Chapter 8

Calculating binding free energies for protein-carbohydrate complexes

Jodi A. Hadden[1], Matthew B. Tessier[1], Elisa Fadda[2], Robert J. Woods[1, 2]

[1]Complex Carbohydrate Research Center, 315 Riverbend Road, University of Georgia, Athens, GA 30602

[2]School of Chemistry, National University of Ireland, Galway, Ireland

1 Introduction

The ability to accurately predict the effects that changes in carbohydrate structure or protein sequence have on binding energy is fundamental to understanding biomolecular recognition (specificity and selectivity). Furthermore, such capability can be exploited in the design and optimization of carbohydrate-based therapeutics, or glycomimetics[1-2].

Considerable effort has been expended on the development of carbohydrate force fields that are capable of reproducing the conformational properties of carbohydrates in solution. Properties such as rotamer populations for glycosidic linkages depend on the relative energies of the rotational states, which in turn arise from a balance of intermolecular interactions between the carbohydrate and solvent molecules. While predicted rotamer populations for carbohydrates free in solution may be entirely force field dependent, the predicted conformations of carbohydrates in protein-bound complexes are relatively insensitive to the choice of force field. This is because when a carbohydrate binds to a protein, its internal flexibility is typically constrained by strong intermolecular interactions, leading to effectively a single bound conformation. That different force fields would predict similar structures is related to the fact that glycosidic linkages typically populate a limited number of conformations, both in solution and bound to proteins, which are also relative energy minima on the gas phase potential energy surface. Since current carbohydrate force fields are generally parameterized so as to reproduce relative quantum-mechanical gas phase rotational energies[3], they tend to predict similar minima.

In contrast, the interaction energies between a carbohydrate and its protein receptor are extremely sensitive to the force field, as these depend on the absolute contribution from Coulomb and van der Waals (vdWs) interactions – properties that are strongly affected by parameter choice in both the protein and carbohydrate. Additional complexity arises from estimating the energetic contributions from binding site and ligand desolvation, as well as entropic changes associated with binding. This latter component is particularly relevant for carbohydrates, which are highly flexible molecules. Thus, it is a far more daunting task to accurately predict the strength of a

carbohydrate-protein interaction than to predict either the conformational properties of the carbohydrate or its 3D structure bound to a protein.

1.1 Component-sum methods

The estimation of binding free energies necessitates the evaluation of contributions from each of the enthalpy-related and entropy-dependent (desolvation, configurational, and conformational entropy) components. Some of these terms may be directly derived by averaging the molecular mechanical (MM) contributions over the course of an MD simulation of the complex, such as values for vdWs and electrostatic energies. Desolvation free energy is typically calculated by applying an implicit solvent approximation, such as the Poisson-Boltzmann (PBSA) or generalized Born (GBSA) method. The resultant values from MM-PB/GBSA analyses may be further augmented by estimates of the entropic changes that occur upon binding. Summing these contributions for complex, receptor, and ligand allow calculation of absolute binding energies (**Figure 1**). Unfortunately, the approximations made in estimating each binding energy component may sometimes mean the total binding energy contains too much error to permit quantitative assessment. Such calculations are however relatively straightforward to perform, require only a simulation of the complex, and are not limited in terms of protein or ligand size. The MM-PB/GBSA approach, examined in the second section of this chapter, is often employed to examine the binding properties of a range of ligands with a single protein receptor.

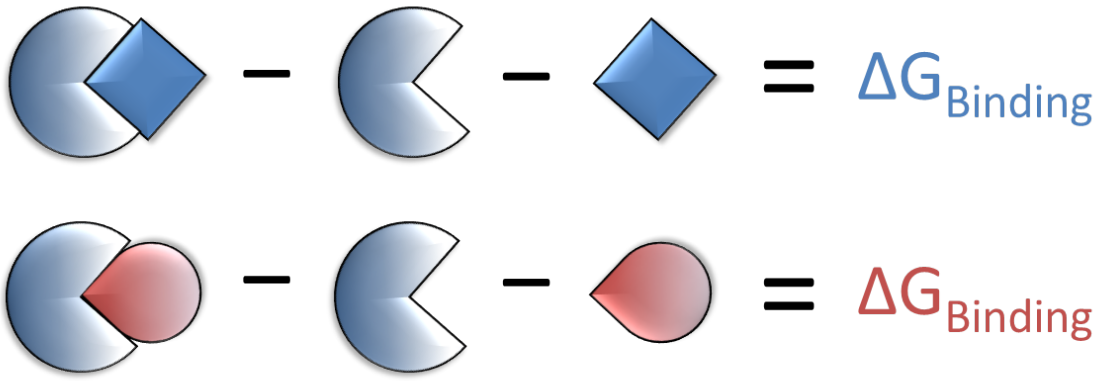

Figure 1. Direct ΔG approaches are analogous to the physical process of binding, with the binding energy computed as the sum of contributions from products minus the sum of contributions from reactants for the binding process.

1.2 Equilibrium methods

In many cases, inaccuracies in direct (or absolute) binding energy estimates may be decreased by performing the calculation such that the energy is derived for one state relative to that of a closely related state. For example, it is generally more accurate to compute the effect of a point mutation in the protein by performing a structural

perturbation during an MD simulation rather than attempting to determine the difference between two absolute binding free energies. While such chemical mutations are physically impossible, they may readily be performed *in silico* and they form the basis of computational approaches such as free energy perturbation (FEP) or thermodynamic integration (TI) (**Figure 2**). In contrast to MM-PB/GBSA approaches, FEP/TI methods directly account for energies arising from solvent or entropic effects. Moreover, the simulations may be performed under constant temperature and pressure conditions, giving rise to Gibbs free energies. However, the ability to maintain thermodynamic equilibrium is approximately inversely proportional to the size of the structural change (number of atoms) being made, as well as essential for computational convergence. Equilibrium perturbation approaches are therefore not well suited to screening libraries of diverse ligands for binding to a receptor, but are ideal for probing the effects of subtle structural changes upon affinity. The TI method is presented in the third section of this chapter.

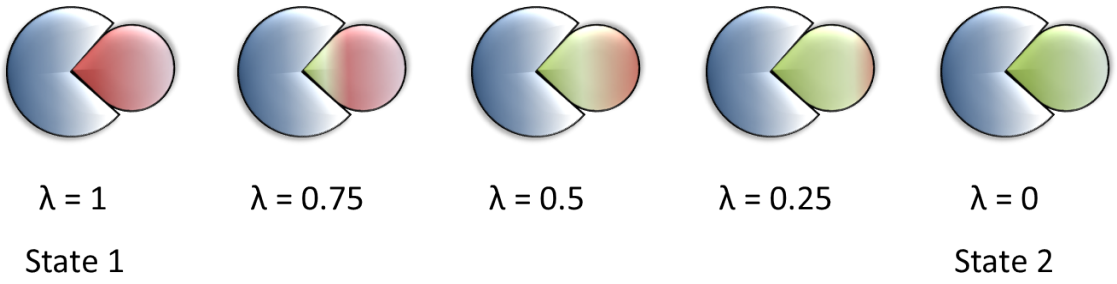

Figure 2. In TI simulations, a mixing parameter λ provides a nonphysical path to alchemically mutate between two closely related states over the course of an MD simulation, allowing computation of the free energy difference between them.

1.3 Non-equilibrium methods

Binding interactions in protein-carbohydrate complexes can also be probed through methods that pull or otherwise force the ligand out of the binding site. An example is steered molecular dynamics (SMD) simulations (**Figure 3**), which bear similarities to both the MM-PB/GBSA and FEP/TI methods. With regard to the former, SMD can be used to compute an absolute binding energy and is not limited in scope, while as in the latter, this approach is performed under conditions of constant temperature and pressure, and therefore does not require auxiliary estimates of solvation or entropic effects. A practical limitation of this method is that the pulling must be performed very gradually, greatly increasing the computational expense. This means SMD, discussed in the fourth section of this chapter, is better suited for studies where a small number of highly accurate binding energies are desired.

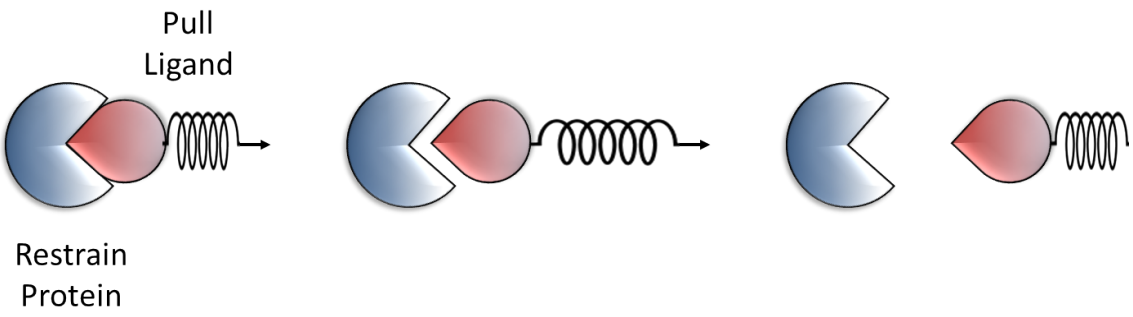

Figure 3. In SMD simulations for calculation of free energy, a spring attached to the ligand in a protein-ligand complex is pulled at constant velocity, pulling the ligand out of the binding site and into the surrounding solvent in order to elucidate the strength of the binding interaction.

1.4 Additional considerations

A further challenge in modeling protein-carbohydrate complexes is presented by the generally limited scope of model systems. *In vivo*, carbohydrates are typically complex polymers (mammalian glycans, polysaccharides), and yet most simulations tend to examine only the interaction between the carbohydrate residues proximal to the receptor protein. In addition, carbohydrates are typically covalently anchored to a protein (glycoprotein) or lipid (glycolipid) that is itself embedded in a membrane. While such considerations may be less critical when developing a small molecule inhibitor to block binding interactions, they are of fundamental importance when attempting to interpret binding data for an ensemble of endogenous glycans. Lastly, and perhaps the most daunting for carbohydrate modeling, biological affinity arises not only from monomeric interactions, but from avidity due to the dense display of ligands (glycoproteins and glycolipids) on cell surfaces, and the multimeric nature of many carbohydrate receptor proteins. Addressing these issues will require both innovative computational approaches, as well as more detailed biophysical characterization.

As in all computational studies, careful attention must be paid to system design and simulation protocols. In addition to reviewing the theory of each of the methods outlined above, this chapter will attempt to provide some practical guidance so that many of the typical pitfalls and omissions in their employment may be avoided.

1.5 Example system: Concanavalin A

Concanavalin A (ConA) is the best thermodynamically and structurally characterized legume lectin[4], and represents an archetype in the study of protein-carbohydrate interactions. ConA is expressed in the jack-bean and while it is known to bind various mannose and galactose containing glycans[5-6], it demonstrates a particularly high affinity for the trimannose Man-α-(1-6)-[Man-α-(1-3)]-Man (referred to as 3MAN)[7-8]. Microtitration calorimetry experiments[8] show that the binding affinity for the ConA-

3MAN complex is on the order of -7.5 kcal/mol, one of the highest binding energy values measured for a lectin-carbohydrate complex[4].

The wide availability and ease of purification of ConA has made it an ideal system for structure determination and binding characterization. Access to several high-resolution crystallographic complexes, the first of which was solved in 1996,[7] has also made it one of the most studied systems for predicting binding free energies in protein-carbohydrate complexes. Over the past decade, a variety of theoretical free energy methods have been applied to the ConA system, including MM-GBSA[4] and TI[9-10]. The more recent studies have focused on elucidating the role of conserved waters in the binding of ConA to 3MAN as well as a synthetic trimannoside mimic (3HET)[9-10]. Throughout this chapter, the crystal structures of ConA bound to 3MAN (PDB ID 1CVN)[7] and 3HET (PDB ID 3D4K)[9] will serve as examples for application of free energy of binding methods to protein-carbohydrate complexes.

2 MM-PB/GBSA

Component-sum methods offer a strategy for calculating free energies of binding in a precise, detailed, and high-throughput manner. They do this by employing computationally inexpensive techniques to estimate the interaction contributions that make up free energy. A key aspect is the simplification of the solvated protein-ligand complex potential energy surface by substitution of explicit solvent with a continuum dielectric model (implicit solvent). The gold standard of these implicit solvent models, the Poisson-Boltzmann (PB) method, is based on the Poisson continuum dielectric model combined with a Boltzmann distribution, which accounts for the presence of ions and other inhomogeneities in the dielectric field. Another popular implicit solvent model is the Generalized Born (GB) method, which is developed as an approximation to the PB method. GB models are designed to be computationally more efficient than PB while making relatively small compromises in precision.

Several recent studies have employed PB[11-12] and GB[4, 13-17] implicit solvents to calculate binding free energies of protein-carbohydrate complexes, with the most common applications in the study of lectin binding[4, 13] and binding of disease related proteins[14-15, 17].

The theory and application of these methods will be discussed with respect to the calculation of absolute (ΔG) and relative ($\Delta\Delta G$) binding free energies of protein-carbohydrate complexes.

2.1 General theory

Implicit solvent models serve to reduce the computational expense associated with explicit solvent simulations by substituting the large numbers of solvent atoms (which can account for more than half the atoms in the system) with a continuum function of high dielectric field, which represents the solvent. The solute exists inside this solvent field as a cavity of comparatively low dielectric field (**Figure 4**). The computational

efficiency of these solvent models has made them popular choices for performing rapid molecular dynamics simulations and high-throughput binding energy calculations.

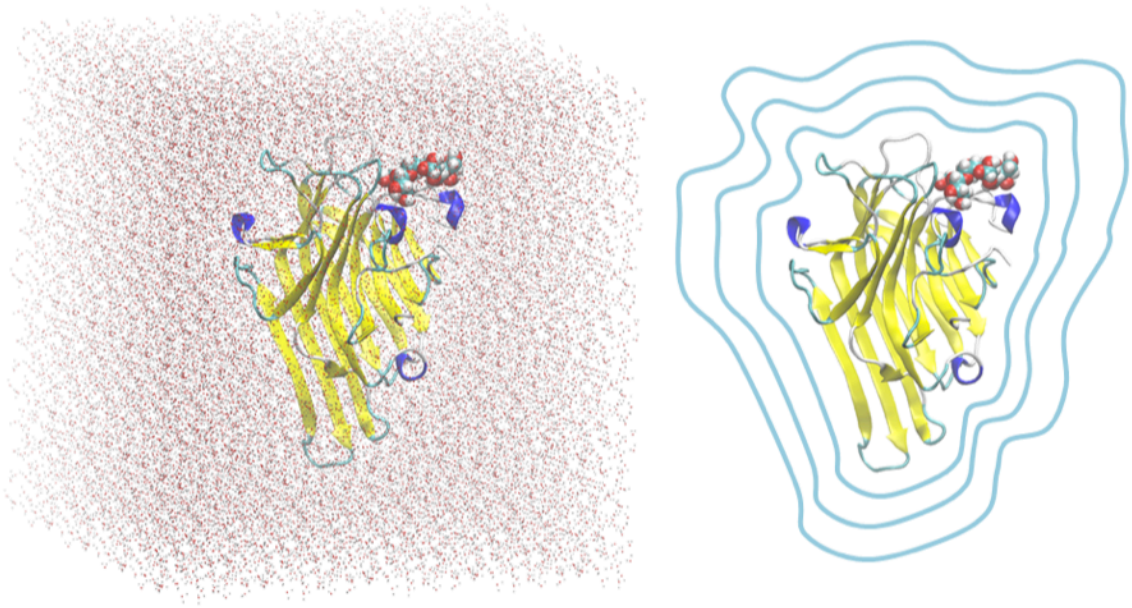

Figure 4. ConA complexed with trimannose (3MAN, PDB ID 1CVN) shown in explicit solvent (left) and an implicit solvent (dielectric field) representation (right).

The potential energy of a system (V_{System}) in implicit solvent is computed as a sum of the enthalpic contribution from the molecular mechanics (MM) force field (H_{MM}), and polar (G_{Polar}) and non-polar ($G_{Non\text{-}polar}$) solvation components from the implicit solvent model:

$$V_{System} = H_{MM} + (G_{Polar} + G_{Non-polar}) \quad (1)$$

The polar solvation component represents the solvent dielectric model, the PB or GB function, discussed later in the section. The non-polar solvation component approximates the cavity formed by the solute surrounded by the solvent dielectric. This term is usually defined as a function of the solvent accessible surface area (SASA) and constitutes the SA model:

$$G_{Non-polar} = \gamma \cdot SASA + c \quad (2)$$

Here, the non-polar contribution is represented by a surface tension coefficient γ with surface area dependence. Additional SA models exist where **Equation 2** is the repulsive component to the non-polar solvation energy and an attractive component is modeled separately[18]. Calculations made using component-sum contributions from molecular mechanics and these implicit polar/non-polar solvation functions are frequently abbreviated as MM-PBSA or MM-GBSA methods.

Calculating free energy of binding for protein-ligand complexes depends on completing a thermodynamic cycle (**Figure 5**). Ideally, the free energy of binding (ΔG_{Total}) would be

computed as the difference between the energetic contribution of the complex and those of the free protein and free ligand:

$$\Delta G_{Total} = G_{Complex} - \left(G_{Protein} + G_{Ligand}\right) \quad (3)$$

Binding energy in solution ($\Delta G_{Binding(soln)}$) is determined by combining the computed energy changes in each step of the thermodynamic cycle. Specifically, these are the energy changes associated with the process of binding in the gas phase ($\Delta G_{Binding(gas)}$), solvating the complex ($\Delta G_{Solv(Complex)}$), and independently solvating the protein and ligand ($\Delta G_{Solv(Protein)}$, $\Delta G_{Solv(Ligand)}$). Combining the MM-PB/GBSA enthalpy contribution with entropy ($T\Delta S_{Conf/Vib}$) gives the full binding energy equation:

$$\Delta G_{Binding(soln)} = \Delta G_{Binding(gas)} + \Delta G_{Solv(Complex)} - \left(\Delta G_{Solv(Protein)} + \Delta G_{Solv(Ligand)}\right) - T\Delta S_{Vib/Conf} \quad (4)$$

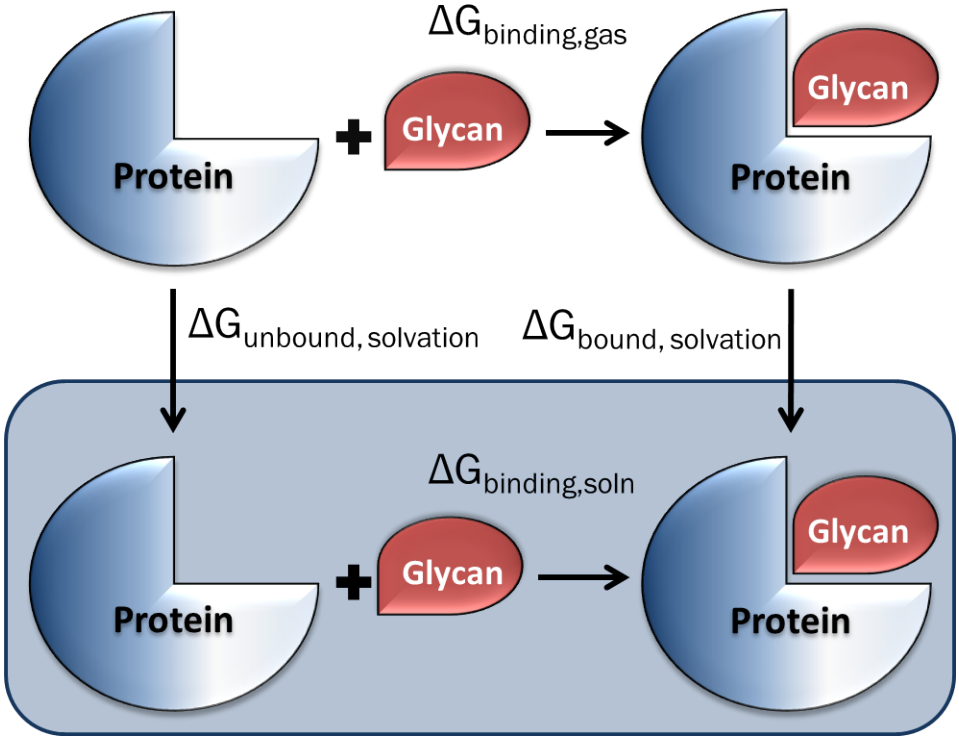

Figure 5. Thermodynamic cycle for calculating free energy of binding for protein-carbohydrate complexes in solution.

In order to apply MM-PB/GBSA methods, an ensemble of protein-ligand complex structures is typically generated from a MD simulation. To facilitate rapid calculation of binding energy from a single trajectory, frames of the complex are partitioned to also generate frames of the free component molecules, working on the assumption that the bound conformations of the protein and ligand are representative of their free conformations. The use of an implicit solvent model further simplifies the calculation by

serving as a solvent approximation for these free protein and ligand structures, as considerable additional simulation time would be required to generate converged thermodynamic data using explicit solvation. In some software implementations, this method can be performed by alternatively using separate trajectories for the complex, protein, and ligand, though data obtained through this approach exhibits inherent noise.

While the use of implicit solvent models reduces overall computational expense of MM-PB/GBSA calculations, their use in place of explicit solvent has an intrinsic downside. Although implicit solvents suitably reproduce the behavior of bulk water, not all water molecules behave as bulk water. Some water molecules play roles as part of the protein or ligand[19], or mediate hydrogen-bond networks between them, which several studies have shown can be especially important in protein-carbohydrate complexes[20-21]. Free energy calculations employing implicit solvent will therefore omit these additional stabilizing effects and underestimate the net binding contribution of these solvent molecules. Considerations should be made to include tightly bound waters as part of the protein-ligand complex, similar to co-factors or coordinate ions.

2.2 Poisson-Boltzmann

The Poisson-Boltzmann implicit solvent model is the standard by which all other implicit solvent models are compared. It is a differential equation based on the Poisson method for calculating the potential of a solute with charge density $p^f(r)$:

$$\nabla\left[\varepsilon(r)\nabla\psi(r)\right] = -p^f(r) \tag{5}$$

Here, ∇ is the divergence operator, $\varepsilon(r)$ is the dielectric at position r, and $\nabla\psi(r)$ is the gradient of the electrostatic potential $\psi(r)$.

When the solute has a non-zero charge density, then a Boltzmann distribution term is included to account for the presence of ions in the solvent:

$$\nabla\left[\varepsilon(r)\nabla\psi(r)\right] = -p^f(r) - \sum n_i^0 q_i e^{\frac{q_i\psi(r)}{kT}} \tag{6}$$

Here, ion i has charge of q_i and density in bulk solution of n_i^0, and kT denotes the Boltzmann constant at temperature T. The functional form of **Equation 6** is linearized based on the assumption of a small solute dielectric relative to the larger solvent dielectric. As the PB equation cannot be logically solved without such linearization, this assumption serves to further improve the computational efficiency of the method. Folgolari, et al.[22] and Wang, et al.[19] provide a more complete history, as well as functional derivations of the Poisson-Boltzmann equation.

2.3 Generalized Born

The Generalized-Born implicit solvent is a continuum solvent model that functions as an approximation to the Poisson-Boltzmann implicit solvent. It is based on the Born

equation[23], which describes the transfer of ion *i* with charge q_i and radius α_i from a medium of dielectric ε_i to another of ε_0:

$$\Delta G_{Born} = -\frac{1}{2}\left(\frac{1}{\varepsilon_i} - \frac{1}{\varepsilon_0}\right)\frac{q_i^2}{\alpha_i} \tag{7}$$

The *generalized* Born form of the equation assumes that a molecule forms a cavity, which is defined by a set of spherical radii, α_i for each atom in the molecule. These radii, called effective Born radii, are analytically solved values that represent the degree to which an atom is buried within the solute cavity. In principle, accurate description of these radii should give reasonable agreement with PB solvation free energies[24]. The general form of the GB model as given by Still, et al.[25] is:

$$\Delta G_{Born} = \left[-\frac{1}{2}\left(\frac{1}{\varepsilon_i} - \frac{1}{\varepsilon_0}\right)\sum_i^n \frac{q_i^2}{\alpha_i}\right]_{self} + \left[-\frac{1}{2}\left(\frac{1}{\varepsilon_i} - \frac{1}{\varepsilon_0}\right)\sum_{i=1}^n \sum_{j=1, j\neq i}^n \frac{q_i\, q_j}{\sqrt{r_{ij}^2 + \left(\sqrt{\alpha_i\, \alpha_j}\right)^2 e^{\frac{-r_{ij}^2}{2\left(\sqrt{\alpha_i\, \alpha_j}\right)^2}}}}\right]_{pairwise} \tag{8}$$

In this case, the *self* term is the familiar Born equation for the transfer of ion *i* between two dielectric fields. The *pairwise* term describes the interaction between ions *i* and *j* separated by distance r_{ij}. The same effective Born radius is used for both. Extensive effort has been expended to calculate these radii for different types of molecular systems[25-29], and to generate suitable radii for universal application[24], but none have been designed specifically for carbohydrates.

GB implicit solvents are ideal for MD simulations because their functional forms are less computationally expensive than that of PB. However, when calculating free energies from existing trajectories, the accuracy of PB may be preferred. Zhang, et al.[30] and Feig, et al.[31] provide informative comparisons of the two methods.

2.4 Entropy

The entropic penalty incurred upon ligand binding is an important factor in the binding free energy. In component-sum methods, entropy is often decomposed into two components, vibrational and conformational, as per:

$$\Delta S_{Total} = \Delta S_{Vibrational} + \Delta S_{Conformational} \tag{9}$$

The vibrational contribution results physically from changes in bond stretching/compressing and angle bending that occur in the protein and ligand upon their binding. In classical mechanics, this contribution is generally derived via normal

modes analysis[32], and is considerably more expensive to compute than the enthalpic contribution in component-sum free energy methods.

Conformational entropy stems from changes in the global configuration of the ligand upon binding. This is particularly relevant for carbohydrates, which are highly flexible in terms of their glycosidic linkages and exocyclic free rotamers. Binding to a receptor protein often causes regions of a carbohydrate to become rigid, restricting its available torsion space, and thus incurring an entropic penalty. This type of conformational entropy has been estimated for glycosidic linkages by comparing torsion rotation profiles of bound and free carbohydrates[4, 14]. The Karplus-Kushick[33] approach allows calculation of conformational relative entropy from the covariance matrices of the glycosidic angles in the bound (σ_B) and free (σ_F) states:

$$\Delta S_{Conformational} = \frac{1}{2} k \ln\left(\frac{\sigma_B}{\sigma_F}\right) \tag{10}$$

Consideration of conformational entropy is especially important if performing free energy calculations for the development of glycomimetics. The removal of torsional flexibility reduces the entropic penalty of binding, and may be exploited in rational drug design.[34]

2.4 Application to ConA

Bryce, et al. previously calculated free energies of binding for the ConA-3MAN complex using the MM-GBSA method on a 1 ns MD trajectory[4] (**Table 1**). Here, this study is extended to include data for both 3MAN and 3HET, employing longer MD timescales.

All simulations were performed using the *pmemd* module of the AMBER 11[35] software package. For consistency with earlier studies[9], the Mn^{2+} ions in the 3MAN and 3HET complexes were changed to Ca^{2+}. All crystallographic waters were retained, and hydrogens were added using the *tleap* module of AMBER 11. These hydrogens were relaxed in a brief minimization (1000 steps steepest descent, followed by 1000 steps conjugate gradient) using GB implicit solvent[29]. The AMBER ff99SB[36] and GLYCAM-06[37] force fields were used for the receptor protein and carbohydrate ligands respectively, employing 1-4 nonbonded scaling values appropriate to each. The relaxed complex was neutralized with Na^+ counterions and solvated with explicit TIP3P[38] water in a box of 10 Å buffer from solute to box edge. This solvated complex was minimized (5000 steps steepest descent, followed by 5000 steps conjugate gradient), heated from 5 K to 300 K over 50 ps, and then submitted to 30 ns of MD at 300 K. Simulations were performed in the NPT ensemble at 1atm, employing a Berendsen-type[39] thermostat and barostat, with temperature and pressure coupling constants of 10 ps and 0.1 ps respectively. Beyond a 10-Å non-bonded cutoff, electrostatics were treated with PME[40] and vdW interactions were truncated. High-frequency motions involving hydrogen atoms were restrained using the SHAKE[41] algorithm. The *ptraj* module of AMBER 11 was used to process the trajectories and extract frames from the converged portions of each simulation (the last 29 ns for 3MAN, and the last 10 ns for 3HET). For MM-PB/GBSA,

1000 evenly spaced frames were extracted. For normal modes analysis, 100 evenly spaced frames were extracted. Binding energy calculations were performed using the *MMPBSA.py* module of AMBER 11, employing GB radii suggested by Onufriev, et al[29].

Data from this study (presented in **Table 1**) shows the GBSA (ΔG_{GB}) energy to be consistent with previous results, but the MM (ΔH_{MM}) and entropic (ΔS_{Vib}) energies are notably divergent. This may be in part due to changes in the force field parameters between GLYCAM-06[37] and GLYCAM-93[4, 42], but is also likely reflective of the far longer timescale of the present simulations. For consistency with Bryce, et al., only vibrational entropy contributions were calculated, though inclusion of conformational entropy would provide an improved estimate to the entropic penalty of binding.

It is notable that the analysis predicts 3HET to be a weaker ligand than 3MAN, by between 1-3 kcal/mol, which compares favorably with the experimental value of 1.2 kcal/mol. The fact that the absolute free energies of binding are lower than experiment, by approximately a factor of 2 (6-8 kcal/mol) can be attributed to several issues. It must be recalled that these are classical force field energies, and do not take into account the effect of polarization of the ligand or protein, which may be expected to occur upon ligand binding. Polarization can account for up to approximately 1/3 of the total electrostatic interaction[43]. To compensate in part for this weakness, classical force fields frequently employ slightly over-polarized partial charges[44], which if not optimal may be expected to lead to an over-estimation of the strengths of the electrostatic interaction energies.

Further, the strength of binding of the carbohydrate to the protein may depend on the presence or absence of explicit lone-pairs on the hydroxyl groups in the carbohydrate, or on appropriate atoms in the protein[45]. By analogy, the interaction energies computed for conserved water molecules differ when using the TIP3P (no lone pairs) or TIP5P[46] (includes lone pairs) water models. The omission of polarization and or lone pairs may also impact the quality of the structures during the MD, which in turn would impact the computed interaction energies.

The omission of conformational entropy from the present analysis likely also impacts both the absolute and the relative interaction energies. However, a simple maximum entropy estimate, based on the structural differences between 3MAN and 3HET results in an estimate of only approximately 1kcal/mol penalty in the binding of 3HET relative to 3MAN. See the ***Application to ConA*** subsection in the TI portion of this chapter for further comments on the relative binding energies calculated here with MM-PB/GBSA.

Table 1: Calculated absolute binding free energies[a] for ConA complexes

Component	**3HET[b]** 30 ns	**3MAN[b]** 30 ns	**3MAN**[4] 1ns
ΔH_{MM}	-93.6 (0.2)[c]	-106.8 (0.3)	-118.2 (8.6)
ΔG_{GB}	56.8 (0.1)	68.1 (0.2)	69.2 (7.3)
ΔG_{PB}	57.6 (0.2)	67.0 (0.2)	---[d]

$-T\Delta S_{Vib}$	23.5 (0.4)	24.3 (0.4)	16.4 (-)
$\Delta G_{Binding(GB)}$	-13.3 (0.4)	-14.4 (0.4)	-32.6 (6.2)
$\Delta G_{Binding(PB)}$	-12.5 (0.4)	-15.5 (0.4)	---[d]
Experimental[20]	-6.4 (<0.1)	-7.6 (<0.1)	-7.6 (<0.1)

[a]kcal/mol. [b]This work. [c]Values in parentheses are errors in the standard deviation of the mean (standard errors). [d]Data not reported.

2.5 MM-PB/GBSA protocol

Initial Setup for Free Energy Simulations

1. The choice of protein-carbohydrate complex should be made such that available computational resources are capable of simulating reasonable timescales and that parameters exist for all molecules relevant to the binding interaction, including any unique glycomimetic drug structures, co-factors, or coordinated ions. Note that modeled coordinate ions may require special restraints in order to maintain the appropriate binding site topography.
2. Gather or generate structures of the protein-carbohydrate complexes for study. Glycosciences.de offers a search tool (http://www.glycosciences.de/database) for identifying protein-carbohydrate complexes available as biomolecular crystal structures through the RSCB Protein Data Bank (http://www.rcsb.org/pdb). Complex structures can also be generated through molecular docking, discussed elsewhere in this book.
3. Prepare simulation files for the complex, including coordinates for any co-factors, coordinated ions, or conserved waters that appear important for the binding interaction. As x-ray crystal structures do not contain hydrogen atoms, these are generally added at this step. The orientations of polar hydrogen atoms are important for maintaining hydrogen bonds from the crystal structure starting geometry, which may be critical to binding, so care should be taken with respect to their placement. Programs like *reduce*[47] can be used to predict hydrogen bond orientations for heavy atom coordinates in PDB structures, or hydrogen positions can be energy minimized with respect to their adjacent heavy atoms to correct their placement. Optimize these polar hydrogen positions before explicitly solvating (see below).
4. Energy minimize the complex to remove any close contacts in the system, and optionally optimize orientations of polar hydrogen atoms.
5. Prepare explicitly solvated simulation files using the coordinates obtained from minimization.
6. Energy minimize the solvated complex to remove any close contacts introduced by explicit solvent.
7. Perform an MD simulation to equilibrate the solvated complex, producing a final structure that represents the converged binding interaction. Note that if the protein-carbohydrate complex was generated through molecular docking or grafting, considerably more simulation time may be required to reach

convergence relative to a complex based on experimentally derived coordinates. The output of this step will provide initial coordinates for free energy simulations.

Calculating Free Energy of Binding With MM-PB/GBSA

8. Using the equilibrated endpoint structure generated in the previous step, perform a production MD simulation of the solvated protein-carbohydrate complex to create an ensemble of representative structures for use in MM-PB/GBSA calculations. These structures must characteristically describe the binding interaction in terms of topography and free energy, which means both aspects must be converged (see below).
9. Evaluate the MD production data to confirm convergence of binding interaction structure and topography. The system should have been simulated sufficiently long to allow the carbohydrate to sample all the conformational space available to it in the protein binding site. If this criterion is not met, the MD simulation should be extended.
    - The glycosidic linkage torsion angles (φ, Ψ, and ω(s)) monitored over the course of the trajectory should show reasonable sampling of states, especially if multiple states are observed.
    - A relative root mean squared deviation (RMSD) of the bound carbohydrate over the course of the trajectory (unfit, based on the ligand heavy atoms, with the trajectory aligned to the protein backbone) should show positional equilibration in the binding site.
10. Evaluate the MD production data to confirm convergence of free energy. A suggested method is to divide the trajectory into a set of twenty of more subsets and use MM-PB/GBSA (see below) to calculate the average free energy for each subset. These representative averages can be plotted versus simulation time to assess convergence. Any portion of the trajectory that has not converged should be thrown out for final MM-PB/GBSA calculations and the MD simulation extended if necessary. See **Figure 6** for an example.
11. Process the MD trajectory prior to performing MM-PB/GBSA calculations to remove any molecules that are not important for the complex binding interaction (i.e. explicit solvent and counterions, but not coordinated ions or waters essential to binding). The trajectory should also be imaged if needed, and any corrupt frames should be removed. Depending on your software implementation, this may also be the step to extract representative structures for the MM-PB/GBSA calculation (see below). Regardless, generating a new trajectory file that contains only these structures may increase the computational efficiency of the MM-PB/GBSA calculation.
12. Generate separate parameter/topology files for each of the free (unsolvated) complex, receptor protein, and ligand. These files are generally required by MM-PB/GBSA calculations and should contain the appropriate PB/GB radii[25-29] for these molecules.
13. Perform MM-PB/GBSA calculations using an ensemble of representative structures extracted from your trajectory. These structures must come from the

converged portion of the trajectory and are typically extracted at an evenly spaced interval. A suitable number of structures necessary to provide a reasonable statistical average should be used (1000 at minimum is suggested). Some software implementations allow decomposition of free energy contributions on a per-residue basis, which may provide valuable information regarding important residues in the binding site.

14. Calculate the vibrational entropy contribution to free energy using normal modes calculations. These are very computationally expensive compared to MM-PB/GBSA, so a smaller subset from the ensemble of representative structures is generally employed (100 at minimum is suggested). For very large protein-carbohydrate complex systems, it may be necessary to truncate non-essential regions of the protein to reduce computational intensity, however this must be done carefully in order to retain residues important to binding and binding site definition.
15. Calculate the conformational entropy contribution to free energy, which results from the loss of flexibility for exocyclic torsions and glycosidic linkages upon binding of a carbohydrate. Quantification of this entropic penalty requires additional converged MD data of the carbohydrate free in solution in order to compare torsion profiles of the bound and free states.[33] This method was applied by Kadirvelraj, et al. in studies of antibody binding to the GBSIII bacterial polysaccharide[14].
16. Calculate free energy of binding by summing enthalpic components and subtracting entropic components, $\Delta G_{Complex-(Protein+Ligand)} - T\Delta S_{Vib/Conf}$, to obtain $\Delta G_{Binding}$. Relative free energies of binding can be obtained by taking the difference in binding energy between two similar states of the complex, $\Delta\Delta G_{AB} = \Delta G_A - \Delta G_B$.

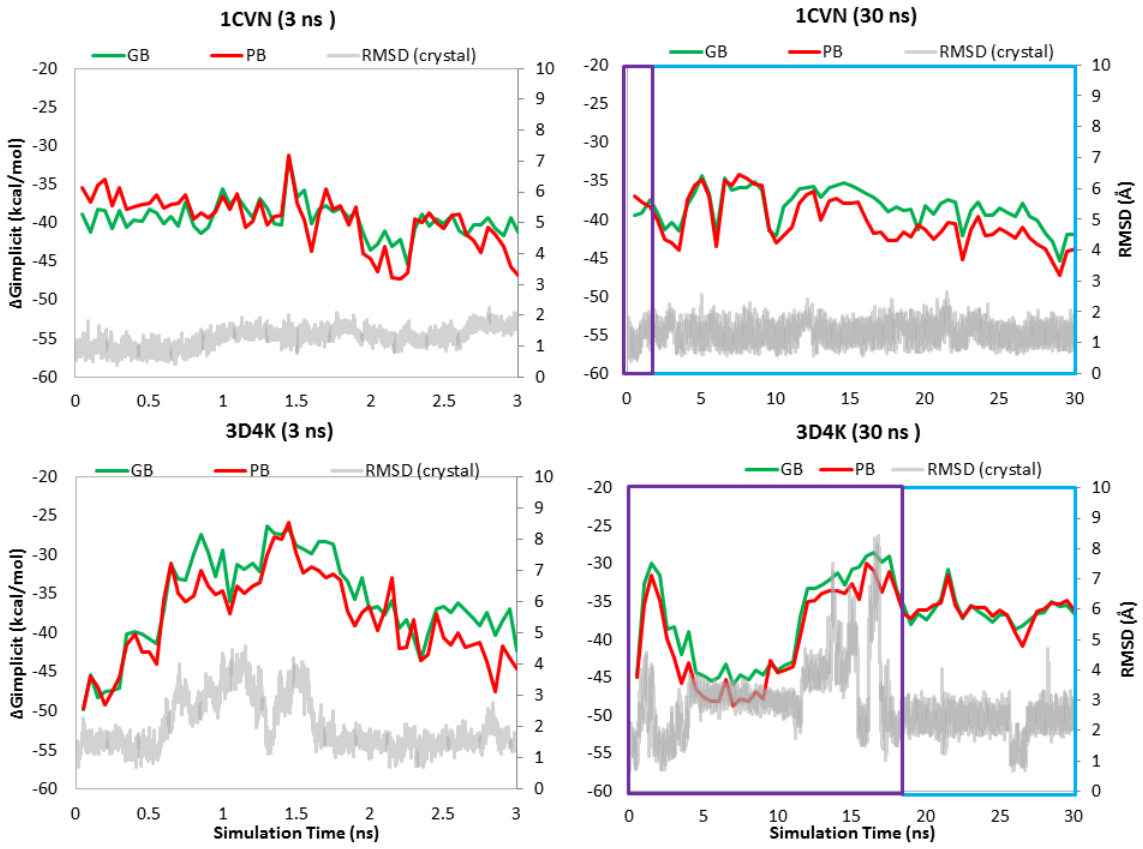

Figure 6. An example of convergence assessment for binding free energies calculated with MM-PBSA (red) and MM-GBSA (green). The ligand RMSD (not including hydrogen atoms, gray) relative to the first frame of the trajectory is included to test for structural convergence. Simulation of ConA with trimannose, 3MAN (PDB ID 1CVN), reaches structural convergence within 1 ns (top left), but shows some instability in free energy calculation over the 2 to 30 ns timescale (top right, blue). Meanwhile, simulation of ConA with the trimannose mimic, 3HET (PDB ID 3D4K), does not reach structural or energetic convergence within 3 ns (top left), but later appears to converge between 20 and 30 ns (top right, blue).

3 Thermodynamic integration (TI)

Thermodynamic Integration (TI) was originally proposed by Kirkwood in 1935[48]. Unlike MM-PB/GBSA, which computes absolute binding energy ($\Delta G_{Binding}$), TI computes relative binding energy, or the difference in binding energy between two states ($\Delta\Delta G_{Binding}$). In terms of protein-carbohydrate complexes, these two states can be either:

1) The complex and the free protein, in which the ligand is annihilated over the course of the TI simulation.
2) Two versions of the complex, in which there are relatively small differences in the protein structures.

3) Two versions of the complex, in which there are relatively small differences in the ligands

The first case gives rise to an absolute ligand binding energy, while the latter two cases can be useful for studying how mutations in a protein affect ligand binding, or how well different ligands bind to the same protein. Though much of the same information can be obtained through MM-PB/GBSA at a lesser computational expense, the TI approach offers the advantage of directly incorporating desolvation and entropic effects. TI has found application in the study of protein-binding interactions in complexes with carbohydrates[9-10, 49-53] as well as glycomimetic drugs such as inhibitors of influenza neuraminidase[54-55].

3.1 General theory

TI is commonly referred to as computational alchemy. While MM-PB/GBSA is a post processing method that utilizes a MD trajectory, data for TI calculations are collected numerically over the course of a MD simulation in which the initial state (state A) is alchemically mutated to the final state (state B). This mutation is accomplished through incorporation of a nonphysical mixing parameter λ, which is used to couple the two states and interpolate between them by mediating their contributions to the potential *V* of the mixed (mutating) system, given by:

$$V(\lambda) = (1-\lambda)V_A + \lambda V_B \tag{11}$$

Values of λ range from λ=0 where the system is wholly state A with no coupling to B, to λ=1, where the system is wholly state B with no coupling to A. At each step in the MD, the potential is calculated as *what it would have been* for both state A and state B. These two potentials are combined via λ to generate the mixed potential, which is then applied to propagate the motion of the mixed system. It is therefore the mixed system, which lies somewhere between states A and B as per the value of λ, that is effectively simulated, evolving according to its mixed potential and propagating as a single set of coordinates.

Generally, simulations are performed at a number of discrete λ windows between λ=0 and λ=1, during which ensemble averaged values of $\langle \partial V(\lambda)/\partial\lambda \rangle$ are collected. The relative free energy change between states A and B can then be obtained via integration over the resulting function as λ 0→1, according to:

$$\Delta\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{12}$$

In order to obtain a relative *binding* free energy, a full thermodynamic cycle must be completed. This means the mutation must be performed in both the bound and free systems, then the relative binding energy is given by the difference of these, as per:

$$\Delta\Delta G_{Binding} = \Delta\Delta G_{Bound} - \Delta\Delta G_{Free} \tag{13}$$

As TI is an equilibrium method, the value of ΔΔG can be calculated in either the forward or reverse directions, using state A or B as the initial state, as indicated by the relationship:

$$\Delta\Delta G_{Forward} = -\Delta\Delta G_{Reverse} \tag{14}$$

3.2 Soft-core potentials

For a system where state A has *N* atoms and state B has *N+1* atoms, the mutation corresponds to appearing the additional atom as λ 0→1. At λ=0 the atom is fully decoupled from the overall system and should have no effect on it, while at λ=1 the atom is fully coupled and should interact normally. The standard Lennard-Jones potential between two atoms *i* and *j* separated by distance r_{ij} is given by:

$$V(\lambda)_{ij}^{LJ} = \lambda 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{15}$$

However, when dealing with decoupled atoms, this functional form has the limitation that as λ→0 a singularity can occur. This is known as the origin singularity or vdW endpoint problem[56]. As the atom becomes decoupled, overlap with coupled atoms becomes possible, allowing r_{ij}→0 where the Lennard-Jones potential becomes undefined. Concomitantly, as r_{ij}→0 the Lennard-Jones interactions become increasingly repulsive, causing instabilities in the potential calculation. This affects the numerical integration of the MD and furthermore the value of $\langle \partial V(\lambda)/\partial\lambda \rangle$ when performing TI.

Various methods have been developed to address these issues, including analytical fitting schemes[57-58], slow growth methods[59], and nonlinear mixing functions for $V(\lambda)$[56, 60-65]. Nevertheless, most modern TI implementations employ soft-core potentials[62, 66] to allow overlap of decoupled particles while avoiding endpoint singularities. The functional form of the soft-core Lennard-Jones potential as given by Beutler, et al.[62] is:

$$V(\lambda)_{ij}^{LJ} = \lambda 4\varepsilon_{ij}\left[\frac{1}{\alpha_{LJ}(1-\lambda)^4 + \left(\frac{r_{ij}}{\sigma_{ij}}\right)^{12}} - \frac{1}{\alpha_{LJ}(1-\lambda)^2 + \left(\frac{r_{ij}}{\sigma_{ij}}\right)^{6}}\right] \tag{16}$$

The supplement of α_{LJ}, a positive constant, prevents the denominator from approaching zero and becoming undefined as λ→0, r_{ij}→0. Furthermore, due to the factor of (1–λ), the contribution of α_{LJ} will only be invoked as λ→0, when decoupling of the atom necessitates soft-core behavior **Figure 7**.

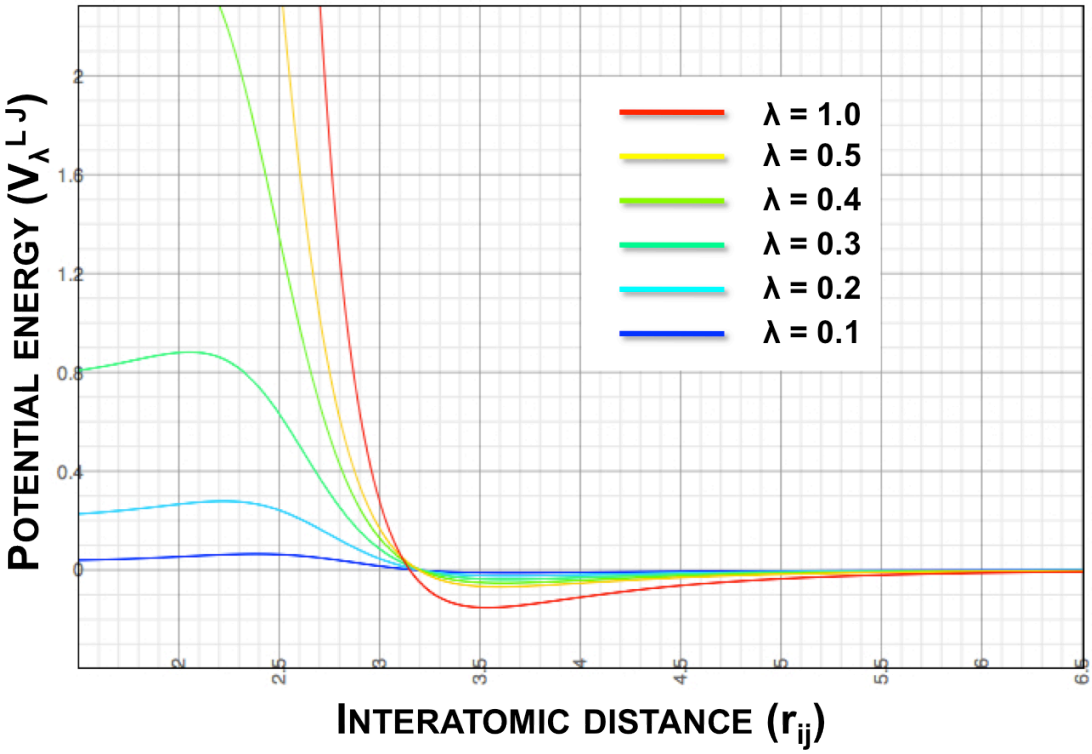

Figure 7. This example shows plots for the Lennard-Jones interaction between two oxygen atoms of the TIP3P water model[38]. The functions were evaluated at a range of λ values according to the Beutler, et al. equation with the suggested α_{LJ} parameter of 0.5[62]. As λ→0, the potential begins to exhibit soft-core behavior and remains defined despite the overlap of partially decoupled atoms with others in the system.

Coulomb interactions have the standard form:

$$V(\lambda)_{ij}^{C} = \lambda\left[\frac{1}{4\pi\varepsilon_0}\frac{q_i\, q_j}{r_{ij}^2}\right] \tag{17}$$

These can also be modeled using soft-core potentials, as in the form given by Beutler et al.[62] as:

$$V(\lambda)_{ij}^{C} = \lambda\left[\frac{1}{4\pi\varepsilon_0}\frac{q_i\, q_j}{\left(\alpha_C(1-\lambda)^2 + r_{ij}^2\right)^{\frac{1}{2}}}\right] \tag{18}$$

Again, the supplement of α_C prevents endpoint singularity and only applies as λ→0.

Note If the atom is being *disappeared* as λ 0→1 and is instead fully decoupled at λ=1, the value of λ must be replaced with (1–λ) and vice versa for the above equations.

When scaling nonbonded properties, care must always be taken that an atomic charge never becomes unshielded by its corresponding vdW radius, as this would lead to a catastrophic collapse in the system. For some TI implementations, this requires dividing the mutation of Lennard-Jones and Coulomb interactions into separate simulation steps **Figure 8**, which increase the computational expense. However, if soft-core potentials are implemented for both Lennard-Jones and Coulomb interactions, mutations can be performed in a single step given suitable optimization of α_{LJ} and α_C[67].

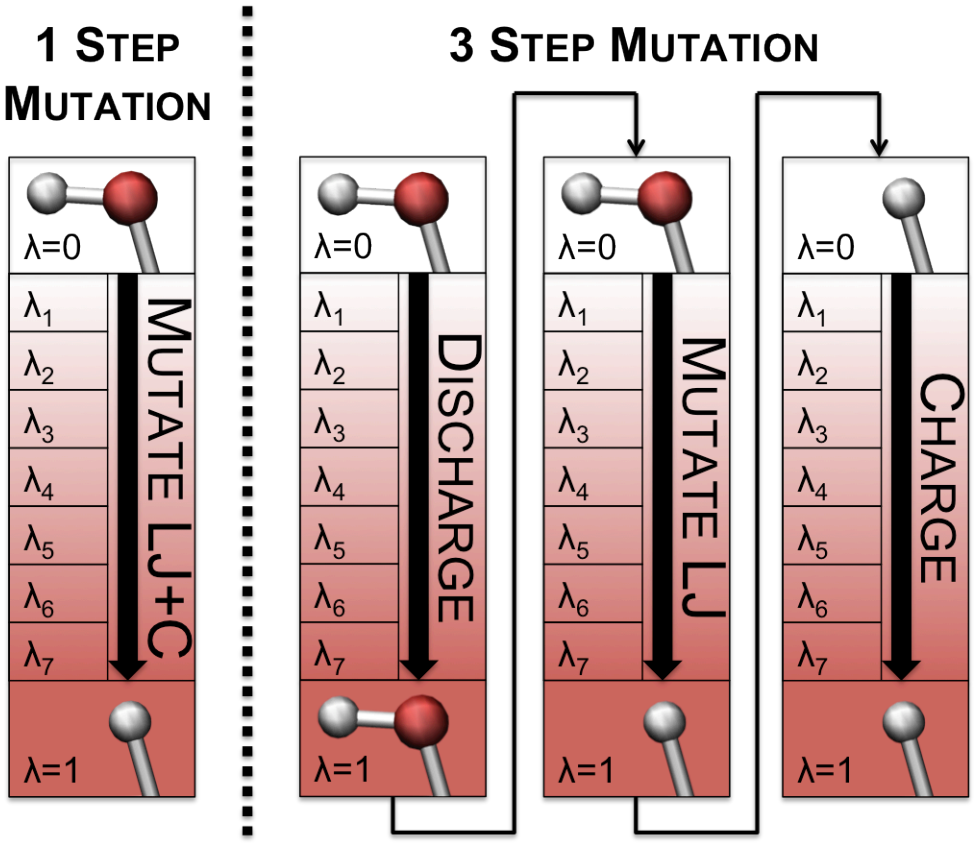

Figure 8. In this example, a hydroxyl group (OH) is mutated into a hydrogen (H) as λ 0→1. When soft-core potentials are used for both Lennard-Jones and Coulomb interactions, TI simulations can be run as a single mutation step with no transition states. However, if soft-core potentials are only employed for Lennard-Jones interactions, Coulomb interactions must be mutated as a separate step in order to avoid the unshielding of atomic charges. The most innocuous way to do this is to perform the vdW mutation in the absence of charge, which requires three mutation steps: **1)** Discharge the mutating region of state A; **2)** Mutate the Lennard-Jones interactions A→B; **3)** Charge the mutating region as state B.

3.3 Application to ConA

A study conducted by Kadirvelraj, et al.[9] in which relative binding affinities of ConA-3MAN versus ConA-3HET were computed using TI is cited here as an example of the method.

All simulations were performed in AMBER 8[68], prior to implementation of soft-core potentials in the AMBER package. In this case, the hydroxyethyl side chain at the C-2 position in the central mannose residue of 3HET was mutated to the native hydroxyl group of 3MAN over two mutation steps. Coulomb interactions were perturbed first, within the shielding of the existing vdW radii in the mutation region. Lennard-Jones interactions were perturbed in the second step, employing a nonlinear scaling scheme for $V(\lambda)$ and small time steps to avoid problems with instability near endpoints. Several sampling times, λ series, and integration methods were examined, and a 21 λ-window splined fit integration protocol with sampling of 100 ps per λ window was found to provide the most accurate computation. The data from this study is presented in **Table 2**, along with experimentally determined values and relative binding free energies calculated with MM-PB/GBSA from the previous section.

The TI values computed by Kadirvelraj, et al.[9] compare reasonably well with the corresponding experimental relative binding affinities from Clarke, et al.[20]. The slight (0.5 kcal/mol) overestimation in the relative energy is attributed to the Arg228 side chain in the ConA binding site, which did not alter its position on mutation, despite the observation that it adopts slightly different orientations in each of the experimentally determined complexes. In TI calculations, the entropy component is implicit and is not necessary to calculate separately.

The relative MM-PB/GBSA energies are consistent with trends observed in the enthalpy, entropy, and overall binding free energy, with MM-GBSA values computed with this implementation of the method being highly accurate compared to experiment for the ConA system. Because the solvent entropy is included implicitly in the MM-PB/GBSA calculation, the magnitude of the predicted enthalpic contribution appears smaller versus experiment. Lower entropy estimates versus experiment can be attributed to the omission of conformational entropy from the MM-PB/GBSA calculation.

While TI is generally a more accurate method, MM-PB/GBSA is still useful to identify experimental trends. Given the considerably greater computational expense of TI, component sum methods are often preferred when a large number of relative binding affinity calculations are required.

Table 2. Relative binding affinities[a] for ConA complexes (3HET → 3MAN)

	MM-PB/GBSA[b]		**TI[9]**	**Experiment[20]**
$\Delta\Delta G_{Binding(GB)}$	**-1.1 (0.6)[c]**	**$\Delta\Delta G_{Binding}$**	**-1.7 (0.8)**	**-1.2**
$\Delta\Delta G_{Binding(PB)}$	**-3.0 (0.6)**			
$\Delta(\Delta H_{MM}+\Delta G_{GB})$	-1.9 (0.4)	ΔΔH	----	-2.3
$\Delta(\Delta H_{MM}+\Delta G_{PB})$	-3.8 (0.6)			
ΔΔS	0.8 (0.6)	ΔΔS	----	1.1

[a]kcal/mol. [b]This work. [c]Values in parentheses are errors in the standard deviation of the mean (standard errors).

3.2 TI protocol

1. See **Initial Setup for Free Energy Simulations** in the MM-PB/GBSA section of this chapter in order to generate equilibrated coordinates for the bound and free versions of state A.
    - If you are studying a mutation in the protein, you will need to perform simulations of the protein-ligand complex and the free protein.
    - If you are studying a mutation in the ligand or are attempting to calculate an absolute binding energy, you will need to perform simulations of the protein-ligand complex and the free ligand.
2. Prepare coordinates of state B from the equilibrated coordinates of state A so that the two have exactly matching per-atom coordinate positions and overlay perfectly in configuration space. This superimposition of initial coordinates is required to facilitate the propagation of a single trajectory through the TI simulations. The unique (mutation) regions of the two states need not have matching coordinates if soft-core potentials are used.
3. Select λ values. It is not necessarily possible to select λs that are well suited to your system without first knowing the shape of the $\langle \partial V(\lambda)/\partial\lambda \rangle$ function. After examining preliminary simulation results, you may find it necessary to alter your λ selections or perform simulations at additional λs (see below). In general, you will need enough λs to adequately accommodate the precision in the shape of $\langle \partial V(\lambda)/\partial\lambda \rangle$.
    - If you plan to integrate using a rule that requires predetermined abscissas, such as a quadrature rule (e.g. Gaussian quadrature), use the corresponding abscissas as λ values.
    - If you plan to integrate using a rule that does not dictate abscissa values, the number of abscissas (e.g. Simpson's rule requires an odd number of points) or the spacing of the abscissas, you may select λ values of your own, which need not be evenly spaced.
4. Run TI simulations at each λ window. This can be done sequentially, feeding the endpoint coordinates of each λ into the simulation for the next λ window, or in parallel, where each λ window is run independently of the others. While the parallel strategy allows TI to be performed faster by running all λs simultaneously, the sequential method has the added benefit that successive λ windows begin with further equilibrated initial coordinates, and may be considered preferable in the absence of time constraints. Regardless, it is important to perform adequate equilibration of the system at each λ window before collecting $\langle \partial V(\lambda)/\partial\lambda \rangle$ for integration (see below).
   Note For TI simulations that involve decoupling always use a thermostat that acts on individual atoms, such as the Langevin thermostat[69].
5. Analyze preliminary simulation data to assess the suitability of selected λ values. A plot of $\langle \partial V(\lambda)/\partial\lambda \rangle$ versus λ should be examined to evaluate its smoothness and

determine if the selected λs adequately describe the shape of the curve. Alter λ values or include simulations at additional λ windows if necessary.

6. Analyze preliminary simulation data to assess convergence. An extensive delineation of convergence checking in MD is beyond the scope of this chapter, but a plot of $\langle \partial V(\lambda)/\partial\lambda \rangle$ over time can be very informative and is a good place to begin. It may be necessary to extend simulations of some λ windows, throwing out data from portions that have not yet reached equilibration. TI simulations are computationally expensive, as they require the potential to be calculated twice at each step. Given current hardware limitations, this often makes long simulations unfeasible, exacerbating methodologically inherent sampling issues. Much work has been done recently to address this shortcoming through development and application of enhanced sampling methods such as Hamiltonian replica exchange accelerated TI.[70-74]
7. Integrate over $\langle \partial V(\lambda)/\partial\lambda \rangle$ as λ 0→1 to calculate ΔΔG. Jorge, et al. have evaluated several commonly used integration techniques for application to TI data[75].
8. Perform TI simulations for both bound and free states of the system, and then subtract $\Delta\Delta G_{Bound} - \Delta\Delta G_{Free}$ to obtain $\Delta\Delta G_{Binding}$.
9. A hysteresis calculation comparing $\Delta\Delta G_{Forward}$ with $\Delta\Delta G_{Reverse}$ can be used to estimate error in the TI[51, 76].

4 Non-equilibrium free energy methods

4.1 General theory

The second law of thermodynamics states that the work *W* done by an external perturbation to bring a system from state A to B is never smaller than the free energy ΔG associated with the transition:

$$\left\langle W_{A\to B} \right\rangle_A \geq \Delta G_{A\to B} \tag{19}$$

The subscript A indicates that the ensemble average of work values is derived from experiments that all begin with the same initial state A. This inequality holds only in cases of a reversible or quasi-static (QS) transformation, where the system is at equilibrium with its environment at all times during the change. In cases of a non-equilibrium, irreversible, or non-QS transformation, the work done by the external perturbation is always *larger* than the free energy difference between states A and B. Furthermore, the work in a non-QS change depends on the chosen path between states A and B, and exceeds the reversible work by an amount equal to the dissipative (non-useful) work.

4.2 Jarzynski's equality

Based on **Equation 19**, the concept of obtaining free energies from a non-QS transformation appears contradictory, or at least non-physical, thus warranting a modified equation for use in non-equilibrium simulations. Hence, in 1997, Jarzynski[77-78]

proposed an equation that allows free energy values associated with the transition A→B to be derived even if the change is performed in the far-from-equilibrium regime:

$$\Delta G_{A\to B} = -kT\ln\left\langle e^{-\left(\frac{W_{A\to B}}{kT}\right)}\right\rangle \tag{20}$$

The exponential average is calculated over states sourced from an equilibrated ensemble of the initial state A. According to Jarzynski's equality, the dissipative work associated with an irreversible transformation is discounted through this exponential averaging. In order for this relationship to hold, the exponential average must be based on an equilibrated canonical ensemble representing state A and must be converged.

The accuracy of free energy values obtained using Jarzynski's equality rests on fulfillment of the convergence criterion, which in turn depends on the degree of irreversibility of A→B. The further from equilibrium a system is, the higher the dissipative work, and the greater the number of states that must be included in the average for it to converge. A practical way to gauge the convergence of the exponential average is to ensure the standard deviation of the work distribution is within a few units of k_BT. However, this is an extremely stringent condition, hardly obtainable in the simulation of large biomolecular systems[79-81]. If the normal distribution of the work is much broader than just a few k_BT, the average will be highly biased by the lower values of work that correspond to rare pulling events.

An alternative approach to avoid problems of convergence is to use the first two terms of the cumulant expansion of **Equation 20**:

$$\Delta G_{A\to B} = \langle W_{A\to B}\rangle_A - \frac{1}{2kT}\left(\langle W^2\rangle_A - \langle W\rangle_A^2\right) \tag{21}$$

Park, et al. have shown that when the probability distribution of the work $P(W)$ is Gaussian (i.e. in the stiff-spring approximation) the free energy is given exactly by these two terms[82-83]. This expression has been used extensively in the literature, as it is not biased by rare events[79, 81-88].

4.3 Steered molecular dynamics (SMD)

In steered molecular dynamics (SMD)[89], a nonphysical spring is used to apply an external force to a molecular system in order to drive some change in the coordinates of that system. This method can be used to impel ion diffusion, steer conformational change, or in the case of protein-carbohydrate complexes, withdraw a ligand from its binding site. The spring, represented by a stiff harmonic potential, is employed to pull the ligand away from its receptor protein at a constant velocity. This allows determination of the steering forces, providing valuable information on the atomistic level regarding the strength of specific interactions involved in binding[82, 89-90]. SMD can be used to reproduce non-equilibrium single molecule experiments such as atomic force microscopy (AFM) and optical tweezers methods[90-92], although the forces are not

quantitatively comparable due to different experimental timescales. Recently, SMD has been applied to study carbohydrate binding[81] and transporting[86] interactions.

To assess binding affinity, the ligand is physically pulled away from the receptor binding site at a constant velocity by means of a stiff harmonic potential of the form:

$$U = k\left[r_0 - \left(r_0 + vt\right)\right] \tag{22}$$

Here, k is the harmonic force constant, v is the velocity at which the distance between the receptor and ligand is increased along the reaction coordinate, and r_0 is the initial distance (at t=0) between the receptor and ligand, usually based on centers of mass. This distance is increased continuously, and can proceed along one or more possible reaction paths. The ideal path is the one that entails the least conformational change between the bound and unbound states.

The total work is obtained through numerical integration of the steering force along the reaction path, according to:

$$W(t) = v\int_0^t F(t)\,dt \tag{23}$$

In the case of a well-converged work distribution, the free energy can be calculated by direct application of **Equation 20** or **Equation 21**.

The main advantage of SMD versus equilibrium free energy calculations depends on the ability to run simulations in the fast pulling regime, which translates into relatively short simulation times.

4.4 Application to ConA

Here, SMD is performed on the ConA-3MAN complex using GROMACS 4.0.5[93], employing the AMBER ff99SB[36] and GLYCAM-06[37] force fields for the receptor protein and carbohydrate ligand respectively. A spring set between the center of mass of 3MAN and the center of mass of Asp16 in the ConA binding site with a harmonic constant of k = 18 kcal/mol·Å^2 was found to satisfy the stiff-spring approximation. The unbinding reaction coordinate was chosen along the vector x=0.5, y=0.5, z=0.0 based on the minimum conformational change of the protein upon unbinding.

An example of the force time series calculated for unbinding of 3MAN from ConA at a pulling rate of 5 Å/ns is shown in **Figure 9**, along with snapshots of the corresponding unbinding events. This profile indicates that the 3MAN unbinding process occurs in two stages. The first stage, taking place between 0 to 1.2 ns at 5 Å/ns, involves the unbinding of the two distal mannose residues (i.e., the terminal mannose connected to the spring and the central mannose) via breaking of contacts with the protein backbone at Asp16, Pro13, and Tyr12, as well as a direct contact with a highly conserved water molecule[7, 10]. The second stage, taking place between 1.3 and 2.5 ns at 5 Å/ns, involves unbinding of the third mannose residue via breaking of several contacts between with

the protein backbone at Lys99, Gly98, and Trp40, as well as a strong ionic interaction with the side chain of Asp208.

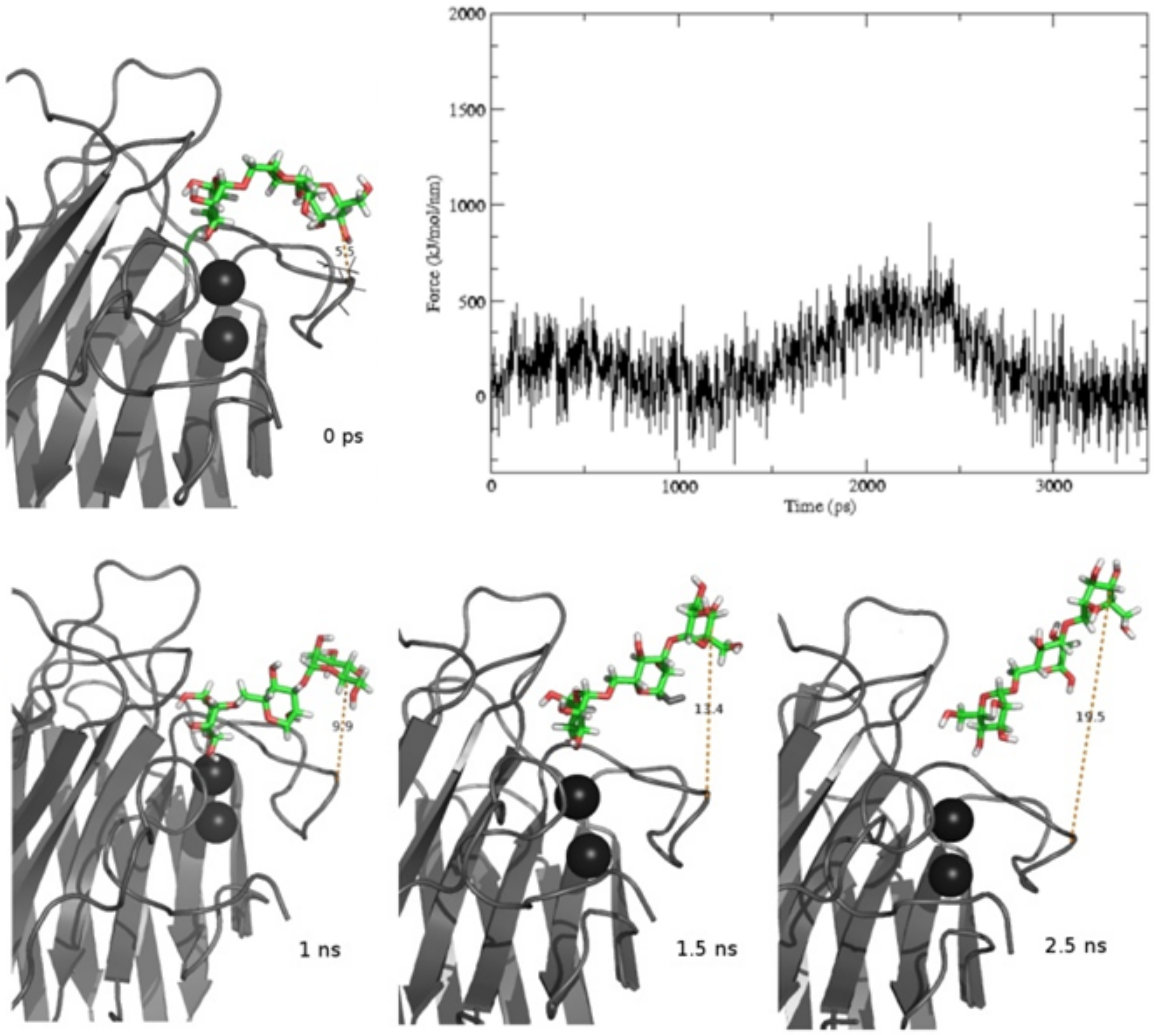

Figure 9. Force time series calculated for the SMD unbinding of 3MAN from ConA. The corresponding unbinding events are shown in ribbon structure, where the Ca^{2+} and Mn^{2+} ions are represented as black spheres.

4.5 SMD protocol

1. See **Initial Setup for Free Energy Simulations** in the MM-PB/GBSA section of this chapter in order to generate an equilibrated ensemble of coordinate frames for the protein-carbohydrate complex of interest.
2. Select harmonic constant *k*. This constant must be large enough to guarantee the validity of the stiff-spring approximation[82-83], which ensures the distribution of the external work is Gaussian. However, excessively large values of *k* generate noise due to fluctuations in the external force applied[82].

3. Select pulling velocity *v*. The choice of an adequate value for *v* is critical for the success of SMD simulations. Data collected from fast pulling (100 Å/ns and above) are very noisy, thus the Jarzynski's equality exponential averaging requires a very large number of trajectories to converge[82, 84]. Conversely, at pull rates lower than 1 Å/ns, the dissipative work is minimal and fewer trajectories are required. Nevertheless, it should be kept in mind that slow pulling is extremely computationally expensive, and may negate the advantages of non-equilibrium free energy methods. Pulling rates in the range of 2-10 Å/ns are generally considered adequate[80, 82-88].
4. Using a series of coordinate frames extracted from the equilibrated ensemble produced in step 2, as well as selected values for *k* and *v*, run SMD simulations. Generate a sufficient number of independent trajectories to reach convergence in the data. The required number of trajectories will depend on the system under study and is sensitive to factors such as system size and ligand flexibility, as well as the value of *v*.
5. If the work distribution is well converged, calculate the free energy by application of **Equation 20** or **Equation 21**. If the work distribution is not Gaussian, methods other than Jarzynski's equality and the cumulant expansion have been tested and reviewed extensively elsewhere[79-80, 87, 92, 94-95].

5 Discussion and conclusions

The ever-present issue of precision versus accuracy in computational simulation is nowhere more obvious and critical than when computing ligand binding free energies. An essential prerequisite for confidence in simulation data is the demonstration that the simulated system has converged on the timescale of relevant molecular motions. In the absence of convergence, the level of precision in the calculation remains uncertain. For MD simulations, longer sampling times may clearly improve convergence, but the use of structural restraints, such as on the protein backbone, can also contribute to this end. In the extreme case of no motional freedom, convergence is trivial. That is, for a single configuration, each method and force field can only result in a single energy. As odd as such an approach may sound in the context of MD simulations, it harkens back to the earliest modeling studies, in which energy minimization was the only method employed. Nevertheless, single-configuration methods provide the current basis for some extremely useful modern applications, such as automated docking.

Once molecular motion is permitted, adequate sampling of all the accessible conformational states is required in order to optimize precision. However, the accuracy of the resulting values is independent of the precision whether data is taken from a single configuration or a converged structural ensemble of a complex. For example, in the case of component sum or direction ΔG calculations (MM-PB/GBSA), convergence may often be reached in less than 100 ns, making this approach relatively high-throughput. Yet, even given a converged data set, the approximations associated with continuum solvent models and entropic estimates currently relegate such calculations to the role of simply ranking the relative binding affinities of carbohydrate ligands.

Correct ranking is nevertheless extremely useful, both in interpreting endogenous glycans binding preferences[13] and potentially in the design of glycomimetics[96-98].

Assuming the methodology is properly implemented in the computational application, the accuracy of the calculated binding energy is a function of the choice of force field for the protein, the carbohydrate, any ions, and the solvent. All too often, the effects of the chosen water model are ignored, with each research group preferring a particular model, often for reasons of familiarity. This is potentially problematic, since some water models behave differently in terms of their strength of interaction with proteins[10] and peptides[99].

Given that the binding free energy represents the preference of the ligand to bind to the protein over remaining in bulk solvent, the water model can be expected to influence the accuracy of certain types of ligand binding calculations. Choice of water model is particularly likely to affect values of *absolute* binding free energies (MM-PB/GBSA, SMD), which, unlike directly calculated *relative* energies (TI), do not benefit from cancellation of systematic errors. In the case of methods employing implicit solvent models, the desolvation free energy contribution can depend heavily on the continuum approximation, for example PB versus GB models. The quality and applicability of continuum solvent models further depend on the set of molecules employed in their parameterization. To date, no continuum solvent model has been developed with a particular focus on reproducing the affinities of protein-carbohydrate interactions. This is a particularly challenging task, given that these interactions depend on a particularly delicate balance of hydrophilic interactions between the carbohydrate and protein/solvent, as well as non-traditional hydrophobic interactions that are common in protein-carbohydrate complexes.

Water models aside, the choice of force field will also affect the accuracy of computed binding free energies. While many current force fields employ similar vdWs terms, each varies significantly in the treatment of electrostatic interactions[2-3, 100]. Some force fields employ partial atomic charges that are transferrable among monosaccharides, while others employ monosaccharide-dependent values. Some employ partial charges derived to reproduce quantum mechanical molecular electrostatic potentials, while others assign charges empirically so as to reproduce bulk condensed phase properties. It is self-evident that employing one charge derivation strategy for the protein and another for the carbohydrate is inconsistent, but is this necessarily inaccurate? There is after all, no such thing as a partial atomic charge. Furthermore, the partial atomic charges employed for the water model may have been developed independently of the protein and carbohydrate models.

One approach to address this question would be to compare the computed binding free energies for a range of protein-carbohydrate complexes. While MM-PB/GBSA methods could be applied, the number of approximations associated with treating desolvation and entropic changes degrades their accuracy. Two alternative approaches remain.

The first is an extension of the TI method, in which the entire ligand is annihilated, or completely decoupled from the remaining system over the course of the simulation.

This is performed for both the bound and free states of the ligand in order to obtain the absolute (standard) binding free energy. In contrast to the case of a minor mutation in either the ligand or the protein, annihilation by double decoupling methods (DDM) requires that the ligand be constrained in the bound state while it is allowed to rotate freely in solution. This results in the need for appropriate entropy-related corrections[101]. Such simulations have been applied to determine the standard binding energies of conserved waters on protein surfaces[10, 76], and have demonstrated that the choice of water model can affect the computed affinities[10]. A limited number of ligands comparable in size to mono- or disaccharides have also been the subject of DDM studies[102], although application to intact carbohydrates may be further complicated by the internal flexibility inherent to many glycans.

SMD offers a second potential approach to test the accuracy of binding energy calculations for a number of combinations of proteins, carbohydrates, and water models. Precise results from SMD require extensive pathway sampling for convergence, which although computationally expensive, are appropriate for parallel implementation.

Lastly, a limitation in binding energy studies arises not from the computational perspective, but from the lack of a large, diverse experimental data set. While the 3D structures of a wide range of protein-carbohydrate complexes have been determined by x-ray crystallography, far fewer uniform data sets exist for the associated ligand binding affinities. Uncertainties therefore arise when comparing experimental data from various methods, such as isothermal titration calorimetry, surface plasmon resonance, and microarray fluorescence. Avidity effects often further compound difficulties in interpreting experimental affinities. That is, many carbohydrate-binding proteins are multimeric, and many glycans contain multiple binding epitopes, thus the measured interaction energies are not necessarily representative of monomeric binding.

References

1. Ernst, B.; Magnani, J. L., From carbohydrate leads to glycomimetic drugs. *Nat. Rev. Drug Discov.* **2009,** *8* (8), 661-677.
2. Fadda, E.; Woods, R. J., Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today* **2010,** *15* (15-16), 596-609.
3. Foley, B. L.; Tessier, M. B.; Woods, R. J., Carbohydrate force fields. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**.
4. Bryce, R. A.; Hillier, I. H.; Naismith, J. H., Carbohydrate-protein recognition: Molecular dynamics simulations and free energy analysis of oligosaccharide binding to Concanavalin A. *Biophys. J.* **2001,** *81* (3), 1373-1388.
5. Rini, J. M., Lectin Structure. *Ann. Rev. Biophys. Biomol. Struct.* **1995,** *24*, 551-577.
6. Bouckaert, J.; Hamelryck, T. W.; Wyns, L.; Loris, R., The crystal structures of Man(alpha 1-3)Man(alpha 1-0)Me and Man(alpha 1-6)Man(alpha 1-0)Me in complex with concanavalin A. *J. Biol. Chem.* **1999,** *274* (41), 29188-29195.

7. Naismith, J. H.; Field, R. A., Structural Basis of Trimannoside Recognition by Concanavalin A. *J. Biol. Chem.* **1996,** *271*, 972-976.
8. Mandal, D. K.; Kishore, N.; Brewer, C. F., Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding on N-linked carbohydrates and ovalbumin to Concanavalin A. *Biochemistry* **1994,** *33* (5), 1149-1156.
9. Kadirvelraj, R.; Foley, B. L.; Dyekjaer, J. D.; Woods, R. J., Involvement of Water in Carbohydrate-Protein Binding: Concanavalin A Revisited. *J. Am. Chem. Soc.* **2008,** *130* (50), 16933-16942.
10. Fadda, E.; Woods, R. J., On the Role of Water Models in Quantifying the Binding Free Energy of Highly Conserved Water Molecules in Proteins: The Case of Concanavalin A. *J. Chem. Theory Comput.* **2011,** *7* (10), 3391-3398.
11. Laitinen, T.; Rouvinen, J.; Perakyla, M., MM-PBSA free energy analysis of endo-1,4-xylanase II (XynII)-substrate complexes: binding of the reactive sugar in a skew boat and chair conformation. *Org. Biomol. Chem.* **2003,** *1* (20), 3535-3540.
12. Liu, Z.; Zhang, Y. Z., Molecular dynamics simulations and MM-PBSA calculations of the lectin from snowdrop (Galanthus nivalis). *J. Mol. Model.* **2009,** *15* (12), 1501-1507.
13. Ford, M. G.; Weimar, T.; Kohli, T.; Woods, R. J., Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity. *Proteins-Structure Function and Genetics* **2003,** *53* (2), 229-240.
14. Kadirvelraj, R.; Gonzalez-Outeirino, J.; Foley, B. L.; Beckham, M. L.; Jennings, H. J.; Foote, S.; Ford, M. G.; Woods, R. J., Understanding the bacterial polysaccharide antigenicity of Streptococcus agalactiae versus Streptococcus pneumoniae. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (21), 8149-8154.
15. Xu, D.; Newhouse, E. I.; Amaro, R. E.; Pao, H. C.; Cheng, L. S.; Markwick, P. R. L.; McCammon, J. A.; Li, W. W.; Arzberger, P. W., Distinct Glycan Topology for Avian and Human Sialopentasaccharide Receptor Analogues upon Binding Different Hemagglutinins: A Molecular Dynamics Perspective. *J. Mol. Biol.* **2009,** *387* (2), 465-491.
16. Yui, T.; Shiiba, H.; Tsutsumi, Y.; Hayashi, S.; Miyata, T.; Hirata, F., Systematic Docking Study of the Carbohydrate Binding Module Protein of Cel7A with the Cellulose I alpha Crystal Model. *J. Phys. Chem. B* **2010,** *114* (1), 49-58.
17. Yao, J. Z.; Nellas, R. B.; Glover, M. M.; Shen, T. Y., Stability and Sugar Recognition Ability of Ricin-like Carbohydrate Binding Domains. *Biochemistry* **2011,** *50* (19), 4097-4104.
18. Tan, C.; Tan, Y. H.; Luo, R., Implicit nonpolar solvent models. *J. Phys. Chem. B* **2007,** *111* (42), 12263-12274.
19. Wang, J.; Tan, C. H.; Tan, Y. H.; Lu, Q.; Luo, R., Poisson-Boltzmann solvents in molecular dynamics Simulations. *Commun. Comput. Phys.* **2008,** *3* (5), 1010-1031.
20. Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G. J., Involvement of water in carbohydrate-protein binding. *J. Am. Chem. Soc.* **2001,** *123* (49), 12238-12247.
21. Kadirvelraj, R.; Grant, O. C.; Goldstein, I. J.; Winter, H. C.; Tateno, H.; Fadda, E.; Woods, R. J., Structure and binding analysis of Polyporus squamosus lectin in complex with the Neu5Ac alpha 2-6Gal beta 1-4GlcNAc human-type influenza receptor. *Glycobiology* **2011,** *21* (7), 973-984.

22. Fogolari, F.; Brigo, A.; Molinari, H., The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.* **2002,** *15* (6), 377-392.
23. Born, M., Volumes and hydration warmth of ions. *Z. Phys.* **1920,** *1*, 45-48.
24. Onufriev, A.; Case, D. A.; Bashford, D., Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.* **2002,** *23* (14), 1297-1304.
25. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990,** *112* (16), 6127-6129.
26. Ghosh, A.; Rapp, C. S.; Friesner, R. A., Generalized born model based on a surface integral formulation. *J. Phys. Chem. B* **1998,** *102* (52), 10983-10990.
27. Dominy, B. N.; Brooks, C. L., Development of a generalized born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B* **1999,** *103* (18), 3765-3773.
28. Tsui, V.; Case, D. A., Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **2000,** *122* (11), 2489-2498.
29. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000,** *104* (15), 3712-3720.
30. Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M., Solvent models for protein-ligand binding: Comparison of implicit solvent Poisson and surface generalized born models with explicit solvent simulations. *J. Comput. Chem.* **2001,** *22* (6), 591-607.
31. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004,** *25* (2), 265-284.
32. Ma, J. P., Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **2005,** *13* (3), 373-380.
33. Karplus, M.; Kushick, J. N., Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981,** *14* (2), 325-332.
34. Chaires, J. B., Calorimetry and thermodynamics in drug design. In *Annual Review of Biophysics*, Annual Reviews: Palo Alto, 2008; Vol. 37, pp 135-151.
35. Case, D. A.; Darden, T. A.; Cheatham, I., T. E. ; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 11, University of California, San Francisco. **2010**.
36. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65* (3), 712-725.
37. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29* (4), 622-655.
38. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

39. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.
40. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
41. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327-341.
42. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraserreid, B., Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides .1. Glycam-93 Parameter Development. *J. Phys. Chem.* **1995,** *99* (11), 3832-3846.
43. Hensen, C.; Hermann, J. C.; Nam, K. H.; Ma, S. H.; Gao, J. L.; Holtje, H. D., A combined QM/MM approach to protein-ligand interactions: Polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem.* **2004,** *47* (27), 6673-6680.
44. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.
45. Tschampel, S. M.; Kennerty, M. R.; Woods, R. J., TIP5P-consistent treatment of electrostatics for biomolecular simulations. *J. Chem. Theory Comput.* **2007,** *3* (5), 1721-1733.
46. Mahoney, M. W.; Jorgensen, W. L., A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000,** *112* (20), 8910-8922.
47. Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999,** *285* (4), 1735-1747.
48. Kirkwood, J. G., Statistical Mechanics of Pure Fluids. *J. Chem. Phys.* **1935,** *3* (300-313).
49. Zacharias, M.; Straatsma, T. P.; McCammon, J. A.; Quiocho, F. A., Inversion of Receptor-Binding Preferences by Mutagenesis - Free-Energy Thermodynamic Integration Studies on Sugar Binding to L-Arabinose Binding-Proteins. *Biochemistry* **1993,** *32* (29), 7428-7434.
50. Liang, G.; Schmidt, R. K.; Yu, H. A.; Cumming, D. A.; Brady, J. W., Free energy simulation studies of the binding specificity of mannose-binding protein. *J. Phys. Chem.* **1996,** *100* (7), 2528-2534.
51. Pathiaseril, A.; Woods, R. J., Relative energies of binding for antibody-carbohydrate-antigen complexes computed from free-energy simulations. *J. Am. Chem. Soc.* **2000,** *122* (2), 331-338.
52. Ganguly, D.; Mukhopadhyay, C., Binding diversity of the two binding sites of ricin B lectin. *Biopolymers* **2006,** *83* (1), 83-94.
53. Bucher, D.; Grant, B. J.; McCammon, J. A., Induced Fit or Conformational Selection? The Role of the Semi-closed State in the Maltose Binding Protein. *Biochemistry* **2011,** *50* (48), 10530-10539.

54. Lawrenz, M.; Baron, R.; McCammon, J. A., Independent-Trajectories Thermodynamic-Integration Free-Energy Changes for Biomolecular Systems: Determinants of H5N1 Avian Influenza Virus Neuraminidase Inhibition by Peramivir. *J. Chem. Theory Comput.* **2009,** *5* (4), 1106-1116.
55. Lawrenz, M.; Wereszczynski, J.; Amaro, R.; Walker, R.; Roitberg, A.; McCammon, J. A., Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy. *Proteins* **2010,** *78* (11), 2523-2532.
56. Simonson, T., Free-Energy of Particle Insertion - an Exact Analysis of the Origin Singularity for Simple Liquids. *Mol. Phys.* **1993,** *80* (2), 441-447.
57. Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R., Thermodynamics of Cavity Formation in Water - a Molecular-Dynamics Study. *Faraday Symposia of the Chemical Society* **1982,** (17), 55-67.
58. Lin, C. L.; Wood, R. H., Free-Energy of Solvation of a Small Lennard-Jones Particle. *J. Comput. Chem.* **1994,** *15* (2), 149-154.
59. Pearlman, D. A.; Kollman, P. A., A new method for carrying out free energy perturbation calculations: Dynamically modified windows. *J. Chem. Phys.* **1989,** *90*, 2460-2470.
60. Mezei, M., Polynomial Path for the Calculation of Liquid-State Free-Energies from Computer-Simulations Tested on Liquid Water. *J. Comput. Chem.* **1992,** *13* (5), 651-656.
61. Resat, H.; Mezei, M., Studies on Free-Energy Calculations .1. Thermodynamic Integration Using a Polynomial Path. *J. Chem. Phys.* **1993,** *99* (8), 6052-6061.
62. Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; Vangunsteren, W. F., Avoiding Singularities and Numerical Instabilities in Free-Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994,** *222* (6), 529-539.
63. Maye, P. V.; Mezei, M., Calculation of the free energy of solvation of the Li+ and Na+ ions in water and chloroform. *Theochem-J. Mol. Struct.* **1996,** *362* (3), 317-324.
64. Pitera, J. W.; Van Gunsteren, W. F., A comparison of non-bonded scaling approaches for free energy calculations. *Mol. Simul.* **2002,** *28* (1-2), 45-65.
65. Steinbrecher, T.; Mobley, D. L.; Case, D. A., Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.* **2007,** *127* (21).
66. Zacharias, M.; Straatsma, T. P.; McCammon, J. A., Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994,** *100* (12), 9025-9031.
67. Steinbrecher, T.; Joung, I.; Case, D. A., Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.* **2011,** *32* (15), 3253-3263.
68. Case, D. A.; Darden, T. A.; Cheatham, I., T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A., AMBER 8, University of California, San Francisco. **2004**.
69. Schneider, T.; Stoll, E., Molecular-Dynamics Study of a 3-Dimensional One-Component Model for Distortive Phase-Transitions. *Phys. Rev. B* **1978,** *17* (3), 1302-1322.
70. Woods, C. J.; Essex, J. W.; King, M. A., The development of replica-exchange-based free-energy methods. *J. Phys. Chem. B* **2003,** *107* (49), 13703-13710.

71. Woods, C. J.; Essex, J. W.; King, M. A., Enhanced configurational sampling in binding free-energy calculations. *J. Phys. Chem. B* **2003,** *107* (49), 13711-13718.
72. Fajer, M.; Hamelberg, D.; McCammon, J. A., Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration. *J. Chem. Theory Comput.* **2008,** *4* (10), 1565-1569.
73. Khavrutskii, I. V.; Wallqvist, A., Computing Relative Free Energies of Solvation Using Single Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. *J. Chem. Theory Comput.* **2010,** *6* (11), 3427-3441.
74. Khavrutskii, I. V.; Wallqvist, A., Improved Binding Free Energy Predictions from Single-Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. *J. Chem. Theory Comput.* **2011,** *7* (9), 3001-3011.
75. Jorge, M.; Garrido, N. M.; Queimada, A. J.; Economou, I. G.; Macedo, E. A., Effect of the Integration Method on the Accuracy and Computational Efficiency of Free Energy Calculations Using Thermodynamic Integration. *J. Chem. Theory Comput.* **2010,** *6* (4), 1018-1027.
76. Hamelberg, D.; McCammon, J. A., Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *J. Am. Chem. Soc.* **2004,** *126* (24), 7683-7689.
77. Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E* **1997,** *56* (5), 5018-5035.
78. Jarzynski, C., Nonequilibrium equality for free energy differences. *Physical Review Letters* **1997,** *78* (14), 2690-2693.
79. Ytreberg, F. M.; Zuckerman, D. M., New methods for approximating free energy differences using Jarzynski's equality. *Biophys. J.* **2004,** *86* (1), 633A-633A.
80. Zhang, D. Q.; Gullingsrud, J.; McCammon, J. A., Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. *J. Am. Chem. Soc.* **2006,** *128* (9), 3019-3026.
81. Bu, L.; Beckham, G. T.; Shirts, M. R.; Nimlos, M. R.; Adney, W. S.; Himmel, M. E.; Crowley, M. F., Probing Carbohydrate Product Expulsion from a Processive Cellulase with Multiple Absolute Binding Free Energy Methods. *Journal of Biological Chemistry* **2011,** *286* (20), 18161-18169.
82. Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K., Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003,** *119* (6), 3559-3566.
83. Park, S.; Schulten, K., Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **2004,** *120* (13), 5946-5961.
84. Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. E., Free energy calculations with non-equilibrium methods: applications of the Jarzynski relationship. *Theoretical Chemistry Accounts* **2006,** *116* (1-3), 338-346.
85. Hwang, H.; Schatz, G. C.; Ratner, M. A., Steered molecular dynamics studies of the potential of mean force of a Na+ or K+ ion in a cyclic peptide nanotube. *J. Phys. Chem. B* **2006,** *110* (51), 26448-26460.
86. Jensen, M. O.; Yin, Y.; Tajkhorshid, E.; Schulten, K., Sugar transport across lactose permease probed by steered molecular dynamics. *Biophys. J.* **2007,** *93* (1), 92-102.

87. Goette, M.; Grubmueller, H., Accuracy and Convergence of Free Energy Differences Calculated from Nonequilibrium Switching Processes. *J. Comput. Chem.* **2009,** *30* (3), 447-456.
88. Ytreberg, F. M., Absolute FKBP binding affinities obtained via nonequilibrium unbinding simulations. *J. Chem. Phys.* **2009,** *130* (16).
89. Grubmuller, H.; Heymann, B.; Tavan, P., Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **1996,** *271* (5251), 997-999.
90. Sotomayor, M.; Schulten, K., Single-molecule experiments in vitro and in silico. *Science* **2007,** *316* (5828), 1144-1148.
91. Dudko, O. K.; Hummer, G.; Szabo, A., Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (41), 15755-15760.
92. Hummer, G.; Szabo, A., Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (50), 21441-21446.
93. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theo. Comp.* **2008,** *4* (3), 435-447.
94. Cossins, B. P.; Foucher, S.; Edge, C. A.; Essex, J. W., Protein-Ligand Binding Affinity by Nonequilibrium Free Energy Methods. *J. Phys. Chem. B* **2008,** *112* (47), 14985-14992.
95. Balsera, M.; Stepaniants, S.; Izrailev, S.; Oono, Y.; Schulten, K., Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophys. J.* **1997,** *73* (3), 1281-1287.
96. Masukawa, K. M.; Kollman, P. A.; Kuntz, I. D., Investigation of neuraminidase-substrate recognition using molecular dynamics and free energy calculations. *J. Med. Chem.* **2003,** *46* (26), 5628-5637.
97. Malaisree, M.; Rungrotmongkol, T.; Decha, P.; Intharathep, P.; Aruksakunwong, O.; Hannongbua, S., Understanding of known drug-target interactions in the catalytic pocket of neuraminidase subtype N1. *Proteins* **2008,** *71* (4), 1908-1918.
98. Wang, P.; Zhang, J. Z. H., Selective Binding of Antiinfluenza Drugs and Their Analogues to 'Open' and 'Closed' Conformations of H5N1 Neuraminidase. *J. Phys. Chem. B* **2010,** *114* (40), 12958-12964.
99. Nerenberg, P. S.; Head-Gordon, T., Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* **2011,** *7* (4), 1220-1230.
100. Woods, R. J.; Tessier, M. B., Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Curr. Opin. Struct. Biol.* **2010,** *20* (5), 575-583.
101. Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A., The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **1997,** *72* (3), 1047-1069.
102. Deng, Y. Q.; Roux, B., Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009,** *113* (8), 2234-2246.