JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

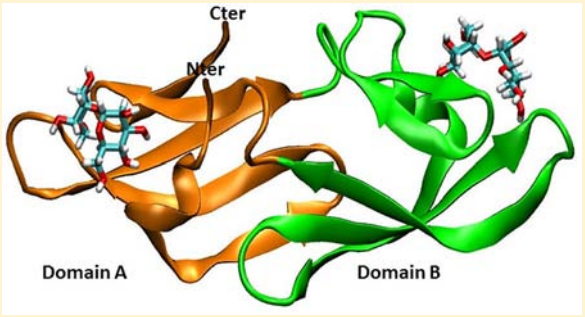
Carbohydrate Recognition by the Antiviral Lectin Cyanovirin-N

Yukiji K. Fujimoto[†,‡] and David F. Green*[,†,‡,§,∥]

[†]Department of Chemistry, [‡]Laufer Center for Physical and Quantitative Biology, [§]Department of Applied Mathematics and Statistics, and [∥]Graduate Program in Biochemistry and Structural Biology, Stony Brook University, Stony Brook, New York 11794-3600, United States

Supporting Information

ABSTRACT: Cyanovirin-N (CVN) is a cyanobacterial lectin with potent antiviral activity and has been the focus of extensive preclinical investigation as a potential prophylactic for the prevention of the sexual transmission of the human immunodeficiency virus (HIV). Here we present a detailed analysis of carbohydrate recognition by this important protein, using a combination of computational methods, including extensive molecular dynamics simulations and molecular mechanics/Poisson−Boltzmann surface area (MM/PBSA) energetic analysis. The simulation results strongly suggest that the observed tendency of wild-type CVN to form domain-swapped dimers is the result of a previously unidentified *cis*-peptide bond present in the monomeric state. The energetic analysis additionally indicates that the highest-affinity ligand for CVN characterized to date (α-Man-(1,2)-α-Man-(1,2)-α-Man) is recognized asymmetrically by the two binding sites. Finally, we are able to provide a detailed map of the role of all binding site functional groups (both backbone and side chain) to various aspects of molecular recognition: general affinity for cognate ligands, specificity for distinct oligosaccharide targets, and the asymmetric recognition of α-Man-(1,2)-α-Man-(1,2)-α-Man. Taken as a whole, these results complement past experimental characterization (both structural and thermodynamic) to provide the most complete understanding of carbohydrate recognition by CVN to date. The results also provide strong support for the application of similar approaches to the understanding of other protein−carbohydrate complexes.



INTRODUCTION

Roughly 25 years ago, Lifson and colleagues identified mannose-specific carbohydrate binding proteins (most notably legume lectins such as concanavalin A) as potential inhibitors of human immunodeficiency virus (HIV) viral−cell fusion.[1] The outer envelope glycoprotein of HIV (gp120) is heavily glycosylated, with between 20 and 28 N-linked glycosylation sites occupied in various natural viral isolates; glycosylation consists of both high-mannose and complex oligosaccharide subtypes, in roughly equal proportions.[2] Antiviral lectins act by binding these carbohydrates and thus interfering with some aspect of cellular recognition and/or membrane fusion; the specific mechanism of inhibition may involve either direct blocking of CD4 receptor or CXCR4/CCR5 coreceptor binding or interference in required conformational changes associated with fusion.[3−5]

More recently, numerous lectins from diverse sources have been found to have antiviral activity, many with much greater potency and specificity than the plant lectins. As a therapeutic target, the carbohydrates of the viral envelope are attractive for several reasons. First, the glycosylation plays an essential role in helping the virus avoid detection by the humoral immune response. The carbohydrates are added by the enzymatic systems of the host cell, and reduced levels of glycosylation have been associated with increased susceptibility to immune system recognition;[6] treatment of infected cells with antiviral carbohydrate binding agents has additionally been associated with evolution of viral strains with reduced glycosylation.[7,8] Because of these conflicting pressures on viral fitness, inhibitors that work through specific interactions with the sugars may be less susceptible to the evolution of viral resistance.[9] Second, the remarkable density of carbohydrates on the surface of gp120 distinguishes it from naturally occurring human glycoproteins, which typically have many fewer sites of glycosylation, and thus nonspecific interactions may be avoided. Finally, while the lack of oral bioavailability of protein therapeutics can make small molecule drugs more attractive, topical application as a prophylactic virucide does not suffer from this concern.[10,11] Preclinical trials for use of lectins as a topical agent to prevent sexual transmission of HIV in simian models have shown significant promise,[12,13] and recent work has demonstrated the prophylactic activity of recombinant mucosal bacteria expressing a lectin virucide.[14,15]

Beyond the inhibition of cellular infection by HIV, protein−carbohydrate interactions play key roles in a vast array of biology. Many human retroviruses are heavily glycosylated in much the same manner as HIV, and in fact, lectins with anti-HIV activity are often active against a diverse range of viruses, including influenza, ebola, and herpes simplex.[16−18] Additionally, bacterial pathogens often display cell surface carbohydrates distinct from

Received: June 13, 2012
Published: October 11, 2012

dx.doi.org/10.1021/ja305755b | J. Am. Chem. Soc. 2012, 134, 19639−19651

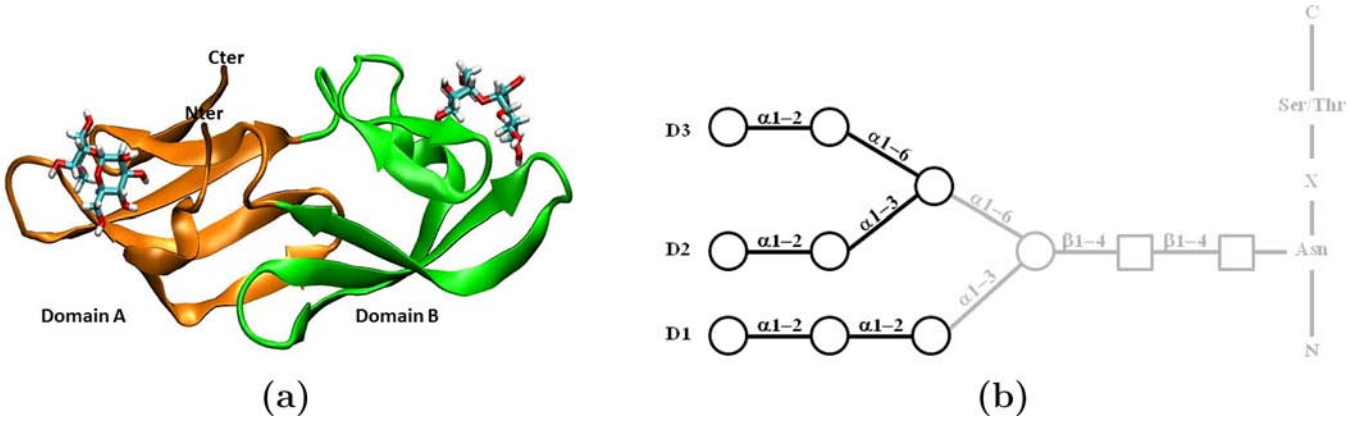


Figure 1. The structures of cyanovirin-N and of high-mannose oligosaccharides. (a) The solution NMR structure of cyanovirin-N bound to α-Man-(1,2)-α-Man is shown. Domain A, consisting of residues 1−36 and 88−101, is colored orange, while Domain B, consisting of residues 37−87, is colored green. (b) The largest of the high-mannose N-linked oligosaccharides, Man9, is shown as a schematic, with circles indicating mannose and squares *N*-acetyl-glucosamine. The three trisaccharides considered in this study (corresponding to arms D1−3) are highlighted in bold.

those on eukaryotic glycoproteins,[19] and many cancer cells are characterized by unique glycosylation patterns.[20] Thus, specific, high-affinity lectins have potential applications as antimicrobial agents as well as for cancer cell targeting.

Despite their importance, the study of protein−carbohydrate interactions has greatly lagged that of other biomolecular complexes. In particular, many computational methods that have shown great success in understanding protein−protein, protein−nucleic acid, and protein−small molecule interactions have seen only limited application to carbohydrate recognition. As a good example, the combination of explicit solvent molecular dynamics (MD) with Poisson−Boltzmann-based calculations of free energies has been demonstrated as a powerful approach for predicting relative binding free energies and for dissecting energies into contributions from individual chemical groups.[21−26] As well as facilitating the understanding of natural systems, computational approaches provide unique opportunities for molecular design; as a perfect example, detailed studies on protein systems, coupled with robust computational models and innovative algorithms, have made the computational design and engineering of proteins and their complexes a reality.[27−32] Robust protocols for the computational analysis of protein−carbohydrate systems thus promise to be an enabling technology, opening up a wide range of potential applications to these important systems. While there have recently been some attempts to characterize the binding energetics of protein−carbohydrate complexes through computational approaches, studies of this type remain rare.[26,33,34]

Among the best characterized of the antiviral lectins is cyanovirin-N (CVN), originally isolated from the cyanobacterium *Nostoc ellipsosporum*.[35] Under physiological conditions, CVN is a small, monomeric protein containing two pseudosymmetric binding domains (Figure 1a); these domains have differing affinities for specific oligosaccharides, but both bind exclusively to sugars containing an α-(1−2)-linked mannobiose substructure.[36−39] As this structure makes up the terminal sugars on all three branches of high mannose oligosaccharides (exemplified by Man9, Figure 1b), CVN exclusively targets relatively underprocessed glycans and shows little affinity for complex N-linked oligosaccharides. Interestingly, at high concentrations CVN has a tendency to form domain-swapped dimers, and this form is the only state found in the crystal phase for wild-type CVN;[40,41] mutations that preferentially stabilize both the monomeric state and the dimer have been identified. Thermodynamic, structural, and *in vivo* efficacy studies have all suggested preferential recognition of the D1 arm of high mannose sugars,[38,39,41] and multivalent interactions appear necessary for antiviral potency.[59] Despite the apparent wealth of data for this system, a complete understanding of the mechanism of CVN's antiviral activity has been hindered by the inherent complexities in both synthetic chemistry and structural biology of carbohydrates. Among the questions that remain to be answered are: What are the determinants of specific carbohydrate recognition by CVN, what is the mechanism by which multivalent interactions by CVN lead to potent antiviral activity, and can CVN be engineered into a better virucidal agent? Here we present a comprehensive answer to the first of these questions, building on a computational framework for modeling protein−carbohydrate interactions that we have previously demonstrated as a promising approach.[26] The success of computational methods in explaining the structural determinants of carbohydrate recognition provides strong motivation for the use of similar approaches to address the remaining questions.

METHODS

Construction of CVN Complexes with Trisaccharides. The initial structure for all simulations originated from the last snapshot taken from an earlier study where the solution NMR structure of CVN bound to α-Man-(1,2)-α-Man (PDB 1IIY) was modified and simulated in a droplet of water.[26] Briefly, the backbone atoms of equivalent residues in each site were superimposed (by minimizing the backbone heavy atom root-mean-square deviation, RMSD), and the coordinates of the sugar in domain A were then replaced with those from the superimposed structure to construct bound state models with increased symmetry.

Structures of three distinct trimannoses representing the three arms of Man9 (α-Man-(1,2)-α-Man-(1,2)-α-Man, α-Man-(1,2)-α-Man-(1,3)-α-Man, and α-Man-(1,2)-α-Man-(1,6)-α-Man) were built by extending our dimannose model by one unit from the anomeric carbon of the reducing sugar. In addition, an alternate structure of α-Man-(1,2)-α-Man-(1,2)-α-Man was constructed by extending the dimannose by one unit from C2 of the nonreducing sugar. All these manipulations were done using the CHARMM software package,[42] and default conformations were used for the newly built portions of each molecule. A short minimization (100 steps) was performed on all the newly built structures to avoid any clashes. Three models were constructed in each case, two 1:1 (protein:sugar) complexes with a single sugar bound to each of the two binding sites as well as a 1:2 complex with both binding sites occupied.

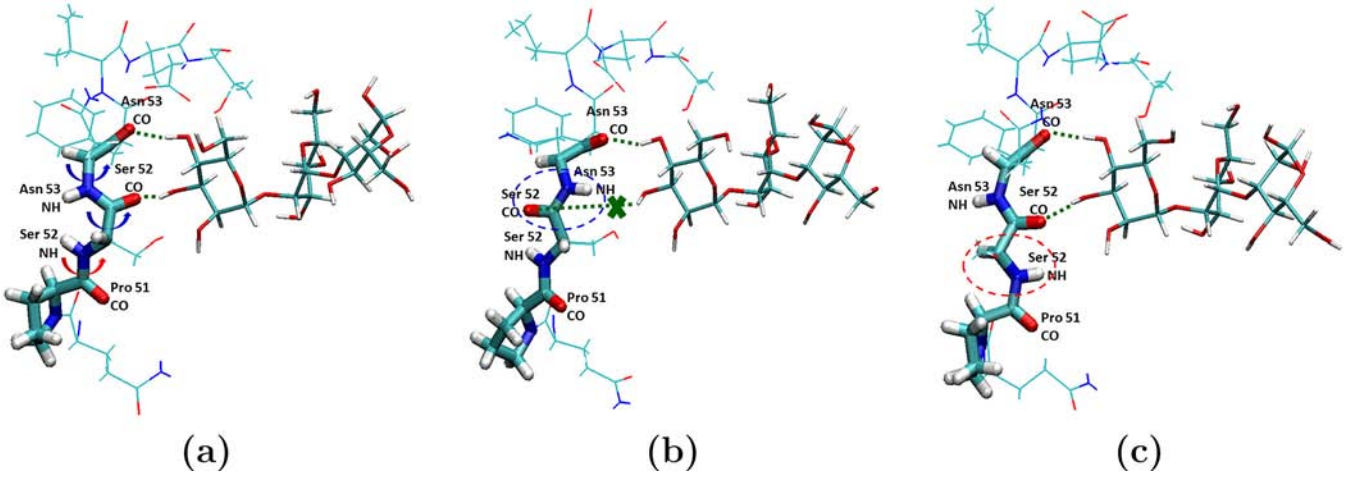


Figure 2. Alternate backbone configurations in hinge region. In the solution NMR structure (a), the amide protons of Ser 52 and Asn 53 clash; this can be alleviated either by a crank-shaft-like motion of the peptide bond between these groups [blue arrows, forming (b)], or by a *trans* to *cis* isomerization of the peptide bond between Pro 51 and Ser 52 [red arrow, forming (c)]. In conformation (b), the stabilizing contact between the sugar and the backbone carbonyl of Ser 52 is lost, while in conformation (c), this contact is maintained.

MD Simulations. Explicit solvent molecular dynamics simulations were performed using the CHARMM[42] and NAMD[43] computer programs, using PARAM22 (protein)[44] and CSFF (carbohydrate)[45] parameter sets and the TIP3P water model.[46] Pre- and postprocessing of all complexes was done with CHARMM, while production simulations were done using NAMD. Each complex was solvated in a box of water with a minimum of 10 Å between any solute atom and the box edge in all directions. Randomly selected water molecules were replaced with sodium and chloride ions to match physiological ionic strength (145 mM) and to obtain a net zero charge for the system (CVN has a formal charge of $-3.0e$); roughly 14 Na^+ and 11 Cl^- ions were included in each case. A total of 200 ns was simulated for each complex (using a 2 fs time step) under NPT ensemble conditions (P = 1 atm, T = 300 K) using periodic boundary conditions and particle-mesh Ewald (PME) for long-range electrostatics. Short-range interactions were cut off at 12 Å, and bonds involving hydrogens were held fixed using SHAKE. Temperature was held constant through the use of Langevin dynamics with a friction coefficient of 5 ps^{-1}, and pressure was regulated using a Langevin piston with an oscillation period of 0.2 ps and a damping time of 0.1 ps.

Calculation of Binding Free Energies. Binding free energies were computed with a molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) model.[47] The first 50 ns of each simulation were excluded from the analysis to ensure adequate equilibration of each system. The explicit solvent MD trajectories were sampled every 100 ps, for a total of 1500 snapshots per trajectory. The total binding energy for each snapshot was computed as the sum of a Poisson–Boltzmann-based electrostatic contribution (ΔG^{elec}), the intermolecular van der Waals energy (ΔG^{vdw}), and a term proportional to the solvent accessible surface area buried on binding. The area was computed with CHARMM, using a 1.4 Å probe radius, and the energetic contribution was given by $\Delta G^{h\phi} = 0.005\Delta A + 0.86$ kcal/mol for each snapshot. These energies were then averaged over all frames, and standard errors of the mean were computed treating all 1500 frames as statistically independent; an autocorrelation analysis of the energetic time series suggested that this was appropriate.

Continuum Electrostatic Calculations. The linearized Poisson–Boltzmann equation was solved using a multigrid finite-difference solver distributed with the integrated continuum electrostatics (ICE) software suite (courtesy of B. Tidor), using standard protocols.[48−50] Charges were taken from the PARAM22 and CSFF parameter sets for consistency with the MD simulations. The dielectric boundary was set as the molecular surface generated with a 1.4 Å radius probe, and a 2.0 Å ion exclusion layer was used; the surfaces were generated using radii optimized specifically for use in continuum electrostatic calculations.[51,52] The internal and external dielectric constants were set to 2 and 80, respectively, and the (monovalent) ionic strength was set to 145 mM. Boundary conditions were computed using a 3-step focusing procedure on a 129^3-unit cubic grid, with the molecule occupying first 23%, then 92%, and finally 184% of the grid. Boundary conditions at each level were taken from the previous calculation, with Debye–Hückel potentials used at the boundary of the lowest level. The highest-resolution grid was centered on the oligosaccharide, and potentials at atoms falling off this grid were taken from the middle-resolution calculation. Electrostatic contributions to the binding free energies were computed as the sum of a desolvation penalty for both the protein and the sugar and a bound-state, solvent-screened interaction.

Component Analysis. Group-wise component analysis was done using the ICE package (courtesy of B. Tidor),[48,49] again using standard protocols.[53−55] Each protein residue was partitioned into three groups: backbone carbonyl, backbone amino, and side chain, and the sugars were partitioned into one group per hydroxyl. For each group, the desolvation penalty and the indirect (intramolecular) and direct (intermolecular) interactions were computed. The sum of these is the mutation energy, equivalent to the difference in energy between the natural system and a hypothetical mutant with that group replaced with a hydrophobic isostere (in the context of all other groups in their natural state). Additionally, the solvent-accessible surface area (and corresponding energy) of each group was computed, as were the pairwise intermolecular van der Waals interaction energies between all groups.

Configurational Entropies of Protein Side Chains. The method used to estimate entropies was a histogram-based approach. To calculate entropy from each simulation, torsion angles from every side chain were extracted every picosecond and grouped into 10° bins. While each residue was treated independently, the torsions of a given side chain were not; for a side chain with N torsional degrees of freedom, a N-dimensional histogram was generated. The frequencies in each bin (P_i) were used to calculate the entropy using the relation $S = -RT\sum(P_i \ln P_i)$, where R is the gas constant and T is temperature (300 K).

RESULTS

Structural Models of Trisaccharide Recognition. Initial models of the three trisaccharides highlighted in Figure 1b were constructed from the structure of CVN in complex with α-Man-(1,2)-α-Man by extending the dimannose by one monosaccharide; the starting disaccharide structures for this procedure were computationally refined models based on the solution (NMR) structure, as previously described.[26] To construct an α-(1,2)-α-(1,2)-linked trimannose with this procedure, one may consider either the (new) monosaccharide or the (existing) disaccharide to be the reducing sugar. The two approaches result in chemically

dx.doi.org/10.1021/ja305755b | J. Am. Chem. Soc. 2012, 134, 19639−19651

Table 1. Asymmetric Recognition of α-Man-(1,2)-α-Man-(1,2)-α-Man[a]

	ΔG^{bind}	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{h\phi}$
Domain A				
Terminal binding mode	−31.7 (0.1)	−7.4 (0.1)	−21.8 (0.1)	−2.5 (0.0)
	−30.5 (0.1)	−6.5 (0.1)	−21.5 (0.1)	−2.5 (0.0)
Internal binding mode	−34.3 (0.1)	−10.0 (0.1)	−21.6 (0.1)	−2.7 (0.0)
	−34.0 (0.1)	−10.6 (0.1)	−20.7 (0.1)	−2.7 (0.0)
Domain B				
Terminal binding mode	−37.5 (0.1)	−10.9 (0.1)	−24.1 (0.1)	−2.6 (0.0)
	−35.7 (0.1)	−9.4 (0.1)	−23.7 (0.1)	−2.6 (0.0)
Internal binding mode	−28.1 (0.1)	−0.9 (0.1)	−24.7 (0.1)	−2.5 (0.0)
	−27.5 (0.1)	+0.2 (0.1)	−25.2 (0.1)	−2.6 (0.0)

[a]All energies are in kcal/mol. For each value, the results of both singly and doubly bound simulations are provided as the first and second row, respectively. $\Delta G^{bind} = \Delta G^{elec} + \Delta G^{vdw} + \Delta G^{h\phi}$.

identical trisaccharides but different bound state geometries; both of these were considered. Each trisaccharide was modeled in the two binding sites both independently and in a doubly bound form, and each model was then subjected to explicit-solvent MD simulation.

Backbone Fluctuations in Domain B Binding Site Loop. In our first set of simulations, domain A behaved stably and gave consistent results (both in structural and energetic terms) in the doubly and singly bound models, while domain B showed a fundamental lack of stability, with the carbohydrate dissociating from the protein in many cases. Visual analysis localized the structural plasticity of domain B to the "hinge" region of residues 50−54. In the starting NMR structure (Figure 2a), the backbone amide protons of Ser 52 and Asn 53 clash (1.56 Å H−H distance), resulting in a highly strained configuration; during the simulations, two alternate conformations were sampled that relieved this strain. In all of the simulations, a crank-shaft-like rotation of the peptide bond between Ser 52 and Asn 53 was observed (Figure 2b: the movement involves a concerted rotation of the ϕ dihedral of Ser 52 and the ψ dihedral of Asn 53). While the energetic barrier between this state and the starting structure was low (as characterized by a 60 ns average lifetime of the initial state), it was not particularly stable; in several cases transitions back to the initial state were observed, and visual analysis linked sugar dissociation to this state.

To further characterize this alternate conformation, a second set of simulations was carried out, beginning with trisaccharide structures modeled into the alternate protein structure. These simulations showed only moderate backbone flexibility in the hinge region, with roughly 60° fluctuations in the ϕ dihedral of Ser 52 and the ψ dihedral of Asn 53, but in every case the sugar disassociated within 100 ns. Visual analysis of the structures provides a clear rationale for this behavior: The backbone carbonyl of Ser 52 makes a hydrogen-bonded interaction with the hydroxyl at position 3 of the nonreducing sugar in the NMR structure, but this is lost (and replaced by an unfavorable interaction with the backbone amide proton) in the alternate configuration.

Identification of a *cis*-Peptide Bond in Domain B. In a single simulation, a second transition, involving a *trans*- to *cis*-isomerization of the peptide bond between P51 and S52, was observed (Figure 2c). Unsurprisingly, the barrier to this transition is very high, and only a single transition was seen in a total of about 3 μs of simulation. However, the structure after the transition remained stable for the duration of the simulation (100 ns). In order to further assess the stability of this conformer, models of all trisaccharides were again constructed in this background and subjected to the same simulation protocol. In every case, the structure remained remarkably stable, with only small thermal fluctuations around a single structure in the hinge region. Additionally, all oligosaccharide ligands remained stably bound throughout the simulations, which have been carried out to beyond 200 ns (over 2.5 μs of combined simulation time).

Calculation of Trisaccharide Binding Free Energies. Simulations of each sugar (doubly bound as well as singly bound in each site) were carried out for 200 ns from a *cis*-peptide-containing starting structure, and rigid-body binding free energies were computed for 1500 evenly spaced (every 100 ps) snapshots using an MM/PBSA model. In order to minimize potential bias from the initial structure, the first 50 ns of each simulation were excluded from further analysis (See Supporting Information, SI, for time dependence of computed energies). When average binding energies computed this way (over the last 150 ns of simulation) were compared to those computed over the entire 200 ns simulation, the effect was less than 1 kcal/mol for 87.5% (14/16) of the simulations (and less than 0.5 kcal/mol in 75% (12/16)); in the remaining two cases (both for domain B), excluding the first 50 ns reduced the computed binding free energies by 1.2 kcal/mol. Overall semirigid binding free energies (ΔG^{bind}) were further broken down into contributions from electrostatics (ΔG^{elec}), from bound-state van der Waals interactions (ΔG^{vdw}), and from nonpolar solute−solvent interactions ($\Delta G^{h\phi}$). The electrostatic term includes contributions from loss of favorable interactions with solvent in the bound state (relative to the unbound state) as well as from solvent-screened Coulombic interactions made in the bound state; the hydrophobic solvation term is directly proportional to the solvent-accessible surface area buried upon binding. For all computed energies, results are presented as ensemble-averaged values over all 1500 snapshots; autocorrelation analysis (SI) suggests that binding free energies computed from these snapshots are statistically independent.

Contributions from individual side chains and backbone amino and carbonyl groups (or groupings of these components), $\Delta\Delta G^{group}$, were similarly computed from ensemble averages of electrostatic, van der Waals, and nonpolar solvation terms. The electrostatic term was computed as the difference in computed binding free energy between the native sequence and a hypothetical mutant in which only the groups under consideration have been mutated to hydrophobic isosteres; this has been termed the "mutation energy" in previous work on continuum electrostatic component analysis.[48,55,56] The van der Waals term is simply the sum of the bound-state (intermolecular) van der Waals interactions made by the atoms of the group of

Table 2. Energetic Determinants of Internal vs Terminal Recognition[a]

	$\Delta\Delta G^{group}$	$\Delta\Delta G^{elec}$	$\Delta\Delta G^{vdw}$	$\Delta\Delta G^{h\phi}$
L1/P51 Backbone Amino				
A (L1) terminal	−0.3/−0.2	−0.1/0.0	−0.2/−0.2	0.0/0.0
A (L1) internal	−1.5/−1.6	−1.4/−1.4	−0.1/−0.1	0.0/−0.1
B (P51) terminal	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0
B (P51) internal	0.0/+0.1	+0.1/+0.2	−0.1/−0.1	0.0/0.0
E101/Q50 Side Chain				
A (E101) terminal	−0.2/−0.4	−0.1/−0.3	−0.1/−0.1	0.0/0.0
A (E101) internal	−4.0/−4.3	−5.5/−6.1	+1.5/+1.8	0.0/0.0
B (Q50) terminal	−0.1/−0.1	0.0/0.0	−0.1/−0.1	0.0/0.0
B (Q50) internal	−0.7/−0.8	−0.5/−0.6	−0.2/−0.2	0.0/0.0
A92/E41 Side Chain				
A (A92) terminal	−0.4/−0.4	+0.1/+0.1	−0.5/−0.5	0.0/0.0
A (A92) internal	−0.8/−0.8	+0.1/+0.1	−0.9/−0.9	0.0/0.0
B (E41) terminal	−2.1/−2.0	−2.6/−2.4	+0.6/+0.5	−0.1/−0.1
B (E41) internal	+3.0/+3.8	+4.8/+6.0	−1.8/−2.0	0.0/−0.2
T25/R76 Side Chain				
A (T25) terminal	−2.4/−2.5	−0.4/−0.4	−1.8/−1.8	−0.2/−0.3
A (T25) internal	−1.2/−1.4	−0.1/−0.2	−1.0/−1.1	−0.1/−0.1
B (R76) terminal	−9.5/−9.3	−4.3/−4.1	−4.8/−4.8	−0.4/−0.4
B (R76) internal	−3.8/−4.5	−0.8/−1.0	−2.9/−3.3	−0.1/−0.2
Subtotal				
A terminal mode	−3.1/−3.2	−0.3/−0.3	−2.6/−2.6	−0.2/−0.3
A internal mode	−5.7/−6.2	−5.1/−5.7	−0.5/−0.3	−0.1/−0.2
B terminal mode	−11.5/−11.2	−6.7/−6.3	−4.3/−4.4	−0.5/−0.5
B internal mode	−0.9/−0.8	+4.2/+5.2	−5.0/−5.6	−0.1/−0.4
Remainder				
A terminal mode	−28.6/−27.3	−7.1/−6.2	−19.2/−18.9	−2.3/−2.2
A internal mode	−28.6/−27.8	−4.9/−4.9	−21.1/−20.4	−2.6/−2.5
B terminal mode	−26.1/−24.5	−4.2/−3.1	−19.8/−19.3	−2.1/−2.1
B internal mode	−27.2/−26.8	−5.1/−5.0	−19.7/−19.6	−2.4/−2.2

[a]All energies are in kcal/mol; for each entry, the two values given are those from the singly and doubly bound simulations, respectively. $\Delta\Delta G^{group} = \Delta\Delta G^{elec} + \Delta\Delta G^{vdw} + \Delta\Delta G^{h\phi}$.

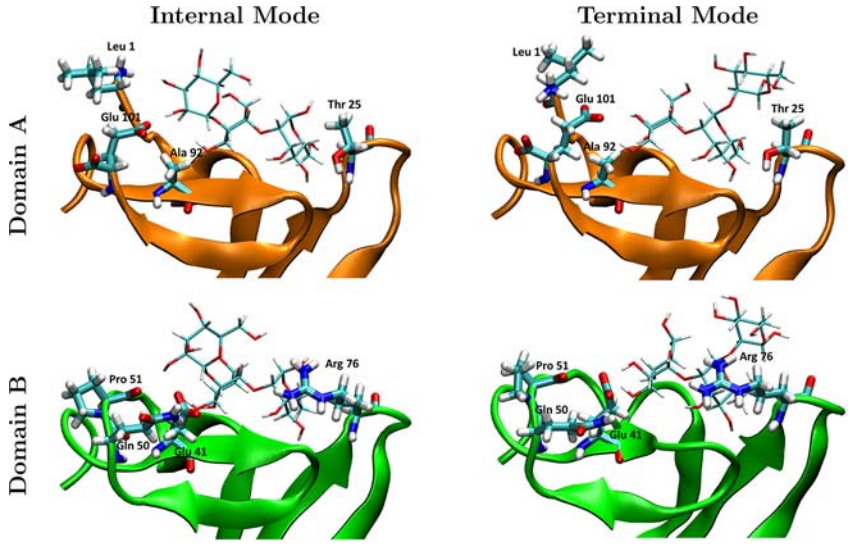


Figure 3. Internal and terminal binding modes of α-Man-(1,2)-α-Man-(1,2)-α-Man. The top row (orange protein chain) shows α-Man-(1,2)-α-Man-(1,2)-α-Man (thin licorice) bound to domain A in both the internal and terminal modes, with key interacting protein side chains shown in thick licorice. The bottom row (green protein chain) shows the same for domain B.

interest, and the hydrophobic solvation term is directly related to the change in solvent-accessible surface area of the group. It should be noted that while the van der Waals and surface-area dependent terms are additive within a group of multiple components, the electrostatic term is not; simply adding the electrostatic terms of individual components would doubly count any electrostatic interactions between them.

dx.doi.org/10.1021/ja305755b | J. Am. Chem. Soc. 2012, 134, 19639−19651

Table 3. Overall Energetics of Trisaccharide Binding[a]

	ΔG^{bind}	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{h\phi}$
Domain A				
α-(1,2), α-(1,2), α-Man_3[b]	−34.3 (0.1)	−10.0 (0.1)	−21.6 (0.1)	−2.7 (0.0)
	−34.0 (0.1)	−10.6 (0.1)	−20.7 (0.1)	−2.7 (0.0)
α-(1,2), α-(1,3), α-Man_3	−29.5 (0.1)	−5.7 (0.1)	−21.4 (0.1)	−2.4 (0.0)
	−30.2 (0.1)	−6.9 (0.1)	−21.0 (0.1)	−2.4 (0.0)
α-(1,2), α-(1,6), α-Man_3	−30.3 (0.1)	−7.3 (0.1)	−20.6 (0.1)	−2.4 (0.0)
	−31.3 (0.1)	−7.2 (0.1)	−21.5 (0.1)	−2.5 (0.0)
Domain B				
α-(1,2), α-(1,2), α-Man_3	−37.5 (0.1)	−10.9 (0.1)	−24.1 (0.1)	−2.6 (0.0)
	−35.7 (0.1)	−9.4 (0.1)	−23.7 (0.1)	−2.6 (0.0)
α-(1,2), α-(1,3), α-Man_3	−33.2 (0.1)	−8.0 (0.1)	−22.7 (0.1)	−2.4 (0.0)
	−32.2 (0.1)	−7.6 (0.1)	−22.2 (0.1)	−2.4 (0.0)
α-(1,2), α-(1,6), α-Man_3	−32.5 (0.1)	−7.6 (0.1)	−22.4 (0.1)	−2.5 (0.0)
	−33.0 (0.1)	−8.2 (0.1)	−22.2 (0.1)	−2.6 (0.0)

[a]All energies are in kcal/mol; for each value, the results of both singly and doubly bound simulations are provided as the first and second row, respectively. $\Delta G^{bind} = \Delta G^{elec} + \Delta G^{vdw} + \Delta G^{h\phi}$. [b]For Domain A, the α-(1,2), α-(1,2)-linked sugar is bound in the internal orientation; all other sugars are bound in the terminal mode.

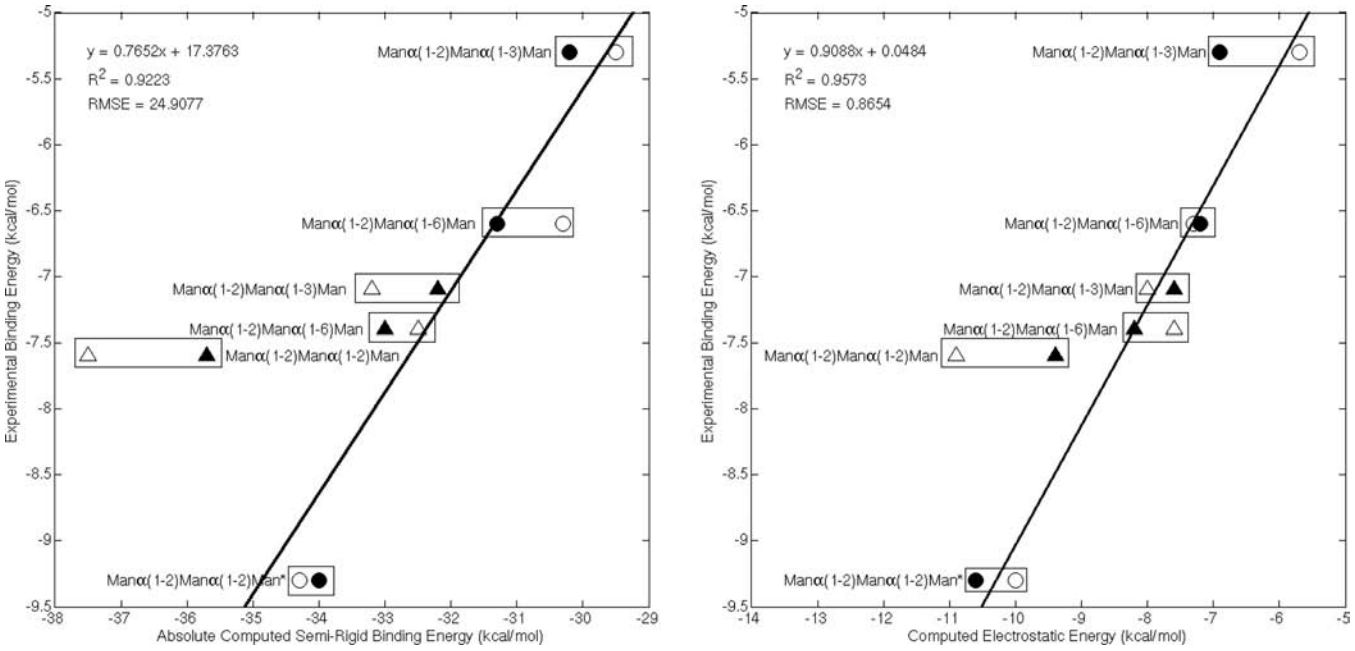


Figure 4. Comparison of computed and experimental binding free energies. The variation of experimentally determined binding free energies (from Bewley et al.)[39] for three trisaccharides in both domain A and B is plotted against the total computed binding free energy (left panel) and against the electrostatic component (right panel). Data for domain A are shown in circles, while those for domain B are shown in triangles; open symbols denote values calculated from simulations with a single site bound, while closed symbols denote the results from doubly bound simulations. With the sole exception of α-Man-(1,2)-α-Man-(1,2)-α-Man in domain B, there is remarkable correlation between the computed and the experimental results; this correlation is enhanced when only the electrostatic component is considered.

Asymmetric Recognition of D1-Arm Trisaccharides in Domains A and B. Table 1 details the computed binding free energies for α-Man-(1,2)-α-(1,2)-α-Man (the trisaccharide corresponding to the D1 arm of high-mannose oligosaccharides) in each of two possible binding orientations discussed above; we term the orientation in which the anomeric hydroxyl of the original disaccharide is unchanged as the "internal" binding mode, and that where this hydroxyl is linked to the reducing sugar as the "terminal" binding mode. In domain A, the internal binding mode is preferred by about 3.0 kcal/mol, and in domain B, on the other hand, it is the terminal mode that is preferred (by upward of 8 kcal/mol). In both cases, the differences result entirely from the electrostatic contribution to the binding free energies.

The determinants of these differences can be narrowed down to a small number of key interactions that differ between the two binding modes in each site and are detailed in Table 2 and Figure 3. In domain A, Glu 101 makes a strongly favorable interaction with the last sugar in the internal binding mode. The charged N-terminus similarly makes a more moderately favorable interaction in this state, while both interactions are absent when the trisaccharide is bound in the terminal mode. In domain B, on the other hand, only slightly favorable interactions with Gln 50 are made in the internal mode, and the free N-terminus is replaced with (noninteracting) Pro 51. Additionally, Glu 41 in

dx.doi.org/10.1021/ja305755b | J. Am. Chem. Soc. 2012, 134, 19639−19651

Table 4. Energetic Determinants of General Affinity for Cognate Sugars[a]

	$\Delta\Delta G^{group}$	$\Delta\Delta G^{elec}$	$\Delta\Delta G^{vdw}$	$\Delta\Delta G^{h\phi}$
Backbone Contributions				
Residues 2−4 (A)	[−6.5/−5.8]	[−6.5/−5.2]	[−1.3/+0.7]	[−0.1/0.0]
Residues 52−54 (B)	[−5.3/−5.1]	[−4.8/−4.4]	[−0.7/−0.4]	[0.0/0.0]
Residues 23−24 (A)	[−4.7/−4.1]	[−3.8/−2.9]	[−1.2/−0.8]	[0.0/0.0]
Residues 74−75 (B)	[−5.0/−3.6]	[−4.3/−2.6]	[−1.1/−0.7]	[0.0/0.0]
Residues 92−95 (A)	[−9.9/−9.5]	[−7.8/−7.2]	[−2.7/−1.8]	[−0.1/0.0]
Residues 41−44 (B)	[−10.5/−8.4]	[−8.1/−5.5]	[−2.9/−2.2]	[−0.1/0.0]
Q6/E56 Side Chain				
Domain A (Q6)	[−1.2/−1.1]	[−0.1/−0.1]	[−1.0/−0.9]	[−0.1/−0.1]
Domain B (E56)	[−1.1/−0.9]	[−0.5/−0.3]	[−0.7/−0.5]	[−0.1/0.0]
T7/T57 Side Chain				
Domain A (T7)	[−3.7/−3.4]	[−1.6/−1.5]	[−2.0/−1.7]	[−0.2/−0.2]
Domain B (T57)	[−3.6/−3.2]	[−1.7/−1.4]	[−1.7/−1.4]	[−0.2/−0.1]
Subtotal				
Domain A	[−24.9/−23.7]	[−17.9/−17.2]	[−7.4/−5.7]	[−0.5/−0.3]
Domain B	[−24.6/−20.6]	[−18.6/−13.5]	[−6.8/−5.3]	[−0.3/−0.2]

[a]All energies are in kcal/mol; for each entry, the range of values seen over all simulations is given. $\Delta\Delta G^{group} = \Delta\Delta G^{elec} + \Delta\Delta G^{vdw} + \Delta\Delta G^{h\phi}$.

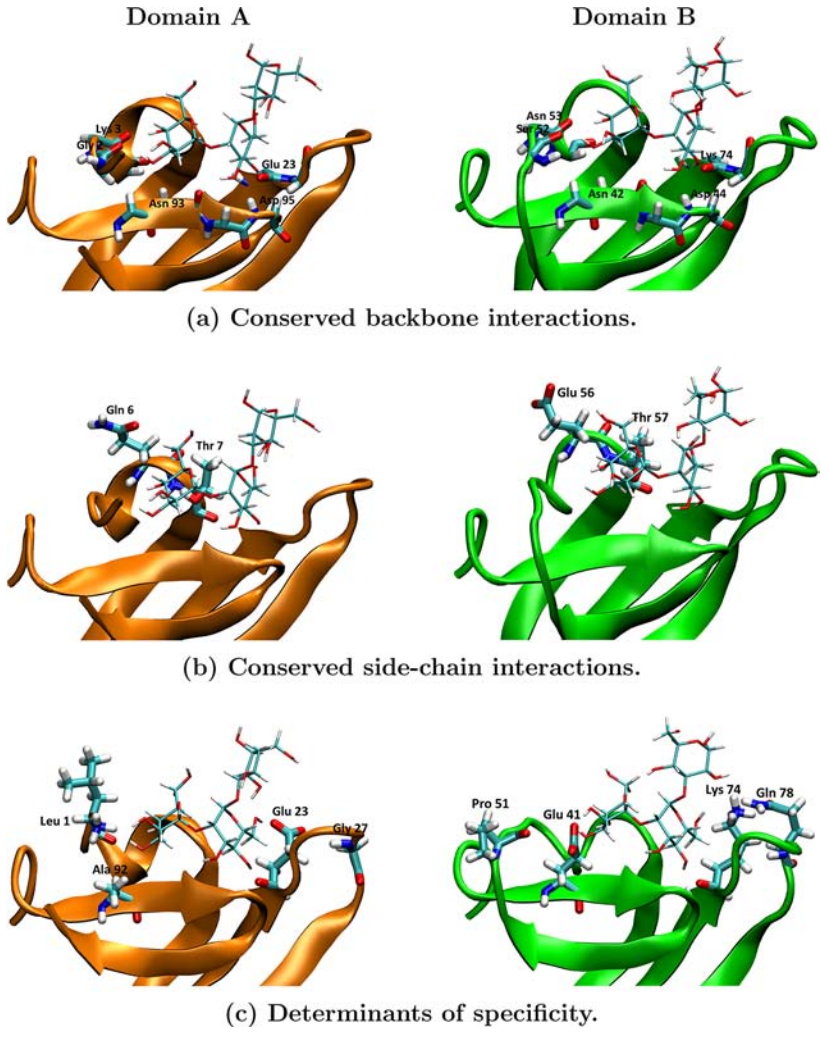


Figure 5. Determinants of general affinity for cognate sugars and of specificity for distinct targets. The location of key protein groups (thick licorice) is displayed with a representative carbohydrate (thin licorice). The top row (a) shows backbone groups contributing to general affinity for all cognate sugars, while the middle row (b) shows the side-chain groups that make similar contributions. The bottom row (c) displays those groups that have been identified as making differential contributions to binding of different ligands in each domain. Domain A (orange protein chain) is on the left, and domain B (green protein chain) is on the right.

dx.doi.org/10.1021/ja305755b | J. Am. Chem. Soc. 2012, 134, 19639−19651

Table 5. Determinants of Differences in Affinity for Terminal Mode[a]

	$\Delta\Delta G^{group}$	$\Delta\Delta G^{elec}$	$\Delta\Delta G^{vdw}$	$\Delta\Delta G^{h\phi}$
A92/E41 Side Chain				
Domain A (A92)	[−0.4/−0.3]	[+0.1/+0.1]	[−0.5/−0.4]	[0.0/0.0]
Domain B (E41)	[−2.2/−1.9]	[−2.7/−2.4]	[+0.5/+0.8]	[−0.1/−0.1]
G27/Q78 Side Chain				
Domain A (G27)	[0.0/0.0]	[0.0/0.0]	[0.0/0.0]	[0.0/0.0]
Domain B (Q78)	[−1.3/−1.2]	[−0.7/−0.5]	[−0.8/−0.4]	[−0.1/0.0]
E23/K74 Side Chain				
Domain A (E23)	[−2.7/−2.5]	[−3.1/−2.8]	[+0.3/+0.5]	[−0.1/0.0]
Domain B (K74)	[−1.9/−1.7]	[−0.9/ −0.6]	[−1.1/−1.0]	[0.0/0.0]
L1/P51 Backbone Carbonyl				
Domain A (L1)	[−1.8/−0.2]	[−1.9/ +0.1]	[−0.3/+0.1]	[0.0/0.0]
Domain B (P51)	[−0.2/−0.1]	[−0.1/ 0.0]	[−0.1/−0.1]	[0.0/0.0]
Subtotal				
Domain A	[−4.8/−3.2]	[−4.9/−2.7]	[−0.5/+0.2]	[−0.1/0.0]
Domain B	[−5.5/−5.1]	[−4.2/−3.9]	[−1.3/−1.0]	[−0.2/−0.1]

[a]All energies are in kcal/mol; for each entry, the range of values seen over all simulations is given. $\Delta\Delta G^{group} = \Delta\Delta G^{elec} + \Delta\Delta G^{vdw} + \Delta\Delta G^{h\phi}$.

Table 6. Interactions of the Arg 76/Asp 44 Salt Bridge and Its Equivalent[a]

	$\Delta\Delta G^{group}$	$\Delta\Delta G^{elec}$	$\Delta\Delta G^{vdw}$	$\Delta\Delta G^{h\phi}$
Arg 76 Side Chain (B)				
α-(1,2), α-(1,2), α-Man_3	−9.5/−9.3	−4.3/−4.1	−4.8/−4.8	−0.4/−0.4
α-(1,2), α-(1,3), α-Man_3	−6.0/−5.1	−1.7/−1.4	−4.1/−3.5	−0.2/−0.2
α-(1,2), α-(1,6), α-Man_3	−4.6/−7.3	−0.7/−2.4	−3.7/−4.4	−0.2/−0.5
Asp 44 Side Chain (B)				
α-(1,2), α-(1,2), α-Man_3	+0.4/0.0	+1.0/+0.4	−0.6/−0.4	0.0/0.0
α-(1,2), α-(1,3), α-Man_3	0.0/−0.8	+0.5/−0.8	−0.5/0.0	0.0/0.0
α-(1,2), α-(1,6), α-Man_3	−0.9/−3.8	−0.8/−4.7	−0.1/+0.9	0.0/0.0
Domain B Subtotal				
α-(1,2), α-(1,2), α-Man_3	−8.2/−8.4	−2.4/−2.8	−5.4/−5.2	−0.4/−0.4
α-(1,2), α-(1,3), α-Man_3	−5.2/−4.7	−0.4/−1.0	−4.6/−3.5	−0.2/−0.2
α-(1,2), α-(1,6), α-Man_3	−4.6/−8.9	−0.6/−4.9	−3.8/−3.5	−0.2/−0.5
Thr 25 Side Chain (A)				
α-(1,2), α-(1,2), α-Man_3[b]	−1.2/−1.4	−0.1/−0.2	−1.0/−1.1	−0.1/−0.1
α-(1,2), α-(1,3), α-Man_3	−2.1/−2.4	−0.2/−0.5	−1.7/−1.7	−0.2/−0.2
α-(1,2), α-(1,6), α-Man_3	−2.4/−2.0	−0.4/0.0	−1.7/−1.8	−0.3/−0.2
Asp 95 Side Chain (A)				
α-(1,2), α-(1,2), α-Man_3[b]	+0.2/+0.3	+0.7/+0.8	−0.5/−0.5	0.0/0.0
α-(1,2), α-(1,3), α-Man_3	+0.5/+0.4	+1.0/+0.9	−0.5/−0.5	0.0/0.0
α-(1,2), α-(1,6), α-Man_3	+0.4/+0.6	+0.9/+1.0	−0.5/−0.4	0.0/0.0
Domain A Subtotal				
α-(1,2), α-(1,2), α-Man_3[b]	−1.1/−1.3	+0.5/+0.4	−1.5/−1.6	−0.1/−0.1
α-(1,2), α-(1,3), α-Man_3	−1.9/−2.3	+0.5/+0.1	−2.2/−2.2	−0.2/−0.2
α-(1,2), α-(1,6), α-Man_3	−2.2/−1.7	+0.3/+0.7	−2.2/−2.2	−0.3/−0.2

[a]All energies are in kcal/mol; for each entry, the two values given are those from the singly and doubly bound simulations, respectively. $\Delta\Delta G^{group} = \Delta\Delta G^{elec} + \Delta\Delta G^{vdw} + \Delta\Delta G^{h\phi}$. [b]For Domain A, the α-(1,2), α-(1,2)-linked sugar is bound in the internal orientation; all other sugars are bound in the terminal mode.

domain B makes favorable interactions in the context of terminal-mode binding but unfavorable interactions (for a net difference of over 5 kcal/mol) in the internal mode, and in domain A, the analogous residue (Ala 92) makes no significant interactions in either mode. Finally, both Thr 25 (domain A) and its corresponding residue in domain B (Arg 76) make more favorable interactions in the terminal model. For Thr 25 this is primarily a van der Waals effect and serves to slightly offset the enhanced stabilization of the internal mode in domain A by Glu 101 and the N-terminus. For Arg 76, on the other hand, a combination of both electrostatic and van der Waals interactions acts to stabilize the terminal mode in domain B, reinforcing the destabilization of the internal mode by Glu 41. It should be noted, however, in the context of larger oligosaccharides, interactions involving Thr 25 and Arg 76 would likely be affected by additional carbohydrates.

When the contributions of these four groups are taken together (Table 2, subtotal), they account almost entirely for the differences between the two binding modes in both sites. While a significant fraction of the overall binding energies (about 27 kcal/mol) come from other groups (Table 2, remainder), the contribution of these remaining groups is near equal in both binding sites and with both binding modes.

dx.doi.org/10.1021/ja305755b | J. Am. Chem. Soc. 2012, 134, 19639−19651

Near Quantitative Prediction of Relative Binding Free Energies. Table 3 shows the computed affinities for each of the three sugars in each binding site; for α-Man-(1,2)-α-(1,2)-α-Man the values given are for the preferred binding mode in each site. The two computed values for each sugar/binding site combination (from singly and doubly bound simulations) are strongly consistent, with the calculations agreeing to within 1.0 kcal/mol for all but one case. As the structure is quite rigid and the binding sites are separated by ∼40 Å, no cooperativity in binding is expected nor has cooperativity been experimentally observed.[39]

Figure 4 shows these same values plotted against the experimentally determined binding free energies of Bewley and co-workers.[39] The computed energies show remarkable correlation with the experimental values for five of the six cases; these data lie along a best-fit line with a slope of 0.77. Interestingly, the electrostatic component correlates even more strongly with the experimental values, with a best-fit slope of 0.91, and deviation of the single outlier from this line is less than for the total energy. This outlier, α-Man-(1,2)-α-(1,2)-α-Man bound in domain B, also shows the least consistency between the values computed for the singly and doubly bound states, with a difference of 1.8 kcal/mol.

Conserved Backbone Interactions Are Primary Source of Affinity. A large fraction of the binding affinity for all three sugars in both sites can be attributed to a common set of interactions, primarily involving backbone contacts; the contributions of these groups are detailed in Table 4 and Figure 5a,b. Overall, these groups account for roughly −20 kcal/mol of overall affinity, with about −16 kcal/mol of this from electrostatics and −4 kcal/mol from van der Waals contacts. In terms of electrostatic contributions, this is a more favorable contribution than the total overall electrostatic contribution to affinity. For van der Waals interactions, on the other hand, these groups contribute a much smaller fraction.

A Small Number of Residues Mediate Affinity Differences between Domains. A small number of groups make consistent interactions within each domain but show significant differences between the two binding sites (Table 5 and Figure 5c). One of these is Ala 92/Glu 41, previously noted to contribute significantly to the differences in the terminal versus internal binding orientation of α-Man-(1,2)-α-(1,2)-α-Man. In domain B, Glu 41 makes strongly favorable electrostatic interactions will all sugars (bound in a terminal orientation), while Ala 92 in domain A contributes almost nothing to the affinity. Similarly, Gln 78 in domain B makes a moderate electrostatic interaction as well as favorable van der Waals contact, for a net favorable contribution of about 1 kcal/mol; the corresponding amino acid in domain A is Gly 27, with no side chain to make interactions with the sugar. On the other hand, Glu 23 in domain A makes much stronger electrostatic interactions (about −3.0 kcal/mol) than does the equivalent residue in domain B (Lys 74, between −0.6 and −0.9 kcal/mol), although these differences are modulated somewhat by opposing differences in van der Waals contact.

One Pair of Interacting Residues Dominate Specific Sugar Recognition in Domain B. A single pair of interacting residues (Arg 76 and Asp 44) in domain B make strong contributions to binding that are not seen in domain A (Table 6). Arg 76 makes hydrogen-bonded contacts with the sugar for approximately 90% of both the singly and doubly bound simulations of α-Man-(1,2)-α-(1,2)-α-Man, while similar contacts were made with α-Man-(1,2)-α-(1,3)-α-Man less than 25% of the time; the α-Man-(1,2)-α-(1,6)-α-Man was somewhat less consistent in the stability of this interaction, with hydrogen bonding observed only 11% of the time in the singly bound simulation but 48% of the time in the other. In all cases, the side chain of Arg 76 interacts with the "third" mannose that extends from the core dimannose binding site.

While visual inspection of these interactions may suggest a largely electrostatic effect, very significant van der Waals interactions (between −3.5 and −4.8 kcal/mol) are also made between Arg 76 and the sugar in all simulations. In fact, roughly 12% of the total protein−carbohydrate van der Waals energy in domain B derives from this single residue.

Arg 76 is also able to form a salt bridge with Asp 44. This interaction seems to partially compete for the interactions with the sugar, although simultaneous formation of the salt bridge and hydrogen bonds to the carbohydrate was observed in some cases (particularly in the α-Man-(1,2)-α-(1,2)-α-Man simulations). Asp 44 is also able to make direct hydrogen-bonding interactions with the sugar, although this was observed less frequently; one exception to this, is the doubly bound simulation of α-Man-(1,2)-α-(1,6)-α-Man, in which a strongly favorable contribution to binding arises from this interaction. In all simulations, states in which Arg 76 and Asp 44 are fully solvated were also observed.

DISCUSSION

Monomeric CVN Contains a *cis*-Peptide Bond in the Hinge Region. One of the most significant results of the simulations presented here is the strong suggestion that monomeric CVN contains a *cis*-peptide bond in the hinge region. As *cis*-peptide bonds are rare, these observations warrant some discussion. First, in the original NMR structure determination, minimal constraints were available in this region, and thus the backbone structure of the hinge was not rigorously defined (C. Bewley, personal communication). It is not surprising that, in absence of explicit restraints, annealing would not sample the *cis*-conformer. Second, the existence of a Pro 51−Ser 52 *cis*-peptide bond provides a structural and energetic explanation for an interesting observation involving a CVN variant containing a Pro to Gly mutation at position 51, previously characterized by Gronenborn and colleagues. This mutation has been shown to preferentially stabilize the monomeric form of CVN; the concentration of guanadinium hydrochloride required to unfold the monomer is increased by 0.7 M,[57,58] and variants containing the P51G mutant crystallize exclusively as monomers.[59,60] Typically, mutations to glycine destabilize proteins, unless they are made in regions requiring violations of the typically allowed region of Ramachandran space, due to increased entropy in the unfolded state. As the P51G variant has additional flexibility that would allow avoidance of the *cis*-peptide bond, we identify in the wild-type monomer that the observed stabilization is explicable; as the *cis*-peptide bond is not present in the domain-swapped dimer, this stabilization should (as it does) be preferentially observed in the monomer. Taken as a whole, these observations lead us to the conclusion that wild-type, monomeric CVN likely contains a *cis*-peptide bond between Pro 51 and Ser 52. It is worth noting that additional mutations in the hinge region (such as ΔQ50 and S52P) can force increased population of the domain swapped dimer,[61,63] highlighting the sensitivity of this portion of the structure to backbone strain.

D1 Arm Is Asymmetrically Recognized by Domains A and B. Oligosaccharide recognition by CVN involves very interesting differences in the preferences of each domain for

binding α-(1,2)-linked mannobiose and a trimannose with two α-(1,2)-linkages. Bewley and co-workers observed that while the disaccharide binds domain B tighter (by roughly 10-fold in affinity) than domain A, the trisaccharide preferentially binds domain A by a similar amount. The simulations presented here provide direct insight into this specificity switch, with an interesting structural mechanism.

The α-(1,2), α-(1,2) trimannose could potentially bind in one of two binding modes—both maintaining the α-(1,2)-linked mannobiose in the same orientation—with the third sugar extending either from the reducing or nonreducing end of the disaccharide. In the context of a larger oligosaccharide, the first would correspond to recognition of the last two sugars on the D1 arm as the mannobiose anchor, and thus we term this the "terminal" binding mode. The second, on the other hand, would correspond to recognition of the second and third sugars from tip of D1 as the anchor, and thus this mode is termed the "internal" orientation. In the crystal structure of Man9 bound to the domain-swapped dimer, only domain A is occupied, and an internal binding mode is observed.[41]

Consistent with the crystal structure, the calculations indicate a strong preference of domain A for the internal binding mode but notably show an even stronger bias against the internal mode in domain B; the net binding energy is 3.5 kcal/mol more favorable for the internal mode in domain A and over 8 kcal/mol less favorable for the internal mode in domain B. The source of these preferences is entirely electrostatic in nature, with differences in buried surface area nearly zero and differences in van der Waals interactions slightly opposing the net difference. In domain A, the internal binding mode makes strongly favorable electrostatic interactions (−10.6 kcal/mol), compared with more moderately favorable interactions (−6.5 kcal/mol) for the terminal mode. In domain B, net electrostatic contributions are reduced from a (favorable) contribution of −9.4 kcal/mol for the terminal binding mode to essentially zero (+0.2 kcal/mol) for the internal mode. The results are entirely consistent between both doubly and singly bound models.

These data were calculated based on semirigid binding free energies, averaging energies over 1500 bound state configurations but with no explicit consideration of any conformational differences between the bound and the unbound states. However, in comparing binding modes for a single ligand, the unbound state is by definition identical, and thus no error arises from this neglect. Additionally missing is an assessment of bound state configurational entropy of both the protein and carbohydrate. While there is likely increased entropy in the less tightly bound states, the magnitude of these differences is unlikely to outweigh the large differences that our computed results suggest.

The origins of this asymmetry in recognition have a clear basis in structure. In domain A, several groups make favorable interactions in the internal binding mode that are absent (or weaker) in the terminal binding mode. Chief among these is the side chain of Glu 101, which makes a strongly favorable interaction with the hydroxyl at position C2 of the third monosaccharide in the internal mode (and makes no interactions in the terminal mode). Additionally, the free N-terminus makes a moderately favorable interaction with the hydroxyl at C3. Slightly modulating these favorable contributions, the backbone carbonyl of Gly 2 makes slightly weaker interactions with the hydroxyl at C3 of the central sugar in the context of the internal binding mode.

Simulations Provide a Structural Map of the Determinants of Binding Affinity and Specificity. Overall, the computed binding free energies show remarkable agreement with the relative energies measured experimentally, and this agreement strongly motivates the use of the computed results to gain deeper insight into the modes of specific carbohydrate recognition by CVN. One of the great advantages of the use of the linearized Poisson−Boltzmann continuum model is a strict pairwise decomposability of the electrostatic contribution to free energy. This allows for a well-defined assignment of the energetic contributions from individual functional groups; solvent-accessible surface areas and Lennard-Jones (van der Waals) energies can be similarly partitioned. Performing such a group-wise component analysis to all the systems studied here resulted in a quite simple partitioning of binding site groups into various functional roles.

First, a large portion of the affinity in both sites (about −24 kcal/mol) derives from backbone contacts with the core α-(1,2)-linked dimannose. Only two side chain groups (Gln 6 and Thr 7 in domain A, and Glu 56 and Thr 57 in domain B) make interactions with the same disaccharide. In this case, the electrostatic energies (about −17 kcal/mol) were strongly dominant, with a more moderate contribution from van der Waals interactions (about −7 kcal/mol).

Overlaying these conserved general affinity determinants, a few residues make consistent interactions within each domain but show distinct differences between corresponding residues in the two domains. For example, Glu 41 and Gln 78 make stabilizing interactions with the carbohydrates in domain B, but the corresponding residues in domain A, Ala 92 and Gly 27, do not make interactions; offsetting this effect, in domain A, Glu 23 makes significantly more favorable interactions than its corresponding residue in domain B, Lys 74.

Finally, a small number of groups seem to make different interactions in the context of different bound oligosaccharides, even when all are bound in a similar orientation (the terminal binding mode). The most obvious of these is the Arg 76/Asp 44 pair in domain B, but some variation is also seen in the corresponding positions in domain A (Thr 25 and Asp 95).

The overall picture that arises, then, is that although there are many differences in sequence between the two domains, the functionally relevant differences are easily mapped to specific positions. As it has recently been discovered that CVN is simply one member of a family of homologous proteins that occurs in some multicellular fungi and plants as well as in numerous microbes,[62] many of the results presented here will likely prove to be transferable to related proteins and thus provide a broad understanding of the sequence−function relationships in the CVN family.

Differences in the Change of Configurational Entropy upon Binding May Explain Outliers in Computed Relative Binding Free Energies. While the results presented here go a long way toward explaining the details of specific carbohydrate recognition by CVN and provide near quantitative reproduction of relative binding free energies in many cases, the results are not perfect, with α-Man-(1,2)-α-(1,2)-α-Man bound in domain B a notable outlier. As discussed above, the computed stabilization of this complex is dominated by contributions from a single amino acid, Arg 76. Arginine is one of the most flexible of the naturally occurring amino acids, with four dihedral degrees of freedom. As a result, the formation of persistent interactions by an otherwise flexible arginine may be expected to be accompanied by significant entropic costs.

In order to estimate whether a lack of accounting for this contribution may explain the computed overstabilization, we computed the difference in side chain configurational entropy between the bound and the unbound states for every residue in the protein (shown as contributions to the binding free energy, $-T\Delta S$, Figure 6). As the vast majority of groups have no role in

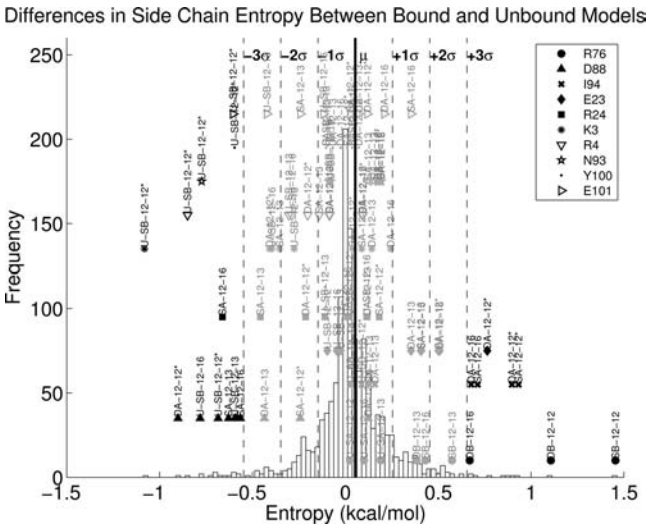


Figure 6. Changes in side-chain entropy on binding. The distribution of differences in computed side-chain entropy between the bound and unbound states is shown as a histogram. The heavy vertical line indicates the mean of the distribution, and the vertical dashed lines correspond to $\pm\sigma$, $\pm2\sigma$, and $\pm3\sigma$ (where σ is the standard deviation of the distribution). All side chains for which any value was outside 3σ are labeled, with the outliers in black and values within 3σ in gray. Points are labeled (U-)[S/D][A/B]-12-yy(*), where S/D denotes results from a singly or doubly bound simulation, A/B denotes the domain, yy denotes the type of linkage for the second glycosidic bond, and * denotes binding in the internal (as opposed to terminal) mode, and a preceding U indicates that the residue is from the unbound domain of a singly bound simulation. Only a few residues show consistent deviations from the bulk distribution, with Arg 76 (domain B) a notable outlier in the context of α-Man-(1,2)-α-Man-(1,2)-α-Man binding.

binding, most of the differences fall in a near-normal distribution centered close to zero; the variance of the distribution is characterized by $\sigma = 0.2$ kcal/mol, and significant outliers may be identified as those points lying beyond $\pm3\sigma$. Only three residues give consistent deviations from the bulk distribution: Arg 76, Asp 88, and Ile 94. Asp 88 shows an increase in bound state entropy (a favorable contribution to binding) for multiple ligands in both domains; as it is not involved in any contacts with any carbohydrates, it is likely that these results are an artifact of an underestimate of the entropy from the single unbound state simulation. Ile 94 (in domain A) pays a slight (≈0.9 kcal/mol) entropic penalty upon internal mode binding of α-Man-(1,2)-α-(1,2)-α-Man and an even smaller penalty (up to 0.7 kcal/mol) upon binding sugars in the terminal mode. The largest entropic penalties by far, however, are for Arg 76 binding α-Man-(1,2)-α-(1,2)-α-Man; a penalty of 1.1 kcal/mol was computed from the doubly bound simulation, and a 1.5 kcal/mol penalty resulted from the singly bound simulation. While the magnitudes of these effects are not quite as large as the deviations from the linear fit of Figure 4, these data strongly suggest that side-chain entropy is a major contribution to the overprediction of the stability of this complex.

These data were computed by assuming that all the side chains behave independently and that backbone configurations do not contribute and without consideration of the flexibility of the carbohydrate. As a result, they should neither be taken as a quantitative prediction of configurational entropic effects nor be directly added to the semirigid binding free energies presented here. Further work will be needed to develop quantitative assessments of these contributions. Similarly, while most of the computed results were highly consistent between singly and doubly bound simulations, certain regions (such as the Arg 76-Asp 44 pair) were seen to be highly dynamic, raising the possibility of imperfect sampling of the various states; longer simulations could help address this question.

CONCLUSION

Through extensive explicit-solvent simulations of the antiviral lectin CVN in complex with a number of oligosaccharide ligands, combined with energetic analysis using a Poisson−Boltzmann-based continuum electrostatic model, we have been able to explain a number of important features of this protein. First among these is that it seems likely that CVN contains a rare *cis*-peptide bond in the monomeric state, which may contribute to its tendency to form domain-swapped dimers under some conditions. Additionally, we have demonstrated that the two domains of CVN, while highly homologous in sequence, recognize certain oligosaccharide targets with distinct binding modes. Finally, we have shown that current computational methods are able to provide a near-quantitative reproduction of relative binding free energies across two binding sites and multiple oligosaccharide ligands. In total, the results demonstrated the importance of careful computational analysis as complement to experimental characterization of structures and binding thermodynamics. As carbohydrate-containing systems present unique experimental challenges, the insights obtained from simulation are particularly meaningful.

ASSOCIATED CONTENT

Supporting Information

Complete time series of computed binding energies for all simulations are provided along with (ensemble averaged) energetic contributions of every protein group.This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

david.green@stonybrook.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research utilized resources at the New York Center for Computational Sciences at Stony Brook University/Brookhaven National Laboratory which is supported by the U.S. Department of Energy under contract no. DE-AC02-98CH10886 and by the State of New York. D.F.G. would also like to thank Bruce Tidor for making the ICE software package available. The work was financially supported by the National Institute of General Medical Sciences (R01-GM086199).

REFERENCES

(1) Lifson, J.; Coutré, S.; Huang, E.; Engleman, E. *J. Exp. Med.* **1986**, *164*, 2101−2106.

(2) Leonard, C. K.; Spellman, M. W.; Riddle, L.; Harris, R. J.; Thomas, J. N.; Gregory, T. J. *J. Biol. Chem.* **1990**, *265*, 10373−10382.
(3) Esser, M. T.; Mori, T.; Mondor, I.; Sattentau, Q. J.; Dey, B.; Berger, E. A.; Boyd, M. R.; Lifson, J. D. *J. Virol.* **1999**, *73*, 4360−4371.
(4) Balzarini, J. *Nat. Rev. Microbiol.* **2007**, *5*, 583−597.
(5) Dey, B.; Lerner, D. L.; Lusso, P.; Boyd, M. R.; Elder, J. H.; Berger, E. A. *J. Virol.* **2000**, *74*, 4562−4569.
(6) Reitter, J. N.; Means, R. E.; Desrosiers, R. C. *Nat. Med.* **1998**, *4*, 679−684.
(7) Balzarini, J.; Van Laethem, K.; Hatse, S.; Froeyen, M.; Peumans, W.; Van Damme, E.; Schols, D. *J. Biol. Chem.* **2005**, *280*, 41005−41014.
(8) Hu, Q.; Mahmood, N.; Shattock, R. J. *Virology* **2007**, *368*, 145−154.
(9) Scanlan, C. N.; Offer, J.; Zitzmann, N.; Dwek, R. A. *Nature* **2007**, *446*, 1038−1045.
(10) McGowan, I. *Biologicals* **2006**, *34*, 241−255.
(11) Klasse, P. J.; Shattok, R.; Moore, J. P. *Annu. Rev. Med.* **2008**, *59*, 455−471.
(12) Tsai, C. C.; Emau, P.; Jiang, Y.; Tian, B. P.; Morton, W. R.; Gustafson, K. R.; Boyd, M. R. *AIDS Res. Hum. Retroviruses* **2003**, *19*, 535−541.
(13) Tsai, C. C.; Emau, P.; Jiang, Y.; Agy, M. B.; Shattock, R. J.; Schmidt, A.; Morton, W. R.; Gustafson, K. R.; Boyd, M. R. *AIDS Res. Hum. Retroviruses* **2004**, *20*, 11−18.
(14) Li, M.; Patton, D. L.; Cosgrove-Sweeney, Y.; Ratner, D.; Rohan, L. C.; Cole, A. M.; Tarwater, P. M.; Gupta, P.; Ramratnam, B. *J. AIDS* **2011**, *58*, 379−384.
(15) Lagenaur, L. A.; Sanders-Beer, B. E.; Brichacek, B.; Pal, R.; Liu, X.; Liu, Y.; Yu, R.; Venzon, D.; Lee, P. P.; Hamer, D. H. *Mucosal Immunol.* **2011**, *4*, 648−657.
(16) Barrientos, L. G.; O'Keefe, B. R.; Bray, M.; Sanchez, A.; Gronenborn, A. M.; Boyd, M. R. *Antiviral Res.* **2003**, *58*, 47−56.
(17) Smee, D. F.; Bailey, K. W.; Hong, M.-H.; O'Keefe, B. R.; Gustafson, K. R.; Mishin, V. P.; Gubareva, L. V. *Antiviral Res.* **2008**, *80*, 266−271.
(18) Tiwari, V.; Shukla, S. Y.; Shukla, D. *Antiviral Res.* **2009**, *84*, 67−75.
(19) Nothaft, H.; Szymanski, C. M. *Nat. Rev. Microbiol.* **2010**, *8*, 765−778.
(20) Dall'Olio, F.; Malagolini, N.; Trinchera, M.; Chiricolo, M. *Front. Biosci.* **2012**, *17*, 670−699.
(21) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.
(22) Archontis, G.; Simonson, T.; Karplus, M. *J. Mol. Biol.* **2001**, *306*, 307−327.
(23) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238−250.
(24) Chocholoušová, J.; Feig, M. *J. Phys. Chem. B* **2006**, *110*, 17240−17251.
(25) Strockbine, B.; Rizzo, R. C. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 630−642.
(26) Fujimoto, Y. K.; TerBush, R. N.; Patsalo, V.; Green, D. F. *Protein Sci.* **2008**, *17*, 2008−2014.
(27) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87.
(28) Sarkar, C. A.; Lowenhaupt, K.; Horan, T.; Boone, T. C.; Tidor, B.; Lauffenbuger, D. A. *Nat. Biotechnol.* **2002**, *20*, 908−913.
(29) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364−1368.
(30) Looger, L. L.; Dwyer, M. A.; Smith, J. J.; Hellinga, H. W. *Nature* **2003**, *423*, 185−190.
(31) Ashworth, J.; Havranek, J. J.; Duarte, C. M.; Sussman, D.; Monnat, R. J., Jr.; Stoddard, B. L.; Baker, D. *Nature* **2006**, *441*, 656−659.
(32) Green, D. F.; Dennis, A. T.; Fam, P. S.; Tidor, B.; Jasanoff, A. *Biochemistry* **2006**, *45*, 12547−12559.
(33) Nesmelova, I. V.; Ermakova, E.; Daragan, V. A.; Pang, M.; Menéndez, M.; Lagartera, L.; Solís, D.; Baum, L. G.; Mayo, K. H. *J. Mol. Biol.* **2010**, *397*, 1209−1230.
(34) Kadirvelraj, R.; Grant, O. C.; Goldstein, I. J.; Winter, H. C.; Tateno, H.; Fadda, E.; Woods, R. J. *Glycobiology* **2011**, *21*, 973−984.
(35) Boyd, M. R.; Gustafson, K. R.; McMahon, J. B.; Shoemaker, R. H.; O'Keefe, B. R.; Gulakowski, R. J.; Wu, L.; Rivera, M. I.; Laurencot, C. M.; Currens, M. J.; Cardellina, J. H., II; Buckheit, R. W., Jr.; Nara, P. L.; Pannell, L. K.; Sowder, R. C., II; Henderson, L. E. *Antimicrob. Agents Chemother.* **1997**, *41*, 1521−1530.
(36) Bewley, C. A.; Gustafson, K. R.; Boyd, M. R.; Covell, D. G.; Bax, A.; Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1998**, *5*, 571−578.
(37) Bewley, C. A. *Structure* **2001**, *9*, 931−940.
(38) Bewley, C. A.; Otero-Quintero, S. *J. Am. Chem. Soc.* **2001**, *123*, 3892−3902.
(39) Bewley, C. A.; Kiyonaka, S.; Hamachi, I. *J. Mol. Biol.* **2002**, *322*, 881−889.
(40) Yang, F.; Bewley, C. A.; Louis, J. M.; Gustafson, K. R.; Boyd, M. R.; Gronenborn, A. M.; Clore, G. M.; Wlodawer, A. *J. Mol. Biol.* **1999**, *288*, 403−412.
(41) Botos, I.; O'Keefe, B. R.; Shenoy, S. R.; Cartner, L. K.; Ratner, D. M.; Seeberger, P. H.; Boyd, M. R.; Wlodawer, A. *J. Biol. Chem.* **2002**, *277*, 34336−34342.
(42) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D.; Nillson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; M., K. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(43) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(44) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(45) Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236−1243.
(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(47) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(48) Green, D. F.; Tidor, B. In *Current Protocols in Bioinformatics*; Petsko, G. E., Ed.; John Wiley & Sons, Inc.: New York, 2003, Chapter 8.3.
(49) Green, D. F.; Kangas, E.; Hendsch, Z. S.; Tidor, B. *ICE — Integrated Continuum Electrostatics*; MIT: Cambridge, MA, 2000.
(50) Altman, M. D.; Tidor, B. *MultigridPBE — Software for computation and display of electrostatic potentials*; MIT: Cambridge, MA, 2003.
(51) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.
(52) Green, D. F. *J. Phys. Chem. B* **2008**, *112*, 5238−5249.
(53) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1999**, *8*, 1381−1392.
(54) Green, D. F.; Tidor, B. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 644−657.
(55) Carrascal, N.; Green, D. F. *J. Phys. Chem. B* **2010**, *114*, 5096−5116.
(56) Hendsch, Z. S.; Jonsson, T.; Sauer, R. T.; Tidor, B. *Biochemistry* **1996**, *35*, 7621−7625.
(57) Barrientos, L. G.; Louis, J. M.; Botos, I.; Mori, T.; Han, Z.; O'Keefe, B. R.; Boyd, M. R.; Wlodawer, A.; Gronenborn, A. M. *Structure* **2002**, *10*, 673−686.
(58) Patsalo, V.; Raleigh, D. P.; Green, D. F. *Biochemistry* **2011**, *50*, 10698−10712.
(59) Fromme, R.; Katilene, Z.; Giomarelli, B.; Bogani, F.; McMahon, J.; Mori, T.; Fromme, P.; Ghirlanda, G. *Biochemistry* **2007**, *46*, 9199−9207.
(60) Matei, E.; Furey, W.; Gronenborn, A. M. *Structure* **2008**, *16*, 1183−1194.
(61) Barrientos, L. G.; Lasala, F.; Delgado, R.; Sanchez, A.; Gronenborn, A. M. *Structure* **2004**, *12*, 1799−1807.

(62) Koharudin, L. M. I.; Viscomi, A. R.; Jee, J.-G.; Ottonello, S.; Gronenborn, A. M. *Structure* **2008**, *16*, 570−584.
(63) Kelley, B. S.; Chang, L. C.; Bewley, C. A. *J. Am. Chem. Soc.* **2002**, *124*, 3210−3211.