significance of carbohydrate−antibody interactions, as well as the extent of the existing structural and thermodynamic data available for this system, it provides a valuable test case for computational methods. We undertook simulations of this system in order to quantify the ability of free-energy perturbation (FEP) methods, employing our GLYCAM parameter set,[27] to reproduce the subtle differences in binding affinities of the known analogues of this carbohydrate antigen. The ultimate goal of these studies is to apply FEP simulations in the prediction of novel high affinity ligands.

For this investigation we selected several analogues, which probe the effects of changes in direct hydrogen bonds, solvent-mediated hydrogen bonds and van der Waals interactions between the antigen and the antibody. By testing the ability of the FEP simulations to reproduce the binding affinities of structures containing these modifications, we have obtained an estimate of the general applicability of these calculations. Three different carbohydrate parameter sets (all variants of the GLYCAM model) and two protein parameter sets (based on the AMBER model) were employed in the simulations.

Scheme 1



Computational Approach

During binding, the ligand, receptor, and solvent all undergo extensive conformational and orientational changes. The complexity of this process severely limits the accuracy of a directly computed ΔG value. Furthermore, biological binding processes occur over time scales beyond the reach of most simulation methods. However, it is frequently only necessary to know the relative energy of binding ($\Delta\Delta G_{12}$ in Scheme 1) to assess the ability of one ligand to compete with another for binding to a common receptor. Due to the complexities associated with the binding phenomenon, $\Delta\Delta G$ values are rarely computed as the difference between the ΔG values for the physical processes. Instead, $\Delta\Delta G$ values are derived from the nonphysical mutation of one ligand into the other. This approach does not require knowledge of the mechanism of binding, but relies only on knowledge of the properties of the bound and free states for each ligand. This protocol relies on the fact that the total free energy for any closed thermodynamic cycle is zero by definition, and that such a cycle may include steps that represent nonphysical processes (see Scheme 1).

From the thermodynamic cycle it can be seen that the relative free energies for the theoretical mutation ($\Delta\Delta G_{34}$) are identical to the experimental relative free energies ($\Delta\Delta G_{12}$). The theoretical free energies can be readily calculated in a FEP simulation, although they are not experimentally measurable. The theoretical $\Delta\Delta G_{34}$ represents the preference for the ligands to bind to the protein, relative to remaining in solution. Over the course of a free-energy simulation, the change in free energy is recorded as state A is slowly converted into state B.[28] The ability to mutate one system into another is clearly a nonphysical process, but is nevertheless feasible in a theoretical model, in which each system is defined by a set of equations. In a stepwise FEP simulation the free-energy difference between states A and B is computed by dividing the states into nonphysical intermediate states. These intermediate states are characterized by a coupling parameter λ, such that $\lambda = 0$ corresponds to State A, $\lambda = 1$ to B, and all intermediate states to a linear combination of A and B. The free energy of the individual states is computed from the Hamiltonian (H).

$$\Delta G_\lambda = -RT \ln\langle \exp(-[H_{\lambda+\Delta\lambda} - H_\lambda]/RT)\rangle_\lambda \quad (1)$$

The total free-energy difference between states A and B is given by the sum of the free energies at each state

$$\Delta G = \sum_{\lambda=0}^{\lambda=1} \Delta G_\lambda \quad (2)$$

Computational Details. All simulations were performed using the Amber 4.1 program,[29] running on an Origin 2000 computer from Silicon Graphics Inc. Initial coordinates for the antibody Fv fragment of murine

(18) Wong, C. F.; McCammon, A. J. *J. Am. Chem. Soc.* **1986**, *108*, 3830−3832.
(19) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. *Science* **1987**, *235*, 574−576.
(20) Reddy, M. R.; Varney, M. D.; Kalish, V.; Viswanadhan, V. N.; Appelt, K. *J. Med. Chem.* **1994**, *37*, 1145−1152.
(21) Weis, W. I.; Drickamer, K.; Hendrickson, W. A. *Nature* **1992**, *360*, 127−134.
(22) Gilliland, G. L.; Quiocho, F. A. *J. Mol. Biol.* **1981**, *146*, 341−362.
(23) Zdanov, A.; Li, Y.; Bundle, D. R.; Deng, S.-J.; MacKenzie, C. R.; Narang, S. A.; Young, N. M.; Cygler, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6423−6427.
(24) Bundle, D. R.; Baumann, H.; Brisson, J.-R.; Gagné, S. M.; Zdanov, A.; Cygler, M. *Biochemistry* **1994**, *33*, 5183−5192.
(25) Bundle, D. R.; Alidés, R.; Nilar, S.; Otter, A.; Warwas, M.; Zhang, P. *J. Am. Chem. Soc.* **1998**, *120*, 5317−5318.
(26) Bundle, D. R.; Eichler, E.; Gidney, M. A.; Meldal, M.; Ragauskas, A.; Sigurskjold, B. W.; Sinnot, B.; Watson, D. C.; Yaguchi, M.; Young, N. M. *Biochemistry* **1994**, *33*, 5172−5182.
(27) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832−3846.
(28) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(29) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. R.; Cheatham, T. E., III; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. In *AMBER 4.1*; Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. R., Cheatham, T. E., III, Ferguson, D. M., Seibel, G. L., Singh, U. C., Weiner, P. K., Kollman, P. A., Eds.; University of California: San Francisco, 1995.

antibody Se155-4 complexed with the trisaccharide antigen: α-D-Gal*p*(1−2)[α-D-Abe*p*(1−3)]-α-D-Man*p*(1-OMe) (**1**) were retrieved from the Brookhaven protein database (pdbid = 1mfa).

The all-atom AMBER force field was employed for the protein,[30] augmented by the GLYCAM parameter set for oligosaccharides and glycoproteins.[27] Water was treated explicitly with the TIP3P model.[31] A cutoff distance of 8 Å was employed for all nonbonded interactions. In principle, the use of a larger nonbonded cutoff distance would be desirable; however, we chose to select a value that offered a reasonable compromise between accuracy and simulation times. A larger value is not practical, given the large size of this system and the desire to compute the binding energies for several ligands. Two features compensate in part for any errors arising from our choice of cutoff distance. First, in the GLYCAM_98R parameter set restrained electrostatic potential (RESP) charges were used. These charges are considerably lower in magnitude than the ESP charges used in GLYCAM_93. The use of RESP charges decreases the distance over which electrostatic interactions are significant. Second, some cancellation of errors may be expected as a result of the fact that we are computing $\Delta\Delta G$ values. A constant dielectric of unity was used throughout. All of the bond lengths were constrained to their initial values with the SHAKE algorithm.[32]

Force Field Parameters. The GLYCAM_93 partial atomic charges were computed by fitting the classical coulomb model to quantum molecular electrostatic potentials computed at the 6-31G* level (ESP-charges). We have shown that this approach is consistent with the TIP3P water model.[27] MD simulations performed with this protocol have been shown to lead to conformational ensembles that are consistent with solution NMR data.[33,34] In contrast, the AMBER PARM91 parameter set for proteins and nucleic acids employed ESP-charges computed from a STO-3G wave function, which results in charges that are lower in magnitude than the corresponding charges computed from a 6-31G* wave function. Independently, each parameter set performs well; however, there is an apparent inconsistency in the treatment of electrostatic interactions when the parameter sets are combined for studying carbohydrate−protein complexes. The precise extent to which this inconsistency influences computed binding free energies is unknown; however, it may be expected to lead to an underestimation of the electrostatic interactions.

After our development of the GLYCAM_93 parameters, a revised version of the AMBER protein parameters was reported, in which the STO-3G charges were replaced by charges computed at the 6-31G* level (PARM94). The PARM94 parameters were also developed for application with a 1−4 scaling factor of 0.833 (SCEE = 1.2), in contrast to a value of 0.5 (SCEE = 2.0) recommended for use with the PARM91 parameters. To be consistent with PARM94, we modified our GLYCAM_93 parameters for use with a 1−4 scaling factor of 0.833. These parameters are referred to as GLYCAM_98.

Although the partial atomic charges for both GLYCAM and PARM94 were derived by fitting to molecular electrostatic potentials computed at the 6-31G* level, during the derivation of the PARM94 charges a restrained fitting procedure was employed leading to so-called RESP-charges.[35] The restrained fitting procedure reduces the excessive polarity often exhibited with ESP-charges. We have performed the FEP simulations both with (GLYCAM_98R) and without (GLYCAM_98 and GLYCAM_93) RESP-charges (charges provided in Supporting Information, Tables S1 and S2). The RESP-charges were generated from the 6-31G* quantum mechanical data by fitting with the RESP module of AMBER,[35] employing a single stage fitting and a restraint weight of 0.01. In contrast to the majority of carbohydrate parameter sets, the GLYCAM model treats every atom uniquely insofar as the partial atomic charges are concerned. This leads to partial charges that reflect both the hydroxyl group configurations and the linkage positions. We believe this treatment provides a more complete description of electrostatic interactions, which are particularly relevant to carbohydrate−protein complexes.

Mutation of the Ligand in the Combining Site of the Protein. A spherical droplet of water with a radius of 18 Å was placed over the complex using the CAP option of AMBER, centered on the center of mass of the ligand. All water molecules within 2.5 Å of any solute atom were removed. This approach was selected because it involved fewer water molecules (∼350) than a full periodic boundary treatment (∼3300) and thus was less computationally intensive. Compared to a test periodic boundary simulation the droplet simulations appeared to be approximately 6-fold faster. Nevertheless, each droplet simulation required approximately 125 h of CPU time. Complete motional freedom was given to the waters, the ligand, and any protein residues within 4.5 Å of any atom in the ligand. This active-site region incorporated 102 amino acid residues. Although the droplet model does not lead to a thermodynamically rigorous treatment of the molecular complex, it has been successfully applied in free-energy simulations.[11,13] The success of this protocol benefits from the fact that the thermodynamic cycle leads to a cancellation of systematic errors.

The energy of the water molecules was minimized with 500 cycles of steepest descent followed by 3500 cycles of conjugate gradient minimization during which all of the atoms in the antibody and ligand were kept frozen. This was followed by energy minimization of the entire system, following the same minimization protocol. The MD simulations employed a 1 fs integration time step. The water was subjected to a simulated annealing sequence in which the water molecules were heated from 0 to 300 K in 10 ps, equilibrated for 10 ps, and then cooled to 0 K in 10 ps. The entire complex was then heated to 300 K in 10 ps and equilibrated for a further 10 ps before beginning the FEP mutation.

The mutations were performed through 11 windows, using a stepwise perturbation with double-wide sampling. A period of 10 ps was allowed for equilibration at each window, followed by a 20 ps period for data collection. The mutations were performed in both a forward and reverse direction. The trajectory at the end of the forward run was equilibrated for 50 ps prior to performing the mutation in the reverse direction. The hysteresis in the energy between the initial state and the state at the end of the reverse mutation was used to estimate the error associated with the computed free energies.

Mutation of the Ligand in Water. The ligand was immersed in a periodic box of 789 water molecules. Minimization and simulated annealing were performed as for the complex. The system was then warmed from 0 to 300 K over 50 ps followed by equilibration at 300 K for 1000 ps. The configuration at the end of this long MD simulation was used as the starting point for the FEP mutations. The mutations were performed in 21 windows using a stepwise perturbation with double-wide sampling. A period of 10 ps for equilibration was allowed at each window, followed by a 5 ps period for data collection. The mutations were performed in both forward and reverse directions. Total simulation time for each mutation was approximately 32 h, on a single R10000 CPU.

Discussion

The antigen specificity of Se155-4 is determined by a 3,6-dideoxy-pyranosyl (abequosyl, Abe) residue (see Figure 1), which occupies a cavity approximately 8 Å deep by 7 Å wide at the interface between the light and heavy chains of the antibody hypervariable region. This interaction exhibits a very high degree of shape and charge complimentarity between the protein and the carbohydrate. On the basis of the reported data, changes in the functionality of the Abe residue have the largest effect on binding, whereas changes within either the galactosyl or mannosyl residues have little impact.[26]

Experimental studies have identified three OH groups that participate in the antibody−oligosaccharide hydrogen bonding

(30) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252.
(31) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 341−345.
(32) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327−341.
(33) Brisson, J.-R.; Uhrínova, S.; Woods, R. J.; van der Zwan, M.; Jarrell, H. C.; Paoletti, L. C.; Kasper, D. L.; Jennings, H. J. *Biochemistry* **1997**, *36*, 3278−3292.
(34) Woods, R. J.; Pathiaseril, A.; Wormald, M. R.; Edge, C. J.; Dwek, R. A. *Eur. J. Biochem.* **1998**, *258*, 372−386.
(35) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

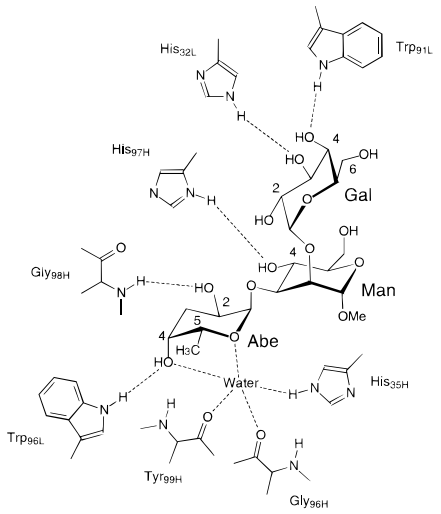


Figure 1. Hydrogen bonds between the wild-type antigen (**1**) and the residues in the combining site of the Fab fragment of Se155-4.[26]

interactions. They are Abe O-4, Abe O-2, and Man O-4. There is a hydrogen bond from Man O-4 to His-97H. Both OH groups of the Abe residue form part of a network of hydrogen bonds involving Trp-96L and a bound water molecule. The bound water molecule acts as a bridge between the antigen and the antibody. Notably, the Gal residue does not contribute significantly to the binding free energy; derivatives with the Gal replaced by a methyl or methoxymethyl group are almost as active as the native trisaccharide.

Antibody Structural Details. Distortion of the binding site could adversely affect the accuracy of the computed free energies. This distortion would manifest itself as a change in the backbone conformation, or a large change in the side chain orientations. Some deviation from the crystal conformation is always expected in a MD simulation, if for no other reason than the fact that the crystal structure contains interdomain contacts that are generally absent in the MD simulation. Moreover, interactions with water molecules can cause the protein side chains to reorient and can cause a displacement of the backbone atoms relative to the solid-phase structure, measured as the root-mean-squared (rms) differences. Large rms values can indicate a significant change in the conformation, which may be a result of improper simulation conditions; particularly possible with a localized droplet treatment of the solvent. To quantify the extent to which the protein remained in a conformation similar to that observed in the X-ray diffraction structure, we computed the rms deviations between the simulated (see Scheme 2) and experimental (see Figure 1) structures. The average rms value, regardless of the bound trisaccharide ligand, was 1.1 Å for the backbone atoms, when computed with the GLYCAM_93 parameters. The corresponding value for the simulations employing the GLYCAM_98R parameters was 0.8 Å. In both cases the rms values for the side chain atoms were approximately 40% higher. Notably, the simulations involving bound monosaccharide ligands led to larger distortions in the protein, of approximately 35%, in both the backbone and side chain positions.

The larger deviations associated with the monosaccharide ligands, may be due to the fact that the crystal structure was determined in the presence of the oligosaccharide. There is no X-ray structure of the complex containing only a monosaccharide ligand. It is reasonable to expect the amino acid side chains to adopt alternative conformations in the absence of the Gal and Man residues, although little change in the backbone

Scheme 2

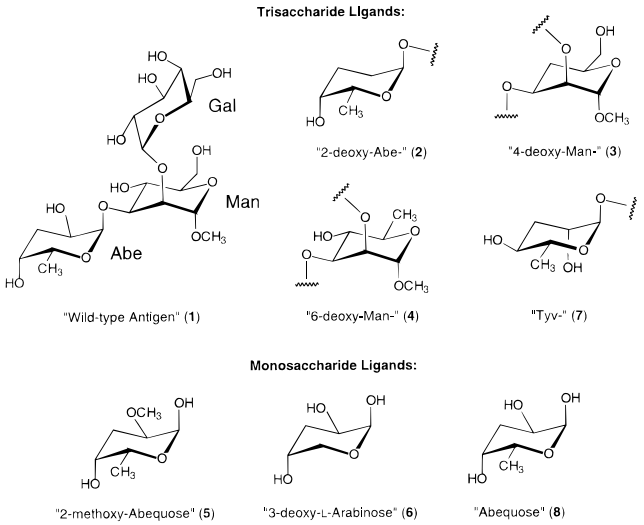


Table 1. Experimental and MD-derived Relative NOE Intensities for the Ligand **1** in Solution

interacting spins		experiment[24]	calculated NOE $\langle r^{-3}\rangle^2$ GLYCAM_93	calculated NOE $\langle r^{-3}\rangle^2$ GLYCAM_98R
Man H-1	Man H-2	100[a]	100	100
	Gal H-5	**94**[b]	**52**	**70**
Gal H-1	Gal H-2	100	100	100
	Man H-2	**139**	**99**	**108**
	Abe H-3	**11**	**26**	**16**
	Abe H-5	**41**	**24**	**52**
Abe H-1	Abe H-2	100	100	100
	Man H-3	**126**	**72**	**112**
Abe H-3	Abe H-4	100	100	100
	Abe H-2	46	77	76
	Abe H-5	46	59	58
	Gal H-1	**14**	**14**	**9**

[a] Estimated relative error ±20%.[24] [b] Inter-residue NOEs are presented in boldface.

conformation would be expected. This is consistent with the computed rms values.

Ligand Conformation. Initial modeling studies of the wild-type ligand (**1**) with the CHARMm force field, performed in vacuo, were unable to identify either a single conformation or an ensemble of conformations that was consistent with the experimental NMR NOE intensities for the trisaccharide in solution.[24] The authors of that work concluded that in solution this trisaccharide is flexible, in contrast to what has been observed for other small oligosaccharides.[24] Using the data from the last 500 ps of a 600 ps MD simulation of the wild-type trisaccharide ligand, we have applied the isolated spin-pair approximation to compute the NOE intensities. Although this is a relatively short trajectory, the ensemble of structures produced during the simulation reasonably reproduced the experimental NMR intensities (see Table 1). Intra-residue NOEs are highly dependent on pyranoside ring conformation but do not contribute to an understanding of the overall conformation. Conversely, the inter-residue NOEs are highly dependent on the orientation between the carbohydrate residues. In the development of the GLYCAM parameters, relatively little attention was focused on their ability to reproduce subtle details of the pyranoside ring geometry, such as ring puckering. The agreement between experimental and computed intra-residue NOEs is reasonable but suggests room for improvement in the parameters. The inter-residue NOEs are well-reproduced by the GLYCAM parameters. The initial CHARMm modeling was

Table 2. Torsion Angles[a] for Bound and Free Ligand **1**

	bound			free	
glycosidic linkage	experiment[23]	GLYCAM_93	GLYCAM_98R	GLYCAM_93	GLYCAM_98R
ϕ (Gal−Man)	−40	−38 ± 11	−50 ± 9	−40 ± 12	−42 ± 10
ψ (Gal−Man)	24	22 ± 10	24 ± 10	43 ± 27	44 ± 14
ϕ (Abe−Man)	−47	−60 ± 10	−51 ± 7	−56 ± 11	−43 ± 15
ψ (Abe−Man)	−15	−34 ± 10	−15 ± 12	−24 ± 26	15 ± 28

[a] In degrees, for the angles defined as: $\phi = \phi_H = H_1-C_1-O_1-C_X$, $\psi = \psi_H = C_1-O_1-C_X-H_X$.

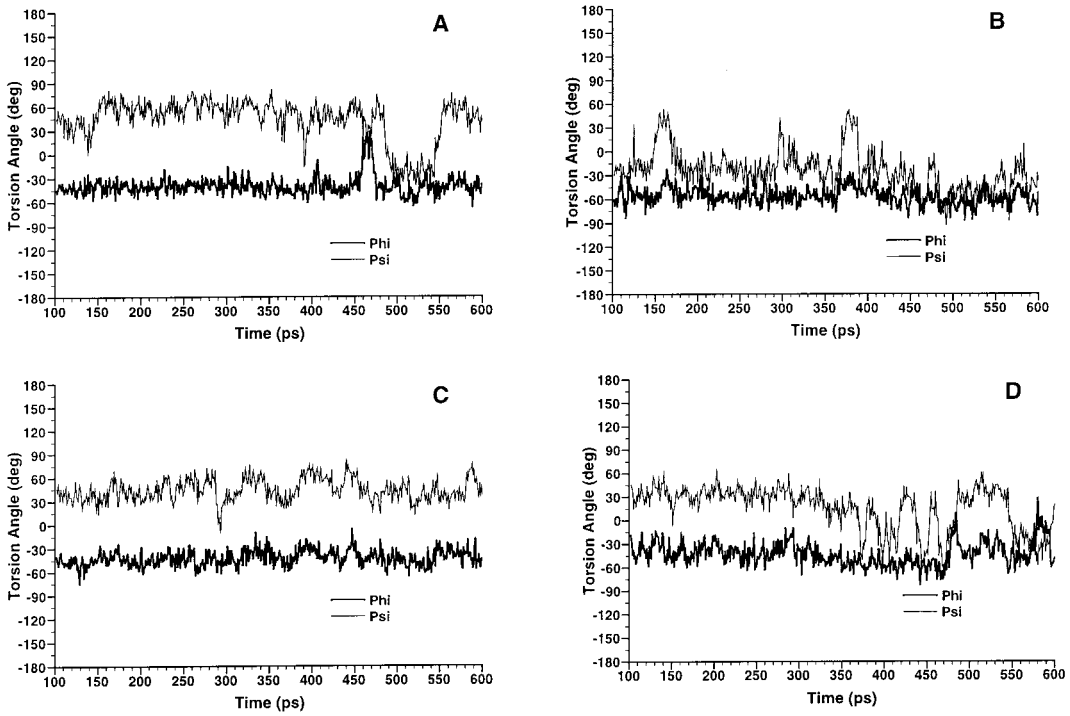


Figure 2. MD trajectories for the wild-type ligand in water. (A) Glycosidic torsion angles for the Gal-α-(1→2)-Man linkage computed with the GLYCAM_93 parameters. (B) Glycosidic torsion angles for the Abe-α-(1→3)Man linkage computed with the GLYCAM_93 parameters. (C) As in (A), but computed with the GLYCAM_98R parameters. (D) As in (B), but computed with the GLYCAM_98R parameters.

performed using energy minimization in the absence of water, and it now appears that inclusion of water is necessary for an accurate description of the conformations and dynamics of this trisaccharide. An analysis of the hydrogen bonds formed during the simulation indicated that a direct inter-residue hydrogen bond between Abe-O2 and Gal-O2 was present for less than 8% of the simulation time, regardless of which parameter set was employed. Water-bridged hydrogen bonds between Abe-O2 and Gal-O2 (50% with GLYCAM_93, less than 5% with GLYCAM_98) or between Abe-O2 and Man-O4 (29% with GLYCAM_93, 43% with GLYCAM_98) were detected. Notably, the direct inter-residue hydrogen bond is present in the Fab−antigen complex, but is replaced by a water-bridged hydrogen bond in the Fv−antigen complex.[23,26]

The values of the glycosidic torsion angles for the Gal-Man and Abe-Man linkages were computed over the course of the trajectory and are presented in Table 2 and Figure 2. Standard deviations of less than 15° are typical of glycosidic angles confined to a single conformation.[34] In these simulations the ϕ-angles adopt conformations close to those seen in the bound state (see Table 2). The ψ-angles show more variability, and the average values are slightly further from those of the bound state. The trajectories indicate the presence of several conformational states in solution, particularly for the ψ-angles. The ψ-angles associated with the Abe-Man linkage show distinctly different preferences between the two carbohydrate parameter sets. The newer parameters based on RESP-charges indicate this angle strongly prefers a value of approximately 30°, whereas the earlier model predicts a preference for approximately −30°. These differences are reflected in the computed values for the inter-residue NOE between proton H-1 in Abe and proton H-3 in Man (see Table 1). A value of approximately 30° brings the Abe-O2 and Man-O4 hydroxyl groups sufficiently close to allow formation of a water-bridged hydrogen bond. Earlier data from modeling performed in vacuo suggested that upon binding the major conformational change occurred about the Gal-Man linkage.[24] Our data suggest that it is the ψ-angle of the Abe-Man linkage that changes most significantly upon binding. Not surprisingly, the bound trisaccharide exhibits reduced motion about each of the linkages, and the ψ-angles no longer display more motion than the ϕ-angles. The data in Tables 1 and 2 indicate that the newer GLYCAM_98R parameters reproduce the ligand geometries and NMR data somewhat better than the earlier GLYCAM_93 parameter set.

Geometry of the Antigen−Antibody Complex. The accuracy of the simulated complex may be judged in part by the degree to which the hydrogen bonds between the antigen and the antibody are reproduced. The MD data were analyzed for the complexes involving ligands **1**−**6** (see Figure 1 and Scheme 2) and the distances between hydrogen bonded atoms are presented in Table 3. Included for comparison are the corresponding distances determined crystallographically for the complex with ligand **1**. Interatomic distance provides a general measure of the strength of a hydrogen bond, with a distance

Table 3. Computed[a] Average Hydrogen Bond Lengths[b] between Ligands and Amino Acid Residues in the Combining Site of Se155-4

hydrogen bond (experiment)[24]	wild-type	4-deoxy-Man (**3**)	6-deoxy-Man (**4**)	2-deoxy-Abe (**2**)	2-methoxy-Abe (**5**)	3-deoxy-L-Ara (**6**)
Abe O-2−Gly_{98H} N	3.0 (0.2)[c,d]	2.9 (0.1)	3.0 (0.2)		5.5 (0.8)	4.9 (0.6)
(2.8)	3.1 (0.3)[e]	3.0 (0.2)	3.1 (0.3)		4.4 (0.4)	3.2 (0.2)
Abe O-2−Gal O-2	4.4 (0.4)	6.0 (0.4)	4.6 (0.7)		*g*	*g*
(4.4)[f]	4.9 (0.5)	3.4 (0.6)	4.4 (0.5)			
Abe O-2−His_{34L} N-ϵ	3.7 (0.6)	3.6 (0.5)	3.0 (0.3)		3.5 (1.1)	3.3 (0.5)
(3.8)	3.8 (0.6)	3.7 (0.5)	3.9 (0.6)		3.4 (0.5)	3.9 (0.4)
Abe O-2−His_{97H} N-δ	4.1 (0.6)	3.9 (0.5)	4.0 (0.6)		4.7 (0.7)	6.9 (1.3)
(3.9)	5.1 (0.6)	4.4 (0.7)	4.9 (0.7)		4.8 (0.8)	6.1 (0.4)
Abe O-4−Trp_{98L} N-ϵ	2.9 (0.1)	2.9 (0.1)	2.9 (0.2)	3.0 (0.1)	6.2 (0.7)	3.0 (0.2)
(3.0)	2.9 (0.1)	2.9 (0.1)	2.9 (0.1)	2.8 (0.1)	3.3 (0.5)	3.0 (0.2)
Abe O-4−His_{35H} N-ϵ	3.7 (0.3)	3.5 (0.3)	3.5 (0.3)	3.5 (0.3)	5.3 (0.4)	3.8 (0.3)
(3.7)	3.6 (0.3)	3.4 (0.2)	3.5 (0.2)	3.5 (0.3)	4.0 (0.5)	3.7 (0.3)
Abe O-5−Trp_{33H} N-ϵ	3.6 (0.6)	3.4 (0.2)	3.8 (0.5)	4.5 (0.4)	4.3 (0.5)	4.0 (0.5)
(3.7)	3.8 (0.4)	3.7 (0.3)	3.6 (0.3)	3.9 (0.3)	3.8 (0.5)	3.4 (0.3)
Man O-4−His_{97H} N-δ	3.8 (1.0)		4.0 (0.6)	3.7 (0.6)	*g*	*g*
(2.7)	4.9 (0.6)		4.4 (1.0)	3.4 (1.0)		
Gal O-2−Trp_{93L} N-ϵ	4.6 (0.3)	4.8 (0.3)	2.9 (0.2)	2.8 (0.1)	*g*	*g*
(2.8)	3.1 (0.3)	4.2 (0.6)	3.1 (0.3)	3.4 (0.6)		
Gal O-4−Asn_{96L} O-δ	6.3 (1.7)	5.9 (0.6)	9.0 (1.3)	9.9 (1.1)	*g*	*g*
(3.6)	3.1 (0.6)	5.8 (0.6)	3.8 (0.9)	6.5 (1.1)		
Wat1−Abe O-5	3.2 (0.4)	3.2 (0.3)	3.2 (0.4)	3.1 (0.3)	5.0 (0.3)	3.3 (0.3)
(3.1)	3.3 (0.3)	3.2 (0.3)	3.1 (0.2)	3.1 (0.2)	3.5 (0.5)	3.0 (0.2)
Wat1−Gly_{96H} O	2.8 (0.2)	2.8 (0.1)	2.7 (0.1)	2.9 (0.2)	2.8 (0.2)	2.8 (0.1)
(2.9)	2.9 (0.2)	2.9 (0.2)	2.9 (0.2)	2.7 (0.1)	3.0 (0.2)	3.2 (0.2)
Wat1−Tyr_{99H} O	3.2 (0.5)	3.1 (0.5)	3.2 (0.5)	3.2 (0.4)	3.4 (0.7)	3.3 (0.6)
(3.1)	3.0 (0.3)	3.0 (0.3)	3.1 (0.3)	3.2 (0.4)	3.0 (0.3)	2.8 (0.2)

[a] Derived from 50 ps trajectories, either prior to (for the wild-type antigen) or after (for the synthetic analogues) the FEP simulation. [b] Distances between heavy atoms in Å. [c] Based on the GLYCAM_93 parameters. [d] Standard deviations in parentheses. [e] Based on the GLYCAM_98R parameters. [f] Bridging water position is only partially occupied in the simulations, and thus only the distance between the heavy atoms in the antigen is reported here. [g] Ligand present only as a monosaccharide in the experiment and in the corresponding simulation.

between non-hydrogen atoms of less than approximately 3.2 Å defining a moderately strong hydrogen bond and 3.2−4.0 being a weak interaction.[36] The MD data are in good agreement with the experiment and indicate that the glycosyl residues near the surface of the protein are more flexible than those buried in the combining site. This is shown particularly well by the magnitude of the standard deviations. In addition to the length of the hydrogen bond, the standard deviation provides some indication of the strength, with deviations of 0.1−0.2 Å characterizing a strong interaction. For the weaker interactions, the GLYCAM_98R parameter set combination performs better than the earlier parameters, suggesting that the RESP-charges are superior to the unrestrained ESP-charges. The enhanced agreement obtained with this parameter combination also reflects an improved description of the protein, resulting from the use of the PARM94 versus PARM91 protein parameters. Most hydrogen bonds between the antibody and wild-type antigen were very well reproduced, even those involving water molecules. In the sc-Fv crystal structure, the N-δ atom of His-97H is involved in a very long hydrogen bond (3.9 Å) with Abe-O2 and a short hydrogen bond with Man-O4. The former was conserved throughout most of the simulations, whereas the latter, stronger interaction surprisingly was not. This general behavior was not dependent on the protonation state of this residue. His-97H is exposed directly to solvent and in some of the simulations underwent a modest conformational change, which caused a disruption of these interactions. Interestingly, other weak hydrogen bonds were consistently reproduced. The hydrogen bonds to the Man and Gal residues were generally predicted to be longer than those found in the crystal structure; however, these residues are located only in a shallow groove on the antibody surface and are more subject to movement than the Abe residue.

Notably, the crystallographically determined distance between the Gal-O2 and Abe-O2 hydroxyl oxygen atoms (4.4 Å) in the Fv−antigen complex was well-reproduced in the simulations. A water molecule in the solid-state structure is located between the Gal-O2 and Abe-O2 atoms; however, during the simulations a water molecule was found to be present in this position for less than 5% of the simulation. This observation may help to explain the two modes of binding observed experimentally for the wild-type ligand, namely that in the Fab−antigen complex Abe-O2 forms a direct hydrogen bond with Gal-O2, whereas in the Fv−antigen complex these two groups form a water-bridged interaction. The MD results suggest that coordination to the water molecule is not necessary for the ligand conformational change.

Binding Free Energies. As we were unable to infer the protonation state of His-97H from hydrogen bond length analysis alone, we performed the FEP simulations with either N-δ, N-ϵ, or both nitrogen atoms protonated (see Table 4). Each protonation state is possible, given that the pK_a value of histidine is approximately 7.0.[37] A charged histidine residue in the binding site has been predicted for other carbohydrate−protein complexes.[13] In the crystal structures of the Fab and Fv fragments of Se155-4, it is the N-δ atom of His-97H that forms a hydrogen bond with the O-4 atom of the mannosyl residue,[24] rather than the more commonly protonated N-ϵ position.[37] If this histidine residue were diprotonated, it could function as a proton donor in interactions with Man-O-4 and with Abe-O-2, which would be consistent with the observed affinity of the 2-methoxy-Abe (**5**) analogue. In the absence of further experimental data, all other histidine residues were treated as nonionized, with the single proton being located at the generally preferred site. The value for the relative binding free energy for ligand **5** computed with a diprotonated His-97H (0.41 ± 1.0 kcal mol^{-1}) was the

(36) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, 1997.

(37) Creighton, T. E. *Proteins. Structures and Molecular Properties*, 2nd ed.; W. H. Freeman: New York, 1993.

Table 4. Comparison of Computed and Experimental $\Delta\Delta G$[a] Values in kcal mol^{-1}

ligand (experiment)[26]	protonation site in His-97H	calculated $\Delta\Delta G$		
		GLYCAM_93	GLYCAM_98	GLYCAM_98R
2-deoxy-Abe-	N-ϵ	0.98 ± 0.4	−0.70 ± 0.9	0.63 ± 0.6
(**2**)	N-δ	0.86 ± 0.4		0.05 ± 1.0
(≫**1.8**)[b]	**N-ϵ, N-δ**	**1.93 ± 0.5**		**1.61 ± 0.5**
4-deoxy-Man-	N-ϵ	0.37 ± 0.5	−0.55 ± 0.6	0.52 ± 0.6
(**3**)	N-δ	−0.58 ± 0.8		0.22 ± 0.2
(**0.85 ± 0.3**)	**N-ϵ, N-δ**	**0.91 ± 0.8**		**0.22 ± 1.0**
6-deoxy-Man-	N-ϵ	0.26 ± 0.6	−0.38 ± 0.5	−0.79 ± 1.1
(**4**)	N-δ	0.85 ± 0.7		−0.46 ± 0.6
(**0.50 ± 0.2**)	**N-ϵ, N-δ**	**1.01 ± 0.2**		**0.00 ± 0.7**
2-methoxy-Abe-	N-ϵ	2.45 ± 2.0	−	1.04 ± 1.0
(**5**)	N-δ	3.16 ± 1.4		1.17 ± 1.5
(−**0.14 ± 0.2**)[c]	**N-ϵ, N-δ**	**−1.18 ± 1.3**		**0.41 ± 1.0**
3-deoxy-L-Ara	N-ϵ	1.27 ± 0.2	1.72 ± 1.3	1.74 ± 0.5
(**6**)	N-δ	1.00 ± 0.4		3.64 ± 0.1
(**1.4 ± 0.7**)[c]	**N-ϵ, N-δ**	**0.50 ± 1.0**		**0.88 ± 1.3**
Tyv-	N-ϵ	8.09 ± 1.0	−	3.50 ± 0.5
(**7**)	N-δ	4.28 ± 1.0		4.68 ± 0.5
(≫**2.4**)[b]	**N-ϵ, N-δ**	**8.15 ± 1.0**		**4.88 ± 0.5**

[a] Relative to a value of 0.0 kcal mol^{-1} for the wild-type ligand. [b] Estimated by immunoassay. [c] Experimental value measured for the α-methyl glycoside.

closest of the simulations to the experimental value (−0.14 ± 0.2 kcal mol^{-1}), however, it differed in sign. Although the GLYCAM_98R simulations did not reproduce the slightly higher affinity reported for 2-methoxy-abequose, the experimental value is very close to zero and falls within our error estimates. As we wish to apply this method in the search for significantly better binding ligands ($\Delta\Delta G < -0.5$ kcal mol^{-1}) slightly underestimating the relative affinity may be acceptable. At this time there are no known stronger binding ligands available for evaluation.

To further assess the protonation state of His-97H we have performed a linear regression analysis of the computed versus experimental data. Although the two data points obtained by immunoassay are somewhat qualitative, they are also very significant due to their relatively large $\Delta\Delta G$ values (≫1.8 and ≫2.4 kcal mol^{-1}, respectively). To incorporate the immunoassay data into the regression analysis we have assumed that the reported values represent lower bounds. The correlation was relatively insensitive to the error estimate assigned to these two points. For the $\Delta\Delta G$ values computed with GLYCAM_93 this analysis produced correlation coefficients between the computed and experimental data of approximately 0.4, 0.2, and 0.8 for protonation at each of the possible sites in His-97, namely, N-ϵ or N-δ or both N-ϵ and N-δ, respectively. The corresponding values for the data obtained from GLYCAM_98R were 0.4, 0.3, and 0.80. It is apparent that both GLYCAM_93 and GLYCAM_98R perform similarly, and both are more consistent with a model in which HIS-97H is diprotonated.

As anticipated from the electrostatic treatment, the GLYCAM_93 carbohydrate parameters with the PARM91 protein parameters performed poorly on mutations involving removal of a hydroxyl group. For the simulations based on the newer AMBER and GLYCAM parameters it was apparent that using unrestrained charges (ESP-charges) for the carbohydrate led to a very poor reproduction of the relative binding energies, and only a few mutations were performed with this parameter combination. In particular, it appeared that the ESP-charge model underestimated the destabilizing effect of forming a deoxy-derivative. Notably, the charge model had little effect on the simulations involving the 3-deoxy-L-Ara (**6**) analogue. This is not unexpected since this mutation involves the loss of the methyl group at the 5-position in Abe. The methyl group is nonpolar and has almost zero net charge in both the ESP (−0.087 au) and RESP (−0.025 au) parametrizations.

The experimental binding free energy of the 4-deoxy-Man (**3**) derivative is lower than that of the wild-type trisaccharide by 0.85 kcal mol^{-1}.[26] This is consistent with the fact that this OH group is involved in a hydrogen bond with the antibody (see Figure 1), and its removal should disfavor binding. The calculated $\Delta\Delta G$ showed only a weak dependence on the protonation state of His-97H, being between 0.22 and 0.52 kcal mol^{-1}. This reflects in part the fact that this hydrogen bond length was consistently overestimated in our simulations. Interestingly, the calculated $\Delta\Delta G$ for the 2-deoxy-Abe (**2**) derivative, which interacts weakly with His-97H, showed a strong dependence on the protonation state. The 2-OH group of the Abe residue is also involved in a water-mediated inter-residue hydrogen bond with Gal-O2 and a direct hydrogen bond with the main chain amide NH of Gly-98H. Immunoassays indicate that ligand **2** is weaker than ligand **1** by more than 1.8 kcal mol^{-1}. The FEP simulations correctly predict that this analogue will not bind as strongly as the wild-type ligand, with the best agreement (1.61 kcal mol^{-1}) being obtained for the diprotonated His residue.

The 6-OH group of the Man residue is not in close contact with the antibody, and experimentally the removal of this hydroxyl group has only a small effect on binding. Again the best agreement (0.0 kcal mol^{-1}) with experiment (0.50 kcal mol^{-1}) was obtained with His-97H being diprotonated. Simulations performed for the other protonation states suggest that the deoxy-analogue would be preferred over the wild-type. It should be noted that this mutation presents a particularly demanding case. In the bound orientation, the hydroxymethyl group of the Man residue extends away from the protein surface into the solvent. Thus, there is no obvious reason for binding to the protein to be disfavored. This example suggests that the FEP simulations are more accurate when the mutation results in a change in a direct interaction between the protein and the carbohydrate. Under the present approximate treatment of solvation and nonbonded interactions, the FEP simulations may not be able to reproduce extremely subtle effects related primarily to differential interactions with solvent.

The contribution of the methyl group at the 5-position in the Abe residue was estimated by measuring the binding energy of

3-deoxy-L-arabinose (**6**) relative to that of abequose (**8**). Experimentally, the replacement of the methyl group of abequose with a hydrogen atom weakens the binding by 1.4 kcal mol^{-1}. The loss of a methyl group naturally leads to a loss of enthalpy arising from van der Waals interactions, and some gain of entropy, from an increase in the motional freedom of the protein residues in the binding pocket. For the diprotonated His-97H simulations, the calculated $\Delta\Delta G$ was 0.88 kcal mol^{-1}. This mutation appeared to be less sensitive to the parameter selection or to the protonation state, which is consistent with its apolar character. The large value for the $\Delta\Delta G$ predicted for the N-δ protonation state was apparently due to a conformational change in the combining site, further suggesting that protonation at N-δ alone is unlikely.

The mutation of the abequosyl residue to tyvelosyl (**7**), equivalent to epimerization at C-2 and C-4, leads to a large relative binding energy. This is particularly relevant because it is epimerization at the C-2 and C-4 positions in the 3,6-dideoxypyranosyl residue that generates the immunodominant sugars that differentiate between *Salmonella* serotypes A, B, and D_1.[38] This differentiation is achieved through very high shape and charge complimentarity in the combining site of the antibody proteins. There is no calorimetrically determined relative binding energy for the ligand **7**; however, immunoassays have estimated that it must be much greater than 2.4 kcal mol^{-1}.[26] Our simulations clearly indicate that this ligand is strongly disfavored, regardless of the protonation state of His-97H.

The negative relative binding free energies for the deoxy-analogues, computed with the GLYCAM_98 parameters with ESP-charges, imply that these analogues have a stronger preference for binding to the protein rather than remaining in the solvent, relative to the wild-type ligand. This is the reverse of the observed experimental behavior. This reversal is consistent with an underestimation of the strength of hydrogen bonds between the carbohydrate and the protein. In terms of the force field, this behavior may mean either that the O−H attraction is too low, or that the O−O repulsion is too high. On the basis of simulations of condensed phase crystals of carbohydrates and by employing several parameter sets, we believe it is the latter that is responsible for the poor performance of the ESP-charge model.[39] The use of lower magnitude partial atomic charges in the RESP model restores the agreement with experiment and is further support for the hypothesis that the ESP-charges lead to excessive O−O repulsions between the protein and the carbohydrate. The current simulations are based on the crystal structure of the sc-Fv antibody fragment, whereas the thermodynamic data were acquired with the Fab fragment. The extent of the discrepancy in binding energies that this difference introduces is unknown; however, the geometric similarity of the binding regions of the sc-Fv and Fab fragments is very high. For the wild-type ligand, the binding energy is essentially identical for the sc-Fv and Fab fragments.[40]

(38) Birnbaum, G. I.; Bundle, D. R. *Can. J. Chem.* **1985**, *63*, 739−744.
(39) Woods, R. J.; Chappelle, R. *THEOCHEM* **1999**, submitted for publication.

Conclusions

In general, the GLYCAM_98R parameters performed comparably to the earlier version GLYCAM_93, both in terms of structures and energies of the antibody−ligand complexes; however ,the former provided an improved model for the wild-type ligand in solution. We found that the use of unrestrained ESP-charges combined with the standard 1−4 scaling value of 0.833 uniformly overestimated the stability of the deoxy-analogues. When RESP-charges were employed for the glycosyl residues, the overall agreement with experiment was restored.

The simulations reported here are consistent with the reported NMR data for the wild-type ligand in solution and with the crystallographic data for the Fv−antigen complex. The simulations indicate that the free oligosaccharide displays a moderate amount of internal flexibility, which manifests itself as oscillations around well-defined average glycosidic torsion angles. The bound conformation is encompassed within the observed conformations for the free ligand. The conformation and intermolecular hydrogen bond patterns computed for the bound ligand are in excellent agreement with the reported crystallographic data for the antibody−ligand complex.

Comparison with the calorimetrically determined binding energies indicates that the average absolute error in the relative free energies computed by GLYCAM_93 and GLYCAM_98R are 0.62 kcal mol^{-1} and 0.55 kcal mol^{-1}, respectively, for the model in which HIS-97 is diprotonated. Although based on a limited number of mutations, the FEP simulations appear to reproduce the very subtle differences in binding among the known ligands, and we believe FEP simulations may be successfully employed in the design of higher affinity ligands. We are pursuing these simulations at present.

The GLYCAM parameters are freely available upon request to rwoods@ccrc.uga.edu.

Acknowledgment. This research was supported primarily by Grant GM-55230 from the National Institutes of Health.

Supporting Information Available: Tables S1a, S1b, and S2 containing the ESP- and RESP-atomic partial charges for each of the ligands (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA9914994

(40) Bundle, D. A., personal communication, 1999.